University of Nevada, Reno

Global Ensemble Streamflow and Flood Modeling with Application of Large Data Analytics, Deep learning and GIS

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Hydrology

by

Sepideh Bahrami

Dr. P. Wigand / Dissertation Advisor

August 2019

THE GRADUATE SCHOOL

We recommend that the dissertation
prepared under our supervision by

SEPIDEH BAHRAMI

Entitled

Global Ensemble Streamflow and Flood Modeling with Application of Large Data Analytics, Deep Learning and GIS

be accepted in partial fulfillment of the
requirements for the degree of

DOCOTOR OF PHILOSOPHY

Peter Wigand, Ph. D., Advisor

Anne Nolin, Ph. D., Committee Member

Thomas Albright, Ph. D., Committee Member

Javad Sattarvand, Ph. D., Committee Member

Behrooz Abbasi, Ph. D., Graduate School Representative

David W. Zeh, Ph.D., Dean, Graduate School

August 2019

ABSTRACT

Flooding is one of the most dangerous natural disasters that repeatedly occur globally, and flooding frequently leads to major urban, financial, anthropogenic, and environmental impacts in the subjected area. Therefore, developing flood susceptibility maps to identify flood zones in the catchment is necessary for improved flood management and decision making. Streamflow and flood forecasting can provide important information for various applications including optimization of water resource allocations, water quality assessment, cost analysis, sustainable design of hydrological infrastructures, improvement in agriculture and irrigation practices. Compared to conventional or physically based hydrological modelling, which need a large amount of historical data and parameters, the recent data-driven models which require limited amounts of data, have received growing attention among researchers due to their high predictive performance. This makes them more appropriate for hydrological forecasting in basin-scale and data-scarce regions. In this context, the main objective of this study was to evaluate the performance of various data driven modeling approaches in flood and streamflow forecasting. One of the significant desires in daily streamflow prediction in today's world is recognizing possible indicators and improving their applicability for effective water management strategies. In this context, the authors proposed an ensemble data mining algorithm coupled with various machine learning methods to perform data cleaning, dimensionality reduction and feature subset selection. To perform the task of data mining, three data cleaning approaches: Principle Component Analysis (PCA), Tensor Flow (TF) and Tensor Flow K-means clustering (TF-k-means clustering) have been used. For the feature selection four different machine learning approaches including: K Nearest Neighbor (KNN), Bootstrap aggregating, Random Forest (RF) and Support Vector Machin (SVM) have been

investigated. Out of twelve different combinations of data mining and machine learning, the best ensemble model was TF-k-means clustering coupled with RF, which outperformed the other methods with 96.52% classification accuracy.

Thereafter, a modified Nonlinear Echo State Networks Multivariate Polynomial (NESN-MP) named in the current study as Robust Nonlinear Echo State Network (RNESN) was utilized for daily streamflow forecasting. The RNESN decreases the size of the reservoir (hidden layer which performs random weigh initialization), reduces the computational burden compared with NESN-MP, and increases the interactions between the internal states. The model is thus simple and user-friendly with better learning ability and more accurate forecasting performance. The method has been tested with data provided by the United States Geological Survey (USGS), Natural Resource Conservation Service (NRCS), National Weather Service Climate Prediction Center (NOAA) and Daymet Data Set from NASA through the Earth Science Data and Information System (ESDIS). Each data set includes the daily records of the local observed hydrological and large-scale weather/climate variability parameters. The efficiency of the proposed method has been evaluated in three regions namely Berkshire County (MA), Tuolumne County (CA), and Wasco County (OR). These basins were designated based upon the wide range of climatic conditions across the US that they represent. The simulation results were compared with NESN-MP and Adaptive Neuro-Fuzzy Inference System (ANFIS). The results validate the superiority of the proposed modeling approach compared to NESN-MP and ANFIS. The proposed RNESN approaches out performs the other methods with an RMSE = 0.98.

For flood forecasting, an Evidential Belief Function (EBF) model, both as an individual model *and* in combination with Logistic Regression (LR) methods, has been proposed to prepare the flood susceptibility map. In in this study we proposed a new ensemble of models of Bootstrap

aggregating as a Meta classifier based upon the K-Nearest Neighbor (KNN) functions including coarse, cosine, cubic and weighted as base classifiers to perform spatial prediction of flood. We first selected 10 conditioning factors to spatial prediction of floods and then their prediction capability using the relief-F attribute evaluation (RFAE) method was assessed. Model validation was performed using two statistical error-indexes and the area under the curve (AUC). Results concluded that the Bootstrap aggregating -cubic KNN ensemble model outperformed the other ensemble models. Therefore, the Bootstrap aggregating -cubic KNN model can be used as a promising technique for the sustainable management of flood prone areas. Furthermore, the AUC results indicated that the EBF, EBF from LR, EBF-LR (enter), and EBF-LR (stepwise) success rates were 94.61%, 67.94%, 86.45%, and 56.31%, respectively, and the prediction rates were 94.55%, 66.41%, 83.19%, and 52.98%. The results showed that the EBF model had the highest accuracy in predicting the flood susceptibility map, in which 14% of the total areas were located in high and very high susceptibility classes and 62% were located in low and very low susceptibility classes.

I would like to dedicate this dissertation to my beloved parents Pari and Mansour, for without their inspiration, coaching and enthusiasm none of this would have happened.

Acknowledgments

I would like to express my deepest appreciation and gratitude to my advisors Dr. Peter Wigand and academic supervisor Dr. Anne Nolin for their continuous support and advice. Not only have they aided in my professional development and academic progress, but also in my personal life. I could not have imagined having better mentors throughout this experience. I am also grateful to my other committee members Dr. Behrooz Abbasi and Dr. Javad Sattarvand and Dr. Thomas Albright for their insightful and thought-provoking comments.

I humbly extend my gratitude to Karin Peternel, a fellow alumni of UNR's Hydrologic Sciences graduate program. Thanks are due to Dr. Bahareh Abdollahi, Dr Amin Chitszan, Dr. Khabat Khosravi for their unconditional support along this study.

And finally, I am deeply indebted to my respected and beloved family for their support, and unconditional love.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Streamflow and Flood Forecasting

1.1.1 Background

Climate and weather-driven natural hazards, such as floods, flash floods, storm surges, and consequent massive mass movements and sedimentation are the most prominent natural disasters all over the world. Guha et al has reported that a total of 57% of natural disaster's victims in 2011 are related to "hydrological disasters", which triggered a total economic loss of more than 70 billion US dollars (e.g. 230%average increase compared to the previous decade) (Guha-sapir, Hoyois, & Below, 2011)

Flood is defined as overflow of the resulting stream-flow from heavy rainfall which ultimately covers the flood plain, and areas that are not covered by water under normal conditions (Kron, 2002). With world population increasing, the need for optimizing allocation plans for water resources and energy production demands has become more important. Therefore, the development of technologically driven solutions for controlling water quality and quantity in river systems is an essential task. According to the United Nations International Strategy for Disaster Reduction and data from insurance companies, the socioeconomic effect of floods is growing. Floods affect more people worldwide than any other natural hazard. Flood risk results from the interplay of a range of processes. For river floods, these are the flood-triggering processes in the atmosphere, runoff generation in the catchment, flood waves traveling through the river network, possibly flood defense failure, and finally, inundation and damage processes in the flooded areas. In addition, ripple effects, such as regional or even global supply chain disruptions, may occur.

Therefore, Floods are not considered as secluded events, as they are profoundly accompanied by other issues such as food supply reduction, disease epidemics and environmental deprivation. With variations in intensity and frequency of future climate, prediction of severe climatic events is becoming a key element to shield the social order and commence timely reaction, hence efficiently dropping socioeconomic damage (Pappenberger et al., 2015). Although flood prediction is critical at the local level, it is likewise vital at the global level. Therefore, the management of the immediate response and assistance for major upcoming disasters must be reformed through action by international organizations at different levels. The earlier the planning phase starts, the more improved preliminary actions, organization and data collection are possible, and as a result limitation of the consequences of social and economic losses. While some countries ,e.g., European countries, have apparatuses in place to mitigate the effects of natural disasters, developing countries often struggle through a much longer recovery process. In this context, flood hazard maps which would be accessible on the national, regional and global level, could be attained to intensify preparedness, (Alfieri, Thielen, & Pappenberger, 2012). However, this set of maps is useful to define flood hazard zones, they do not integrate daily changes, which require a real-time system observation(Bui et al., 2019).

A basic flood map can be quickly developed using available datasets, including:

- Flood surfaces (generally from flood model outputs) for a range of flood magnitudes
- Digital Elevation Models (DEM) often from LiDAR
- Forecast floods in the catchment (at storm surge levels)

Additional datasets can be incorporated to improve the complexity and skill of the system to provide perception into the behavior of a flood, such as:

- Actual and forecast streamflow, along with catchment conditions
- A complete GIS dataset of assets, building and roads, infrastructure etc.
- Real-time hydrologic modelling

The best information provided by a flood map is restricted to classifying the flooded area and the potential of flooding for infrastructure and buildings. Taking this a step further, a map of a flood depth surface can be formed by integrating the water surface forecasts and streamflow level onto a Digital Elevation Model (DEM). The depth of the surface not only provides information on the depth of flood and likely flood extent, but also on the strictness of flooding (Bui et al., 2019)

However, for a better result a flood mapping tools can be created by integrating the forecast flood surface and streamflow level with GIS datasets. For instance, a GIS dataset might comprise the low points on evacuation routes to determine what roads are passable, or what facilities require managed evacuation, for example a nursing home. Moreover, it can provide the base levels on all transformers for electricity sub-stations in a given region then by incorporating the flood surface into the GIS, a signal of power availability to the region can be fast determined during flood events. In addition to peak forecast information, time-based flood forecasts complement additional detail to a flood map, and afford information on at risk areas along with time of happening (Bui et al., 2019).

The remote sensing data, such as satellite imagery, can produce summaries of affected areas and improve the management plan(Proud, Fensholt, Rasmussen, & Sandholt, 2011).

Although, various research institutes and national hydrometeorological services use operational flood forecasting tools to surge the attentiveness for floods and any water-related challenges, they often focused on specific river basins or climatic region, which is restricted to national boundaries(Alfieri, Salamon, Pappenberger, Wetterhall, & Thielen, 2012)

Recently various machine learning and data mining algorithms for the flood forecasting have been studied. These approaches include: logistic model tree (Chapi et al., 2017), Nave bayes tree (NBT) (Khosravi et al., 2018b), support vector machine (SVM) (Khosravi et al., 2018b) and hybrid of adaptive neuro-fuzzy inference system (ANFIS) with cultural algorithm (Tien Bui et al., 2018) and bees algorithm or with imperialistic competitive algorithm (ICA) and firefly algorithm (FA) (Bui et al., 2018).

Although these algorithms showed a reasonable ability in the prediction of flood susceptibility mapping, Khosravi et al. (2018c)) (Khosravi et al., 2018a) stated that there isn't a universal guideline to evaluate a model performance under different condition. While every model has advantages and disadvantages, in under different condition, diversemodels must be applied and eventually the best performance has to be selected for the future analysis (Bui et al., 2019).

However, some research shows that bivariate statistical models demonstrate better predictive power than both machine learning and data mining algorithms (Rahmati and Pourghasemi 2017) this is due to the fact that machine learning and data mining algorithms are more complex and require an expert to perform accurate simulations, thus, bivariate

models, which are very simple to run with similar or sometimes better predictive power, can be used as adequate substitutes.

Furthermore, numerous flood forecasting modeling approaches are based on the observed river stage, while future values are deduced through river routing models or by coupling observed rainfall data into hydrological models. Nevertheless, only few attempts have been made so far to move towards operational systems with coupled hydro-meteorological models producing streamflow predictions at the global scale (Sperna Weiland, Van Beek, Kwadijk, & Bierkens, 2010); To the best authors' knowledge, none of these runs operationally with ensemble predictions. Indeed, real-time hydrological modeling requires a large amount of information, including not only static maps describing the surface and sub-surface basin features, but also data assimilation techniques or a long-term balance of streamflow forecasts to give an estimate of the initial conditions, from which the forecast is run (Bui et al., 2019).

With the occurrence of heavy precipitation across most of the U.S., it can be found that streamflow levels are increasing as well. A Climate Central analysis of streamflow data performed at more than 2,100 active gauges across US demonstrated that the number of days with high stream flow (the top 25 percent of readings) has risen over the past 30 years in the largest rivers of the U.S., including the Ohio, Missouri, and Mississippi. ((2017, May 10). High Streamflow is Increasing, Raising Flood Risks. Retrieved from https://www.climatecentral.org/gallery/maps/high-streamflow-is-increasing-raising-flood-risk)

This streamflow analysis agreed with the National Climate Assessment and previous Climate Central analyses findings about rise in heavy precipitation in the Northeast and Midwest, as a result of global warming. Heavy precipitation is the key element driving streamflow and flooding, accompanied with urbanization, the expansion of impermeable surfaces and failure in engineering of dams and levees. The increasing number of days with high streamflow indicates that the risk for stream and river flooding is also on the rise. Additional data from the National Climate Assessment shows this is already happening, as the magnitude of flooding is increasing in the Mississippi and Ohio Valleys, and the Northeast. In addition to heavy rain, spring snow melt can also play a role in streamflow. During spring, the largest increases in high streamflow days occur in the Upper Mississippi River Valley and the Northwest as a result of snowmelt. Therefore, with the knowledge of streamflow contribution in flooding, we have to improve the streamflow prediction accuracy (Bahrami, 2018a)

However, a variety of hydrological models has been recently implemented for forecasting streamflow, in the last decades data-driven models have gained significant interest among researchers. Because, data-driven models are capable of handling highly non-linear, non-additive hydrological processes numerically with no need of understanding underlying physical processes involved. These models include artificial neural network (ANN) (Prakash, Sudheer, & Srinivasan, 2014), recurrent neural networks (Chen, Chang, & Chang, 2013)support vector machines, genetic programming approach(Nayak, Sudheer, Rangan, & Ramasastri, 2005), and neuro-fuzzy (Nayak et al., 2005). Among the mentioned data driven modeling approaches, no single modeling approach consistantly beats the

others. Unlike physic-based models, data-driven models principally depend upon historical observational data, which takes into account watershed characteristics and the physical processes involved. However, the operation of physically based models often necessitates severe computation, user expertise, and parameter identification through field measurement, which can be very challenging. In addition, physic-based models sometimes cannot be adjusted to a minor change in watershed response where the boundary condition changes (Alvisi & Franchini, 2011). All these physic-based models factors inspire the application of data-driven models instead, which mainly aim to yield accurate predictions while disregarding the complicated underlying physical processes.

Most recently, the wavelet-based data-driven modeling approach has grown significantly due to its power in capturing both the periodic and chaotic behavioral trend of time series data(Adamowski & Sun, 2010). The wavelet decays the original signal into several different resolutional levels to extract the useful information and hence raises the model performance (Nourani Vahid A4 - Komasi, Mehdi A4 - Mano, Akira, 2009). The application of wavelet based neural network (WNN) for hydrologic modeling was initially introduced by (Wang et al., 2011), who indicated that the combination of wavelet techniques and ANN could enhance the model accuracy, especially in the long lead-time prediction. Since then, many studies have performed WNN for both the long and short lead-time stream flow forecasting. They have verified that WNNs frequently yield more consistent and accurate results when compared to the traditional ANNs (Adamowski & Sun, 2010) Moreover, several studies have counted uncertainty in wavelet-based flood forecasting to improve the model reliability . In streamflow forecasting, the model

accuracy, in general, deteriorates with the increase in the lead-time, which can be attributed to the weak dependence between the modeled variable and input(s). The prediction error is generally classified into three main categories including phase/temporal error, amplitude error, and shape error (Prakash et al., 2014). The phase error is related to lag in timing of the simulated hydrograph, which is perilous in flow forecasting(Prakash et al., 2014).The amplitude error is mainly caused by the noisy input data or due to the uncalibrated model structure. This would lead to either overfitting or underfitting (Shamseldin & O'Connor, 2010). The shape error is evaluated by the rate of flow change in the rising and falling limbs of a hydrograph. Moreover, the evaluation of the model performance has been shown by the instinctive graphical representation as well as statistical measures such as root mean square error (RMSE), Nash–Sutcliffe coefficient (ESN) and WI, which are objective and quantitative in nature but without considering the temporal dimension of time series (i.e., hydrograph). In a few studies, the temporal error, but only restricted to the peak flow rather the entire hydrograph, has been engrossed to improve the prediction accuracy of peak flow in terms of both magnitude and/or timing(Liu, Brown, Demargne, & Seo, 2011). In such a situation, the model may not warrant the best possible solution when there are multiple peaks as the model performance often biases toward a particular peak flow. Furthermore, several researchers have made effort to reduce the phase error by modifying the modeling approaches. For instance, to diminish the phase error, Abrahart et.al. (Abrahart, Heppenstall, & See, 2007) used correction factor for calibrating ANN models; however, they concluded that this method is only applicable for a short lead-time forecasting. Besides, the model accuracy degrades as the lead-time of forecasting increases(Kasiviswanathan, Cibin, Sudheer, & Chaubey, 2013). In the long lead-time

forecasting, several possible reasons such as the decrease of the interconnection between input(s) and output, over parameterization, and uncalibrated model structure could lead to the increase of prediction uncertainty. However, most studies have focused on conferring the model accuracy, but not the model precision except few (Alvisi & Franchini, 2011)(Kasiviswanathan, He, Sudheer, & Tay, 2016) Additionally, Kasiviswanathan et.al, conducted a research to evaluate the performance of WNN data driven models in forecasting high flow event. Their result indicated that the model they used is not capable of capturing high flow event while pretty accurate forecasting of low flow event.

In view of the above, the primary objective of this research is to identify the robust modeling approach from coupling data analytics with data-driven methods, namely data mining and deep learning algorithm, for the long lead-time streamflow and flood forecasting, through assessing both modeling accuracy and precision using large scale climate variability indices. One of the significant desires in daily streamflow prediction in today's world is: recognizing possible indicators and improving their applicability for effective water management strategies. In the context, the aim of this study is to assess the feasibility of an ensemble streamflow and flood forecasting and early warning system at the global scale, built up with a data mining algorithm to investigate the most important factors in governing streamflow and decrease the amount of input variables. Therefore, the model can evaluate the system performance in its initial stage, where no model parameter has been specifically calibrated.

1.1.2 Objectives

The principal objectives of the current study are as follow:

a) In the current study application of large data analytics has been evaluated. The daily data has been acquired from 1980 to 2018 for 18 parameters in 5 different case study across the United State.

b) Proposing a new ensemble method of data mining using three different data cleaning and dimensionality reduction approaches including: Principle Component Analysis (PCA) Tensor Flow (TF) and Tensor Flow coupled with k means clustering (TF-K-means clustering) for data cleaning. The proposed data preprocessing method is then capable of handling the impact of bad data and reducing uncertainty on streamflow forecasting. Data are firstly screened through data mining algorithm then fed into a high nonlinear deep learning approach, which yields more reliable and accurate forecasting results in the presence of missing or corrupted data. Our goal is to confirm the consistency of the proposed method in ungauged basins with limited observed data and to verify the robustness of the simulation results in different climatic regions.

c) The produced cleaned subset then will be evaluated through four different feature importance algorithms including: K Nearest Neighbor (KNN), Support Vector Machine (SVM), Bootstrap aggregating and Random Forest (RandF) for the following four-fold:

    1. Presenting a better understanding of the underlying processes that affect the streamflow
    2. Reduces Overfitting: Less redundant data means less opportunity to make decisions based on noise.
    3. Increasing the prediction performance
    4. Introducing faster and more cost-effective input variables.

5. In many cases, in application contexts where the search for the best feature set is still an active research topic, the classification accuracy achieved with cleaned data sets is often significantly better than with the full data set.

Data cleaning often leads to insight into the nature and severity of error-generating processes. The result can then give methodological feedback to the modeler to improve study validity and precision of outcomes. It may be necessary to amend the study protocol, regarding design, timing, observer training, data collection, and quality control procedures. In extreme cases, it may be necessary to restart the study. Programming of data capture, data transformations, and data extractions may need revision, and the analysis strategy should be adapted to include robust estimation or to do separate analyses with and without remaining outliers and/or with and without imputation (Van den Broeck et.al 2005).

Using large data entry for the modeling approach including the local observed information, interannual and seasonal climate variability to see the large-scale climatic indices impact on governing streamflow. As this modeling approach is not limited by the amount of data entry, it can combine various parameter sets with different resolutions. This behavior results in:

1. Capturing all possible inducing parameters in streamflow generation
2. It is a promising tool in predicting high flow event (up to 1800 m3/s).
3. It is not highly parametrized which in turn reduces the uncertainty in the model structure.

4. Daily streamflow forecasting is conducted, which is essential for effective operations of inflow to a given reservoir requiring active regulation of water storage for optimum use of available resources.

d) Introducing a robust modeling approach for streamflow forecasting called Robust Nonlinear Echo State Network (RNESN). RNESN is a modified nonlinear echo state network with more accurate forecasting results and less computational time compared to Nonlinear Echo State Networks Multivariate Polynomial (NESN-MP) which was developed by Bahrami et al. (Bahrami et.al 2018). In RNESN, the reservoir size (hidden layer which performs random weigh initialization) is considerably decreased compared with NESN-MP to diminish the computation load. The efficiency of the proposed method lies on long lead-time daily streamflow forecast up to 50 days which is helpful for an early flood warning system.

e) To perform spatial prediction of flood, we also aim to use a novel ensemble models of Bootstrap aggregating as a Meta classifier based on the K-Nearest Neighbor (KNN) functions including coarse, cosine, cubic and weighted as base classifiers at Haraz watershed in the northern Iranian province of Mazandaran. Ten conditioning flood factors and their prediction capability using relief-F attribute evaluation (RFAE) method has been established. Models validation will be performed using two statistical error-indexes and the area under the curve (AUC).

f) Another contribution of the present study is evaluating the performance of evidential belief functions (EBF) method which is rarely applied for flood analysis, but it has been used for other categories of natural disaster such as landslide susceptibility assessment, Land subsidence, and to predict groundwater potential zones. The main purpose of this

research is to generate a flood susceptibility map using the EBF method, as EBF has rarely been used for floods and has shown high accuracy in previous studies involving another natural hazard mapping. The results of this method are compared to EBF-LR (enter method), EBF-LR (stepwise method), and EBF from LR methods. The results of the current study will be useful for land-use planning and management for future flood mitigation studies.

All the contributions are elaborated in detail and evaluated in different case studies. The process is repeated by considering representative basins from different climatic and land use scenarios from different regions in the United States and Iran.

This dissertation is based upon the following papers, of which I am the primary author of three published papers, and one submitted paper for the section 1 and section 2. The section 3 and section 4 are based upon one published and one submitted paper of my research in collaboration with other universities.

1. S. Bahrami, M.A. Chitsazan, M.S. Fadali, O.Kisi, P.Wigand "Global Ensemble Streamflow and Stormflow Forecasting for Flood Early Warning with Application of Large data Analytics and Deep Learning Using Large-Scale Climate variability Indices", Submitted to the Journal of Hydrology,
2. S. Bahrami, P. E. Wigand, "Daily Streamflow Forecasting Using Nonlinear Echo State Network", International Journal of Advanced Research in Science, Engineering and Technology, 2018, pg no: 6720-6727.
3. S. Bahrami, P. E. Wigand, "Bad Data Analysis on Streamflow Forecasting Using Nonlinear Echo State Network", International Journal of Advanced Research in Science, Engineering and Technology, 2018, pg no: 7054-7060.
4. S. Bahrami, P. E. Wigand, "Sensitivity Analysis on Daily Streamflow Forecasting", International Journal of Advanced Research in Science, Engineering and Technology, 2018, pg no: 7312-7317.
5. S. Bahrami, D. Tien Bui, K. Khosravi, Shahabi, H. "A Novel Flood Modelling Based on K-Nearest Neighbor Classifier ensembled with Data Mining Algorithm.", Journal of Remote Sensing, Submitted to Remote Sensing Journal.
6. Tien Bui. D.; Bahrami, S.; Khosravi, K.; Shahabi, H.; Daggupati, P.; Adamowski, J.F.; Melesse, A.M.; Thai Pham, B.; Pourghasemi, H.R.; Mahmoudi, M.; Flood spatial modeling in northern Iran using remote sensing and gis: A comparison between evidential belief functions and its ensemble with a multivariate logistic regression model. Remote Sens. 2019, 11, 1589.

Organization of the Reaserach:

- In the First chapter a comprehensive introduction focused upon the importance of flood and streamflow forecasting has been presented.
- In the second chapter the methods of data mining, machine learning and deep learning in daily streamflow forecasting has been investigated.
- In the third chapter the ensemble method of data mining and machine learning in flood susceptiblity mapping has been proposed
- In the forth chapter Spatial Flood modeling using remote sensing and GIS with the application of evidential belief functions for feature selection has been evaluated.
- In the fifth chapter the comprehensive conclusion regarding daily streamflow forecasting and flood spatial mapping has been presented.

2 Data analytics Integration into Daily Streamflow forecasting

2.1 Global Ensemble Streamflow Forecasting for Flood Early Warning with Application of Large data Analytics and Deep Learning Using Large-Scale Climate Variability Indices

Abstract— Streamflow and flood forecasting can provide important information for various applications including optimization of water resource allocations, water quality assessment, cost analysis, sustainable design of hydrological infrastructures, improvement in agriculture and irrigation practices. Compared to conventional or physically based hydrological modelling, which need a large amount of historical data and parameters, the recent data-driven statistic hydrological models, require only a limited amount of data, and as a result, have received growing attention among researchers due to their high predictive performance. This makes them more appropriate for hydrological forecasting in basin-scale and data-scarce regions. The generated cleaned data subset will then be evaluated using four different feature importance algorithms including: K Nearest Neighbor (KNN), Support Vector Machine (SVM), Bootstrap aggregating and Random Forest (R and F). Therefore, in this paper the authors propose a data mining approach for feature selection and data cleaning followed by a novel data driven method for daily streamflow forecasting. In this context, in order to decrease the number of input variables in the modeling process, a feature selection approach using ensemble Random Forest and Tensor Flow k means clustering has been proposed for the preprocessing phase. Thereafter, a modified Nonlinear Echo State Networks Multivariate Polynomial (NESN-MP) named in the current study as Robust Nonlinear Echo State Network (RNESN) has been utilized for the forecasting phase. The RNESN decreases the size of the reservoir (hidden layer which performs random weigh initialization), reduces the computational burden compared with NESN-MP, and increases the interactions between the internal states. The model is thus simple and user-friendly with better learning ability and more accurate forecasting performance. The proposed method does not need complex optimization, parameter tuning, or extensive

training process. Furthermore, the model can cope with corrupt and missing data better compared to NESN-MP. The method is tested with data provided by the United States Geological Survey (USGS), Natural Resource Conservation Service (NRCS), National Weather Service Climate Prediction Center (NOAA) and Daymet Data Set from NASA through the Earth Science Data and Information System (ESDIS). Each data set includes the daily records of the local observed hydrological and large-scale weather/climate variability parameters. The efficiency of the proposed method has been evaluated in three regions namely Berkshire County (MA), Tuolumne County (CA), and Wasco County (OR). These basins were designated based on the wide range of climatic conditions across the US that they represent. The simulation results are compared with NESN-MP and Adaptive Neuro-Fuzzy Inference System (ANFIS). The results validate the superiority of the proposed modeling approach compared to NESN-MP and ANFIS.

2.1.1 Introduction

The number and diversity of water-related challenges such as floods, prolonged droughts, glacier dynamics, and economic to population growth are steadily increasing (Dehghani et al., 2014). To tackle these challenges, new hydrologic measurements, new approaches for modeling hydrological process, and new methods for analyzing hydrological data are required (Kirchner, 2006). In the last decades, hydrological models, particularly seasonal hydrological forecasting models, have been successively developed p to explore solutions for sustainable water management (Song et al., 2015). The models are developed to forecast hydrological processes like streamflow up to a few months ahead (Sene et al., 2018). Streamflow forecasting is essential in terms of hydroelectric power programming, flood mitigation, agricultural and domestic water supplies, and irrigation management, which involves dynamic regulation of reservoir storage for optimal use of available water resources (Viel et al., 2016). Therefore, developing an optimal streamflow forecasting model as a stochastic property of environmental modeling is crucial (Tree-Ring Society. et al., 2001). The existing dynamicity, intrinsic complexities, and disordered geographies in the spatial and temporal expansion of the streamflow prediction models may obstruct the accurate prediction process (Bayazit, 2015). Note that the performance of any hydrological model is highly correlated to the set of model parameters. To decrease the large uncertainty in the application of hydrological models, all feasible parameter sets in the generation of the hydrological process must be considered. There are various parameter sets that affect streamflow generation, such as, local micrometeorological conditions, soil standard (e.g. soil temperature and soil moisture at different zone), SNOTEL (e.g. snow water equivalent, high and low temperature, precipitation increment,

and precipitation accumulation), zonal and meridional winds, relative humidity (vapor pressure), and atmospheric circulation on snow budget (solar radiation) (Brown et al., 1996). Note that large-scale climate variability patterns are imperative factors in streamflow forecasting that have been usually neglected. The Pacific North American pattern (PNA) serves as the main contributor on snow cover variability over North America (at higher PNA index there is a lower snow cover). Similarly, the El Niño-Southern Oscillation (ENSO) has a significant impact on accumulated snow cover in the Western US that changes streamflow level considerably (Rasouli et al., 2011). The impact of ENSO on streamflow timing has been widely reported (Kennedy et al., 2009). It results in lower winter precipitation for El Niño and higher winter precipitation for La Niña episodes in the Pacific Northwest and has the opposite effects in the desert Southwest USA. Consequently, in US coastal regions, most of the streamflow comes from early spring snowmelt and seasonal high flow events. Furthermore, due to change in atmospheric pressure, the North Atlantic Oscillation (NAO) has considerable influence on Arctic Oscillation (AO), accumulated snow, and wind patterns across the eastern US that affects streamflow levels (Coulibaly et al., 2005). Therefore, streamflow generation patterns have deteriorated due to global climate change pattern over the globe.

Furthermore, uncertainty analysis prior to the model calibration is key to the effective implementation of the hydrologic model. The major application of sensitivity analysis is to indicate the uncertainties in the input parameters of the model, which could affect model performance. There are different optimization algorithms that have been developed and can be applied in the hydrologic model. They can be performed with different objective functions to calibrate and quantify the uncertainties in the system. The first purpose of this

study was to evaluate the model calibration performance and sensitivity of parameters using TensorFlow k means clustering and Random Forest algorithm for predicting daily streamflow. Thereafter, we proposed a powerful streamflow forecasting engine with unique features, while incorporating the interannual and seasonal climate variability impacts into streamflow forecasting.

2.1.1.1 Forecasting Models

The daily development of hydrology has mainly been formed by the need for solving practical problems, such as managing water resources, accurate forecast of flood or drought events, and manipulating water supply infrastructure. Hence, many hydrologists are developing practical predictive models for operational purposes that are of paramount importance for flood risk reduction.

Although physically-based models have been used widely in the past, their application is limited because they suffer from an implicit upscaling principle. Thus, they do not extrapolate properly because their underlying premises restrict their applicability to a limited range of parameter values. For example, physically based models calibrated on a one-time interval often perform poorly on another time interval with different patterns of rainfall and runoff. Furthermore, they suffer from some oversimplifying assumptions that do not yield accurate forecasting results. In general, physically-based models are highly repetitive and preserve repeating themselves throughout their boundary condition (Beven and Binley, 1992). Physical models utilize the average of the state variables like water flux, volumetric water content, and hydraulic potential over the whole study area, which result in obscuring the heterogeneity of the subsurface.

Seasonal streamflow forecast models rely on various parameter sets including land surface attributes, initial hydrological conditions, soil properties, groundwater, snowpack, and the current streamflow. Initial hydrological conditions (IHCs) was the preliminary step for the development of the ensemble streamflow prediction approach in the 1970s (Arnal et al., 2017). This approach suffers from several issues such as uncertainties associated with the process of calibration, complicated model operation, and physical limitations (Dettinger et al., 2004). Various statistical, physical, and conceptual models have been proposed for streamflow forecasting (Partington et al., 2012). Statistical models such as regression-based models do not appropriately indicate the relationship between variables (Chua and Wong, 2011). However, conceptual hydrological models (e.g., soil and water assessment tool (SWAT)) (Arnold et al., 1998) incorporates mathematical formulations with various hydrological processes to enhance the forecasting accuracy (Peng et al., 2017). Slater et al., (2015) conducted research to forecast streamflow in deterministic and probabilistic terms for all initialization months, flow quintiles, and seasons. Their results showed relatively accurate streamflow forecasts from low to high flows. However, the accuracy of their models is not reliable due to the heterogeneous hydrogeological characteristics of the watershed system. Note that large data input, overparameterization, high dependency on some necessary values may limit the application of comprehensive simulation models (Carquex et al., 2018). Besides, streamflow is influenced by many factors such as evapotranspiration, rainfall, atmospheric circulation, and temperature which cause nonlinear and time-variable process. Because all the mentioned models assume a linear or near linear relationship between the input and output series, in the last two decades researchers have focused on alternative data-base predictive methods (Pagano et al., 2009).

Artificial intelligence (AI) models have been presented as a soft computing tool and data-driven based engine for streamflow forecasting and corresponding floods. These models are based on artificial neural networks (ANN) (Asadi et al., 2019), fuzzy network sets (Ahmed and Sarma, 2007), genetic programming (Barge et al., 2016), generalized regression neural network (GRNN) (Yaseen et al., 2018), support vector machine (SVM) (Kisi, 2015), relevance vector machine (RVM) (Xu et al., 2016), and nonparametric methods (Bhuiyan et al., 2017). Among these methods, fuzzy network models depend on user expertise, while the efficiency of the others depend upon the model structure to infer the relationship between the input and output variables. Asadi et al., (2019) demonstrated that SVM or empirical mode decomposition support vector machine (EMD-SVM) yields the most accurate results compared to autoregressive moving average (ARMA), ANN, and multiple linear regression (MLR) (Kalra et al., 2013). However, Lutz et al., (2014) indicated that the forecasting accuracy of the annual streamflow throughout the state of Utah using ANN and SVM are the same.

Because ANN models have some drawbacks including over-fitting and under-fitting, slow learning speed, the curse of dimensionality, and slow convergence to a local optimum, they perform poorly in the processing of complex hydrological phenomena (Duan et al., 1992). Typically, their disadvantages also : (1) They suffer from increasing complexity along with processing time; (2) They need elaborate parameter tuning and optimization task; (3) They call for nonconvex optimization to trap in local optima; (4) As complex structures of hydrological models with many parameters, the optimization choice of parameters is a difficult and time-consuming task; (5) The results obtained from previous studies are inconsistent due to the differences in the study areas (by considering small

catchment), input datasets, and the selected structures for the models; (6) In the context of hydrological modeling, in a larger catchment, there is more heterogeneity, which needs a more spatially distributed representation of the area.

Furthermore, many studies have applied the original streamflow time series as the input variables in their forecasting model, which results in missing some features of different resolution (Guo et al., 2013). A few studies incorporate the exogenous effects of all other hydrological variables such as climate variability or ocean-atmosphere circulation impacts on the streamflow fluctuations (Makkeasorn et al., 2009). During the last decade, there is growing interest in incorporating atmospheric circulation variability and outputs of numerical weather prediction models in streamflow forecasting (Rasouli et al., 2011). Due to existing complex weather conditions all over the globe, climate variability has a large influence on hydrological processes. In this context, large-scale climate indices should be considered in streamflow forecasting for an effective operational strategy [Kashid, 2010]. Li et al. [35] investigated the relationship between large-scale ocean-atmosphere patterns, which effect the annual maximum flood in the Wangkuai Reservoir watershed in China. Based on their results, three main climate variability indices including: North Pacific Oscillation (NPO), North Atlantic Oscillation (NAO), and Atlantic Oscillation (AO) are major contributors in flood peak generation. Furthermore, Marohasy [34] applied ANN to estimate the impact of climate indices in the prediction of rainfall in Australia. They also performed an optimization analysis to indicate the best optimal choice of inputs including climate indices to deliver improved performance in monthly rainfall forecasting. Kashid et al. [31] examined the impact of lagged rainfall, Equatorial Indian Ocean Oscillation indices, El Niño Southern Oscillation (ENSO) indices, and longwave radiation on weekly

rainfall using genetic programming. They suggested that information about large-scale atmospheric circulation patterns can be successfully used for prediction of weekly rainfall. Additionally, Mekanik et al. [33] conducted an investigation to predict long-term spring rainfall. They used an ANN model integrated with multiple linear regression analysis to the inputs of Indian Ocean Dipole (IOD) indices and lagged ENSO. They found that a combination of lagged ENSO-IOD indices can improve forecasting accuracy.

All previous studies offer extrapolative information on large-scale climate conditions under complex weather modes. However, the validated climate indices are such an important input variable for long-term prediction, they primarily have an inadequate emphasis on exploring the relationship between large scale climate indices and streamflow mechanisms. As their models cannot incorporate these input variables in combination. In the respect, this study aims to recognize highly-influenced climate variability indices along with local observed information, and yield a suitable and reliable model which delivers the best predictive performance through application of all contributing indices.

One of the significant goals in Daily streamflow prediction in today's world is: recognizing possible indicators and improving their applicability for effective water management strategies [36]

In this paper, a powerful streamflow forecasting engine with unique features is proposed. The forecasting engine is reliable, accurate, and user-friendly. Furthermore, we aim to incorporate the interannual, seasonal, and climate variability impacts in different regions in the US into streamflow forecasting.

However , recently, Echo State Networks (ESN) has been presented by Jaeger. The ESN's architecture is divided into two parts: a dynamical reservoir consisting of recurrent

topology of nonlinear processing elements (PEs), Otherwise stated "echo states", and a readout which is a memoryless linear network. The outputs of internal PEs are imported into the readout which reads the reservoir's output and then will create the network result. The most stimulating assets of ESNs is that the training process is only happening in the readout while the recurrent topology of the PEs weight is fixed.

Sacchi et.al 2007, conducted a research to evaluate the ESN performance in the monthly water inflow forecasting. In this study the performance of the ESN is compared with ANFIS, Self-Organizing Nonlinear Auto- Regressive model with exogenous input (SONARx), and the Radial Basis Function (SONARX-RBF) network. The results demonstrated that the ESN affords more accurate results for one-step ahead water inflow.

N. J. de Vos 2012 applied ESN in rainfall-runoff forecasting. The results display that the ESN yield better results compared to feedforward networks and old traditional recurrent networks. However, the recent studies prized a valuable finding for the application of ESN in water flow forecasting, but they all suffer from three important shortcoming: First they used just one input variable to predict the output and they thus are highly dependent on the water flow accuracy. Second their model also will create outputs from a memoryless linear network (readout) which can not capture the high nonlinear behavior in the water flow, and third to update the errors they are performing gradient decent algorithm

In the current study we aim to propose a new architecture of ESN which does not have the problem with adapting the system inputs as we consider various input variables. Moreover, to count for nonlinearity, the recommended training algorithms in the readout, uses various multivariate polynomial state of internal weights. This will also decrease the computational complexity and instability in the internal state through fastening the process of training. Besides, to eliminate the gradient decent over the time and topology, the new structure of ESN applies Recursive Least Square (RLS) approach to update the errors.

2.1.1.2 Uncertainty Analysis

As an inherent symptom of any modeling task, all hydrologic models suffer from some degree of uncertainty with regards to input data, initial or boundary conditions, forcing

data, and model structure along with poor knowledge of hydrological process mechanism (Doherty and Johnston, 2003). Therefore, model uncertainty is an important issue when developing a modeling system (Farmer and Levin, 2018). Many streams in the world do not have accurate observed streamflow data or their data are hard to access or even corrupted. The National Research Council paying increased attention to minimizing the impacts of bad data among investors for uncertainty assessments of hydrologic prediction (Saltelli, 2008). Bad data can be produced because of ungauged basins, potentially inaccurate measurement, incomplete data collection, uncertain estimates, "fat-fingered" data entry, policy concerns, mis-categorization, etc. (Doherty and Johnston, 2003). Given the hydrological process complexity, using an adaptation of globalized or regionalized uncertainty is optimal. Furthermore, regarding bad data analysis, there are several methodological studies for predicting streamflow response in ungauged basins with bad data, which utilize deterministic physically based models to calculate streamflow. They are based on distributed hydrologic parameters and statistical regionalization using regression models to transfer hydrologic information from gauged to ungauged basins. Moreover, the distributed hydrologic parameters approach focuses on dispersing errors into measurement, parameter and structural uncertainty. The uncertainties are then disseminated toward the model output.

Statistical regionalization is a challenging task in hydrological science (Cybenko, 1989) due to poor streamflow data, which is normally calibrated. Moreover, the results have been usually examined on different basins, while every catchment characteristic varies from one case to another. Consequently, there is no universal method for regionalization, and the uniqueness of the watersheds and the obscurity of parameters

brings major uncertainty in the ungauged basins' simulations (Samuel et al., 2011). International organizations, such as the United Nations Development Program (UNDP) and the World Bank, are seeking a precise approach to the development and management of freshwater resources.

Artificial intelligence (AI) methods are popular developments in several hydrological areas due to their ability to incorporate a tried-and-true model with no need for prior knowledge of the existing functional or nonlinear input-output relationship (Valizadeh et al., 2017). Shu and Ouarda, (2008) considered the homogeneous region characteristics to find similar hydrological sites for predicting flood quantile in ungauged basins. Chen et al., (2010) applied hydrological records of nearby catchments with similar homogenous characteristics to predict streamflow of ungauged catchments. They concluded that the spatial characteristics and temporal distribution considered in the applied model reflect most of the rainfall-runoff behavior in nature. Some researchers also used regression trees and model tree ensembles to predict a complete flow-duration curve (FDC) for streams (Schnier and Cai, 2014). Senent-Aparicio et al., (2019) combined machine learning with the soil and water assessment tool (SWAT) to estimate instantaneous peak flow (IPF) in areas where sub-daily observational data are scarce. The results of their study can contribute to the superior ability of extreme learning machine (ELM) to estimate IPF, thereby reducing uncertainties associated with IPF estimations.

All previous studies for ungauged estimation have applied the homogeneous nearby basin parameters, which give inaccurate results because every catchment is unique in its characteristics. Hence, a direct transfer of model parameter values from gauged to

ungauged basins may not be appropriate. Thus, there is a need to produce a more accurate estimation of daily streamflow for ungauged basins (Buytaert and Beven, 2009). More importantly, the limited data accessible for physical parameterization heavily rely on model calibration, which sometimes results in parameterization schemes that conflict with the physical understanding of the region's hydrology.

Several investigations depended on empirical relationships, for example, curve numbers and the Hargreaves equation produced for temperate regions (Singh and Goyal, 2017). These limitations are probably going to bring considerable uncertainty into model projections, especially in situations where climatic or environmental conditions vary from those experienced in the calibration period. Therefore, most forecasts still reflect considerable uncertainty that develops with time and restrains the predictability of observed events beyond a lead time of a few weeks. These computationally costly developments are not always achievable and modelers have to balance the tradeoff between the costs and profits of improving all model aspects (Flato, 2011). The proposed model provides a valuable complement to physical models, especially in data-scarce regions with little data accessible for model parameterization. It also has powerful performance in dealing with bad data due to its high learning capability that yields accurate forecasting.

2.1.1.3 Feature Selection and Data mining

Data mining is a ground-breaking technology, developed with database and artificial intelligence. It is a processing overture of action of extracting trustworthy, novel, useful and understandable patterns from a database. Currently, data mining has been used in

business management, production control, electronic commerce, market analysis and scientific research and many other fields to explore a wide range of applications.

Clustering is the process of grouping the data into classes or clusters, so that objects within a cluster have high similarity in agreement to one another but are very dissimilar to objects in option clusters. A cluster of data objects can be treated collectively at one time as one group, and may be considered as a class of data compression. Unlike classification, clustering is an effective means for partitioning a set of data into groups based upon data similarity and then ascribe labels to the relatively small number of groups. Clustering is an unsupervised learning, as it does not rely on predefined classes and class labelled training examples. For this reason, clustering is a form of learning by observation, rather than learning by examples. As shown in Figure 1, three clusters are formed containing data points based upon center position. A cluster having a greater number of points is cluster of good quality [2][7].

During the current study we perform the feature selection and data mining for classification and clustering to aid in solving some issues arising during modeling. The data mining algorithm cluster is run and reforms large datasets with little value into a small datasets with high value. Smaller input data sets are desirable, because not only do they decrease the model complexity, they are also more user friendly.

With the application of feature selection, we will be able to recognize the most effective and meaningful inputs in governing streamflow, because a noisy dataset makes it more problematic to realize the patterns. Even if the input data set is not an issue, performing feature selection is essential, because unnecessary and redundant inputs can degrade the

accuracy of the model. Tolosi et al. (2011) claimed that numerous commonly used classification algorithms can produce misleading feature importance rankings if the training datasets are comprised of large clusters of correlated features. This can misperceive explanation of the model, because large groups of predictive features can be pre-screened and even appear unrelated.

Based upon the distribution of the features in the training dataset, there are many different methods, for example, principal component as cluster centroid [as in Huang et al. (2003a)] or even several representatives [as in Jäger et al. (2003)] that may produce better results. Tolosi et al 2011, indicated that with a high related feature input, classical model selection algorithms including Random Forest or Penalized Logistic Regression are unstable.

2.1.1.4 TensorFlow K-means Clustering

In the context of above study, we proposed Tensor Flow k means clustering algorithm for identifying and cleaning interrelated features. In data science, cluster analysis (or clustering) is an unsupervised-learning method that can help to understand the nature of data by grouping information with similar characteristics. The clusters of data can then be used for creating hypotheses on classifying the data set. The k-means algorithm is one of the clustering methods that proved to be very effective for the purpose. The k-means algorithm starts with the choice of the initial centroids, which are just random guesses of the actual centroids in the data. After starting with some guesses for the centroid locations, the k-means algorithm then updates those guesses based upon the data. The process is to assign each sample a cluster number, representing the centroid it is closest to. After that,

the centroids are updated to be the means of all samples assigned to that cluster (Clustering and k-means. Retrieved from https://databricks.com/tensorflow/clustering-and-k-means) Initially, we performed ensemble Tensor flow k-means clustering to generate a good quality dataset (evenly distributed with least amount of missing data) which works as follows:

1. Generate samples from initial centroids
2. Randomly choose initial centroids
3. Associate each sample to its nearest centroid
4. Update each centroid to be the mean of the samples associated to it
5. Remove outliers which are the values most distant from the range of the values allowed for that feature. Their inclusion can lead to a bad fit later while building a model.

Then we applied Random Forest (RF) as a feature importance and classification approach to obtain the most effective inputs identified for each region. TF can assign similar weights to interrelated features and thus recover model stability and interpretability. Feature selection can be reliable when it has been performed on an evenly distributed data, which can present the whole data set the best. Because, the data set is high-dimensional with lots of missing data, a tensor flow as a data mining algorithm has been conducted to clean the missing data and generate an appropriate data set for the task of feature selection.

2.1.1.5 Random Forest

Random forest is an ensemble learning approach, which uses decision trees as base learners. The “ensemble learning” produces various classifiers and combines their results by boosting (see, e.g., Shapire et al., 1998) and Bootstrap aggregating Breiman (1996). The

base learners in ensemble learning are defined as high variance, low bias models and each single base learner acquires a different aspect of data through both row and column sampling. For the task of classification, the combination is performed by taking a majority vote. Then a specific feature that maximizes information gain (I.G.) or reduction in Gini impurity, which is more computationally parsimonious, will split the data. The feature Data is then divided amongst its children according to the value of a splitting feature. If the feature is categorical, data belonging to each category of splitting feature goes to a separate child. In the case of a numerical feature, the best threshold value of the feature (the one used to decide in favor of this feature to be used as splitting feature) is used to split data into two parts, each going to one child. Scikit-learn's random forest model has a feature importance_ attribute that gives the value of Gini impurity reduction caused by each feature across all levels normalized across trees. The only hyperparameter of interest here is the number of base learners. A grid search is performed which gives 33 base learners as an optimal value. Random forests seldom overfit, usually, they saturate with an increasing number of base learners, increasing computational overhead without deteriorating performance.

The objective of performed feature selection in the current study is three-fold as follows:

1. Increasing the prediction performance
2. Introducing faster and more cost-effective input variables and reducing overfitting as it causes less opportunity to make decisions based upon noisy inputs.
3. Presenting a better understanding of the underlying process that affect streamflow

2.1.2 Contributions

The principal contributions of the paper are as follow:

1) A powerful streamflow forecasting engine called robust nonlinear echo state network (RNESN) is proposed. RNESN is a modified nonlinear echo state network with more accurate forecasting results and less computational time compared to NESN-MP. In RNESN, the reservoir size is decreases considerably compared to NESN-MP to diminish the computation load. Several nonlinear relations between the internal states are added to the read-out in RNESN structure to increase the learning capability. Therefore, the interactions of the internal states are increased by modifying the readout functions, which results in better learning capability and more accurate forecasting. Furthermore, the feedback matrix is removed because it has an insignificant impact on forecasting results, which further reduces the computational time.
2) A modified streamflow forecasting engine is developed to minimize the impact of bad data on streamflow forecasting. The readout in RNESN is modified to improve the performance of the proposed method in dealing with bad data. Multivariable polynomial functions of the internal states are added in the readout to increase the nonlinear interactions, which yields more reliable and accurate forecasting results in the presence of missing or corrupted data. Moreover, the robustness and reliability of NRESN are evaluated with limited data. Our goal is to confirm the consistency of the proposed method in ungauged basins with limited observed data and to verify the robustness of the simulation results in uncertain inputs.

3) We aim to incorporate the local observed information, interannual and seasonal climate variability impacts into streamflow forecasting in different regions in the US. As this modeling approach is not limited to the amount of data entry, it can combine various parameter sets with different resolutions. This behavior results in capturing all possible inducing parameters on the streamflow generation. This model is not limited to a specific region and is applicable to various hydrological regions. Furthermore, daily streamflow forecasting is conducted, which is essential for effective operations of inflow to a given reservoir that requires active regulation of water storage for optimum use of available resources. Unlike similar studies in the literature, our forecasting engine presents long-term streamflow forecasting.

The effectiveness and robustness of the proposed methodology are evaluated. All the contributions are elaborated in detail and evaluated in different case studies. The process is repeated by considering representative basins from different climatic and land use scenarios from different regions in the United States. The simulation results of RNESN are compared with those for NESN-MP and ANFIS (Jang, 1993) to validate the proposed method. As the ANFIS model reflects inputs with preferences to cope with model uncertainties and imprecision, we have been considering this Modeling approach as an additional validation approach.

2.2 Study Area

Catchment characteristics are descriptors of the landscape, which forms catchment behavior by manipulating how catchments store and allocate water among other sources. There is a rising recognition that a large sample of catchments can provide intuition that

cannot be gained from a small sample (Gupta and Raman, 2014). The large-sample data set utilized in this paper provides local observed information and interannual and seasonal climate variability (Table 1, Brown et al., 1996; Thornton et al., 2012) . Numerical weather forecasts have been provided by the NOAA (National Oceanic and Atmospheric Administration) (Hamill et al., 2006). The Niño 3.4 index represents the impact of ENSO as it measures the Pacific sea surface temperature (SST) anomalies. Besides the variables shown in Table 1, daily index values for PNA, AO, and NAO were also obtained from the climate prediction center (CPC) and NOAA. Furthermore, soil standard parameters in higher elevations are not usually available. Therefore, to capture all the physiographic and local features, streamflow is considered as one of the input data. This in turn indirectly incorporates the topographic conditions like slope, steepness, land use practices, plant coverage, and snow budget into daily streamflow forecasting. The selected catchments have two years of continuous data records from 2016 to 2018 and are also minimally impacted by anthropogenic effect (Newman et al., 2015). Climatic indices were obtained from Daymet meteorological forcing data (Addor et al., 2017), which indicate that the annual precipitation cycle is apparently strongest over the Pacific coast (high during winter in Cascade area) and California along the Sierra Nevada and is also weakest along the Atlantic coast (Addor et al., 2017). The data sets are already available to the public to enable users to assess their reliability. To indicate the model applicability and ensure its spatial consistency in different climatic regions, three different basins distributed across the US are chosen. In Massachusetts basin, Berkshire County as the snow-dominant watershed is selected (Case Study 1). In California, we chose Tuolumne County as a mixed pluvial-nival (Case Study 2). In Oregon, a basin in Wasco

County is chosen as the rain dominate watershed (Case Study 3) shown in Fig. 1. Table 2 shows a summary of the selected watersheds (USGS.gov).

Table 1**.** The input variables for streamflow prediction modeling

Variables	Description
Average daily precipitation (mm/day)	N15* - Daymet
Average max daily temperature (C°)	Provided by the NOAA
Average min daily temperature (C°)	Provided by the NOAA
Average daily vapor pressure (Pa)	Provided by the NOAA
Average daily solar radiation (W/m^2)	Provided by the NOAA
Average daily streamflow (m^3/s)	Provided by the USGS
Average daily precipitable water (m)	NRCS-SNOTEL
Average daily accumulated precipitation (m)	NRCS-SNOTEL
Average daily SWE(m)	NRCS-SNOTEL
Average daily precipitation increment (m)	NRCS-SNOTEL
Average daily soil moisture (%)	NRCS-soil standard
Average daily Soil temperature (%)	NRCS-soil standard
Average daily wind amplitude (m/s)	NOAA- GFS
Average daily relative humidity(%)	NOAA- GFS
NAO	NOAA- Clim*
AO	NOAA- Clim*
PNA	NOAA- Clim*
ENSO-Niño 3.4 index, central equatorial Pacific sea surface temperature (SST) anomalies	NOAA- Clim*

N15=daily meteorological forcing data from Daymet (Thornton et al., 2012)

NOAA- GFS =numerical weather forecasts by the NOAA Global Forecasting System (GFS) model (Hamill et al., 2006), NOAA-Clim*= NOAA/National Weather Service via ftp://ftp.cpc.ncep.noaa.gov/cwlinks/. (Burn, 2008); NRCS-SNOTEL =Natural Resource Conservation Service, Snow Telemetry (SNOTEL e.g. snow water equivalent, high and low temperature, precipitation increment, and precipitation accumulation)

Table 2. Study area characteristic

Region	Berkshire County, Massachusetts	Tuolumne County, California	Wasco County, Oregon
Watershed	USGS 01333000 Green River at Williamstown, MA	USGS 11284400, Big C AB Whites Gulch NR Groveland CA	USGS 14096850 Beaver Creek below Quartz Creek, NR Simnasho, OR
Hydrologic unit code	2020003	18040009	17070306
Coordinates	Latitude: 42°42'32", Longitude: 73°11'50" NAD27	Latitude: 37°50'31", Longitude: 120°11'02" NAD27	Latitude: 44°57'32", Longitude: 121°23'35" NAD27
Drainage area (Meter Sq.)	110.34	42.47	375,548,276
Gage datum (Meter)	186.85 meter above NAVD88	780.50 meter above NGVD29	688.85 meter above NGVD29

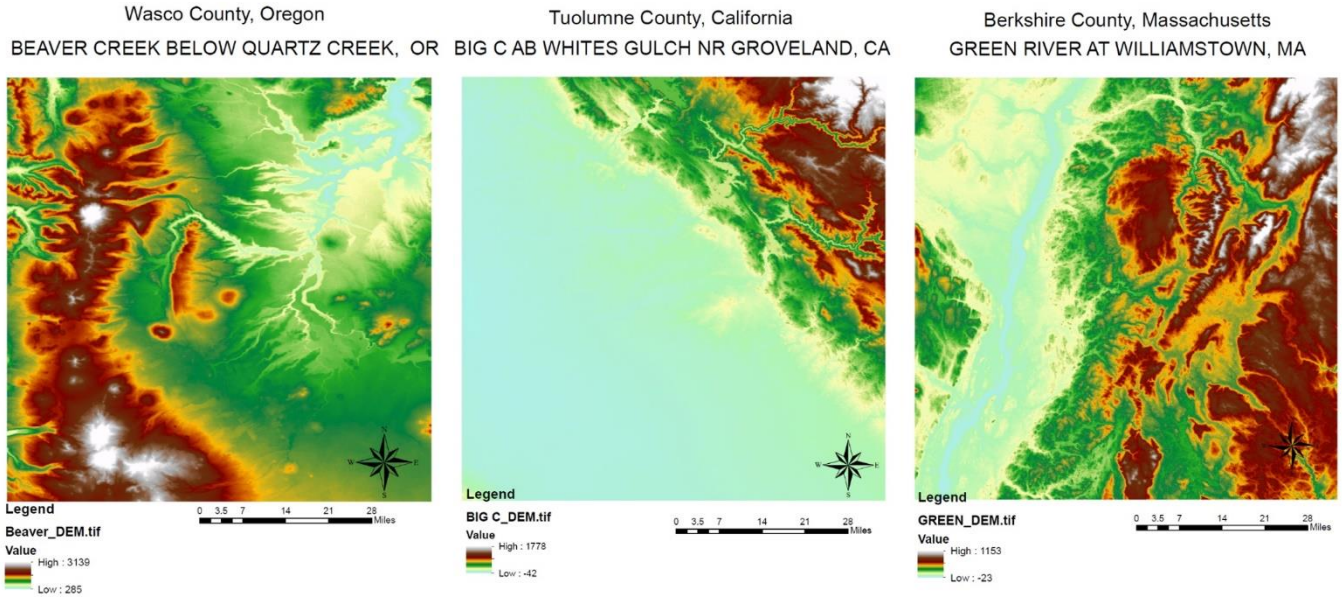


Figure 1. Case study regions

Methodology

In the current study we focused upon performing the feature selection and dimensionality reduction as it aids to solve some issues arising during modelling. It cleans up the existed large amount of dataset with little value to a small amounts of dataset with high value.

Smaller input data set is desirable because not only it decreases the model complexity, it is also more user friendly.

We would be able to recognize the most effective and meaningful inputs in governing streamflow because a noisy dataset makes it more problematic to realize the patterns. Even if the input data set is not an issue, performing feature selection is essential, because unnecessary and redundant inputs can degrade the accuracy of the model. In this study, the features were calibrated using the observed daily streamflow data. Initially, we performed RF without data cleaning to obtain the most sensitive inputs identified for each region. Afterwards, we readjusted the parameters from the first run results then coupled the RF with tensor flow to clean up for the missing data. Feature selection can be reliable when it has been performed on evenly distributed data, which can present the whole data set the best. Because, the data set is high-dimensional with lots of missing data, a tensor flow as a data mining algorithm has been conducted to clean the missing data and generate an appropriate data set for the task of feature selection.

2.2.1 Data Cleaning and Dimensionality Reduction

It has been shown that numerous commonly used classification algorithms can produce misleading feature importance rankings if the training datasets comprise large clusters of correlated features. This can misperceive explanation of the model, because large groups of predictive features can be pre-screened and even appear unrelated.

Based upon the distribution of the features in the training dataset, there are so many different methods, for example, principal component as cluster centroid [as in Huang et al.

(2003a)] or even several representatives [as in Jäger et al. (2003)] may produce better results.

Tolosi et al 2011, indicated that with a high related feature input, classical model selection algorithms including Random Forest or Penalized Logistic Regression are unstable.

We proposed Tensor Flow algorithm for identifying and cleaning interrelated features. TF can assign similar weights to interrelated features and thus recover model stability and interpretability. In the context, we show that the measures of feature relevance corresponding to the above-mentioned methods are biased such that the weights of the features belonging to groups of correlated features decrease as the sizes of the groups increase, which leads to incorrect model interpretation and misleading feature ranking.

We investigated the Random Forest algorithm to perform feature selection for the following reasons:

1. Presenting a better thoughtful of the underlying process that affect the streamflow
2. Reduces Overfitting: Less redundant data means less opportunity to make decisions based on noise.
3. Increasing the prediction performance
4. Introducing faster and more cost-effective input variables

In the current study, for dimensionality reduction, we studied Principle Component Analysis(PCA), Tensor Flow(TF) and Tensor Flow K-means clustering. The differences of three variants (denoted by PCA, TF and TF-Kmeans clustering) has been examined to find the subsequent classification accuracy. Based upon these clustering procedures first a mean shift of all subsets such that the mean for each subset becomes 0 has performed.

Which results in a subsets comprising min (features, instances) linear combinations of the original attributes ((Gansterer and Ecker, n.d.)

PCA: Each feature of M is normalized by its standard deviation (i. e., z-scored). These normalized values are used for computing eigenvalues and eigenvectors (i. e., there is no difference between the covariance and the correlation coefficient) and also for the computation of the new attributes.

Table 3.The propose algorithm for PCA, Calculate the M matrix of the desired subset

Input: *PC's = new attribute of data set
Output: Feature instances as M matrix
performed a mean shift of all features
Set the mean for each feature equals zero
Calculate the PCA with min (features, instances) linear combinations of the original
Return PCA subset
Compute the eigenvalues and eigenvectors of the covariance matrix of M
Calculate the Cov (M)
Multiply M with the eigenvectors of Cov (M).
Return M

*PC's = The PCs are (i) linear combinations of the original attributes, (ii) orthogonal to each other, and (iii) capture the maximum amount of variation in the data.

TF: TensorFlow is a machine learning library developed by Google and released as open source on November 2015 [1]. It provides an interface for expressing and executing machine learning algorithms. These algorithms are described by a directed graph composed of a set of nodes, which represent the instantiation of an operation and have zero or more inputs and zero or more outputs. The graph represents a dataflow computation, with extensions for allowing some kinds of nodes to maintain and update persistent state and for branching and looping control structures within the graph in a manner similar to Naiad [18]. One of the key features of TensorFlow is its ability to run on multiple CPU and other

devices, reducing the computation times for complex problems. Additionally, it could use CUDA extensions for general-purpose computing on graphics processing units. The library is currently used in dozens of commercial Google products such as speech recognition, Gmail, Google Photos and Google Search [27]. It is available as a Python API as well as a C/C++ API. In this work we used the Python API provided by the SkFlow project, which allows building DNN using the interface of Scikit Learn [21]. Specifically, we used a 3 layers DNN with 5 hidden units per layer. The Adagrad algorithm was used as optimizer. In the current paper we used the original Tensoflow prepared by Python library.

TF-k means clustering: In this paper, we have implemented a centroid selection approaches in k-means clustering for improving the recommendation process for recommender systems. We have applied these selection approaches along with traditional k-means for comparing their performance. The algorithms present the centroid selection procedure for k-means clustering. After selecting k seeds, next steps are followed as per Algorithm 1, to accomplish k-means clustering ((2017, December 19). Clustering using K-means algorithm. Retrieved from https://towardsdatascience.com/clustering-using-k-means-algorithm-81da00f156f6)

Table 4.The propose algorithm for TF-k-Means clustering, Choose k users from the dataset, as centroids

Input: U= users in training set; k= total number of clusters;
Output: k centroids = {c_1; c_2; . . . ; c_k}
1: Define desired number of clusters, k with generating random data points with a uniform distribution and assign them to a 2D-tensor constant.
2: Randomly choose initial centroids from the set of data points
3: Find element-wise subtraction of points and centroids that are 2D tensors.
4: Expand points and centroids into three dimensions.
5: Use the broadcasting* feature of subtraction operation, pairs of arrays on an element-by-element basis
6: Calculate the average pair wise distance between points and centroids and determine the cluster assignments
7: Compare each cluster with a cluster assignments vector
8: Get points assigned to each cluster, calculate mean values which are refined centroids
9: Update the centroids variable with the new values.
9: Return updated centroid values {c1; c2; . . . ; ck} along with the cluster assignments values. Algorithm

*Broadcasting: In the context of data mining, we use some less conventional notation. We allow the addition of matrix and a vector, yielding another matrix: C = A + b, where $C_{i,j} = A_{i,j} + b_j$. In other words, the vector b is added to each row of the matrix. This shorthand eliminates the need to define a matrix with b copied into each row before doing the addition. This implicit copying of b to many locations is called broadcasting

2.2.2 Machine Learning Methods

For evaluating the classification performance of the reduced feature sets we used four different machine learning methods. For detailed information about these methods, the reader is referred to the respective references given.

Experiments were performed with a support vector machine (SVM) based on the sequential minimal optimization algorithm using a polynomial kernel with an exponent of 1 (Platt,1998); a k-nearest neighbors (kNN) classifier using different values of k (1 to 9)

(Cover and Hart, 1995); a Bootstrap aggregating ensemble learner using a pruned decision tree as a base learner (Breiman, 1996); a random forest (RandF) classifier using a forest of random trees (Breiman, 2004); and K-means clustering (s. Zahra et al. / Information Sciences 320 (2015) 156–189), PCA (Jolliffe, 2002)

2.2.3 Deep Learning Algorithms

An ESN is a discrete time recurrent neural network. Fig. 2 illustrates an ESN with *K* input states, *N* internal states, and *L* output states (see Chitsazan et al., 2019 for details).

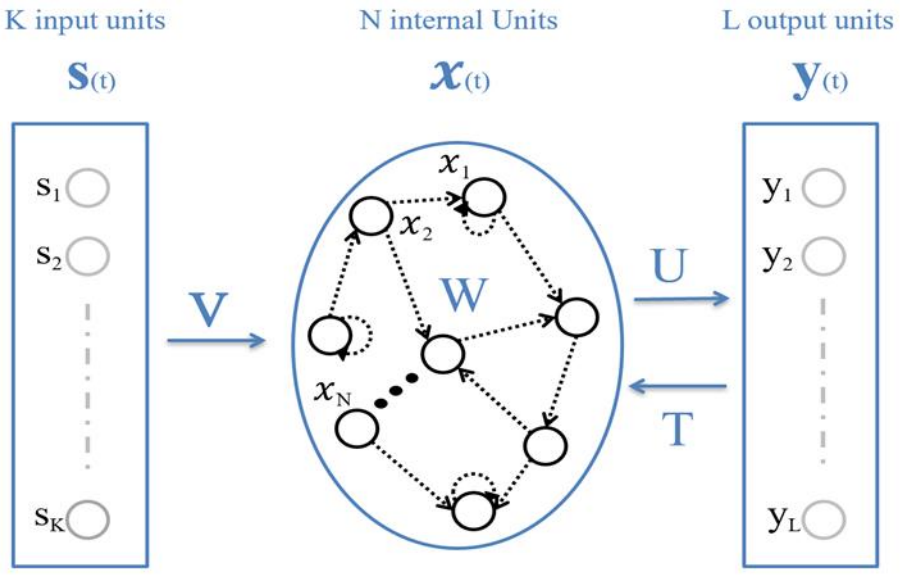


Figure 2. The topological structure of Echo State Network. Input layer denotes the time series input variables, internal state presents a randomly generated weight matrix and the output layer generates the predicted value of streamflow

.

The internal states $\boldsymbol{x}_{(t)}$ and the output states $\boldsymbol{y}_{(t)}$ are calculated in a process as follows:

$$\boldsymbol{x}_{(t+1)} = \mathbf{f}\left(V \times \boldsymbol{s}_{(t+1)} + W \times \boldsymbol{x}_{(t)} + T \times \boldsymbol{y}_{(t)}\right) \quad (1)$$

$$\boldsymbol{y}_{(t+1)} = U \times \boldsymbol{x}_{(t+1)} \quad (2)$$

After choosing the activation function (**f**), the recursive least squares (RLS) algorithm is utilized to compute U where X is full rank.

$$U = \left((X^T X)^{-1}(X^T D)\right)^T \tag{3}$$

The two types of activation function used in the literature are the logistic function and the hyperbolic tangent. The logistic function has the mathematical form:

$$f(x) = \frac{e^x}{1 + e^x} \tag{4}$$

It takes a real-valued number as the input and squashes it between 0 and 1. It aims to introduce nonlinearity in the input space where the network learns to capture complicated relationships. A large negative number passed through the logistic function approaches 0 and a large positive number of approaches 1. Due to this property, the logistic function can be interpreted as an indicator of the firing rate of the neuron; from not firing at all (0) to fully-saturated firing at an assumed maximum frequency. However, logistic activation functions have gradually become less popular due to two major drawbacks called killing the gradients and nonzero-centered outputs (King and Zeng, 2001).

The hyperbolic tangent activation function has the mathematical form:

$$Tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{5}$$

It takes a real-valued number and squashes it between -1 and $+1$ and saturates at large positive and negative values. However, its output is always zero-centered, which helps because the neurons in the later layers of the network receive inputs that are zero-centered.

The activation functions used in this study to translate input signals to output signals is the hyperbolic tangent (tanh). The choice of this function is based upon a large range of nonlinear positive and negative input data, which can be mapped appropriately by this function. With few exceptions, tanh helps the gradient descent converge faster because it makes more uniform steps through the feasibility space of the error function. In some cases, it helps zero-center the data and avoids zigzagging during gradient descent optimization. When the data are not zero centered, gradient descent can only optimize weights of the same node in zigzag. One of the potential problems using tanh is a flat error surface near the origin. Because the saturation zone occurs (gradient close to zero) at very high values

near -1 or 1 and at those high values the gradient/slope will have very small values (Bengio et al., 1994).

As the network updates weights to learn during backpropagation, the changes exclusively depend upon the gradients of output. It performs backpropagation to modify the weights through gradient descent such that the output is minimized. If that gradient is very small, the update in weights is small and thus learning becomes slow. This is known as the vanishing gradient, which was a major obstacle to the success of deep learning (Hochreiter et al., 2001).

In the current work, to overcome this drawback, multiple different techniques are proposed. The V matrix is used to transposes the input variables into the data range that the activation functions lies in (tanh [-1, 1]). Second, weight initialization (avoid initialization with very small weights) to evade early saturation the weights are initialized between [-1,1] because, if the initial weights are too large, then most neurons will get saturated and hence the network will hardly learn, and third using Recursive Least Square (RLS) functions to modify errors and train weight function (U matrix). Because gradient descent back propagation error has the problem of vanishing gradient at a higher value of the input, the RLS is applied. The training process is explained in detail in Chitsazan et al. (2019).

NESN-MP is also proposed to decrease the number of internal states considerably compared to classical ESN by utilizing a cubic multivariable polynomial. This radically diminishes the computational burden. Fig. 3 demonstrates the NESN-MP with a total of

$2p + p^2$ units (Chitsazan et al., 2019).

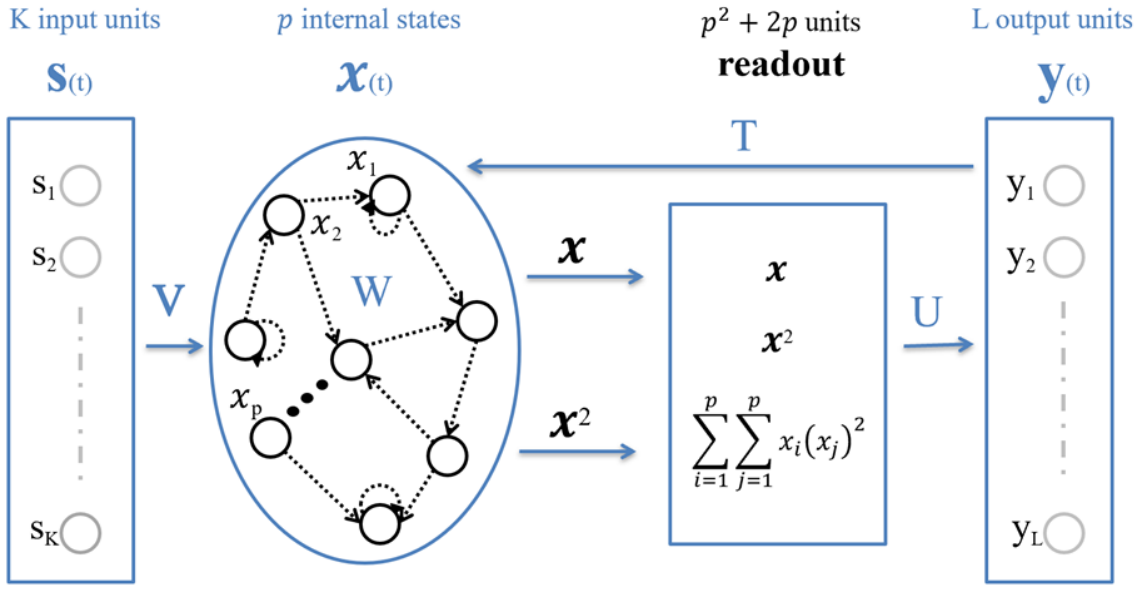


Figure 3. The topological structure of Nonlinear Echo State Network-Multivariate Polynomial. In this schematic, the readout denotes the layer performing multivariate polynomial calculations

In this work, RNESN is proposed. RNESN decreases the number of internal states compared to NESN and minimizes the computational burden and maximizes the interaction between the internal states. Fig. 4 demonstrates the NESN with a total of $3p + 3p^2 + p^3$ units shown in Table 2. Furthermore, each unit in the readout may be created by three internal states compared to two in NESN-MP (shown in eq. (13) and Table 3). This in turn increases the learning capability that yields more accurate forecasting results. W is changed from $\left\lceil \frac{N}{p+2} \right\rceil \times \left\lceil \frac{N}{p+2} \right\rceil$ in NESN-MP to $\left\lceil \frac{N}{p^2+3p+3} \right\rceil \times \left\lceil \frac{N}{p^2+3p+3} \right\rceil$ which $\left\lceil \frac{N}{p^2+3p+3} \right\rceil = p$. Additionally, V varies from $N \times K$ to $p \times K$. The weight matrices (W and V) are utilized to compute the reservoir's internal states. The following equation is used for updating the vector of internal states

$$\boldsymbol{x}_{(t+1)} = \mathrm{f}(W.\boldsymbol{x}_{(t)} + V.\boldsymbol{s}_{(t+1)}) \tag{6}$$

and the readout vector is

$$\bar{\boldsymbol{x}}_{(t+1)} = [\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} \boldsymbol{x}_i^F \, \boldsymbol{x}_j^{2S} \boldsymbol{x}_k^{3T} \, , F = 0 \; or \; 1, S = 0 \; or \; 1, and \; T = 0 \; or \; 1,] \qquad (7)$$

where $\boldsymbol{x}^2{}_{(t+1)} = \left[x^2_{1_{(t+1)}}, x^2_{2_{(t+1)}}, \ldots, x^2_{p_{(t+1)}}\right]$, and $\boldsymbol{x}^3{}_{(t+1)} = \left[x^3_{1_{(t+1)}}, x^3_{2_{(t+1)}}, \ldots, x^3_{p_{(t+1)}}\right]$ p is the number of internal states $\left[\frac{N}{p^2+3p+3}\right]$, $\boldsymbol{s} \in R^{K\times 1}$ (input vector), $\boldsymbol{x} \in R^{p\times 1}$ (internal state vector), $\bar{\boldsymbol{x}} \in R^{(p^3+3p^2+3p)\times 1}$ (readout vector), and $\boldsymbol{y} \in R^{L\times 1}$ (output states).

Table 5. The number of variables in RNESN. F=First order, S= Second order, T= Third order

F	**S**	**T**	**Number of units in** $\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} \boldsymbol{x}_i^F \, \boldsymbol{x}_j^{2S} \boldsymbol{x}_k^{3T}$
1	1	1	p^3
1	1	0	p^2
1	0	1	p^2
1	0	0	p
0	1	1	p^2
0	1	0	p
0	0	1	p
0	0	0	0

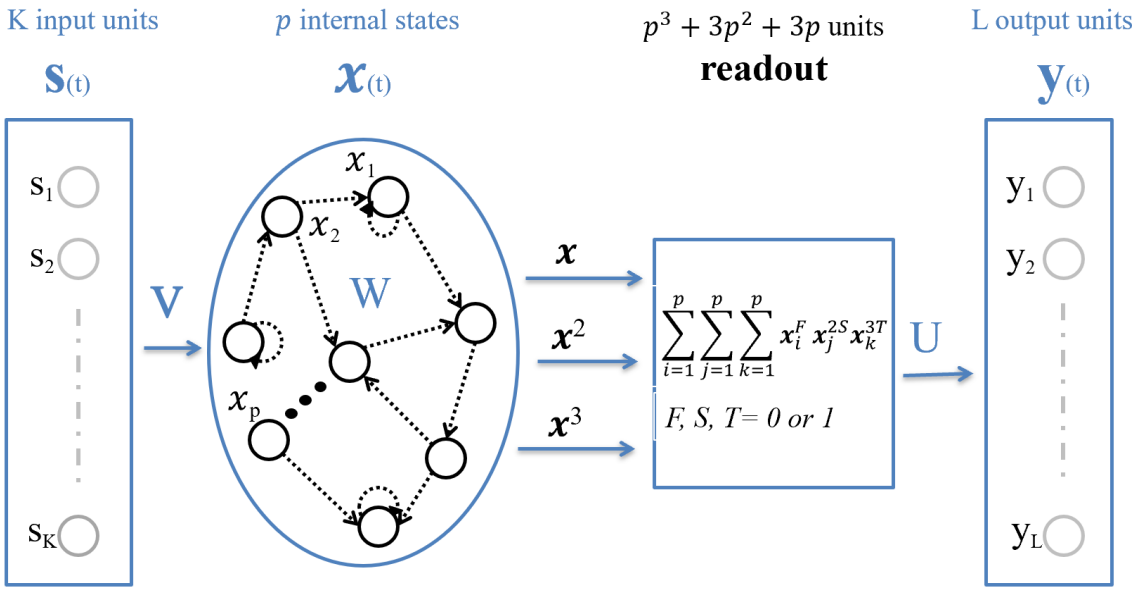


Figure 4. The topological structure of Robust Nonlinear Echo State Network. In this schematic,

Fig. 5 shows an example to compare the order of metrices in ESN, NESN-P, NESN-MP, and RNESN. As can be seen, the number of elements is decreased from 10^8 in ESN to just less than 500 in RNESN. Furthermore, the matrix T is removed in RNESN which improves the computational efficiency considerably.

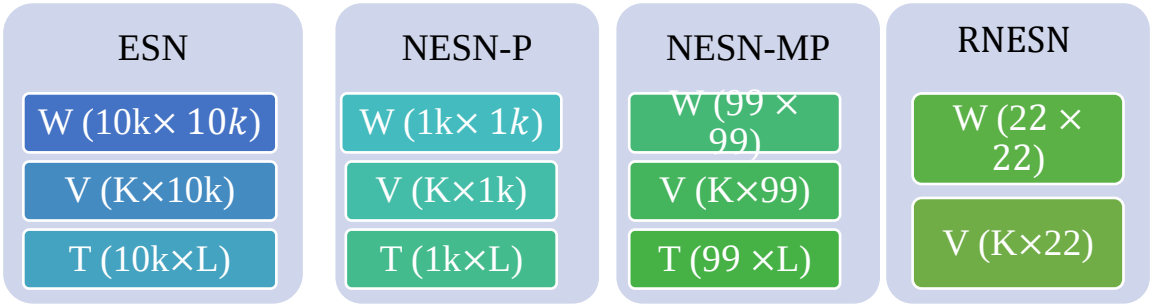


Figure 5. An example to compare the order of metrices in ESN, NESN-P, NESN-MP, and RNESN

2.3 Simulation Results

2.3.1 Data Preprocessing Simulation Results

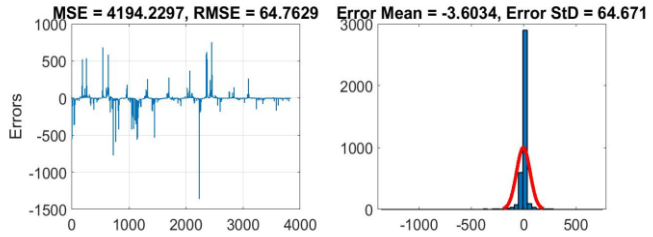
For evaluating the classification performance of the preprocessed data sets we used four different machine learning methods as shown in the table .

Table 6. Classification accuracy (in %)

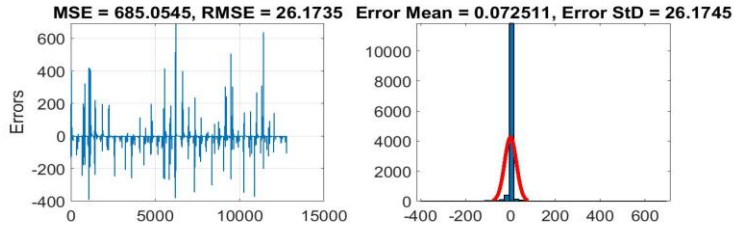
	KNN	Bootstrap aggregating	RandF	SVM
PCA	81.43	83.21	84.75	80.49
TF	88.12	87.06	**96.59***	86.16
TF-K means clustering	90.53	89.53	97.98	88.52

The table shows the overall classification accuracy for all performed dimensionality reduction and data cleaning algorithms. A very interesting observation from the results is that the TF-K means clustering subsets clearly outperform the other created subsets (Figure 6).

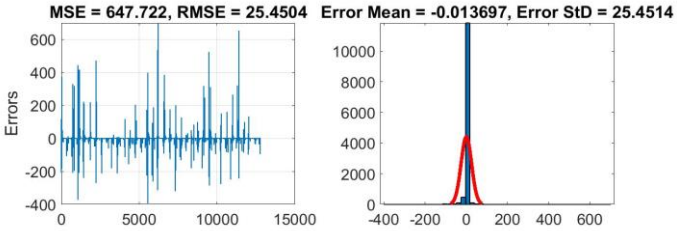
a)PCA-KNN



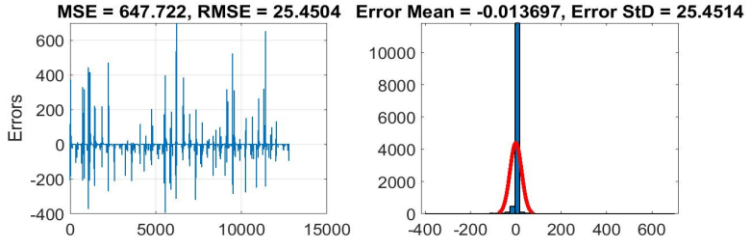
b)PCA-Bootstrap aggregating



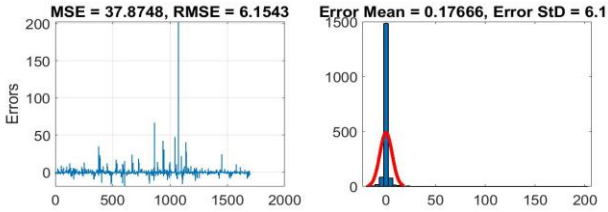
c)PCA-RandF



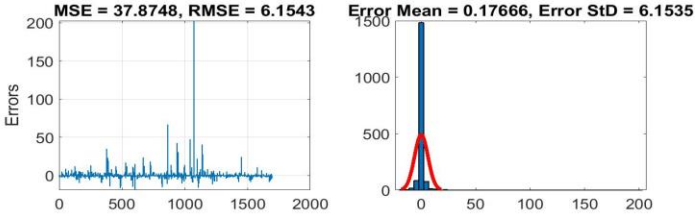
d)PCA-SVM



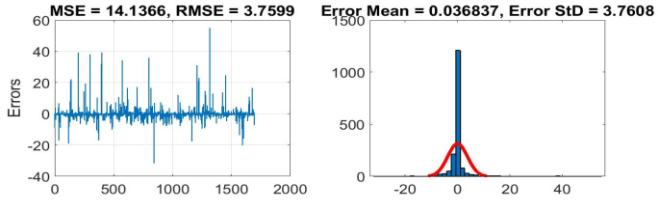
e)TF-KNN



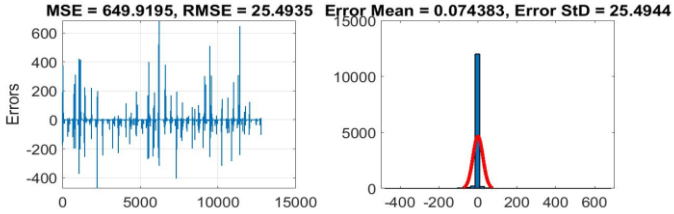
f)TF-Bootstrap aggregating



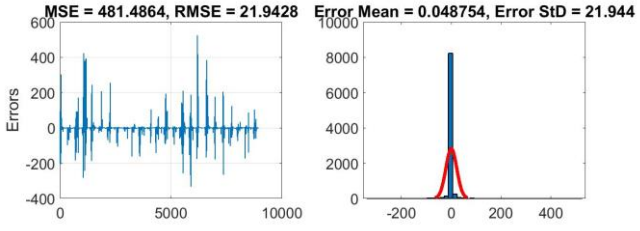
g)TF-RandF



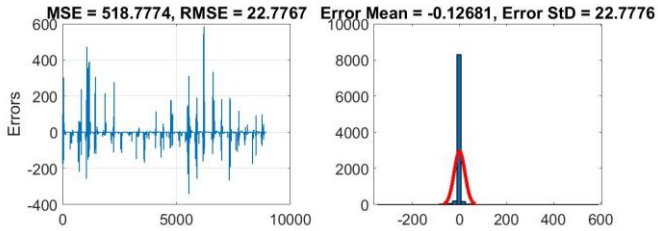
h)TF-SVM



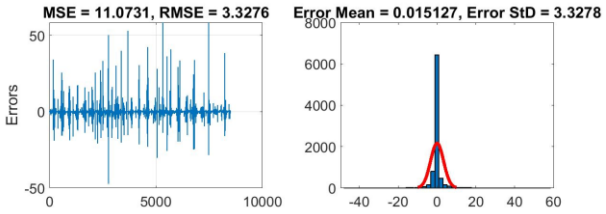
i)TF-Kmeans-KNN



j)TF-Kmeans-Bootstrap aggregating



k)TF-Kmeans-RandF



l)TF-Kmeans-SVM

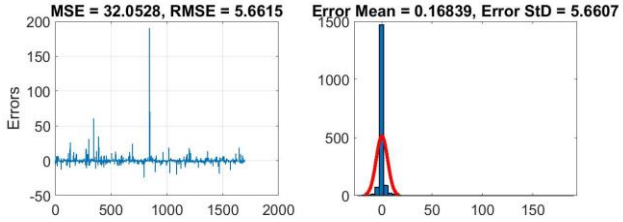


Figure 6. a to l. The graphs of various performed data mining algorithm in the three different subsets of PCA, TF, TF-K means clustering

We have investigated the relationship between various data mining algorithms (feature importance, data cleaning and missing data recovery as well as dimensionality reduction) and the resulting classification performance.

TF-k-means clustering clearly outperforms the other approaches, and also shows acceptable classification accuracy in combination with the Random Forest classification algorithm.

Among the machine learning algorithms investigated, the SVM accuracy was surprisingly low with regards to the PCA subsets. Even though SVMs performed very well on TF-k-means clustering subsets of the original features and it has been commonly reported as a accurate feature importance methods, they achieve only the lowest accuracy for the PCA subsets. A possible reason might be because of the sensitivity of feature importance strategy on the type of performed dataset. In many cases, especially in application contexts where the search for the best feature set is still an active research topic, the classification accuracy achieved with cleaned data sets is often significantly better than with the full uncleaned data set.

It has also been illustrated that the percentage of the total inconsistency of the data captured in the investigated data cleaning approaches is not necessarily associated with the subsequent classification accuracy.

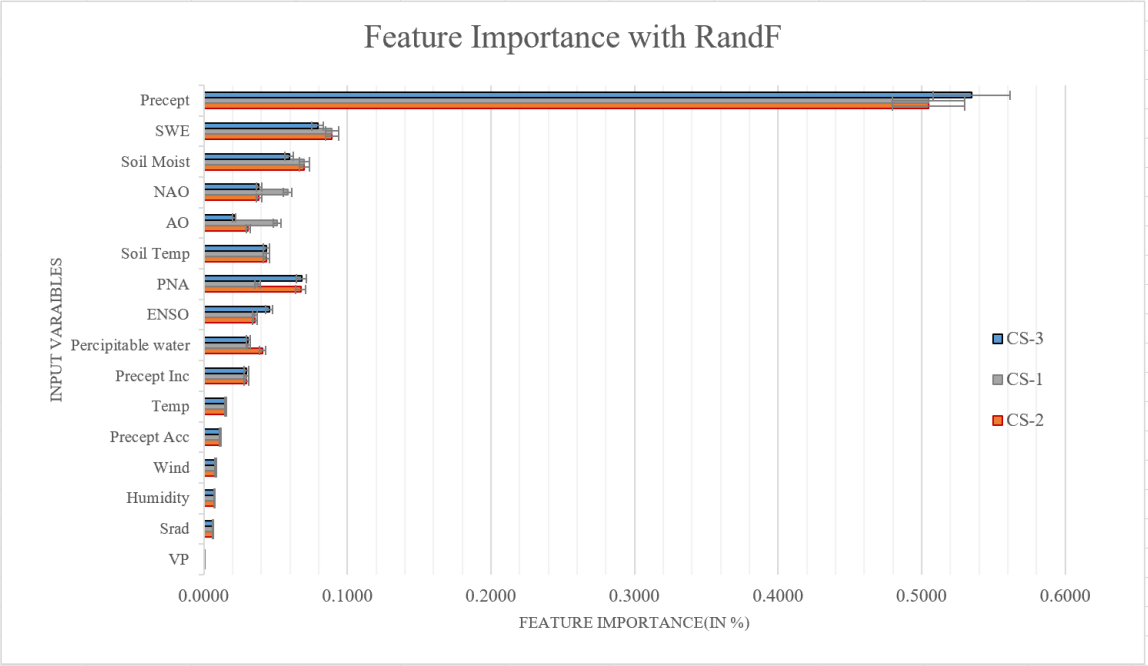


Figure 7. Plot of the feature importance for a study area performed by Random Forest (Error bar in %)

As the figure illustrates, the most effective factor in streamflow governing is precipitation followed by SWE, soil moisture , NAO, AO, soil temperature and so on. The feature importance has been implemented in three case studies and it is clear there are slight differences between contributing factors. The only difference is the climate variability indices. In case study 1 the most important climate variability indices are NAO, AO and PNA respectively, because this region is located in the northeastern US and is affected by AO and NAO much more than the other case studies. There is not much difference between case study 2 and 3, as they both are under effect of PNA and not much influenced by AO and NAO.

Based on the outputs of feature importance algorithm, the input factors including: vapor pressure, solar radiation, wind amplitude and humidity has not been utilized for the forecasting phase. Because their contributing percentage is pretty low it can be ignored as it doses not have any significant impact on forecasting accuracy.

2.3.2 RNESN Simulation Results

The performance of the proposed RNESN was tested using the data described in Section II. Data were split into two disjoint parts, one for training and the other for testing, with their lengths indicated as L_{train} and L_{test}, respectively. The correlation coefficient (R), root mean squared error (RMSE), mean absolute error (MAE), Willmott's index of agreement (WI), and Nash-Sutcliffe coefficient (E_{NS}) shown in (8-12) were used to evaluate the performance of the proposed methods.

The correlation coefficient (R) is characterized as the covariance of the variables partitioned by the product of their standard deviations. R is expressed as:

$$\mathrm{R} = \frac{\sum_{i=1}^{\mathrm{n}}(y(i) - \bar{y}(i))(\hat{y}(i) - \hat{\bar{y}}(i))}{\sqrt{\sum_{i=1}^{\mathrm{n}}[y(i) - \bar{y}(i)]^2}\sqrt{\sum_{i=1}^{\mathrm{n}}[\hat{y}(i) - \tilde{y}(i)]^2}} \tag{8}$$

where n is the quantity of samples, $\boldsymbol{y}$ are the actual values of output, $\overline{\boldsymbol{y}}$ is the average of $\boldsymbol{y}$ over the entire target set, $\tilde{\boldsymbol{y}}$ is the average of $\hat{\boldsymbol{y}}$ over the entire target set, and $\hat{\boldsymbol{y}}$ is the simulated output values.

RMSE is presented as:

$$\mathrm{RMSE} = \frac{1}{n_{max}}\sqrt{\sum_{i=1}^{\mathrm{n}}[y(i) - \hat{y}(i)]^2} \tag{9}$$

MAE is defined as

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y(i) - \hat{y}(i)| \tag{10}$$

WI (Willmott, 1981) is calculated as:

$$\mathrm{WI} = 1 - \frac{\sum_{i=1}^{\mathrm{n}} \big(y(i) - \hat{y}(i)\big)^2}{\sum_{i=1}^{\mathrm{n}} (|y(i) - \bar{y}(i)| + |\hat{y}(i) - \bar{y}(i)|)^2} \tag{11}$$

The Nash–Sutcliffe (Moriasi et al., 2007) is defined as:

$$E_{NS} = 1 - \left(\frac{\sum_{i=1}^{\mathrm{n}} \big(y(i) - \hat{y}(i)\big)^2}{\sum_{i=1}^{\mathrm{n}} \big(y(i) - \bar{y}(i)\big)^2} \right) \tag{12}$$

In this work, the performance of the proposed method was evaluated in the three case studies described in Section 2. The forecasting results of RNSEN were compared to those of ANFIS and NESN-MP. Furthermore, bad data analysis was conducted to show the robustness and stability of the proposed method. Streamflow forecasting is tested for 50 days ahead ($L_{train} = 730$, $L_{test} = 50$ days).

To validate the performance of the proposed method, two more scenarios are defined. In the first scenario, the performance of the proposed method in the presence of bad data is conducted. The input data in the fourth month of the training is increased by 100%. They decreased by 10% by the seventh month and increased by 25% the 14th month. It is shown that although severe changes in the input data are considered, the forecasting accuracy is only insignificantly lowered.

In the second scenario, the input data for only three months are used for the training process. The goal in this scenario is to achieve acceptable forecasting accuracy with far fewer data. This in turn decreases the cost and the computational load considerably. This is one area where the proposed methodology offers clear advantages. The simulation results clearly validate the high learning capability and forecasting accuracy of RNESN in a scarce data operation condition. Note that NESN-MP outperforms ANFIS due to its unique nonlinear structure. However, RNESN outperforms NESN-MP because it has a more complex nonlinear structure. Figs. 6-8 show the simulation results for the three different regions, Berkshire County (MA), Tuolumne County (CA), and Wasco County (OR), mentioned above.

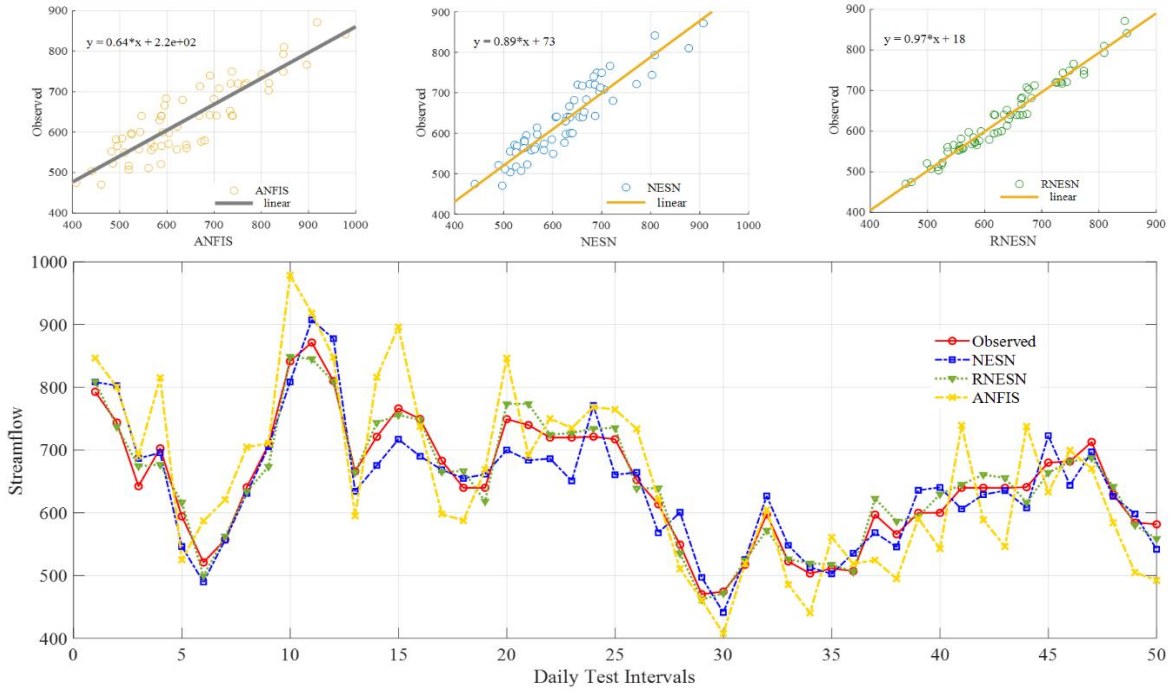


Figure 8. Scatter plot of observed and simulated streamflow (m3/s) using RNESN with 13 different input variables in case study 1 for testing period (2008-1011), CS= Case Study

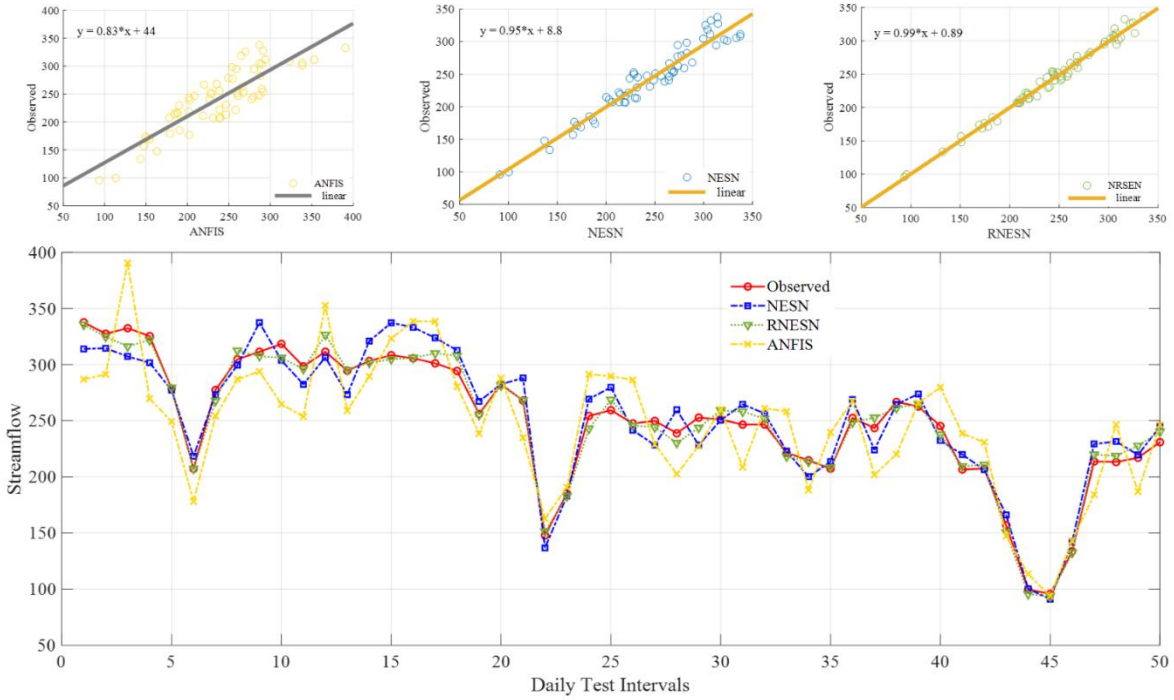


Figure 9. Scatter plot of observed and simulated streamflow (m3/s) using RNESN with 13 different input variables in case study 2 for testing period (2008-1011), CS= Case Study

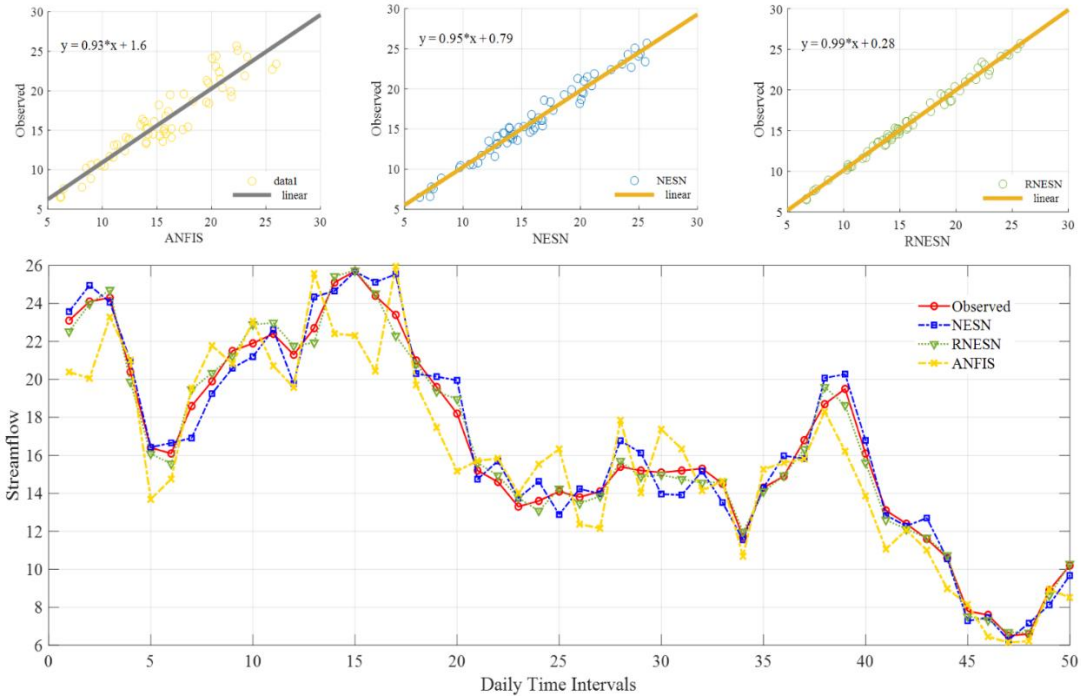


Figure 10. Scatter plot of observed and simulated streamflow (m3/s) using RNESN with 13 different input variables in case study 3 for testing period (2008-1011), CS= Case Study

As clearly observed from the time variation graphs, the forecasts of the proposed RNESN model are closer to the corresponding observed streamflows compared with the other two models in all three case studies. Considerable under- or over-estimations are seen for the ANFIS model. It is evident from the scatterplots that the RNESN has less scattered streamflow forecasts and its fit line is closer to the exact line (y = x).

Table 7.Forecasting results using evaluation indices for the three case studies: R, RMSE, MAE, WI, and E_{SN}

		Streamflow Forecast					Bad Data					Limited Data				
		R	WI	E_{SN}	MAE (cfs)	RMSE (cfs)	R	WI	E_{SN}	MAE (cfs)	RMSE (cfs)	R	WI	E_{SN}	MAE (cfs)	RMSE (cfs)
Case	RNESN	0.98	0.99	0.96	14.8	2.29	0.95	0.95	0.92	15.3	2.45	0.96	0.98	0.92	14.95	2.34

	NESN	0.93	0.96	0.84	31.1	4.65	0.86	0.86	0.81	31.5	5.3	0.88	0.93	0.78	31.9	5.0
	ANFIS	0.78	0.87	0.47	57.4	8.43	0.72	0.64	0.45	72.6	8.57	0.58	0.70	0.39	63.85	8.73
Case Study 2	RNESN	0.99	0.99	0.98	5.38	0.86	0.92	0.94	0.92	5.42	0.92	0.94	0.96	0.95	5.44	0.89
	NESN	0.97	0.98	0.93	12.33	1.88	0.96	0.96	0.83	12.4	2.22	0.89	0.97	0.92	13.25	2.01
	ANFIS	0.87	0.93	0.72	25.98	3.77	0.74	0.75	0.68	27.1	5.15	0.68	0.82	0.69	33.6	4.65
Case Study 3	RNESN	0.99	0.997	0.99	0.39	0.06	0.95	0.96	0.92	0.4	0.06	0.97	0.96	0.96	0.39	0.06
	NESN	0.98	0.992	0.96	0.69	0.11	0.93	0.87	0.82	0.73	0.11	0.93	0.91	0.93	0.7	0.12
	ANFIS	0.94	0.964	0.86	1.52	0.23	0.69	0.79	0.85	1.8	0.26	0.81	0.83	0.68	1.71	0.23

Table 7. provides the evaluation indices for the test stage covering 50 days. In case study 1, RNESN provides an MAE of 14.8 m^3/s which is 52.4% and 74.2% lower than the MAE given by NESN-MP and ANFIS, respectively. In the case of E_{NS}, the RNESN gives the respective value of 0.96, which is well below the E_{NS} of 0.84 and 0.47 for NESN-MP and ANFIS, respectively. This improvement can be seen more clearly in the presence of bad data and in sensitivity analysis.

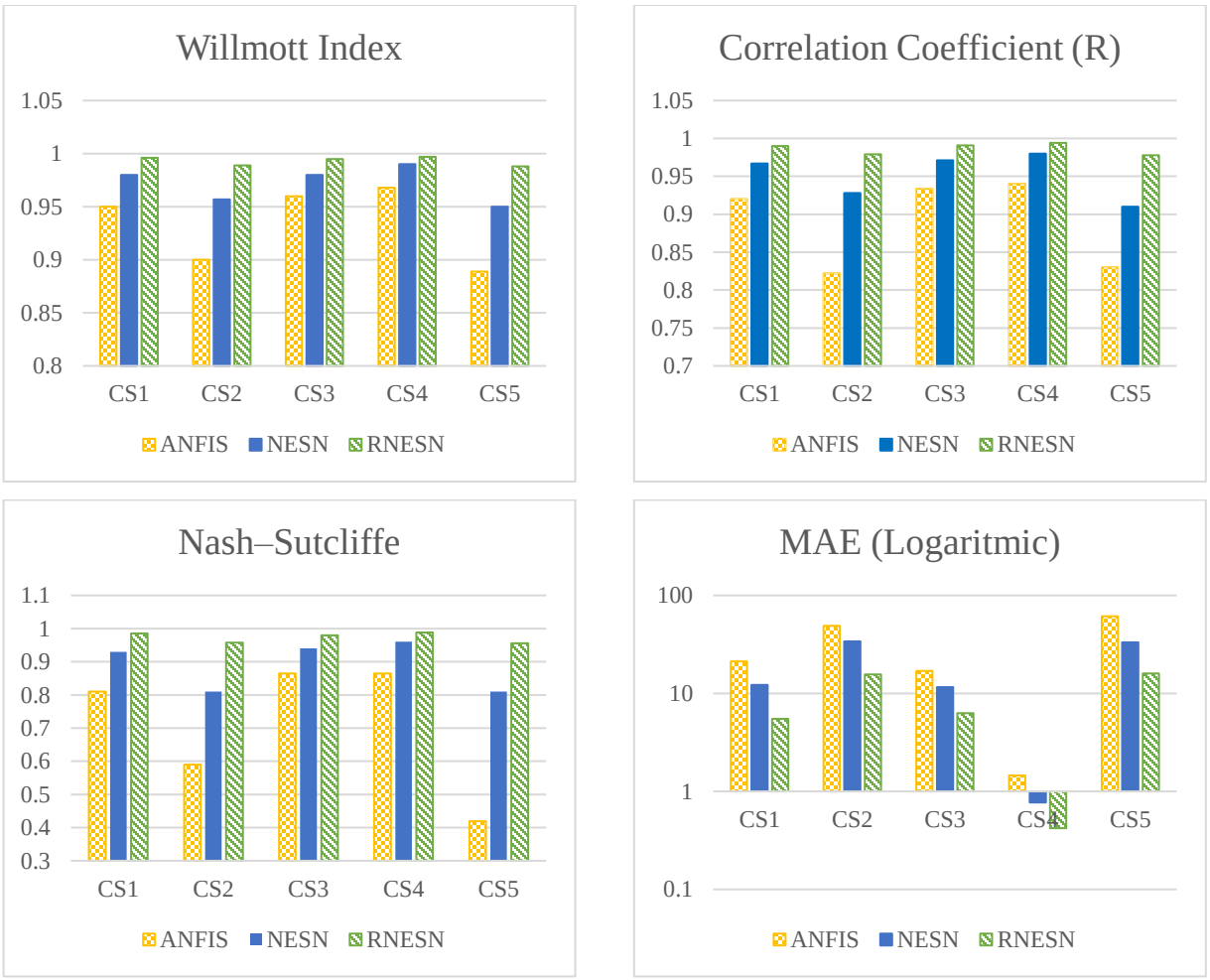


Figure 11. Bar chart of Evaluation indices for three case studies using three models: ANFIS, NESN, RNESN

Note that the forecasting accuracy with bad data is not easy attainable due to inaccurate measurements or misplaced gauging stations. RNESN impressively outperforms ANFIS and NESN-MP in this respect. For example, WI in RNESN varies from 0.99 to 0.95 while WI in NESN-MP and ANFIS drops considerably from 0.96 and 0.87 to 0.86 and 0.64 respectively. As the WI is more sensitive to oscillations in the data, it shows the most change in the presence of the bad data. The same situation is expected in the sensitivity analysis. In the case of limited data, due to the very high learning capability in both NESN and RNESN, they outperform ANFIS. As expected, RNESN provides better forecasting results because it has a more complex structure and considers more complicated interactions between the internal states.

For validating prediction capability and universality of the recommended method with various streamflow profiles, five case studies based on the three study areas are conducted in different seasons shown in Fig. 9. The simulation results for RNESN outperform those for the NESN and ANFIS. The forecasting results may deteriorate if the watershed is snow dominate as in Case study 1. The interannual variability of the accumulated snow in the watershed causes high flow events. Therefore, possible snowmelts due to rapid temperature changes and "rain on snow events" increase the data uncertainty, which affects the training process. In Case Study 2, as the watershed is a mixed pluvial-nival, both the rainfall and snowmelt are driving variables in the generation of streamflow. Therefore, the streamflow is present for almost the whole year while the streamflow value does not change drastically compared to snow or rain regions. This in turns provides better forecasting results compared to Case Study 1. In Case Study 3, which is a rain dominate watershed, the RNESN shows the most accurate agreement with the observed streamflow, because the watershed is not influenced by "winter snow cover increase" to affect streamflow. Note that this case study is designed for warm seasons in which more evaporation results in low streamflow. In such a watershed, there is less accumulated snow in the mountains and that yields lower streamflow. As shown in the simulation results, RNESN performs better in low streamflow.

Note that, to validate the model application at the global scale, the modeling process has been repeated on five different regions across US and the simulation results are presented in appendix.

2.4 Conclusion

Hydrological models frequently comprise uncertainties with negative effects on the estimated results, and thereby on the model reliability and robustness. The majority of published models do not recognize the input data uncertainties and utlize the raw time series inputs for modeling approach. In the current study the proposed data mining algorithm for data preprocessing which has been performed through ensemble TF-k means clustering and RF is capable of eliminating input uncertainties by missing data recovery and data cleaning . The utilized deep learning algorithm (RNESN) for streamflow

forecasting is also a high nonlinear approach in prediction of time series data as it considers all possible interactions between the internal states of inputs through performing different orders of multivariate polynomial. Besides, its training approach minimizes the direct impact of each input data on the output, otherwise stated, due to the structure of the RNESN, the sensitivity of each input parameter is almost the same as the others because the interactions of the input parameters and the internal states are used to train the output matrix and the input parameters are not directly involved in the process of training. More importantly, the robustness of the model can reduce bias estimates of water availability and uncertainty in forecasting of potential future climate changes. The proposed calibration process (RLS) enhance the learning capability through reducing variance. Besides, to improve the forecasting ability and reducing the computational load, the number of the internal states has been significantly decreased through applying modified time series analysis in readout. Moreover, the recommended method has a simple structure, far less computation, and does not need parameter tuning, optimization task, and complex training. Simulation results show the dominance of RNESN over NESN and ANFIS. RNESN provides expressively better values that those provided by NESN and ANFIS for R, MAE, Wilmott index, RMSE, and Nash–Sutcliffe efficiency.

The performance of the proposed method is evaluated with bad data or limited data. Due to high learning capability and low sensitivity, the simulation results for RNESN illustrate the significant improvement in forecasting results in the presence of bad data or training with limited data. This in turn validates the robustness and reliability of the model. Real data is collected at United States Geological Survey (USGS), Natural Resource Conservation Service (NRCS), National Weather Service Climate Prediction Center (NOAA) and Daymet Data Set from NASA through the Earth Science Data and Information System (ESDIS) to validate these observations. As future work, we aim to propose a tradeoff between installing necessitate gauging station and removing useless gauging stations. Our current hydrological measurement networks have designed flaws. For instance, most rainfall observations are made around cities and steep terrain and topography effects are inherently underrepresented. As another example, stream gauges in California’s Sierra Nevada mountains are located downstream of dams, making them

virtually useless for understanding catchment processes. The proposed method can detect and remove costly or inaccurate measurements gauges. Streamflow forecasting can be conducted just by the accurate data acquired by the easy maintenance and inexpensive gauges.

2.5 Daily Streamflow Forecasting Using Nonlinear Echo State Network

ABSTRACT: The prediction of streamflow is an important issue in hydrologic engineering and hydropower reservoir management. Several approaches including statistical, physical or conceptual models have been investigated to forecast streamflow. Most of the methods assume a linear relationship between the input and output series. However, they ignore the nonlinear information hidden in the streamflow series. In this paper, various time series inputs including: day length, precipitation, solar radiation, maximum and minimum temperature per day, and vapor pressure have been used. An advanced and powerful forecast engine called Nonlinear Echo State Network using Multivariable Polynomial (NESN-MP) is used to predict the behaviour of the streamflow. The forecasting is conducted under different climatic conditions to indicate the model's applicability. Furthermore, to demonstrate the efficiency of the proposed method, it is compared with Adaptive Neuro-Fuzzy Inference System (ANFIS). The results of the new method compare favourably with ANFIS.

2.5.1 Introduction

State estimation and forecasting of streamflow have always been general concerns for engineers. State estimation is applied in all energy management systems to identify the present operating state of a system [1-2]. Forecasting is also an important and necessary aid to planning and planning is the backbone of effective operations. In hydrology, streamflow forecasting is vital for water resources engineers, reservoir operators and water managers who strive to balance a range of competing objectives to support their decisions about hydroelectric power programming, flood mitigation, agricultural and domestic water supplies, irrigation management as well as maintenance of environmental flows [3].

Therefore, developing an optimal streamflow forecasting model as a stochastic property of environmental modelling is crucial. However, the existing dynamicity, inherent complexities and chaotic feature in the temporal and spatial expansion of the model may obstruct the accurate and reliable prediction process [4].

Different statistical, physical or conceptual models have been evolved to forecast streamflow [5]. Statistical models, such as regression-based models [6] are extremely simplistic and suffer from a functional form between variables prior to the analysis. Therefore, they do not properly account for the relationship between the dependent and observed explanatory variables. Physically based numerical models, typically, simulate the streamflow generation process through a governing equation employing limited boundary conditions, which need precise data input to enable parameter calibration [7]. Conceptual hydrological models consider different processes of the hydrological cycle along with mathematical formulation to improve the forecasting accuracy [8] such as: the Soil and Water Assessment Tool (SWAT) as a semi-distributed conceptual model [9]. Louise J. et al 2017 [10] conducted research to forecast streamflow in deterministic and probabilistic terms for all initialization months, flow quintiles, and seasons. The result showed a relatively accurate streamflow forecasts from low to high flows, but their model could not decrease uniformly with initialization time.

However, taken as a group, the accuracy of these models is not reliable due to heterogeneous hydrogeological characteristics within the watershed system in nature with respect to time and space. In addition, large data input, large number of parameters, and broad range of necessary values may limit the application of comprehensive simulation

models [11]. Furthermore, all of these models assume that the relationship between the input and output series is linear or at worst near linear. They thus ignore the nonlinear information hidden in the streamflow series which result in a poor model performance. Furthermore, streamflow is under the influence of many factors such as evapotranspiration, rainfall, atmospheric circulation and temperature, and its generation process is nonlinear and time-variable. Therefore, in the last two decades, researchers have focused on alternative data-base predictive methods. Several studies have been presented on developing soft computing tools with Artificial Intelligence (AI) models. Several computer models have been recently conducted to forecast streamflow and corresponding runoff. Some of these models are based on the Artificial Neural Networks (ANN), fuzzy network sets, genetic programming, regression algorithms, support vector machine; and nonparametric methods such as K-Nearest Neighbour (KNN) Regression [12-13]. Among all mentioned methods, the fuzzy network sets depend on the user expertise, while the efficiency of others depends on the model ability to find out the relationship between input and output variables. It has been demonstrated that Support Vector Machine (SVM) yields the most accurate results compared to Auto Regressive Moving Average (ARMA), ANN, and Multiple Linear Regression (MLR) [14], and (something missing?)[12]. On the other hand, Shrestha (2014) indicated that the result of annual predicted streamflow using ANN and SVM, throughout the State of Utah, are the same [15]. Yong Liu et. al (2016) compared the RVM and SVM for long term streamflow forecasting. They found that RVM produces better results for annual streamflow forecasting within a specified climatic condition [16]. Bharti et al. (2017) indicated that in forecasting process of monthly runoff, ANN results

surpasses the Least Square- Support Vector Regression (LS-SVR) results, while LS-SVR results exceed ANN results for monthly sediment prediction [17].

ANNs are the most popular artificial intelligence (AI) techniques used in variety of fields especially in time series forecasting. Successful prediction results of ANN application in hydrological process such as rainfall-runoff modelling, streamflow prediction, reservoir inflow forecasting, rainfall forecasting, and river sediment modelling have been recently published. Although different feedforward neural network models have been well documented, the selection basis of these models has thus far received limited attention [18-19]. Kerh and Lee (2006) introduced ANNs to predict flood discharge at downstream stations with data scarcity, using information at upstream stations of the Kaoping River [20]. Their model demonstrated that back-propagation of the ANN model performs better than the conventional Muskingum method. Due to chaotic behaviour in hydrological time series, one of the most important steps in constructing an ANN model for streamflow forecasting is determining the best inputs. Zhao, X. at. al (2017) used the Phase Space Reconstruction (PSR) method as an alternative approach to select relevant and important input variables for ANN models. They built two different ANN models using the time-lagged records of precipitation and temperature. They indicated that ANNs predict daily streamflow in the adjacent ungauged basins as accurate as in the gauged basin [21]. Zealand et al (1999) used the ANN trained with back-propagation algorithm to predict streamflow 1-week-ahead [22]. However, ANN models have some lapses including over-fitting and under-fitting, slow learning speed, and curse of dimensionality and convergence to local minimum. Therefore, in processing of complex hydrological phenomena, they

betray a poor performance [23-25]. Typically, their disadvantages include the following [26]:

1. *High complexity and long processing time.*
2. *High dependence on parameter tuning and optimization.*
3. *The requirement for nonconvex optimization that can yield suboptimal results and trap in local optima.*

In this paper, NESN-MP has been used as a forecasting engine. The network consists of a reservoir including linear internal states and a readout including nonlinear functions of the internal state. The nonlinear relations between the internal states increase the learning capability, which results in high forecasting accuracy while ensuring that the quality of forecasting does not deteriorate significantly with time. Furthermore, the performance of the forecasting engine is improved by decreasing the number of internal states, and the orders of the weight matrices, which reduces the computational load considerably. Furthermore, in all previous research, the results obtained from these studies are inconsistent due to difference in study areas, input data sets, and the selected structures for each of the models [27]. Many studies have applied the original streamflow time series as the input variables in their forecasting model, which results in missing some features of different resolution [28]. Using just one resolution component could not reflect the internal mechanism of streamflow. Therefore, daily data is preferred, because it is not significantly affected by external factors such as meteorological pattern and anthropogenic activities in the data [29]. However, research has been conducted on evaluation of annual, seasonal or

monthly streamflow for one-time scale condition [25]. Moreover, daily streamflow forecasting at different time scales has not been addressed in the literature.

Based on the outline above, the study in this paper has developed a modified model of ANN to forecast daily streamflow based on various time series forcing-data input including daily precipitation, precipitation duration, solar radiation, temperature and vapor pressure. The remainder of this study is as follows. Section II provides an overview of the NESN. Simulation results are given in Section III, and conclusions are summarized in Section IV.

2.5.2 Nonlinear Echo State Network

This powerful method is simple, effective, with far fewer computations [30]. NESN-MP provides a total of $2p + p^2$ units; p internal states; p squares of the internal states; and p^2 units gained by multiplying the internal states and squares of the internal states. Therefore, the order of weight matrices is decreased radically. The weight matrices (W, T, and V) are then used to calculate the internal states of the reservoir. The vector of internal states is updated using

$$\boldsymbol{x}_{(t+1)} = \mathrm{f}(W.\boldsymbol{x}_{(t)} + V.\boldsymbol{s}_{(t+1)} + T.\boldsymbol{y}_{(t)}) \tag{1}$$

and the readout vector is

$$\bar{\boldsymbol{x}}_{(t+1)} = [\boldsymbol{x}_{(t+1)}, \boldsymbol{x}^2{}_{(t+1)}, \sum_{i_1=1}^{p}\sum_{i_2=1}^{p} \boldsymbol{x}_{i_1(t+1)}.\boldsymbol{x}^2{}_{i_2(t+1)}] \tag{2}$$

where $\boldsymbol{x}^2{}_{(t+1)} = \left[x^2_{1(t+1)}, x^2_{2(t+1)}, \dots, x^2_{p(t+1)}\right]$, p is the number of internal states $\left\lceil \frac{N}{p+2} \right\rceil$, $\boldsymbol{s} \in R^{K\times 1}$ is the input vector, $\boldsymbol{x} \in R^{p\times 1}$ is the internal state vector, $\bar{\boldsymbol{x}} \in R^{(p^2+2p)\times 1}$ is the readout vector, and $\boldsymbol{y} \in R^{L\times 1}$ denotes the output states.

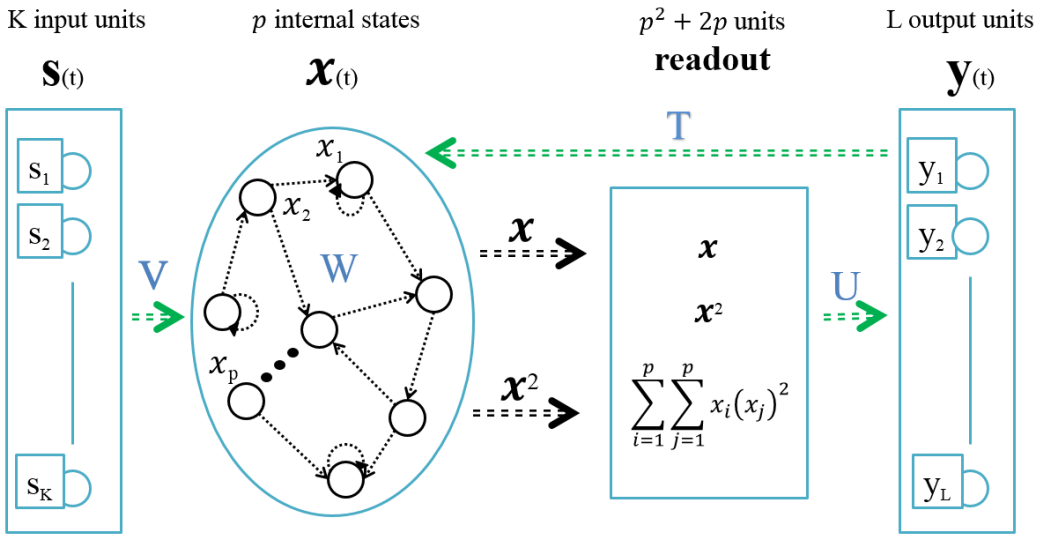


Figure 12. Schematic of NESN-MP.

The matrix $W \in R^{p\times p}$ defines the internal state interconnections within the reservoir. The values in W are fixed values generated randomly over a symmetric interval.

$$W = (w_{ij})_{p\times p} \;;\; w_{ij} \in (-1,1)(i,j = 1,2,\dots,p) \tag{3}$$

Matrix $V \in R^{p\times K}$, containing randomly chosen fixed values, defines the connections of the input with the internal states of the reservoir.

$$V = (v_{ij})_{p\times k} \;;v_{ij} \in (-1,1)(i = 1,2,\dots,p, j = 1,2,\dots,k) \tag{4}$$

The output feedback matrix, $T \in R^{p\times L}$ is

$$T = (t_{ij})_{p\times L} \;;\; t_{ij} \in (-1,1)(i = 1,2,\dots,p, j = 1,2,\dots,L) \tag{5}$$

The output matrix, $U \in R^{L\times(p^2+2p)}$ is

$$U = (u_{ij})_{L\times(p^2+2p)} \;;\; u_{ij} \in (-1,1)(i = 1,2,\dots,L, j = 1,2,\dots,2p+p^2) \tag{6}$$

where K is the number of inputs, p is the number of internal states, and L is the number of outputs.

2.5.3 Simulation Results

The performance of the NESN-MP is tested using climatic observation data (day length, precipitation, solar radiation, maximum and minimum temperature per day, and vapor pressure) with a time interval of 24 hours used to train and test the proposed methods. Each data set is divided into two separate parts for training and testing, with their lengths denoted as L_{train} and L_{test}, respectively. The MSE, root mean squared error (RMSE), normalized root-mean-square error (NRMSE), normalized mean-absolute error (NMAE), and mean absolute error (MAE) shown in (7-11) were used to evaluate the performance of the proposed methods.

NRMSE is often expressed as a percentage and calculated as

$$\text{NRMSE} = \sqrt{\frac{\sum_{i=1}^{n_{max}} \left|\left|y(i) - \hat{y}(i)\right|\right|^2}{\sum_{i=1}^{n_{max}} \left|\left|y(i) - \tilde{y}\right|\right|^2}} \times 100\% \tag{7}$$

where $||\bullet||$ indicates the Euclidean norm, $\boldsymbol{y}$ are the actual output values, $\tilde{y}$ is the average of $\boldsymbol{y}$ over the whole target set $y\,(1), y\,(2), \ldots, y\,(n_{max})$, $\hat{\boldsymbol{y}}$ is the predicted output, and $n_{\max}$ is the number of sample points. Lower values in NRMSE indicate less residual variance. In many cases, especially for smaller samples, the sample range is likely to be affected by the size of sample, which would hamper comparisons.

MSE measures the average of the squares of the errors, which is always non-negative, and values closer to zero are better. Taking the square root of MSE yields RMSE, which has the same units as the estimated quantity. MSE and RMSE are calculated as

$$\text{MSE} = \frac{1}{n_{max}} \sum_{i=1}^{n_{max}} [y(i) - \hat{y}(i)]^2 \tag{8}$$

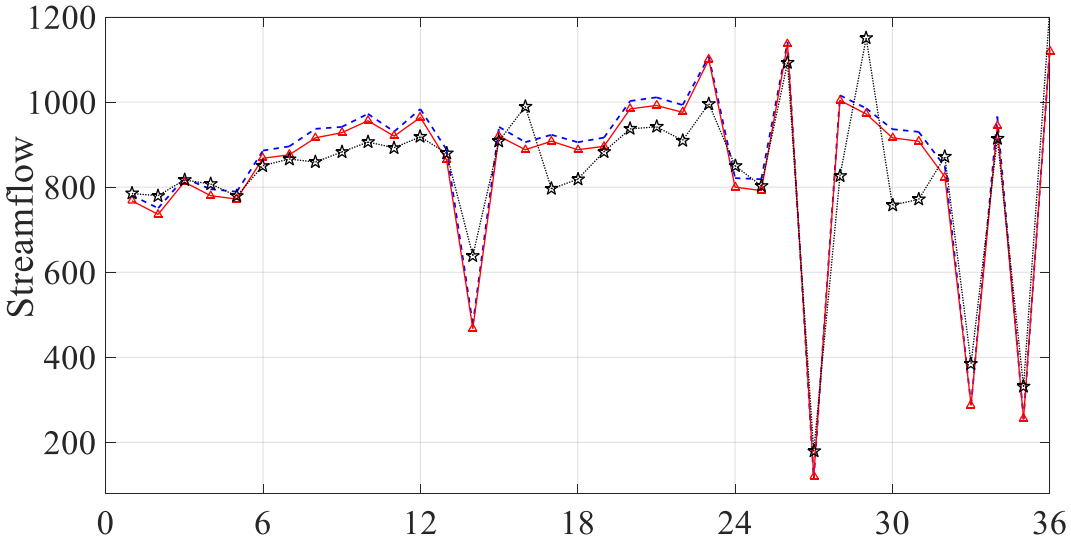
$$\text{RMSE} = \frac{1}{n_{max}} \sqrt{\sum_{i=1}^{n_{max}} [y(i) - \hat{y}(i)]^2} \tag{9}$$

MAE calculates the average magnitude of the errors in a set of predictions without considering their direction. It is the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight. NMAE normalizes MAE by the range of available rating values. MAE and NMAE are defined as

$$\text{MAE} = \frac{1}{n_{max}} \sum_{i=1}^{n_{max}} |y(i) - \hat{y}(i)| \tag{10}$$

$$\text{NMAE} = \frac{1}{y_{max} . n_{max}} \sum_{i=1}^{n_{max}} |y(i) - \hat{y}(i)| \tag{11}$$

where $y_{\max}$ is the maximum value of output. Generally, RMSE and MAE are regularly employed in model evaluation studies [26].



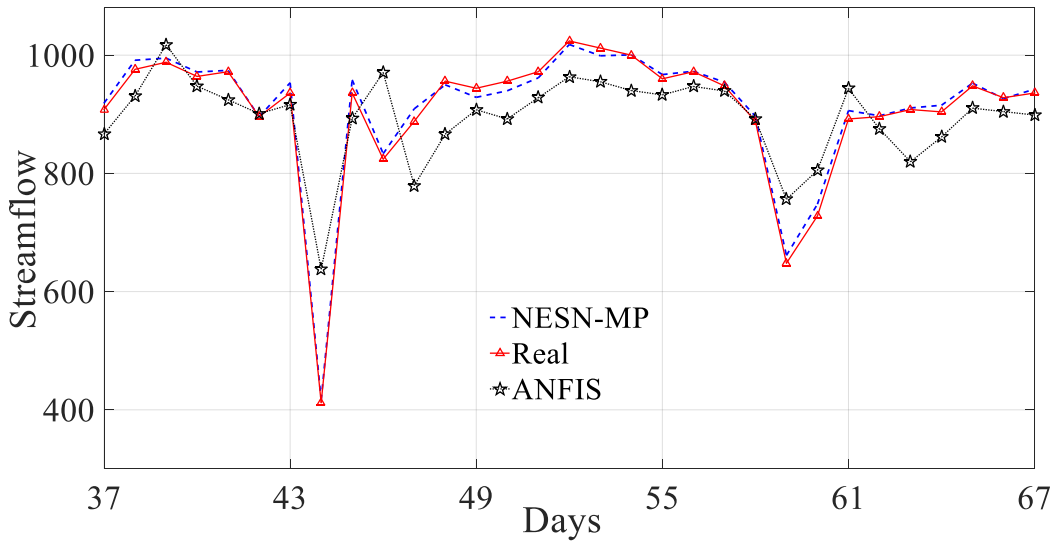


Figure 13. 67 days prediction in case study in 1995.

The streamflow forecasting is carried out for 67 days ahead. $l_{train} = 200$, $l_{test} = 67$ with no overlap and with the test data starting immediately after the training data. Fig. 2 shows the prediction for 67 days ahead for NESN-MP and ANFIS. NESN-MP provide an MAE of 4 for the first 10 days ahead which is significantly below the MAE given by ANFIS. This improvement can be seen on the second 10 days more obviously, where MSE and MAE in ANFIS results increased considerably while those in the proposed methods remain almost constant. The results clearly show that the proposed NESN-MP outperform ANFIS.

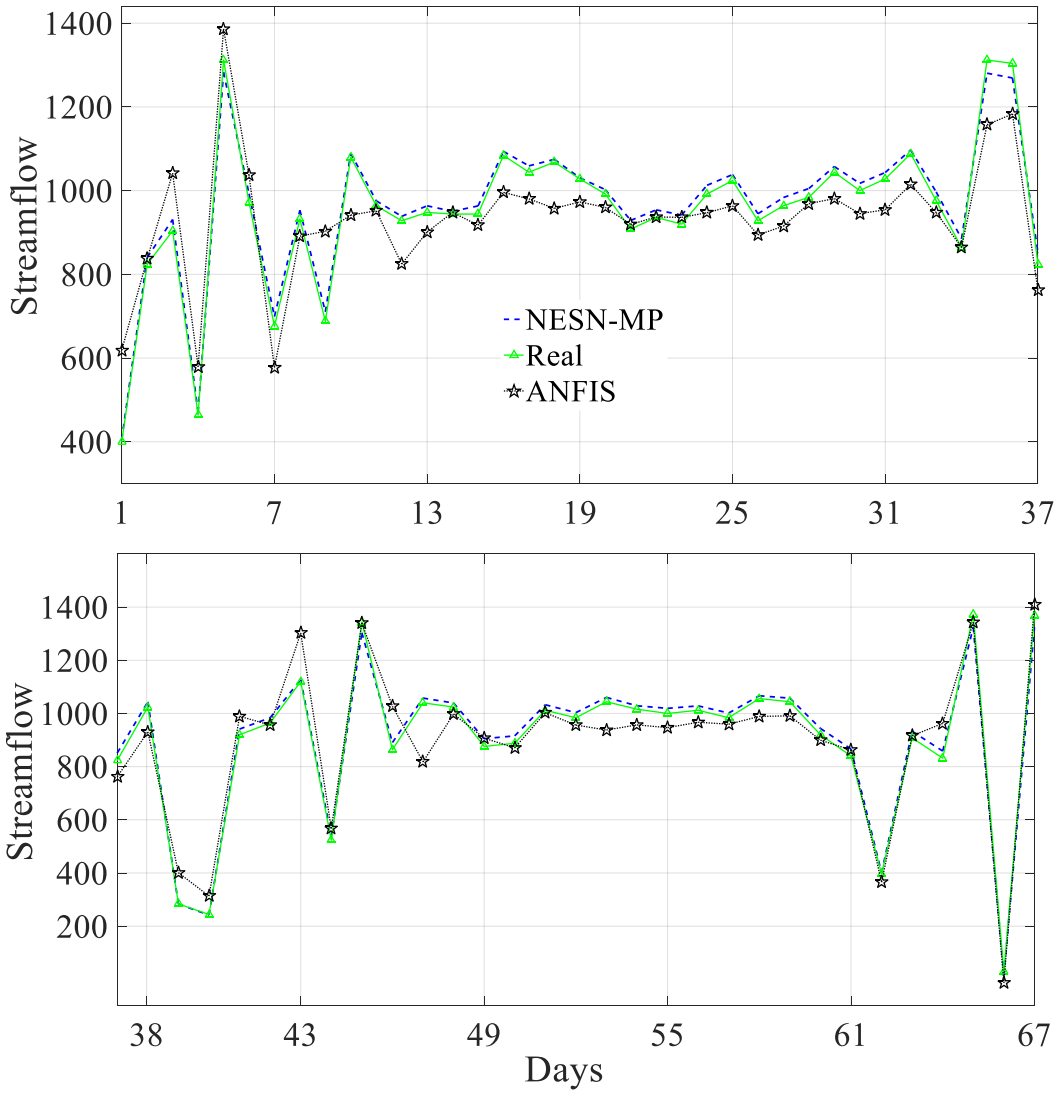


Figure 14. 67 days prediction in case study in 2005.

To validate the prediction ability and universality of the proposed methods with different climatic parameters, 67 days streamflow forecasting in 2005 is shown in Fig. 3. It is shown that the NESN-MP provide an MAE of 9 for the first 10 days ahead forecasting which is 92.8% below the MAE for ANFIS, respectively. In case of RMSE, NESN-MP gives the respective value of 11.21 which are well below the RMSE of 101.2 for ANFIS. Table 1 shows the error indices for both methods for different days.

Table 8. Error indices for case study 2005.

		Days 1-10	*Days 11-20*	*Days 21-30*	*Days 31-40*	*Days 41-50*	*Days 51-60*
ANFIS	**MSE**	10235	3437	1665	7600	2347	763
	RMSE	101.2	58.62	40.8	87.18	48.45	27.63
	NMAE%	16.22	10.98	7.84	15.93	8.8	5.55
	NRMSE %	9.2	5.329	3.7	7.92	4.4	2.51
	MAE	73	49.4	35.3	71.7	39.6	25
NESN-MP	**MSE**	125	43.2	219	264	94.5	80.8
	RMSE	11.21	6.57	14.8	16.26	9.7	8.98
	NMAE%	2	1.2	2.95	2.7	1.89	1.82
	NRMSE %	1.02	0.6	1.34	1.47	0.88	0.81
	MAE	9	5.4	13.3	12.2	8.5	8.2

2.5.4 Conclusion

This study presents daily streamflow forecast based on various time series forcing-data inputs including daily precipitation, precipitation duration, solar radiation, temperature and vapor pressure. A novel echo state networks called NESN-MP has been used as forecasting engine. The nonlinear relations between the internal states increase the learning capability, which results in high forecasting accuracy while ensuring that the quality of forecasting does not deteriorate significantly with time. Furthermore, the daily values for different parameters which affect the streamflow provide accurate forecasting. Simulation results validate the performance of the proposed method and demonstrate its superiority over ANFIS. NESN-MP provides significantly lower values than those given

by ANFIS for MAE, NMAE, MSE, RMSE, and NRMSE. Future work will compare the proposed method with the classical methods.

2.6 Bad Data Analysis on Streamflow Forecasting Using Nonlinear Echo State Network

ABSTRACT: A correct estimation of the stream flow is crucial to reduce the consequences of flash floods. Hydrologic prediction or simulation, especially in ungauged basins, is essential for responsible and sustainable water resource management. In the current study, we develop a framework on a study area including twelve gauged watersheds spanning across different climatic settings in the US. In this work we will propose a novel approach of Nonlinear Echo State Network using Multivariable Polynomial (NESN-MP) to forecast daily stream flow in ungauged basin with bad data. This work aims to demonstrate the ability of NESN-MP to solve a simulation task in comparison with ANFIS. Publicly available climate and US Geological Survey streamflow records are used to train and test the model. The model inputs include time-lagged records of precipitation, solar radiation, day length, vapor pressure and temperature. Furthermore, recurrent feedback loops allow ANN streamflow estimates to be used as model inputs. The successful of these flow prediction approach indicates that the NESN-MP can predict streamflow with bad data entry as accurately as good data set entry in the basins on which they were trained.

2.6.1 Introduction

State estimation and forecasting have always been general concerns for engineers. State estimation is applied in all energy management systems to identify the present operating state of a system [1-2]. Forecasting is also an important and necessary aid to planning and planning is the backbone of effective operations. In hydrology, streamflow

forecasting is vital for water resources engineers, reservoir operators and water managers who strive to balance a range of competing objectives to support their decisions about hydroelectric power programming, flood mitigation, agricultural and domestic water supplies, irrigation management as well as maintenance of environmental flows. Accurate streamflow prediction and developing an optimal streamflow forecasting model, as a stochastic property of environmental modelling, is one of the most important component of watershed planning and sustainable water resource management [3]. The streamflow is under influence of various factors such as evapotranspiration, rainfall, atmospheric circulation and temperature which makes its generation process nonlinear and time-varying. The magnitude and locality of extreme streamflow events due to climate change and anthropogenic factors can end up to damaged infrastructure, degraded surface water quality, loss of agricultural lands, phosphorus diffusion, and sediment pollutants [4]. Therefore, Accurate and timely predictions of high and low streamflow events at either gauged or ungauged watershed will provide required information to make strategic decisions as following; (1) Ensure sustainable watershed planning; (2) Define the dilution potential of catchments; (3) Set ecological streamflow limits; (4) Allocate water resources.

Due to poor data availability greatly compounds with accurately forecasting daily streamflow, water managers must rely on the streamflow estimates from various prediction models [5]. There are four different streamflow forecasting models: conceptual, metric, physics based, and data-driven. The first three mentioned models assume that the relation between the input and output series is linear or even near linear. They thus ignore the nonlinear information hidden in the streamflow series. In contrast to these models, data-driven methods focus on using nonlinear relation between inputs and outputs. However,

they have some disadvantages including high complexity along with high processing time and high dependence on parameter tuning and optimization [6-7].

To overcome these drawbacks, application of the NESN-MP (called NESN in this paper) forecasting engine in stream flow forecasting is presented in [8-9]. It has been shown that this model works well for the circumstance that there is precise observed stream flow data. However, there are many streams all over the world which do not have accurate observed streamflow data, or the data could exist only in a form that is extremely difficult to access while some other data should be kept secret due to policy concern which will produce bad data. Poor decision being made due to poor data. Therefore, reasonable forecasting of any hydrological process is the call of the time and valuable to responsible and sustainable water resources management. In fact, the National Research Council has noted growing attention to minimize the impacts of bad data among stakeholders for uncertainty assessments of hydrologic prediction [10] which can be because of ungauged basins, potentially inaccurate measurement, incomplete data collection, uncertain estimate, "fat-fingered" data entry, policy concerns, mis-categorization, etc [11].

There are some methodological studies for predicting streamflow response in ungauged basins with bad data which utilized deterministic physically based models to calculate streamflow. They performed based on distributed hydrologic parameters, and statistical regionalization which uses regression models to transfer hydrologic information from gauged to ungauged basins. The distributed hydrologic parameters approach, focusses on dispersing errors into measurement, parameter and structural uncertainty, the produced uncertainties are then disseminated toward model output. The statistical regionalization, is a challenging task in hydrological science [12] due to poor streamflow data, which is

normally calibrated [13]. Moreover, the obtained results have been usually examined on different basins, while every catchment characteristics is different from one case to another [14]. Subsequently, there is no universal method for regionalization. While this is a broadly accepted procedure, uniqueness of the watersheds and the obscurity of parameters bring major uncertainty in the ungauged basins' simulations.

As an inherent symptom of any modelling task, all hydrologic models will suffer from some degree of uncertainty [15]. Movement away from methods grounded in traditional statistics toward conceptual, process-based models has blurred our understanding of model uncertainty to the point that most models are considered as almost purely deterministic tools [16]. Qamar et. al. [3], use non-parametric distance-based method to assess streamflow duration curve in ungauged basins. Their work acquires a more robust model with better global performance even if the extension of the selected model to the whole workspace may be less optimal [17]. Given the hydrological process complexity, using an adaptation of globalized/ regionalized uncertainty is optimal [18].

Some international organizations such as the United Nations Development program (UNDP) and World Bank are concerning to generate a precise approach for development, and management of freshwater recourses. Artificial intelligence (AI) methods are recent developments in several hydrological areas due to their ability to incorporate a tried-and-true model with no need to prior knowledge of the existing functional or nonlinear relationship between input and output [19] One of the most common AI methods to predict stream-flows in ungauged catchments is to identify the train model with homogenous nearby basins to forecast the stream-flow with different climate input [20]. Shu and Ouarda (2008) [21] considered the homogeneous region characteristics to find similar hydrological

sites for predicting flood quantile at ungauged basins; their result showed that the ANFIS approach had more capability compared with the other techniques examined in general, however in sites under 1000 m3/s flood quantile, ANN yields better results. Chang Shian Chen et al. (2010) [22] tried to employ the available hydrological record of nearby catchments with similar homogenous characteristics to estimate ungauged catchments. They concluded the temporal distribution and spatial characteristics considered in the model, reflect most of the behaviour of rainfall–runoff in nature.

A method of random forest models and an ensemble of artificial neural networks, has been used to predict several components of streamflow [23]. Some researchers used regression trees and model tree ensembles to predict a complete flow-duration curve (FDC) for streams, [24]. Senent-Aparicio et al [25] Combined machine learning with Soil and Water Assessment Tool (SWAT) to estimate instantaneous peak flow (IPF) in areas where sub-daily observational data are scarce. The results of this study can contribute to superior ability of extreme learning machine (ELM) to estimate IPF, thereby reducing uncertainties associated with IPF estimations. All previous studies for ungauged estimation have applied the homogeneous nearby basin parameters which result in inaccurate results Because every catchment is unique in its characteristics hence a direct transfer of model parameter values from gauged to ungauged basins may not be appropriate. Therefore, there is yet a need to produce a more accurate estimation of daily streamflow at ungauged basins [26].

However, all These studies granted prized baseline application of machine learning to streamflow perdition, their model performance could not be compared due to one unique accurate data set used for every individual research. To circumvent the above challenges, we focus this paper on developing a novel method on streamflow forecasting with bad data

set input. the primary objective of this research is developing a model to yield valuable estimation, in (1) problems corrupted by noise (2), complex systems that, may not be dittoed, and (3) circumstances where input is incomplete or ambiguous by nature producing bad data (Chitsazan et.al, 2018)

In this proposed method, the model employs the concepts of ANN in an iterative procedure to produce bad data set. The generated bad data set derived from original accurate data set of the gauged basins, are then applied to develop a new input. Subsequently, this new input is used for evaluating the model applicability, which in turn is used for generating ensemble simulations in the ungauged basin. To test the generality of the method, twelve different watersheds across the United States are considered. While all the basins considered in this study were gauged with precise data input, the current study assumed some basins to be un-gauged or producing bad data to evaluate the effectiveness of the proposed methodology. This algorithm will always converge, with no need to stochastic training, and is also applicable to any ungauged basins. Recurrent feedback loops are added to this algorithm, allowing future predictions to be based on time-lagged predictions not time-lagged measurements. To evaluate the effectiveness of the proposed methodology in ungauged basin prediction, we compare our result with ANFIS. The process was repeated by considering representative basins from different climatic and land use scenarios as ungauged. The results of the study indicated that the ensemble simulations in the ungauged basins with NESN were closely matching with the observed streamflow and yield better result comparing to ANFIS. The remainder of this study is as follows. Section II provides an overview of the NESN. Simulation results and discussion are given in Section III, and conclusions are summarized in Section IV.

2.6.2 Nonlinear Echo State Network

In the most of practical circumstances, where the main concern is generating accurate predictions with no insight on the internal structure of the process involved, the authors believe NESN approaches can provide appropriate and accurate solutions. As it has been pointed out in the literature, this novel method is easy, effective, with less computations [27]. NESN provides a total of $2p + p^2$ units; p internal states; p squares of the internal states; and p^2 units gained by multiplying the internal states and squares of the internal states. This process will minimize the order of weight matrices radically. The weight matrices (W, T, and V) are then applied to calculate the internal states of the reservoir. The vector of internal states is updated using

$$\boldsymbol{x}_{(t+1)} = \mathrm{f}(W.\boldsymbol{x}_{(t)} + V.\boldsymbol{s}_{(t+1)} + T.\boldsymbol{y}_{(t)}) \quad (1)$$

and the readout vector is

$$\bar{\boldsymbol{x}}_{(t+1)} = [\boldsymbol{x}_{(t+1)}, \boldsymbol{x}^2{}_{(t+1)}, \sum_{i_1=1}^{p}\sum_{i_2=1}^{p} \boldsymbol{x}_{i_1(t+1)}.\boldsymbol{x}^2{}_{i_2(t+1)}] \quad (2)$$

where $\boldsymbol{x}^2{}_{(t+1)} = \left[x^2_{1_{(t+1)}}, x^2_{2_{(t+1)}}, \ldots, x^2_{p_{(t+1)}}\right]$, p is the number of internal states $\left\lceil \frac{N}{p+2} \right\rceil$, $\boldsymbol{s} \in R^{K\times 1}$ is the input vector, $\boldsymbol{x} \in R^{p\times 1}$ is the internal state vector, $\bar{\boldsymbol{x}} \in R^{(p^2+2p)\times 1}$ is the readout vector, and $\boldsymbol{y} \in R^{L\times 1}$ denotes the output states.

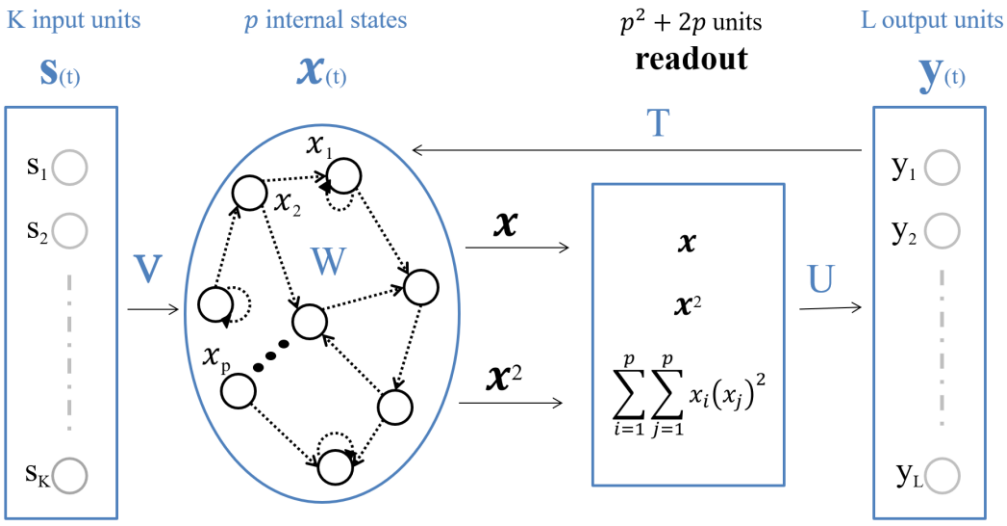


Figure 15. Schematic of NESN.

The matrix $W \in R^{p\times p}$ defines the internal state interconnections within the reservoir. The values in W are fixed values generated randomly over a symmetric interval.

$$W = (w_{ij})_{p\times p} \;\; ; \; w_{ij} \in (-1,1)(i,j = 1,2,\dots,p) \tag{3}$$

Matrix $V \in R^{p\times K}$, containing randomly chosen fixed values, defines the connections of the input with the internal states of the reservoir.

$$V = (v_{ij})_{p\times k} \;\; ; v_{ij} \in (-1,1)(i = 1,2,\dots,p, j = 1,2,\dots,k) \tag{4}$$

The output feedback matrix, $T \in R^{p\times L}$ is

$$T = (t_{ij})_{p\times L} \;\; ; \; t_{ij} \in (-1,1)(i = 1,2,\dots,p, j = 1,2,\dots,L) \tag{5}$$

The output matrix, $U \in R^{L\times(p^2+2p)}$ is

$$U = (u_{ij})_{L\times(p^2+2p)} \;\; ; \; u_{ij} \in (-1,1)(i = 1,2,\dots,L, j = 1,2,\dots,2p+p^2) \tag{6}$$

where K is the number of inputs, p is the number of internal states, and L is the number of outputs.

2.6.3 Simulation Results

The performance of the NESN in presence of bad data is tested using climate (day length, precipitation, solar radiation, maximum and minimum temperature per day, and vapor pressure) and US Geological Survey streamflow data with a time interval of 24 hours used to train and test the models. For the purpose of testing and training, each data set has been divided into two separate parts with their lengths denoted as L_{train} and L_{test}, respectively. To evaluate the performance of the proposed methods the MSE, root mean squared error (RMSE), normalized root-mean-square error (NRMSE), normalized mean-absolute error (NMAE), and mean absolute error (MAE) have been compared. The streamflow forecasting is carried out for 67 days ahead. $l_{train} = 200$, $l_{test} = 30$ with no overlap and with the test data starting immediately after the training data. Fig. 2 shows the prediction for 30 days ahead for NESN and ANFIS without bad data.

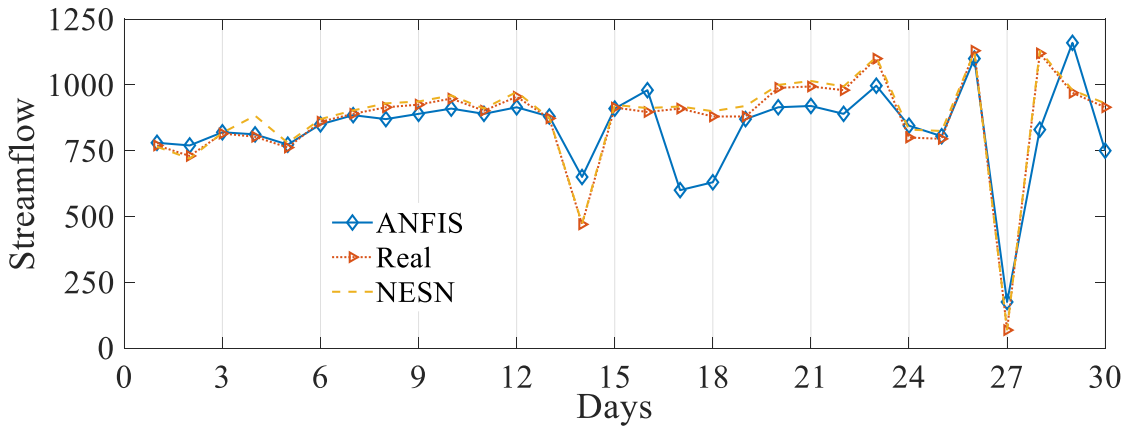


Figure 16. Streamflow forecasting without bad data.

To validate the performance of the proposed method in presence of bad data, severe changes have been made in the input data. The changes vary between 10% and 100% of

the initial values. Fig. 3 and Fig. 4 show the comparison between the streamflow forecasting with and without the bad data for NESN and ANFIS respectively. Table 1 also shows the error indices for different forecasting results shown in Fig. 3 and Fig. 4.

It is shown that the NESN provides the MAE of 15 and 20 with and without bad data which are 80% and 76.7% below those for ANFIS, respectively. In case of RMSE, NESN gives the respective values of 21 and 26 which are well below the RMSE of 115 and 144 for ANFIS with and without bad data respectively. It is shown that the bad data provide uneven impact on the prediction. Therefore, the changes in the predicted results vary during different days.

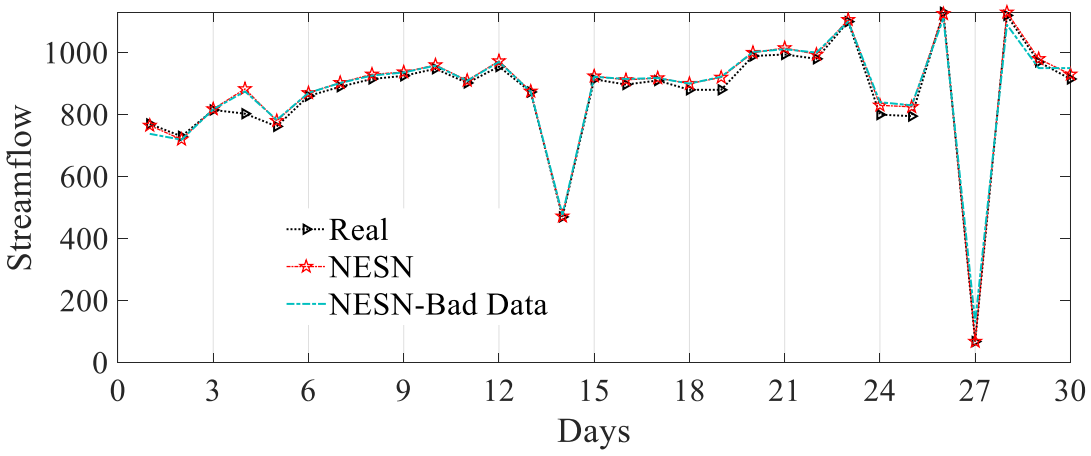


Figure 17. The performance of the NESN in presence of bad data in streamflow forecasting.

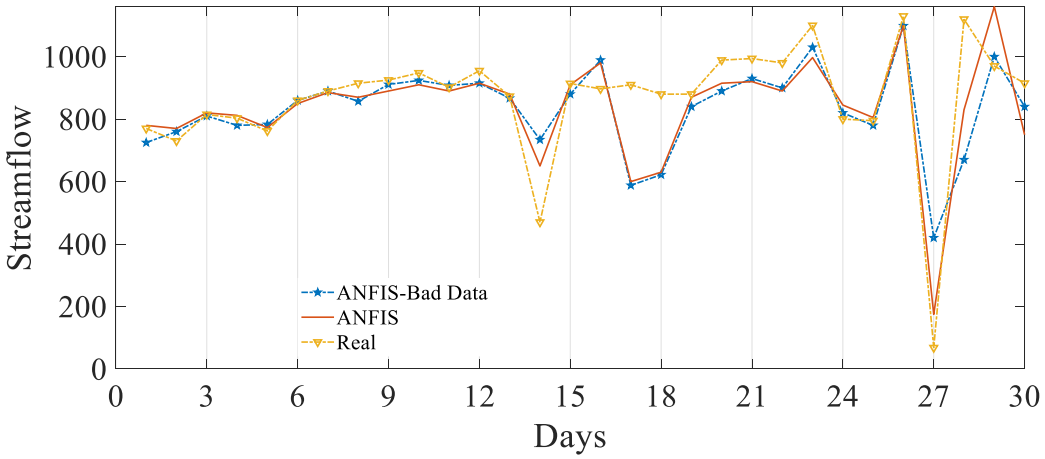


Figure 18. The performance of the ANFIS in presence of bad data in streamflow forecasting.

Table 9. Error indices for ANFIS and NESN with and without bad data.

	ANFIS	ANFIS-Bad Data	NESN	NESN-Bad Data
MSE	13270	20687	449	690
RMSE	115	144	21	26
NRMSE	0.595	0.743	0.110	0.136
MAE	76	86	15	20
NMAE	0.067	0.076	0.013	0.018

2.6.4 Conclusion

This paper addressed the task of predicting daily stream flows with bad data input for water resource purposes, comparing NESN to ANFIS. The purpose of this study was twofold: (1) we aimed to find an effective ANN procedure able to predict mean daily streamflow with bad data input, and (2) we highlighted pros and cons of the two different modelling approaches. Results confirm that NESN can stand the comparison with a ANFIS procedure, producing good performances if correctly trained and appropriately supplied with a good amount of well-chosen information. NESN provides significantly lower values than those given by ANFIS for MAE, NMAE, MSE, RMSE, and NRMSE. However this

kind of approaches suffer from physical interpretability, they can still be considered as a promising tool for predicting stream flow levels, if user is interested in missing data recovery and predicting the streamflow level . NESN seem to be a useful option; however, if a physical interpretation of the process is needed, then the ungenerous conceptual/ANN models would be preferred. Future work will run a sensitivity analysis to explore the most important affecting factor on streamflow forecasting in different circumstance of good data input and bad data input.

2.7 Sensitivity Analysis on Daily Streamflow Forecasting

Abstract: The most forecasting methods still reproduce substantial uncertainty that increases with time and confines the predictability of observed events beyond a few weeks of lead time. Sensitivity analysis (SA) refers to the determination of the contributions every uncertain input data to the uncertainty in the outputs and is a fundamental approach to identify the most significant and sensitive parameters. It helps us understand complex hydrological models particularly for time-consuming distributed flood and streamflow forecasting models based on complicated theory with numerous parameters. SA is increasingly being used in environmental modelling for a variety of purposes, including uncertainty assessment, model calibration, diagnostic evaluation, dominant control analysis, and robust decision-making. This paper aims at delivering an introduction to SA for non-specialist readers, as well as practical advice with best practice examples from the literature. Moreover, as an example, two powerful forecasting engines called Nonlinear Echo State Network (NESN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) are tested. It is shown that NESN is a powerful tool in streamflow forecasting which serves as a robust engine and does not need complex SA and precise observational data input. The SA is conducted under different climatic conditions. The simulation results demonstrate the efficiency of the NESN. The simulation results of the NESN compare favourably with ANFIS.

2.7.1 Introduction

The most hydrological models suffer from uncertainties regarding input data, initial or boundary conditions, forcing data and model structure. These uncertainties may be caused by bad data entry along with poor knowledge of the hydrological process mechanism. Therefore, the model uncertainty is an important issue when developing a modelling system [1]. A useful task to overcome these uncertainties and enhance the model accuracy is to set the values of the model parameters in which the simulation data closely meet observation data. The common approach to achieve this purpose generally called Sensitivity Analysis (SA). SA investigates how the variation in the output of a numerical model can be attributed to variations of its input [2-4]. Within this broad definition, the level of complexity and purposes of SA vary quite significantly depending on the modelling domain and the specific application aims. Depending on whether output variability is obtained by changing the inputs around a reference value, or across their entire feasible space, SA is either referred to as local or global.

In general, SA methods can be broadly categorized into two main classifications, local SA and global SA. The difference between these two approaches laid on their characteristics, scope and applicability [5]. Partial derivatives or finite differences are used as sensitivity indices in the context of local approaches [6]. The local approach does not consider any existing interaction between inputs. Because local SA consider model parameters as varying inputs and aim at assessing how their uncertainty impacts model performance, i.e. how model performance changes when moving away from some optimal or reference

parameter set. Therefore, when we estimate the model parameters, irrelevant or insensitive parameters must be locked at a fixed value to enable more effective SA. In contrast, Global SA applications may consider model parameters and other input factors of the simulation procedure, like the model's forcing data or its spatial resolution simultaneously [7]. Global SA is used for diverse purposes, like verification, supporting model calibration, diagnostic evaluation or simplification [8]; and supporting robust decision-making [9-10]. Besides the local and global SA, several SA methods such as qualitative or quantitative methods, refined or screening methods have been broadly used in different fields, like economics, complex engineering systems, social sciences, and the physics [11-12]. However, there are large differences among these methods in terms of their applicability, sampling schemes, algorithm structures. Given the extensive range of available SA methods, it is very imperative that a practitioner has a clear thoughtful of the appropriate approaches for a specific application. These approaches include choosing an efficient SA method, fitting the method to existing models, and presenting and construing the results. In the context, different types of sensitivity indices can be applied, ranging from correlation trials between inputs and output to statistical properties of the output dispersal. However, analytical computation of all these indices is unbearable for the most models, sensitivity indices can usually be approximated from a sample of inputs and output evaluations [13]. More importantly, the limited data available for physical parameterization of the SA approaches required a substantial dependence on model calibration with a large amount of data input [14]. This dependence occasionally ended in parameterization schemes that are uneven with a physical characteristic of the hydrology of region [15]. Therefore, these limitations are expected to present considerable uncertainty into model projections, particularly in

situations where climatic or environmental conditions differ from those experienced in the calibration period. However, several studies relied on empirical relationships, like curve numbers and the Hargreaves equation, which developed for the moderate regions [16], there are a few studies from these regions to develop a modelling approach which does not rely on complex SA. The complex SA approaches are enabled by the constant progress in various area including computing capabilities, a better understanding of the physical processes and their relations throughout all compartments of the Earth system and the availability and use of more and better observational data which is scarce in the ungauged region. The present rapid development has commanded our systems to be ever more data hungry, thereby growths in model complexity. These computationally expensive developments are not always achievable; hence, model developers must be creative and regularly balance the costs and benefits of improving one aspect over another including increasing the complexity, different parameter selection or fluctuating the model's resolution [17]. However, the Various selection of the parameters will encourage a large variety of simulation results; while, considering that the most existing hydrological models hold complex structures with a large number of parameters, the optimization choice of parameters is a difficult and time-consuming task. Therefore, sensitivity analyses must be easily reproducible to be effective in supporting each new model, and the results should easily be applied to establish a ‘‘continuous learning process’’ [7]. In other words, a sensitivity analysis should be a simple, tractable tool for addressing a complex system.

This is the motivation for the use of NESN-MP [18] (called NESN in this paper) engine in streamflow forecasting to guide future developments for accurate daily streamflow

prediction and is the basis for this paper [19]. The strategy of the proposed forecasting method is to move toward more accurate modelling and forecasting approaches, which dose need an accurate data entry, and beyond that, complex data pre- processing and Sensitivity Analysis. This paper addressed the robustness of the model proposed by Bahrami et.al in which the daily stream flows have been predicted in whether gauged or ungauged basins in different climatic and geographic region [20]. The input data consist of various time series forcing-data including daily precipitation, precipitation duration, solar radiation, temperature and vapor pressure as well as daily streamflow. The nonlinear relations between the internal states increase the learning capability, which results in high forecasting accuracy while ensuring that the quality of forecasting does not deteriorate significantly with time. Our goal is to verify the consistency of the model behaviour and to assess the robustness of the simulation results in uncertain inputs or model assumptions.

The proposed forecasting method appears more than ever as a computer programming tool to establish priorities in improving accurate predictions. Its application is simple, as such it does need an accurate data entry or a large amount of data at the time, and beyond that, complex and computationally expensive data pre-processing along with SA. More importantly, This novel is a user-friendly model such that the user can run the model without prior knowledge about input interaction. This novel model is a powerful and valuable tool to support the examination of uncertainty and predictability across spatial and temporal scales. It can be used for various applications such as accurate and timely predictions of high and low daily streamflow events at either gauged or ungauged watershed without using statically regionalization up to 4 months ahead of the lead time. It

can provide truthful insights into the potential benefits of efforts to provide a forecasting system to managers with prior knowledge of their costs at various activities, including finding minimum data standards, determining model structure, creating priorities for updating forecasting systems, designing field operations. [21-22].

2.7.2 Nonlinear Echo State Network

NESN structure is shown in Fig. 1. In NESN, the networks consist of a reservoir including linear internal states and a readout including nonlinear functions of the internal state. The nonlinear relations between the internal states increase the learning capability, which results in high forecasting accuracy while ensuring that the quality of forecasting does not deteriorate significantly with time. Furthermore, the performance of the forecasting engine is improved by decreasing the number of the internal states, and the orders of the weight matrices which reduces the computational load considerably. Moreover, the proposed methods have simple design, far less computation, and do not require extensive training, parameter tuning, or complex optimization. The formulations are explained in detail in [18], and [23].

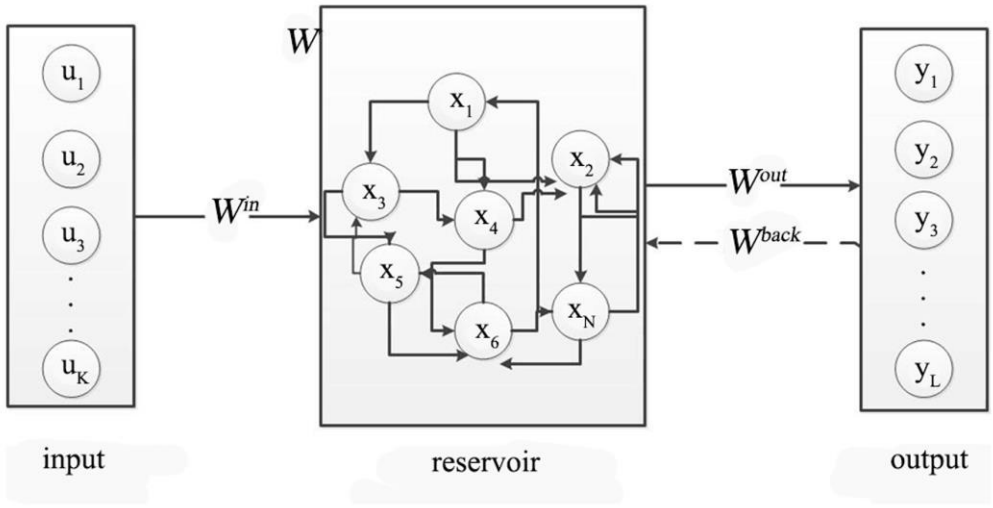


Figure 19. Schematic of NESN-MP.

2.7.3 Sensitivity Analysis for NESN and ANFIS

The sensitivity analysis for two different areas with NESN and ANFIS is explained in this section. Two error indices including the MAE, and RMSE are used to show the sensitivity of the forecasting results for the forecasting engines for specific changes in the input data. The error indices are shown as following:

$$\mathrm{RMSE} = \frac{1}{n_{max}} \sqrt{\sum_{i=1}^{n_{max}} [y(i) - \hat{y}(i)]^2} \tag{1}$$

$$\mathrm{MAE} = \frac{1}{n_{max}} \sum_{i=1}^{n_{max}} |y(i) - \hat{y}(i)| \tag{2}$$

where $\boldsymbol{y}$ are the actual output values, $\hat{\boldsymbol{y}}$ is the predicted output, and $n_{\max}$ is the number of sample points. The input data is set for three changes including the real data, 10 % and 30% of the real data. Table 1 and 2 show the simulations results of sensitivity analysis

using ANFIS and NESN for three case studies. As shown in Table 1, the RMSE for case study I is calculated as 115 and 21 for ANFIS and NESN respectively. The RMSE for case study II and III are also calculated as 98 and 127 for ANFIS and 16 and 11 for NESN respectively. The changes in difference parameters including the precipitation, temperature and vapor pressure are applied for sensitivity analysis.

It is shown that the NESN outperform ANFIS considerably. In case study I, the RMSE changes just by 4 in NESN compared to 51 in ANFIS when there is a 30% change in precipitation. The RMSE in second case study, changes just 1 in NESN while it increases by 13 in ANFIS when the temperature is increased by 10%. In case study III, as expected, NESN shows its robustness during changes while ANFIS could not track the real output in significant changes in the input data. The MAE for both forecasting engines are calculated based on the different changes in the input data. It is shown that NESN is a powerful tool in streamflow forecasting which helps as a robust engine and does not need complex SA and accurate observational data input.

Table 10. RMSE for sensitivity analysis

		ANFIS			*NESN*		
		Precipitation	**Temperature**	**Vapor pressure**	**Precipitation**	**Temperature**	**Vapor pressure**
Case Study I	**real**	**115**	**115**	**115**	**21**	**21**	**21**
	10%	**124**	**122**	**117**	**23**	**22**	**22**
	30%	**166**	**136**	**122**	**25**	**22**	**23**
Case Study II	**real**	**98**	**98**	**98**	**16**	**16**	**16**
	10%	**111**	**126**	**105**	**17**	**15**	**18**
	30%	**135**	**174**	**117**	**21**	**18**	**19**
Case Study III	**real**	**127**	**127**	**127**	**11**	**11**	**11**
	10%	**145**	**122**	**136**	**12**	**11**	**13**
	30%	**198**	**101**	**156**	**14**	**12**	**12**

Table 11. MAE for sensitivity analysis

		ANFIS			**NESN**		
		Precipitation	**Temperature**	**Vapor pressure**	**Precipitation**	**Temperature**	**Vapor pressure**
Case Study I	**real**	**76**	**76**	**76**	**15**	**15**	**15**
	10%	**81**	**80**	**79**	**17**	**16**	**16**
	30%	**88**	**82**	**82**	**21**	**17**	**14**
Case Study II	**real**	**54**	**54**	**54**	**11**	**11**	**11**
	10%	**59**	**62**	**59**	**12**	**12**	**11**
	30%	**72**	**68**	**63**	**14**	**13**	**13**
Case Study III	**real**	**83**	**83**	**83**	**18**	**18**	**18**
	10%	**88**	**87**	**79**	**19**	**16**	**19**
	30%	**94**	**90**	**78**	**22**	**19**	**21**

2.7.4 Conclusion and Future Work

Hydrological models regularly comprise uncertainties which have negative effects on the estimated results, thereby the model accuracy. Therefore, to acquire more accurate model estimates, we better to assess and improve models using different approaches like sensitivity analysis (SA), parameter optimization, operative management, design space exploration, and uncertainty analysis.

In the context, this paper aims to address an introduction on sensitivity analysis for streamflow forecasting.

The strength of the NESN forecasting engine is also evaluated. The robustness and ease of operation in NESN method is high and appear even more than ever as a computer programming tool to establish priorities in improving Accurate predictions. This user-friendly modelling approach does need an accurate data entry or large data entry at the time, or even computationally and complex expensive Sensitivity Analysis along with data pre-processing. the user can apply the model with no need of existing input interaction. This model is a valuable and powerful tool to support the uncertainty and predictability in various spatial and temporal scales (Shortridge et al., 2015).

Moreover, this model robustness can contribute to biased estimates of water availability and uncertainty in forecasting sensitivity to potential future climate changes. Thorough consideration of this accuracy and robustness is important any time that models are used for water planning and management, but especially crucial when using this model to generate insights about future streamflow levels. By considering its predictive accuracy,

error structure, and uncertainties, this method can provide an empirical assessment of watershed behavior and generate useful insights for water management and planning. This makes them a valuable complement to physical models, particularly in data-scarce regions with little data available for model parameterization and warrants additional research into their development and application. Accuracy performance indicated that, for this case study, the use of more information and data did not improve the prediction performance. Our goal is to verify the consistency of the model behaviour and to assess the robustness of the simulation results to uncertain inputs or model assumptions. (Shortridge et al., 2015)

The approach utilized through this study is extendable to similar water projects, which enable reservoir operators to save water as the major contributor for environmental demands, agricultural demands, and hydropower energy production. The developed modelling approach, along with accurate daily predicted values with no need for complex, expensive and time-consuming SA provided a sound basis for the optimal integrated operation of water shed in CONUS and led to minimum evaporation loss by choosing appropriate storage volumes in any related reservoirs resulting in minimum total surface area, and therefore minimum amount of evaporation. As such, the results would provide accurate prediction to model developers; especially those interested in using time series and artificial intelligence-based prediction models; those interested in applying intelligent models in real environments, particularly policy-makers on water and energy resources.

3 A Novel Flood Modelling Method Based On K-Nearest Neighbor Classifier Ensembled With Data Mining Algorithm

*For the reference this part of study is based upon a cumulative research which has been done in collaboration with authors from other universities as follow:

Dieu Tien Bui, Himan Shahabi, Khabat Khosravi ,Sepideh Bahrami, Kayvan Ghaderi, Binh Thai Pham, Ebrahim Omidvar,"Novel Hybrid Intelligence Approach of Bagging Ensemble Based on K-Nearest Neighbor Classifier for Flood Modelling", submitted to Remote sensing journal , Submission date: 22-May-2019 04:07AM (UTC+0800), Submission ID: 1134028179

Abstract

Determining flood prone areas is one of the most important issues for manager in land and disaster managements. We in this study proposed new ensemble models of Bootstrap aggregating as a Meta classifier based on the K-Nearest Neighbor (KNN) functions including coarse, cosine, cubic and weighted as base classifiers to spatial prediction of flood at Haraz watershed in the northern province of Mazandaran, Iran. Although several ensemble models have been developed for this region; however, the KNN has not been earlier explored for flood ensemble modelling over the world. We first selected 10 conditioning factors to spatial prediction of floods and then their prediction capability using relief-F attribute evaluation (RFAE) method were assessed. Models validation was performed using two statistical error-indexes and the area under the curve (AUC). Results concluded that the Bootstrap aggregating-cubicKNN ensemble model outperformed the other ensemble models. Additionally, this model (AUC=0.800) could more decrease the over-fitting and variance problems between the training dataset and eventually well enhance the prediction accuracy of cubicKNN model (AUC=0.660). Therefore, the Bootstrap aggregating-cubicKNN model can be used as a promising technique for the sustainable management of flood prone areas.

3.1 Introduction:

Frequency of flood occurrence is one of the most perilous natural hazard that has increased from pats two decade by over 40% due to deforestation, land-use changes, climate change, poor watershed management and so on (Hirabayashi et al., 2013; Khosravi et al., 2016a; Khosravi et al., 2018b). Flood is defined as overflow of the resulting stream-flow from heavy rainfall which finally cover flood plain, and even areas that was not covered by water during a normal condition (Kron, 2002). Flood occurrence has a direct and in-direct loses and can cause huge damages on life and property including transportation, agricultural sector and garden areas, environmental ecosystem, pollution of surface water through transfer of chemicals and other hazardous industrial wastes and can spread all sort of epidemic disease (Chapi et al., 2017; Messner,Meyer, 2006; Sarhadi et al., 2012; Yu et al., 2013). It has been stated that annually more than 20,000 lives all over the world are lost (Tien Bui et al., 2018) for instance from 1995 to 2015 approximately 109 million people injured by flood while the direct damage was calculated about USD 75 billion per year (Alfieri et al., 2017).

Many Asian countries like Iran do not apart from annual heavy flood. The heavy and divesting flood events are happening every year in northern part of Iran where has been considered as a flood event hot spot. Various flood events have been distinguished in the two northern province of Mazandaran and Golestan; such as Agh Ghala (2019), Gonbad-

E-Kavous (2019), Noshahr (2012), Neka (2013), Behshahr (2013), Sari City (2015, 2019). Moreover there was a sequence a destructive flood event which happened from March 25 to April 8, 2019, in Iran and affected more than 12 provinces all over the country. The most effective parameter which made the event even more destructive was intense rainfall with short duration, poor watershed management and flood warning systems.

Flood warning systems can use forecasts of flood generated by physically-based models to decide about whether floods warnings must be issued to the public or whether previous warnings must be retracted. The physically-based models only can predict discharge therefore their produced map will still remain as a big challenge and weakness to these models (Tehrany et al., 2014). However, it has been shown that geographic Information System (GIS) and Remote Sensing (RS) have brought a new insight to the hydrology science and they can tackle the challenges of flood mapping. For instance, many different data-driven models have been applied to produce the flood susceptibility map such as bivariate models of frequency ratio (Khosravi et al., 2016a; Rahmati et al., 2016a), wrights of evidence (Tehrany et al., 2014), Shannon entropy (Khosravi et al., 2016b), multivariate models including logistic regression (Al-Juaidi et al., 2018; Pradhan, 2010) or Multi-Criteria Decision-Making (MCDM) such as Analytic Hierarchy Process (AHP) (De Brito,Evers, 2015; Kazakis et al., 2015; Rahmati et al., 2016b), Analytic Network Process (ANP) (de Brito et al., 2018),Vlse Kriterijuska Optamizacija I Komoromisno Resenje (VIKOR) and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)(Khosravi et al., 2019). Flood is a highly non-linear process therefore its modeling in a watershed scale is complex and can not be predicted using these simple and

non-linear models. However there are some other methods such as AHP which are nonlinear but their result are not reliable as they require an expert knowledge which can produce high degree of bias and error (Khosravi et al., 2018c).

Recently various machine learning and data mining algorithms for the flood forecasting have been studied. These approaches include: logistic model tree (Chapi et al., 2017), Nave bayes tree (NBT) (Khosravi et al., 2018b), support vector machine (SVM) (Khosravi et al., 2018b) and hybrid of adaptive neuro-fuzzy inference system (ANFIS) with cultural algorithm and bees algorithm (Tien Bui et al., 2018) or with imperialistic competitive algorithm (ICA) and firefly algorithm (FA) (Bui et al., 2018).

Although these algorithms showed a reasonable prediction power in the prediction of flood susceptibility mapping, but Khosravi et al. (2018c) (Khosravi et al., 2018a) stated that there isn't a universal guideline to evaluate a model performance in different condition. While every different model have advantages and disadvantages, in a different condition, dissimilar models must be applied and finally the best performance has to be selected for the future analysis. Although several ensemble models have been developed to overcome this drawbacks, to the best knowledge of authors, the K nearest neighbor (KNN) has not been studied yet for flood ensemble modelling. In the present study two new models of KNN (Coarse KNN, Cosine KNN, Cubic KNN and Weighted KNN**)** and Bootstrap aggregating tree models and also their hybrid model have been applied. The main contribution of the present study is a novel method of KNN, Bootstrap aggregating and their hybrid ensemble which have not been investigated so for in the area of natural hazards, especially flood modeling.

3.2 Description of Study Area

Haraz watershed located in northern part of Iran and in Mazandaran province, has been indicated as one of the most hazardous flood prone area as shown in figure 1. It is affected by destructive floods at the area of 4015 km^2 every year. The watershed Topographic high points is between 51 43′ to 52 36′ E and 35 45 to 36° 22 N. It has been located in extremely mountainous area where the altitude varies from 328 m to 5595 m. The main climate of the study area has been reported as moderate cold climate in fall/winter and mild humid climate in spring/summer. The mean annual rainfall is about 430 mm. The main land cover is Mesozoic formation (about 56.4% of the study area) and rangeland shielded most of the study area (92%).

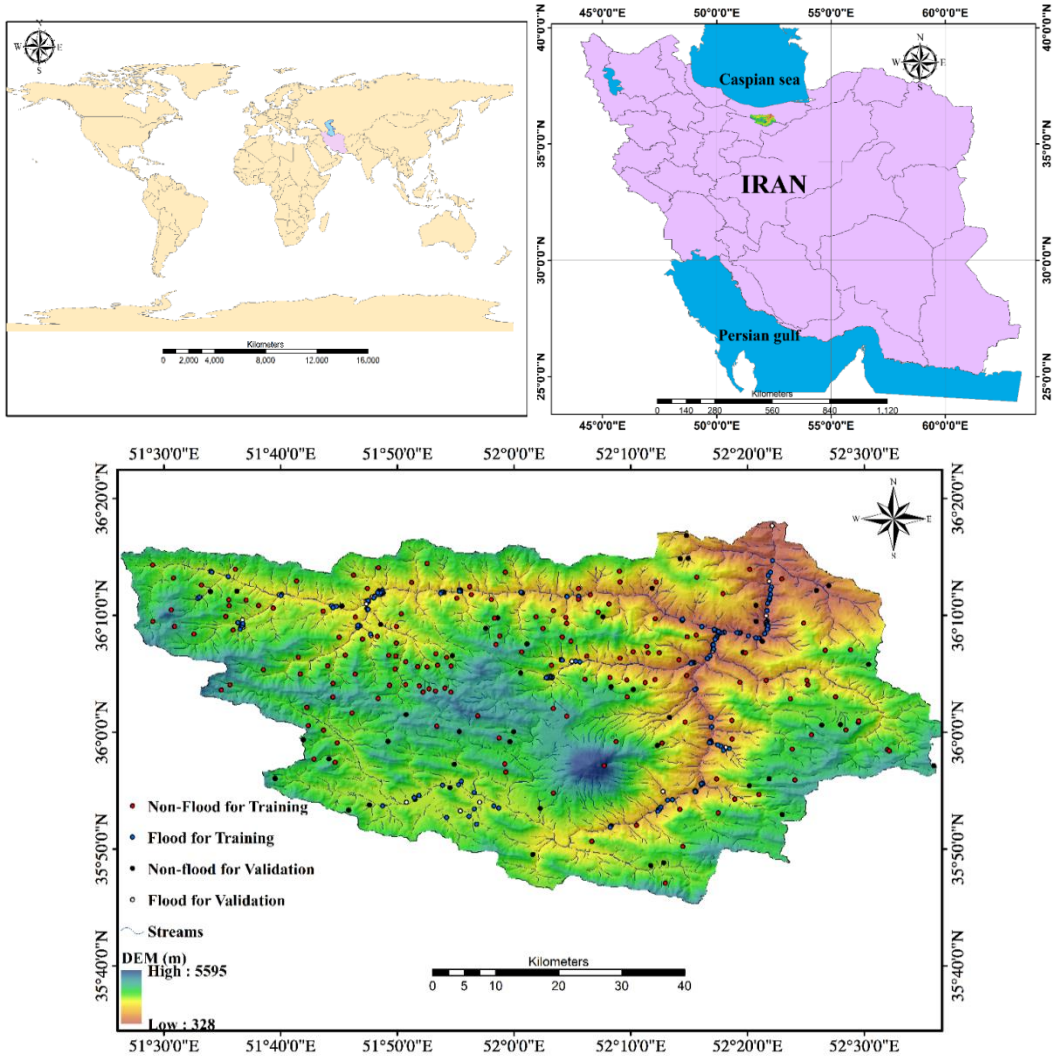


Figure 20. Floods location map of Haraz Catchment in Iran

3.2.1 Data Acquisition:

- **Flood Inventory Map**

Flood inundation locations in 2004, 2008, and 2012 were recorded using field surveys and available documents. A flood inundation inventory map was prepared for the study area which includes 201 flood points. This data set was randomly divided into two groups: training (70%, 141 flood points) and validation (30%, 60 flood points). In flood

susceptibility modelling, both flooded and non-flooded points are needed. According to the literature, an equal balance 1:1 (i.e., 201 non-flood points) was considered to select non-flood points (Rahmati,Pourghasemi, 2017; Tehrany et al., 2014).

- **Flood Conditioning Factors**

However, there are no universal guidelines to select flood-affecting factors, in flood susceptibility modelling the different flood conditioning factors should be considred (Tehrany et al., 2014). According to the literature, 10 flood conditioning factors were selected: altitude, slope, curvature, stream power index (SPI), topographic wetness index (TWI), lithology, rainfall, land use, river density, and distance to river (Rahmati et al., 2016a). A digital elevation model (DEM) of the study area was extracted from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). All topo-hydrological factors were produced using the DEM file. Land use and lithology maps of the study area were obtained from Iranian Department of Water Resources Management (IDWRM). All flood conditioning factors were generated with a spatial resolution of 30 m and then were classified according to the previous studies which have done in this study area (Bui et al., 2018) (Table 1).

Table 12. Flood database for flood hazard mapping

Factor	Variable type	Format	Description
Altitude	Independent variable	Grid	Altitude layer was extracted from a digital elevation model (DEM) and can characterize topography-related processes.

Slope	Independent variable	Grid	Slope layer was produced using the DEM layer. Slope plays an important role in hydrological factors such as flooding.
Curvature	Independent variable	Grid	Curvature layer was generated based on the DEM layer. It can influence on flood inundation situation in large scale.
Stream power index (SPI)	Independent variable	Grid	SPI factor was created based on topographical data which directly reflects the flood potential in a given pixel.
Topographic wetness index (TWI)	Independent variable	Grid	TWI is a topo-hydrological factor which is produced based on the DEM layer. It is commonly used for describing hydrological behavior and soil water/wetness conditions at the catchment scale.
Lithology	Independent variable	Grid	Lithology layer was produced based on geological database of the Geological Surveys of Iran (GSI). It significantly affects natural disasters.
Rainfall	Independent variable	Grid	Rainfall layer was generated based on meteorological databases and it can significantly influence on flooding.
Land use	Independent variable	Polygon	Land use layer was extracted the land use map of Iran. It can clearly reflect the role of human activities to use the land. This factor shows the potential of infiltration.
River density	Independent variable	Grid	River density has an important role in exhausting surface run-off in watersheds. Therefore, this factor has a conceptual relationship with flood inundation.
Distance to river	Independent variable	Grid	The distance to river is a common factor to analyze the flood potential of each point in a given watershed. This factor has useful information related to the flood hazard and vulnerability.

Flooded locations	Dependent variable	Point	Flooded points were recorded in this watershed using the GPS device. These points include unique information regarding flood potential of each pixel.

3.3 Methodology

3.3.1 K-Nearest Neighbor Pattern Classification (KNN) :

It is a non-parametric lazy learning algorithm which does not make any assumptions on the primary data set. This matters, when modeling hydrological process; such as flood and stream flow where there is little or no prior knowledge about the data distribution which maketheir generation process highly nonlinear (Bahrami,Wigand; Wettschereck et al., 1997). As in the real world, most of the existing practical data set does not conform the distinctive theoretical assumptions like linear regression models which has been vastly used (Bahrami,Wigand). K-nearest neighbor stores all contributing cases and classify new cases based on a similarity measure called distance function. Then the case will be classified by a majority votes for its neighbor classes. Thereafter, it would be allocated to the greatest common class between the existing K nearest neighbors (Wu et al., 2008). The optimal choice of *K* depends on the metric. However, a general rule of thumb which is square root of the number of samples can be applied to choose K value, it makes the parameter tuning difficult for diverse applications. The K value depends on the chosen data set and won't be same in various data set even if there is same conceptual model. Based on an empirical rule-of-thumb introduced by the "Pattern Classification" book by Dude (Duda et al., 2012; Guo et al., 2003), the value of K is equal to the square root of the number of occurrences(general rule of thumb). Although there are some other popular methods such

as K-fold cross validation (CV) or Leave-one-out cross validation (LOOCV) and bootstrapping, the K-Fold Cross Validation is preferred amongst all for the following reasons (Liu et al., 2010):

i. There are typically only a few probable choices of K (e.g. from 3 - 10 or 50 - 100);

ii. The model performance is slightly intonation

In this algorithm, the training phase is short and fast. All the training data set are required during the testing phase to prevent any task of generalization. However, this causes the training phase to be time and memory consuming, it helps to make the decision based on the best subset of the entire training data set. This method has various application in diverse problems such as large data classification, pattern recognition, ranking models, computational geometry and vision to proteins etc (He,Wang, 2007).

This algorithm applies a vector as an input with the k training dataset. Then it uses the most common class to classify the k nearest neighbors. During The training phase, the neighbors are defined based on their distance from the test dataset and a in the testing phase the class of test dataset are determined [4]. Then to identify the best k-NN algorithm performance, the number of neighbors (K) can be reformed. There are three k-NN classifiers introduced by MATLAB that are categorized based on different distances which work as follows (Hu et al., 2016):

Coarse k-NN: In this method the number of neighbors is 100 and it is defined as the nearest neighbor between all other classes.

Cosine k-NN: In this method the cosine distance metric is defined as a nearest neighbor classifier. It is generally used as a metric for measuring distance when the magnitude of

the vectors does not matter. The following equation is applied to measure the distance between two vectors u and v (Hu et al., 2016):

$$1 - \frac{u.v}{|u|.|v|'} \quad \text{eq-1}$$

Cubic k-NN: In this method the number of neighbors are 10 and the cubic distance metric (Wu et al., 2008) is defined as a nearest neighbor classifier. The following equation is applied to measure the distance between two n-dimensional vectors u and v:

$$\sqrt[3]{\sum_{i=1}^{n} |u_i - v_i|} \quad \text{eq-2}$$

Weighted k-NN: In this method the number of neighbors is 10 and a distance weight is defined as a nearest neighbor classifier. The following equation is applied to measure the weighted Euclidean distance between two n-dimensional vectors u and v:

$$\sqrt{\sum_{i=1}^{n} w_i (x_i - y_i)^2} \quad \text{eq-3}$$

Where $0 < w_i < 1$ and $\sum_{i=1}^{n} w_i = 1$.

3.3.2 Bagged Tree Ensemble Algorithm

To produce better predictive performance, Ensemble methods applies various decision trees instead of employing only one decision tree. This will produce a strong learner as it combines the weak learners. Two most common techniques to perform ensemble models are (Dietterich, 2000):

- Bootstrap aggregating
- Boosting

For producing strong learner, the weak learner's prediction results will be combined via methods as fallows (Bauer,Kohavi, 1999):

- average/ weighted average
- higher vote prediction Selection

Bootstrap aggregating (Bootstrap Aggregation) has been designed to improve the precision and constancy of machine learning algorithms used in regression and statistical classification. The task of Bootstrap aggregating is to decrease variance while retaining the bias of a decision tree and prevent over-fitting problem. The Bootstrap aggregating Tree can randomly generate multiple sets of input data from training samples with replacement (Maclin,Opitz, 1997). Then the chosen subset data is used to train the assigned trees and will generate various models. Subsequently the average of all the predictions from these trees are used to make the final decision with higher robustness degree. In the context, the accuracy of a single tree will increase by using multiple copies of the trained subset of data.

Boosting is a useful ensemble model in high bias condition. In this technique, predictors are trained sequentially with a simple early training models and then the data are analyzed for errors. At every step, the net error is calculated from the prior successive decision tree (Maclin,Opitz, 1997).

In high bias dataset, when an input is not well classified by a hypothesis, its weight is amplified so that next hypothesis will classify it properly.

The concept of a heuristic originates from the perceptive science domain to make a judgment. A design heuristic is a swift that boosts exploration of a various ideas during product ideation (e.g. idea generation) to treasure a balance in complicated specific instance (train dataset). To find a balance, the averages of various subset selections of observations (several thumb rules) are extracted out from the original dataset.

In the current study, to categorize the dataset into two probable classes, an algorithm of a continues classifiers (H_m, $m = 1,\ldots, M$) $Hm: Dm \rightarrow R$ on the domain of a training set (Flood collection) D, has been generated. The generated classifiers are then grouped into a composite classifier which its resulting prediction is specified as a weighted grouping of individual classifier as fallow:

$$H(d_i) = sign(\sum_{m=1}^{M} \alpha_m H_m(d_i)) \quad \text{eq-4}$$

The eq-4 describes a voting procedure. By the given function, an example d_i is classified based on the majority of classifiers' vote (Giacinto,Roli, 2001; Kamali et al., 2014).

Parameters α_m, $m = 1,\ldots, M$ are identified to indicate the impact of more accurate classifiers on the final result versus less accurate classifiers. The H_m are called weak classifiers as their accuracy is somehow higher than the accuracy of other random classification methods (Waske et al., 2010).

In the presented study the following Bootstrap aggregating algorithm has been experimented (Liu et al., 2014):

1. Training set D initialization

2. Range selection for m = 1, ..., M

2.1. Random selection of the set D to create a new set D_m

2.2. Machine-learning application on the base of $D_m$to train a classifier $\mathrm{Hm}: \mathrm{Dm} \rightarrow \mathrm{R}$.

3. Creation of composite classifier H from $\mathrm{H_m}$, $\mathrm{m} = 1, \ldots, \mathrm{M}$

3.1. d_i classification based on c_i classes, depending on the number of votes gained from H_m.

$$H(d_{i,}c_i) = sign(\sum_{m=1}^{M} \alpha_m H_m(d_{i,}c_i)) \quad \text{eq-5}$$

Note that to achieve a better performance and decrease the classification error, the H_m values can be reformed, while α_m values are persistent.

3.3.3 Flood Factors Selection Using Relief-F Attribute Evaluation (RFAE) Technique

Factor/feature selection in the supervised machine learning algorithms can detect the best factors to accurately classify the example of data and also enhance the efficiency of the training process (Shirzadi et al., 2018). Indeed, the main aim of the feature selection is to enhance learning efficiency of modelling process, robustness of predictive accuracy, reducing complexity, noise and over-fitting problems by eliminating the factors that are irrelevant or they have no predictive information (Ramaswami,Bhaskaran, 2009). There are some methods and techniques for feature selection that all of them are categorized based on the distance, information, dependency, consistency and classifier error rate measures (Dash,Liu, 1997). In this study we selected a distance based-measure relief, attribute

evaluation (RAE) technique, to check the role of conditioning factors on the flood classification performance.

The RAE is one of the distance-based attribute/factor ranking methods which was introduced by Kira and Rendell (1992)(Kira,Rendell, 1992), and then it was enhanced by (Kononenko, 1994)Kononenko (1994) (Hall,Holmes, 2002)(Hall, and Holmes, 2003). A main idea of RAE is to compute the quality of each attribute based on the distance between the instance and its nearest neighbors. Firstly, instances in the training dataset were randomly selected (R_i in line 3). Then, relief searches for k of its nearest neighbors from the same class, and rom each of the different classes, called nearest hit Hj (line 4), and nearest miss M_j (C), lines 4 and 6, respectively. The RAE depending on the average values of R_i, H_j, and M_j (C) (lines 7, 8 and 9), updates the quality estimation W[A] for all attributes. The quality estimation W [A] will be decreased when instances Ri, and Hj have different values of the attribute A. as a result, the attribute A will be separated into two instances with the same class values which this result is desired. Unlike, if Ri and Mj (C) have different values of the attribute A resulting in separating attribute A into two instances with different class values. The prior probability for each class of the misses, P (C), is calculated based on the training dataset. The P (C) is a symmetric in which it ranges from 0 and 1 for hits and misses. When the class of hits is missing in the sum, each probability weight divided with factor 1- P (class (R_i)). It depicts that the sum of probabilities for missing class. This process will be repeated m times. Figure 2 show a pseudo code of relief algorithm (Robnik-Šikonja,Kononenko, 2003).

Algorithm ReliefF

Input: for each training instance a vector of attribute values and the class value

Output: the vector W of estimations of the qualities of attributes

1. set all weights $W[A]$: = 0.0;
2. for i: = 1 to m do begin
3. randomly select an instance R_i;
4. find k nearest hits H_j;
5. for each class C ≠ class (Ri) do
6. from class C find k nearest misses $M_j(C)$;
7. for A:= 1 to a do
8. $W[A]: = W[A] - \sum_{j=1}^{k} \text{diff}(A, R_i, H_j)/(m.k) + \sum_{C \neq class(R_i)} [\frac{P(C)}{1-P(class(R_i))} \sum_{j=1}^{k} \text{diff}(A, R_i, M_j(C))]/(m.k);$
9. end;

Figure 2. Pseudo cod of the basic relief-F algorithm

3.3.4 Evaluation and Comparison

The performance of a new developed model should be tested and evaluated to ensure the performance and to propose the model for other regions (Bui et al., 2018). To check the performance of the models there some metrics that all of them are computed based on the different between observed and estimated values, defined as a forecasting error (Tien Bui et al., 2018). In this study, for this purpose we used of MSE, RMSE and AUC as statistical metrics.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (F_{est.} - F_{obs.})^2 \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (F_{est.} - F_{obs.})^2} \quad (2)$$

where $F_{est.}$, $F_{obs.}$ and n are flood estimated, flood observed (actual floods) and number of floods for modelling process.

In addition to the MSE and RMSE, we used of ROC and AUC as a standard tool for more validation of the prediction capability of the models which has been used in some flood modelling studies (Ahmadlou et al., 2018; Bui et al., 2018; Chapi et al., 2017; Shafizadeh-Moghadam et al., 2018). It is plotted by two statistical metrics including specificity on the x-axis sensitivity on the y-axis (Shirzadi et al., 2019). Specificity and sensitivity are defined as the number of incorrectly and correctly floods classified, respectively (Hong et al., 2018). The AUC can be computed as follows:

$$AUC = \frac{\sum TP + \sum TN}{M + N} \quad (3)$$

where TP (true positive) and TN (true negative) are the number of flood correctly classified pixels as floods and non-flood pixels, respectively. M and N are the number of total flood and non-flood pixels, respectively (Shahabi,Hashim, 2015).

3.4 Result and Analysis:

3.4.1 Selection the Most Important Factors for Flood Modelling

The results of factor selection by RFAE technique were shown in Figure 3. The average merit (AM) values ranging between 0.002 and 0.198 indicated all conditioning factors were qualified to flood susceptibility modelling (AM > 0). Distance to river with the highest average merit (AM = 0.198) was expectantly the most important factor. It is because most of flood points in the study area were naturally located beside the river network. It is followed by slope (AM = 0.186), curvature (AM = 0.160), drainage density (AM = 0.150), elevation (AM = 0.135), TWI (AM = 0.124), lithology (AM = 0.059), rainfall (AM = 0.053), SPI (AM = 0.043) and land use (AM = 0.002).

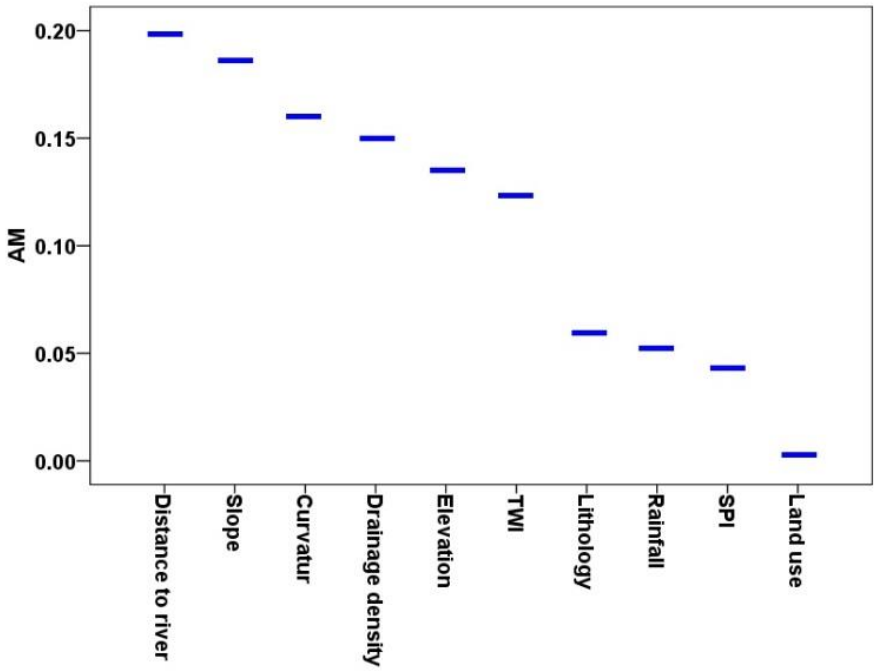


Figure 21. Flood important factors selection by relief-F attribute evaluation (RFAE) technique

3.4.2 Flood Modelling Process

The intelligence approaches of Bootstrap aggregating tree and modified k-nearest neighbor classifiers (Cubic-KNN, Coarse-KNN, Cosine-KNN and Weighted-KNN) were performed to flood modelling. The models were trained and tested with the ratio of 70 and 30 percent of dataset, respectively. The accuracy criteria's of models were calculated according to comparing between the training/test dataset as a target and predicted flood pixels as the output. Figurers 4-… shows the targets and outputs, and some accuracy criteria such as MSE, RMSE, mean and standard deviation (SD) in the training and test steps. In the training step, the MSE of Cubic-KNN, Coarse-KNN, Cosine-KNN, Weighted-KNN and Bagg-Tree models was 0.0568, 0.0575, 0.0504, 0.000 and 0.0072, respectively, while the RMSE was 0.2383, 0.2399, 0.2244, 0.0000 and 0.0848, respectively. Accordingly, The Weighted-KNN had the best performance in the training step (Mean = 0 and SD = 0). In the test step, the MSE and RMSE of Cubic-KNN, Coarse-KNN, Cosine-KNN, Weighted-KNN and Bagg-Tree models were respectively 0.0396 and 0.1989, 0.0682 and 0.2611, 0.0682 and 0.2611, 0.0568 and 0.2384, and 0.0454 and 0.2132. These results suggested the Cubic-KNN as best performed model in the test step (Mean = -0.0324 and SD = 0.1966).

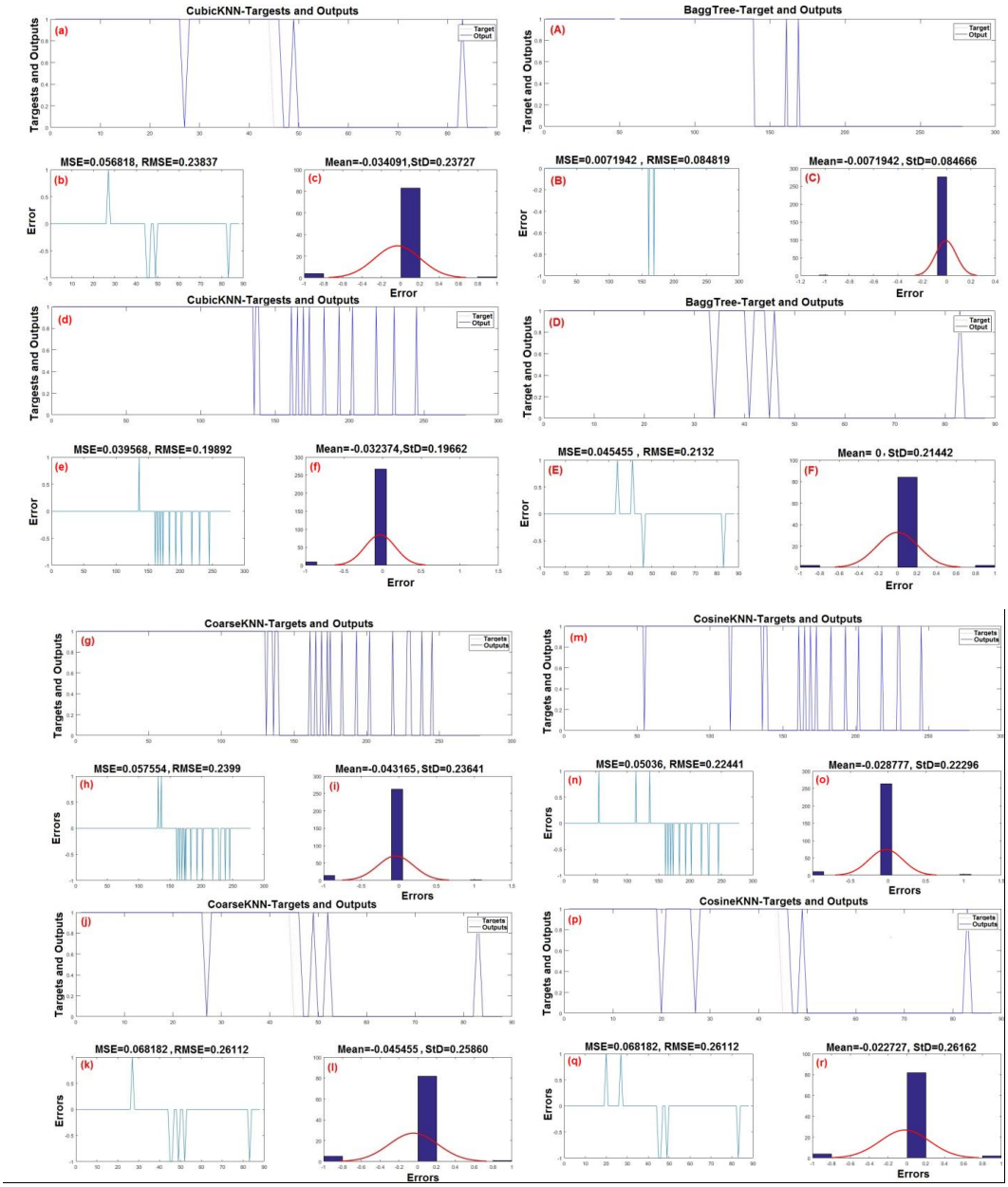


Figure 22. Modelling process using Cubic-KNN:

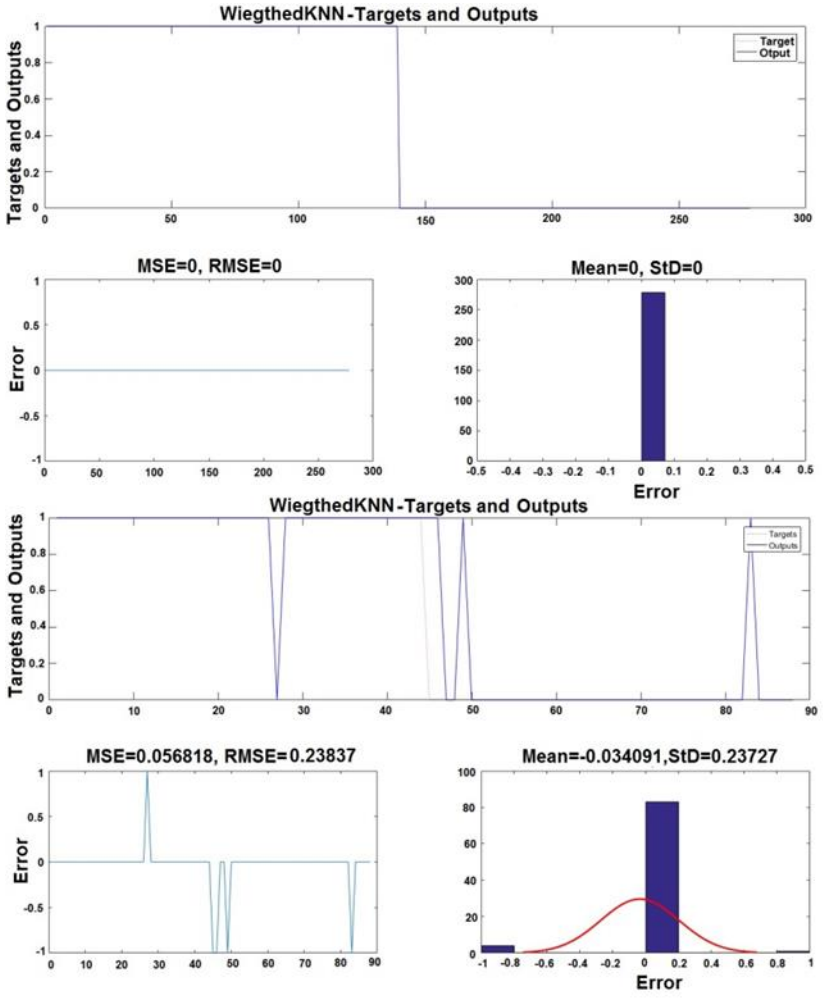


Figure 23. Modelling process using Cubic-KNN

A comparative assessment between the proposed models was considered to evaluate the accuracy of KNN classifier functions in the modelling process. Table 1 present the optimal parameters for the best accuracy of these models. The results indicated that the Cubic-KNN model has the highest accuracy value (96.4%), following by the Cosine-KNN (92.8%), Weighted-KNN (92.1.4%) and Coarse-KNN (92.1%) models.

Furthermore, the hybrid models of Bootstrap aggregating Tree based on KNN classifiers were built, and their optimal parameters values were obtained based on the highest accuracy. Table 2 showed the optimum parameter values of hybrid models. Comparison of

the accuracy was showed that the highest accuracy was gained for the hybrid model of Bagg Tree-Coarse KNN (98.6%), followed by Bagg Tree -Weighted KNN (97.1%), Bagg Tree-Cosine KNN (96.6%) and BaggTree -Cubic KNN (94.3%), respectively.

Table 13.KNN functions used for spatial prediction of flood in the modeling process

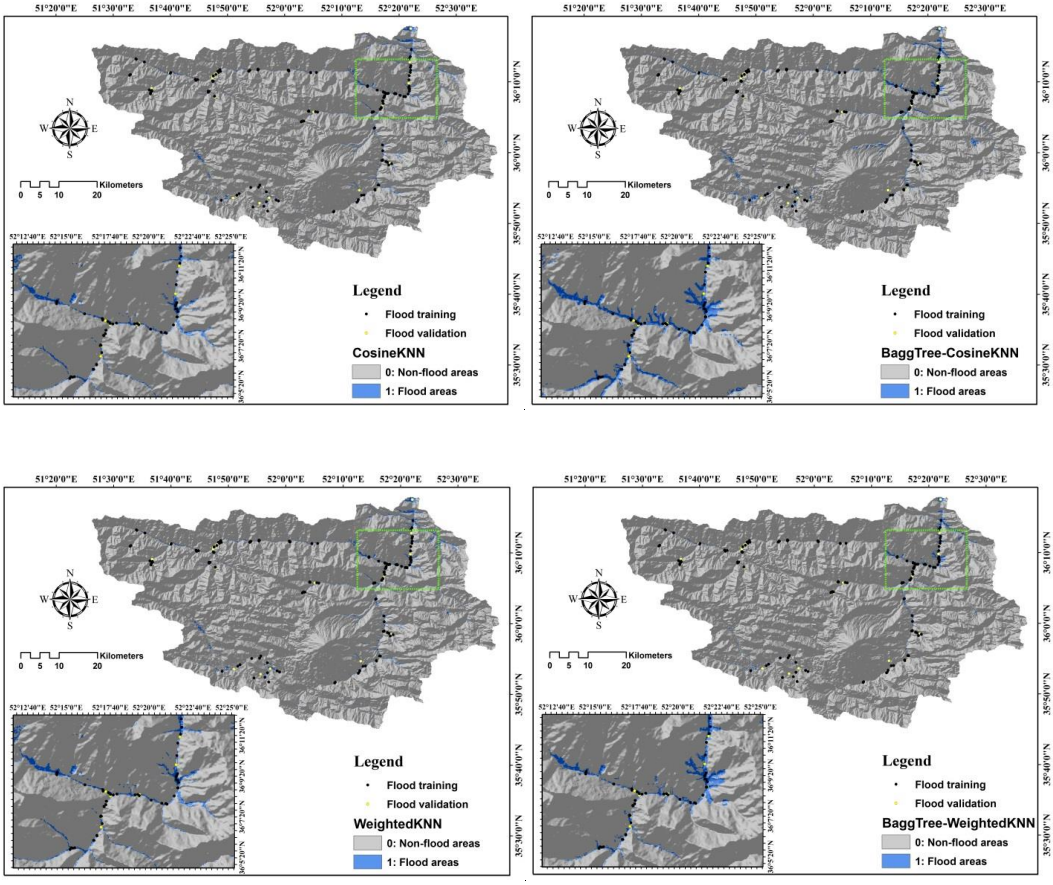
Title	Description			
Classifier Preset	Coarse KNN	Cosine KNN	Cubic KNN	Weighted KNN
Accuracy	92.1%	92.8%	96.4%	92.1%
Distance metric	Euclidean	Cosine	Minkowski (cubic)	metric Euclidean
Distance weight	Equal Standardize	Equal Standardize	Equal Standardize	weight Squared inverse Standardize
Number of neighbors	100	10	10	10
Prediction speed (obs/sec)	~27000	~22000	~15000	~29000
Time Training (Secs)	0.255	0.282	0.293	0.211

Table 14.BagTree ensemble on KNN and its functions used for spatial prediction of flood in the modeling process

Title	Description			
Classifier Preset	BaggTree-Coarse KNN	BaggTree-Cosine KNN	BaggTree -Cubic KNN	BaggTree -Weighted KNN
Accuracy	98.6%	96.6%	94.3%	97.1%
Learner type	Decision tree	Decision tree	Decision tree	Decision tree
Number of learners	30	30	30	30
Ensemble method	Bag	Bag	Bag	Bag
Prediction speed (obs/sec)	~2200	~3900	~5100	~5800
Time Training (Secs)	0.375	0.737	0.693	0.761

3.4.3 Development of Flood Susceptibility Maps

The hybrid models were performed for calculation of the flood susceptibility index (FSI) which was assigned to all pixels of the study area to create flood susceptibility maps. Each pixel of the study area was firstly assigned to a unique FSI; then, these indices were exported in Arc GIS 10.3 format and used to provide the final flood susceptibility maps. Calculated FSI's were classified in two classes of susceptibility including non-flood area and flood area. The flood susceptibility maps of study area according to the Bootstrap aggregating Tree Ensemble Based on Modified K-Nearest Neighbor Classifiers are shown in Figure **24** a-h**.** The maps illustrated that flood susceptible areas in the basin are located around the river network with lower elevation and slope. In comparison to the nearest neighbor models, the hybrid models predicted a higher area of basin as flood susceptible, so that, the most portion of study area were predicted as flood susceptible by hybrid Bagged Tree-Cubic KNN model (Figure 24-b).


51°20'0"E 51°30'0"E 51°40'0"E 51°50'0"E 52°0'0"E 52°10'0"E 52°20'0"E 52°30'0"E
36°10'0"N
36°0'0"N
35°50'0"N
35°40'0"N
35°30'0"N
N
W
E
S
0 5 10 20
Kilometers
52°12'40"E 52°15'0"E 52°17'40"E 52°20'0"E 52°22'40"E 52°25'0"E
Legend
Flood training
Flood validation
CosineKNN
0: Non-flood areas
1: Flood areas
Legend
Flood training
Flood validation
BaggTree-CosineKNN
0: Non-flood areas
1: Flood areas
Legend
Flood training
Flood validation
WeightedKNN
0: Non-flood areas
1: Flood areas
Legend
Flood training
Flood validation
BaggTree-WeightedKNN
0: Non-flood areas
1: Flood areas

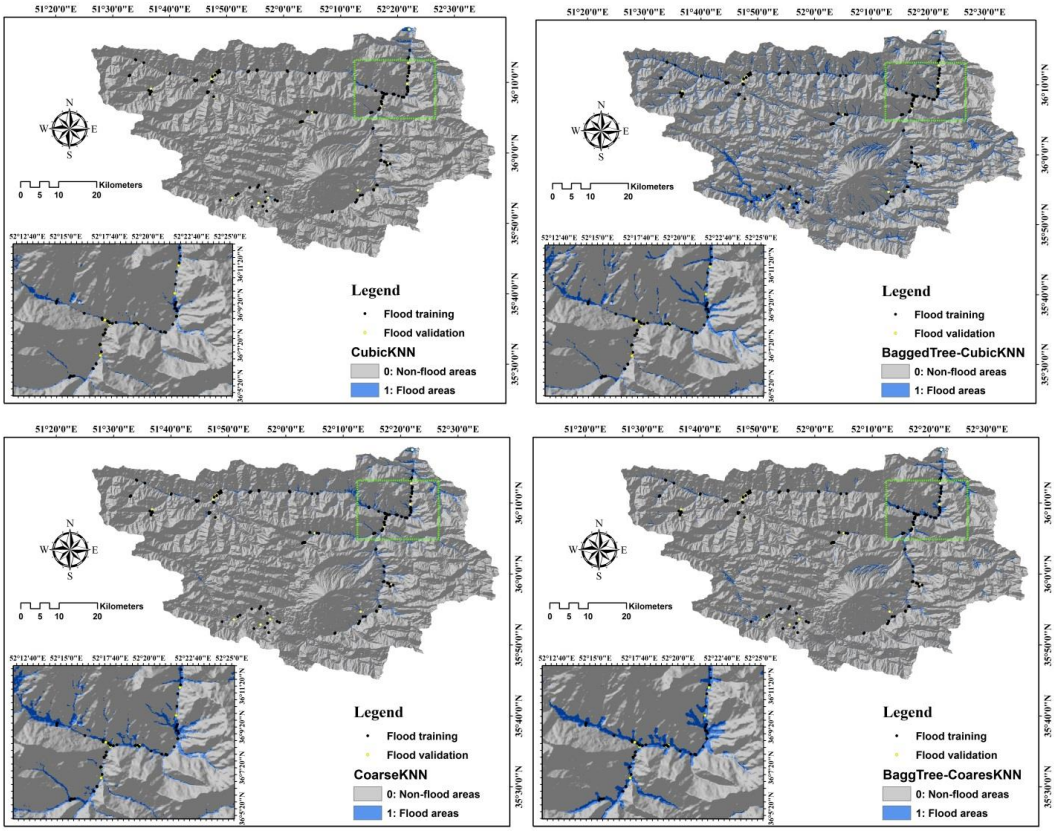


Figure 24. Flood susceptibility maps in the basin extracted from Cubic-KNN (a), Bagg Tree-Cubic KNN (b), Coarse-KNN (c), Bagg Tree-Coarse-KNN (d), Cosine-KNN (e), Bagg Tree-Cosine-KNN (f), Weighted-KNN (g), and Bagg Tree-Weighted-KNN (h)

3.4.4 Evaluation and Comparison

The performance of new hybrid of Bootstrap aggregating Tree based on KNN models in prediction of flood susceptibility were compared with KNN models using area under receiver operating characteristic (AUROC) curve. Figure 25 presented the ROC curves which were produced based on datasets of flood susceptibility maps in training and test steps. The result of ROC curves for KNN classifiers shows the Coarse-KNN model is the most suitable model in training and test steps with AUC of 0.795 and 0.790, respectively,

following by the Weighted-KNN (AUC = 0.719 and 0.710), Cosine-KNN (AUC = 0.692 and 0.690), and

the Cubic-KNN (AUC = 0.662 and 0.660) model, respectively (Figures25- a & b). The validation of hybrid models by ROC revealed that Bagg Tree-Cubic KNN model had highest performance in both of training and test steps with the AUC of 0.811 and 0.800, respectively. It is followed by Bagg Tree-Coarse KNN (AUC = 0.762 and 0.740), Bagg Tree-Weighted KNN (AUC = 0.722 and 0.710), and Bagg Tree-Cosine KNN (AUC = 0.659 and 0.640), respectively (Figures 25 c & d). Therefore, the hybrid models had higher performance comparing to KNN classifier models. Thus, the Bagg Tree-Cubic KNN can be used as a desirable approach to flood susceptibility modelling. According to classification of (Kantardzic, 2011)Kantardzic (2011) the Bagg Tree-Cubic KNN model showed good performance, while the others performed moderate.

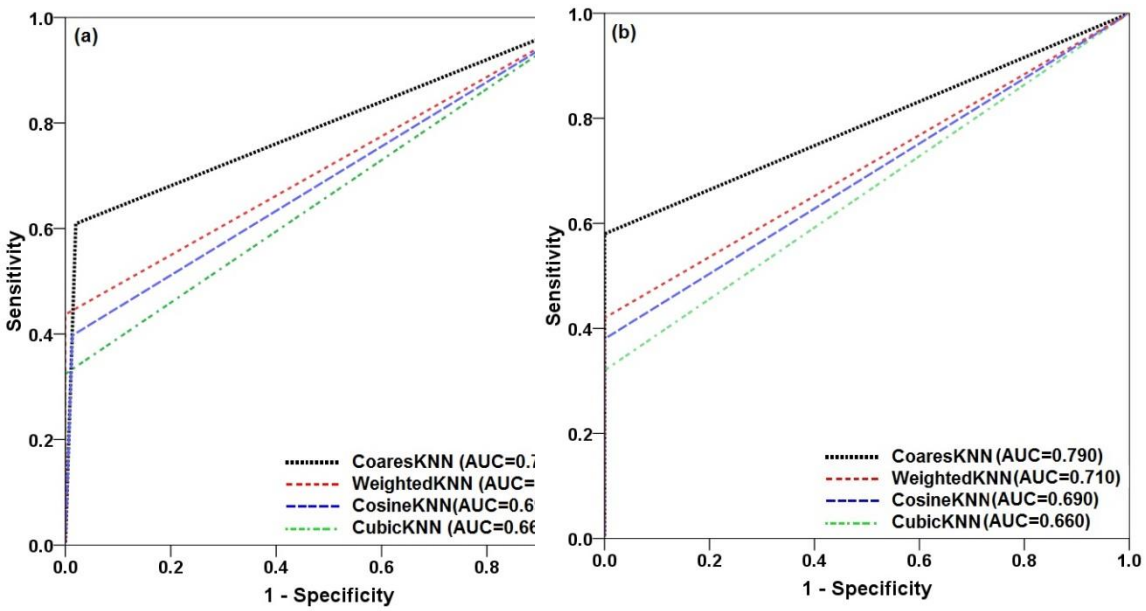


Figure 25. flood models evaluation using AUC: (a) KNN-individual classifiers by training dataset, (b) KNN-individual classifiers by validation dataset, (c) BaggTree-KNN ensembles by training dataset, and (d) BaggTree-KNN ensembles by validation dataset

.

3.5 Discussion and Conclusion

Flood mapping is a useful tool provided conductive information for decision-makers and hazard disaster management in flood-prone areas. In the modelling process by machine learning algorithms, there is still a challenge to specify an algorithm/classifier with maximum attainable accuracy for a given dataset due to data complexity (Fernández-Delgado et al., 2014). On the other hand, due to learner's limitation and also data limitation some classifiers should be tested and evaluated for achieving a reliable result. Therefore, developing a new model with a high prediction accuracy based on the training dataset in each study area is the main goal of researchers. Basically, in this research, we proposed a new intelligent hybrid model (BaggTree -Cubic KNN) for modeling flood, which is a combination of Bootstrap aggregating ensemble technique and the four functions of KNN classifier. The model was tested at Haraz watershed, the Mazandaran province, Iran. Ten flood conditioning factors, including slope angle, elevation, curvature, stream power index (SPI), topographic wetness index (TWI), land use, rainfall, drainage density and distance to the river were selected for flood modelling. The information gain ratio (IGR) was used to optimize the most important conditioning factors for the flood model. The result of IGR showed that all the factors were significant in the model training; however, distance to a river was the most important factor, followed by slope, curvature. The obtained result is in agreement with (Ahmadlou et al., 2018; Bui et al., 2018; Khosravi et al., 2018b; Shafizadeh-Moghadam et al., 2018). On the other hand, most of floods in the study area taken place during the heavy rainfall and as a result of overbanking the rivers. This process leads to inundation the areas adjacent the rivers, flood plains, during the heavy rainfall where are distinguished in the prepared flood susceptibility maps.

The KNN as one of the most attended neighborhood classifiers belong to the weak classifiers in which they are very simple and highly efficient in some fields of studies (Hassanat, 2014). It is remarkable that memory requirement and time complexity are the main limitations of performance ability of KNN classier that they completely depends on every example in the training set (Hassanat et al., 2014). To tackle these limitations and enhancing the performance of KNN, we used of Bootstrap aggregating Meta classifier. The combination of the Bootstrap aggregating ensemble technique and the KNN classifier can provide a solution to build a flood model. It has been shown that, the area under the receiver operating characteristic curve (AUC=0.800) value of the proposed BaggTree -CubicKNN model has the best performance. It could significantly enhance the prediction accuracy of CubicKNN classifier as a base classifier. Chapi et al., (2017) tested and evaluated the Bootstrap aggregating ensemble to improve the power prediction of logistic model tree (LMT) classifier as a new proposed model, Bootstrap aggregating-LMT, for flood mapping also at Haraz watershed. Their result concluded that the Bootstrap aggregating well enhanced the power prediction of base classifier, LMT, during the flood modelling. General speaking, the ensemble model can outperform the basic classifier because the ensemble model integrates the advantage of each classifier (Tien Bui et al. 2016; Shirzadi et al. 2018; Dou et al. 2019). The novel proposed model is a better alternative for flood model, thus, it is recommended as an appropriate method for flood hazard disaster management.

Assessing flood model is a complex procedure, which is linked with various uncertainties. Whereas machine learning approaches are able to proficiently handle this uncertainty problem, which requires historical flood spatial distribution inventory maps. The proposed

ML model can enable decision-makers for a less-expensive early field survey of the district they are meant to cope with high reliability. Thus, the achievement of this research could assist the managers to recognize the flood-prone zones in watershed more accurately. After the assessment of the frequent susceptible areas, decision-managers could prepare more accurate and more useful data related to these regions (such as rainfall and river data) using proposed models to produce accurate flood maps for mitigating further damage. The produced flood maps are fundamental for further analyses, like hazard and risk disaster management and mapping. The model can be used in other parts of the world.

4 Spatial Flood Modeling Using Remote Sensing and GIS With The Application of Evidential Belief Functions for Feature Selection

For the reference this part of study is based upon a cumulative research which has been done in collaboration with authors from other universities as follow:

Tien Bui. D.; Bahrami, S.; Khosravi, K.; Shahabi, H.; Daggupati, P.; Adamowski, J.F.; Melesse, A.M.; Thai Pham, B.; Pourghasemi, H.R.; Mahmoudi, M.; Flood spatial modeling in northern Iran using remote sensing and gis: A comparison between evidential belief functions and its ensemble with a multivariate logistic regression model. Remote Sens. 2019, 11, 1589.

Abstract

Flooding is one of the most dangerous natural disasters that repeatedly occur in the northern region of Iran, and flooding frequently leads to major urban, financial, anthropogenic, and environmental impacts in the area. Therefore, developing flood susceptibility maps to identify flood zones in the catchment is necessary for improved flood management and decision making The main objective of this study was to evaluate the performance of an Evidential Belief Function (EBF) model, both as an individual model *and* in combination with Logistic Regression (LR) methods, to prepare the flood susceptibility map in Haraz Catchment, Mazandaran Province, Iran The spatial database consists of flood inventory, altitude, slope angle plan curvature, Topographic Wetness Index (TWI), Stream Power Index (SPI), distance from river, rainfall, geology, land use, and Normalized Difference Vegetation Index (NDVI). After obtaining the required information from various sources, 151 of 211 recorded flooding points were used for model training and preparation of the flood susceptibility maps. For validation, the results of the model were compared to the 60 remaining flooding points. The Receiver Operating Characteristic curve was drawn and the

Area Under the Curve (AUC) was calculated to obtain accuracy of the flood susceptibility maps prepared by success rates (the training data) and prediction rates (validation data). The AUC results indicated that the EBF, EBF from LR, EBF-LR (enter), and EBF-LR (stepwise) success rates were 94.61%, 67.94%, 86.45%, and 56.31%, respectively, and the prediction rates were 94.55%, 66.41%, 83.19%, and 52.98%. The results showed that the EBF model had the highest accuracy in predicting the flood susceptibility map, in which 14% of the total areas were located in high and very high susceptibility classes and 62% were located in low and very low susceptibility classes. These results can be used for the planning and management of areas vulnerable to floods in order to prevent flood-induced damage; the results may also be useful for natural disaster assessment.

4.1 Introduction

A natural disaster is a major adverse event resulting from natural processes of the earth including floods, hurricanes, tornadoes, volcanic eruptions, earthquakes, tsunamis etc. A natural disaster can cause loss of life or property damage and typically leaves some economic damage in its wake, the severity of which depends on the affected population's resilience, or ability to recover and also on the infrastructure available. Among natural disasters, flooding is considered to be one of the most devastating (Youssef et al. 2011), and an accurate assessment of its risks is hampered by a lack of data and knowledge about flood losses at different scales (Grahn and Nyberg 2017). During the course of a heavy rainfall event, the amount of flow discharge in a river will increase rapidly and the water level will exude from its normal bed, covering the flood plain and the surrounding

areas (Chapi et al. 2017). Life-threatening water overflow in residential areas is a common occurrence in Iran after earthquakes (http://www3.irna.ir/fa/NewsPrint.aspx?ID=214943), and catastrophic flooding events happen annually in Mazandaran, Guilan, and Golestan Province in Northern Iran. In order to prevent loss of life and property during flooding events, and due to the high flood-frequency in this region, areas with a high risk of flooding must be recognized according to flood susceptibility maps (Bubeck et al. 2012) The primary difference between flood susceptibility and flood inundation maps is that flood susceptibility maps (FSM) only show the areas that have a high *potential* for flooding, but flood inundation maps can identify flood prone areas with flood depth.

On a global scale, floods are the most destructive national disasters and the cause of the highest number of deaths and damage; in fact, Opolot (2013) noted that almost 99 million people around the world were affected by floods between 2000 and 2008. In the most recent decade, the occurrence of repeated flood events in the northern part of Iran reached its historical maximum, and the rate and extent of damage increases every year Examples of recent floods in Mazandaran Province and its cities include: 1995, 2003, and 2012 in Noshahr City, 1999 in Neka City, 2013 in Behshar City, and 2013 in Sari City. The flood in Neka City caused more than 1,000,000 worth of damage to agricultural sectors such as wheat, barley and rice fields. The arable lands of about 28 villages with an area of 80 ha were also destroyed in Chahardangeh District, Sari; and 150 ha of agricultural lands were destroyed in Klijan Restagh in Sari City. Hence, the evaluation and identification of sensitive areas is essential in order to prevent and mitigate flood damages and losses (Bubeck et al. 2012).

Hydrologists have used several models to prepare flooding maps, but many of these models are data intensive or their calibration is difficult; however, some of the models are needed to understand of the physical processes within the catchment (Varoonchotikul 2003).

In recent years, a lot of statistical and probabilistic models have been tested to prepare flood susceptibility maps (Lee et al. 2012; Levy et al. 2007). Geographic Information System (GIS) has been used as an effective tool for spatial analysis and data manipulation, because of its ability to handle large amounts of spatial data (Oh and Pradhan 2011); the combination of statistical and probabilistic models with Remote Sensing (RS) and GIS has been widely used by different researchers (Tien Bui et al. 2018b; Youssef et al. 2011)Also, some scientists and researchers have studied natural disasters, specifically FSM, with the help of RS and GIS, using different models such as Decision-Tree (DT) (Khosravi et al. 2018b; Tehrany et al. 2013), Support Vector Machine (SVM) (Tehrany et al. 2015a; Tehrany et al. 2015b), Frequency Ratio (FR) (Khosravi et al. 2016a; Rahmati et al. 2016a), Evidential Belief Function (EBF) (Althuwaynee et al. 2012; Nampak et al. 2014; Tien Bui et al. 2018c), EBF-AHP (Analytical Hierarchy Process) (Althuwaynee et al. 2014), Logistic Regression (LR) (Pradhan 2010), Shannon's entropy and weights-of-evidence (Haghizadeh et al. 2017), Artificial Neural Networks (ANN) (Haghizadeh et al. 2017), AHP (Haghizadeh et al. 2017; Rahmati et al. 2016b), Random Forest (Chapi et al. 2017; Rahmati and Pourghasemi 2017), and Adaptive Neuro-Fuzzy Inference System (ANFIS) (Ahmadlou et al. 2018; Termeh et al. 2018).

Khosravi et al. (2018a) stated that (1) every model has some advantages and disadvantages, (2) model performance depends on the data, accuracy and model structure and (3) there isn't a universal guideline specifying which model should be applied in any given scenario,

therefore several models should be applied and the best of them used for further analysis. According to the literature, many machine learning and data mining algorithms have recently been applied in the field of natural hazards assessment, but there is no consensus among researchers with regard to which model is best. Some research shows that bivariate statistical models demonstrate better predictive power than both machine learning and data mining algorithms (Rahmati and Pourghasemi 2017); this is due to the fact that machine learning and data mining algorithms are more complex and require an expert to perform accurate simulations, thus, bivariate models, which are very simple to run with similar or sometimes better predictive power, can be used as adequate substitutes.

At present, the EBF method is rarely applied for flood analysis, but it has been used for other categories of natural disaster such as landslide susceptibility assessment (Althuwaynee et al. 2014; Jaafari et al. 2019; Pham et al. 2019), Land subsidence (Pradhan et al. 2014; Tien Bui et al. 2018d), and to predict groundwater potential zones (Chen et al. 2019a; Nampak et al. 2014).

The main purpose of this research is to generate a flood susceptibility map using the EBF method, as EBF has rarely been used for floods and has shown high accuracy in previous studies involving other natural hazard mapping. The results of this method are compared to EBF-LR (enter method), EBF-LR (stepwise method), and EBF from LR methods. In general, river flooding is a common natural disaster in the southern Caspian Sea, especially in Haraz catchment (Sadeghi-Pouya et al. 2017); the results of the current study will be useful for land-use planning and management for future flood mitigation in the Haraz Watershed.

4.2 Study Area

Haraz Catchment is located between longitudes of 51° 43´ to 52° 36´E, and the latitude of 35° 45´ to 36° 22´N. The study area lies south of Amol City, Mazandaran Province, Iran (Fig. 1). This catchment has an area of 4,014 km^2. The minimum and maximum elevations are 300 and 5,600 m, respectively. The average annual rainfall for 2006 and 2012 was 723.07 and 831.38 mm, respectively. The important residential centers of Haraz Catchment are Polur, Tashal, Tiran, Rineh, Kandovan, Abasak, Gaznak, Baladeh, and Noor. The main causes of flooding in these areas include: high-intensity rainfall over a short time-period, land-use changes in rangelands (e.g. deforestation or conversion of agricultural land and orchards to residential areas), and the lack of essential actions to prevent flooding. The maximum rainfall occurs in January, February, March, and October; October is the wettest month with an average rainfall of 160 mm.

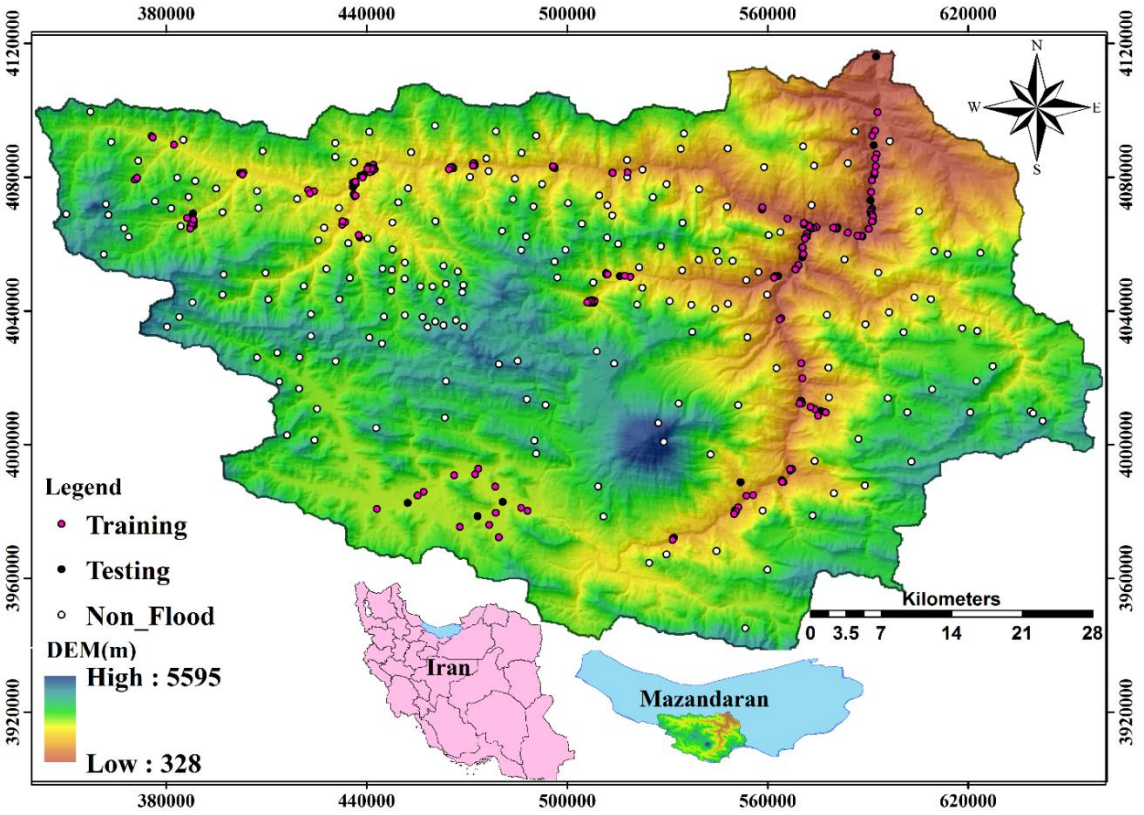


Figure 26. Flood location map with hill-shaded map of Haraz Catchment, Iran

4.3 Methodology

4.3.1 Data Used

4.3.1.1 Flood Inventory Map

In order to estimate the incidence of future floods in the study area, it is essential to analyze the occurrence of previous floods (Manandhar et al. 2010). The accuracy of historical flood data has a profound influence on the legitimacy of predictive flood-potential mapping (Merz et al. 2007). In this study, the flood inventory map was prepared based on the data from flood events occurring in years 2004, 2008, and 2012 (Mazandaran Regional Water Organization); this data was obtained using historical documentation analysis, aerial

photographs, and field surveys. For this study, 211 flooding locations and 211 non-flooding locations were selected. A ‘Non-flooding location’ refers to an area with relatively high elevation (e.g. hills and mountains) that is not affected by flooding; these points were identified using a topographic map and Google Earth (Tehrany et al. 2014b). The flood inventory map was then divided into two parts for training (151 flood locations, 70% of sampled land area) and validation (60 flood locations, 30% of sampled land area) (Pourghasemi and Beheshtirad 2015). Selected photos of floods in the northern part of the country are shown in figs. 2(a-c).

Figure 27. Flood in Surkh Rod and Mahmoud Abad, on April 12, 2015 (a), flood in Neka on September 12, 2012 (b), flood in Behshahr on September 16, 2013 (c).

4.3.1.2 Multi-Collinearity Diagnosis

After determining the flood conditioning factors, it is important to consider potential problems associated with multi-collinearity between independent variables (Pourghasemi and Beheshtirad 2015). If multi-collinearity in the regression equation is high, it means that

there is a high correlation between independent variables (inter-dependent) and, despite a high R^2, the model will not have a high validation accuracy. The exponent of the linear relationship between independent variables of the model is measured by an index that is called tolerance, which is between 0 and 1. Variance Inflation Factor (VIF) and tolerance are the two most important indicators to consider when attempting to detect multi-collinearity (O'brien 2007). Tolerance less than 0.2 and / or VIF greater than 5 indicate multi-collinearity problems (O'brien 2007).

4.3.1.3 Flood Conditioning Factors

In susceptibility mapping for floods or other natural disasters, it is necessary to define a series of conditioning factors (Shafizadeh-Moghadam et al. 2018; Yalcin et al. 2011). In this study, ten conditioning factors were applied; these conditioning factors are: altitude, slope angle, plan curvature, TWI, SPI, distance from river, rainfall, geology, land-use, and Normalized Difference Vegetation Index (NDVI). These data were collected from various sources and then converted into 20 m*20 m cell sizes based on a Digital Elevation Model (DEM).

4.3.1.4 Topographic Factors

A DEM with spatial resolution of 20 m*20 m was used to provide primary and secondary features such as altitude (Fig. 3a), slope angle (Fig. 3b), and plan curvature (Fig. 3c) in ArcGIS 10.1. Altitude, slope angle, and plan curvature factors were classified into nine, five, and three classes, respectively (Althuwaynee et al. 2012), based on a natural break classification scheme (Bednarik et al. 2010). Flat areas with low slope and low altitude classes have a higher potential for flooding.

4.3.1.5 Water Related Factors

Factors such as topographic wetness index (TWI) and stream power index (SPI) were prepared using DEM in the SAGA-GIS 2.8 software. TWI is the accumulation of flow at any location in the catchment, with consideration for downstream flow trends due to gravity (Gokceoglu et al. 2005). Equation 1 was proposed by Moore *et al*. (1991) to calculate the TWI:

$$TWI = \ln(A_S / \tan\beta) \quad (1)$$

where A_S is the specific area of catchment in m^2/m of the catchment and β is the slope angle in degrees. The TWI map is shown in Fig. 3d. Fig. 3e shows the SPI, which is the measurement of the erosive power of water flow, which is shown in Equation 2 (Moore et al. 1991):

$$SPI = (A_S \times \tan\beta) \quad (2)$$

To provide the layer of distance from the river, the digital map of the river was edited using the multi-ring buffer command in ArcGIS10.1; this layer was subsequently divided into six classes: 500, 1,000, 1,500, 2,000, 2,500, and > 2,500 m (Fig. 3f). Distance from river (or distance of measurement points from the river) has a major role in the distribution and magnitude of floods in the area (Glenn et al. 2012). In the northern part of Iran – as a result of insufficient infiltration and percolation due to changes in soil characteristics, vegetation coverage, and ground surface slope – these high-intensity rainfall events generate large amounts of runoff in the vicinity of the nearby river, causing catastrophic flood events in the areas downstream with lower topographic gradients (Kia et al. 2012).

Twenty years of rainfall data (1991-2011) from 17 stations inside and outside of the area was used to generate an annual rainfall map. Several interpolation methods such as kriging (simple and ordinary), inverse distance weighting (IDW) with power of 1 to 5, radial function with kernel functions of completely regularized spline, and spline with tension were compared to find the best method for mapping the rainfall data. The simple kriging method was selected as the best method because it produced the lowest Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) (Khosravi et al. 2016a,b). The rainfall map of the study area was ultimately divided into nine classes (Tehrany et al. 2014a) (Fig. 3g).

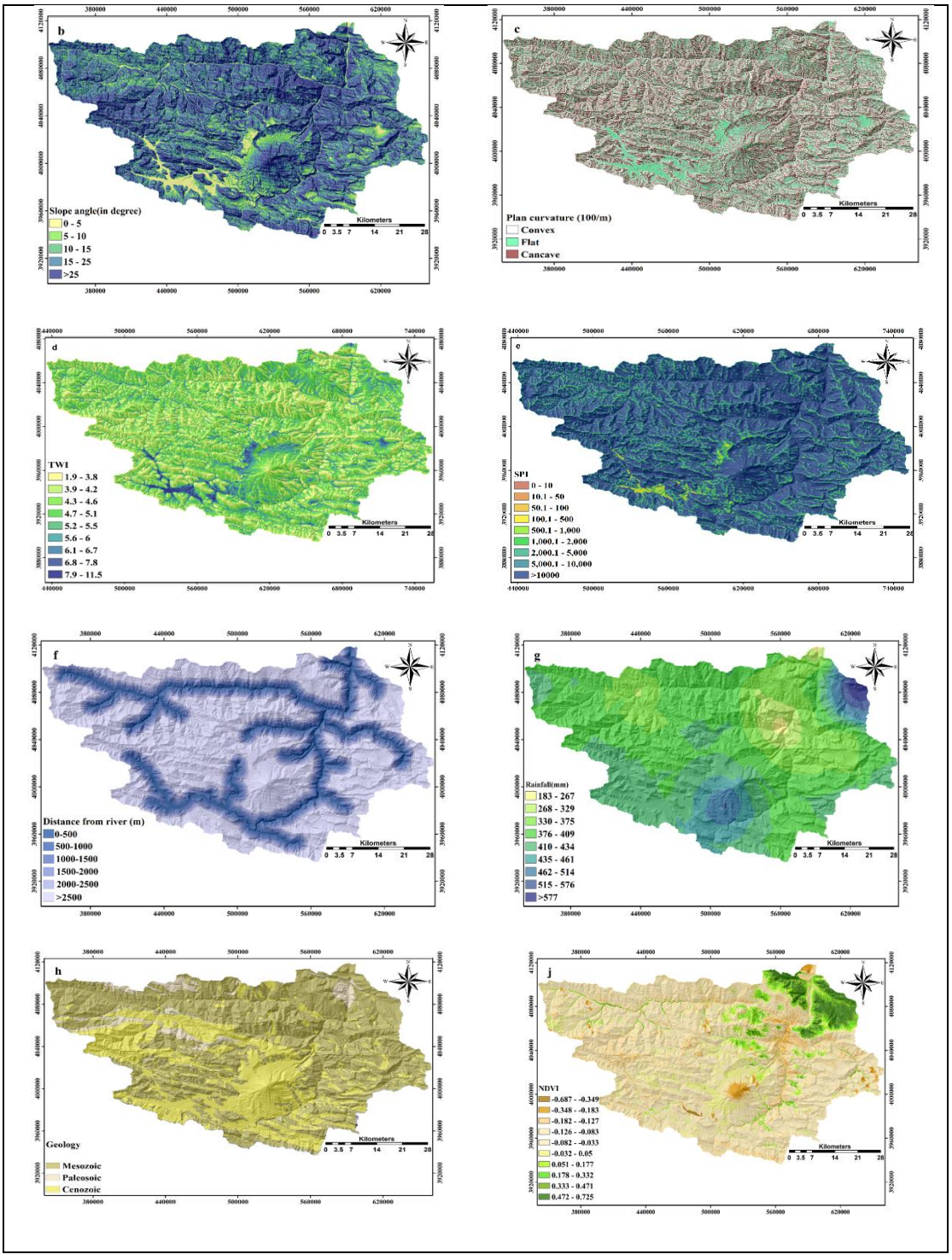


Figure 28. Flood conditioning factors of the study area: a. Altitude, b. Slope angle (In degree) , c. plan curvature, d. topographic wetness index (TWI), e. Stream power index (SPI), f. distance from river, g. rainfall, h. geology, i. land-use, j. NDVI

4.3.1.6 Physical Factors

In general, geology plays a major role in flood susceptibility due to the variable sensitivity of lithological units. Areas with hard and resistant rocks as well as highly-permeable soils have low channel densities (Çelik et al. 2012; Srivastava et al. 2014). The geology layer in GIS shape file format was obtained from the Mazandaran Regional Water Organization, and was originally prepared by the Geological Survey Department in Iran. The geologic composition of Haraz Catchment consists of 38.85% Cenozoic, 56.41% Mesozoic, and 4.73% Paleozoic Era. Finally, the geological map of the study area was classified into three groups (Bui et al. 2018; Khosravi et al. 2016a,b) and is presented in Fig. 3h and Table 1.

Table 15. Lithology of the Haraz Catchment.

Lithology	Formation	Code	Geological age (Era)
Undifferentiated Quaternary deposits	-	Q	
Loose alluvium in the river channels	-	Qal	
Rock fall	-	Qf	
Landslide	-	Qlan	
Moraines(glacial deposits)	-	Qm	
Scree, talus fans	-	Qs	
Scree	-	Qsc	
Marine conglomeratic terraces	-	Qt	
Old terraces	-	Qt1	
Younger gravel fans and terraces	-	Qt2	
Loess	-	Qt2c	
Sinter deposits	-	Qtr	
Basalt and olivine	-	Qv1	
Agglomerate and andesite tuff	-	Qv2	
Damavand magma, Andesite trachyte	-	Qv3	Cenozoic
Quartz-Diorite	-	Qd	
Rhyolite	-	Rh	
Trachyte	Eilka	TR	
Dacite	-	da	
Marl	Mila	E-m	
Pyroclastic and andesite	Karaj	EA	
Conglomerate.	Fajan	Ef	
Gypsum	Kand	Eg	
Tuff, volcanic rock	Karaj	Ek	
Nummulitic limestone	Ziarat	Ez	
Dacite, Pyroclastic with	-	Ezt2	
Calcareous and siliceous shales	Karaj	Eksl	
Middle tuff member:pyroclastics	Karaj	Ekt2	
Trachyandesite, trachybasalt, basanite	Karaj	Evkt2	
Santonian limestone	-	K2l	
Limestone, marl, limestone, silty marl	-	K2plm	
Cast locally with the layers of dolomitic limestone and shale	-	Kgl	
Limestone and marl	-	Kl-m	
Orbitolina Limestone	Tiz Kuh	Klt	
Basalt	-	Kv	
Basalt, diabase, pyroclastic rocks	-	Kv1	Mesozoic
And pyroclastic volcanic rocks are not separated	-	Kv1 2	
Tuff, basaltic andesite pyroclastic	-	Kv2	
Cut from multiple sources, Paland formation	Paland	TRb	
A thin layer of cream limestone, shale, limestone, Elika	Elika	TRe1	
Thick dolomite to form massive, limestone	Elika	TRe2	
Mafic volcanic rocks	-	TRv3	
Gabbro	-	Tgb	
Rocks, marl, conglomerate, Paleocene	Fajan	Pecf	Paleozoic

Sandstone, shale, limestone, quartzite, make peace	Fajan	Pefvmc
Rocks, marl, conglomerate, Paleocene	Dorud	Pd
Limestone, marly and sandy shales	Nesen	Pn
Fusulina limestone, dolomitic limestone	Ruteh	Pr
Basic flow, Pyroclastics	-	Pv

The land-use and NDVI maps were prepared using satellite images of OLI related to Landsat 8 (taken from the National Geography Organization of Iran) for 2013. The land-use map was classified into seven categories, consisting of: rangeland, bare land, forest, garden, irrigated land, residential areas, and water-bodies; this was done using Neural Network Algorithm (ANN) and supervised classification in Environment for Visualizing Images (ENVI 5.1) software (Fig. 3i).

The NDVI map shows the surface coverage of live vegetation; the vegetation index exists as a proportional output between -1 and +1. The NDVI values (Figure 3j) were calculated using the following equation (Pradhan *et al*. 2010):

$$NDVI = (NIR - VIS)/(NIR + VIS) \quad 3$$

Where VIS and NIR are the spectral reflectance measurements acquired in the red and near-infrared bands of the electromagnetic spectrum, respectively.

All of the aforementioned flood conditioning factors were converted to a grid comprised of 20x20 m cells.

4.4 Application of Models

Flood susceptibility analysis is one of the most important studies undertaken in the field of river hydrology (Tehrany et al. 2014b; Tien Bui et al. 2018a); in this study, analysis was

carried out using EBF, EBF from LR, EBF-LR (Enter) and EBF-LR (Stepwise) methods. Preparation of a flood susceptibility map is composed of four main steps: 1) Collecting data and creating a spatial database for conditioning factors associated with flood occurrences, 2) determining the relationship between conditioning factors and flood locations, 3) preparing flood susceptibility maps using different algorithms, and 4) validating the results using success rate and prediction rate curves and visual interpretation. The methodology of this research is presented as a flowchart in Fig. 29.

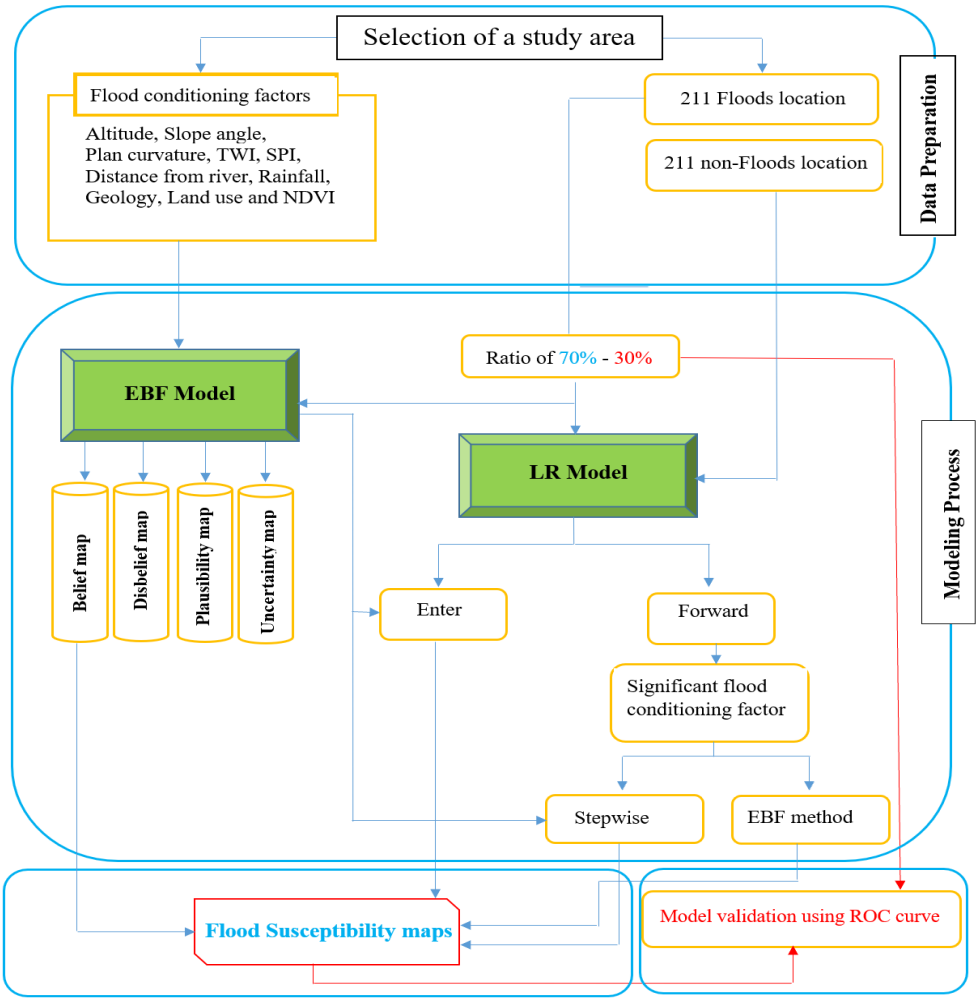


Figure 29. Methodological flow chart adopted in this research for Haraz Catchment.

4.4.1 Evidential Belief Function (EBF) Model

The EBF model is based on the Dempster-Shafer Theory of Evidence (Dempster 1969; Shafer 1976). To use this model, first the layer of the conditioning factors should be transformed into evidential data layers, then it can be integrated using the knowledge of the spatial relationship between the flood occurrences and factors influencing the flooding to generate a predictive Flood Susceptibility Index map (FSI). One of the advantages of this model is that both the predicted flood *and* flooding zone outputs exist within the same degree of uncertainty (Park 2011). The EBF model is composed of four functions, namely: Bel (degree of Belief), Dis (degree of Disbelief), Unc (degree of Uncertainty) and Pls (degree of Plausibility) (Althuwaynee et al. 2012; Carranza and Hale 2003). Four maps of Bel, Dis, Pls, and Unc were used for the assessment of ten factors influencing flooding. Each map shows the probability of flood occurrences. The integration of all the factors shows the level of prediction accuracy. EBF model was performed using the following steps (Chen et al. 2019b). Equations 5 and 6 show how to achieve results of Bel as below:

$$\lambda(Tp) = N / D = [N(L \quad Eij) / N(L)] / [N(Eij) - N(L \quad Eij) / (N(A) - N(L)] \tag{4}$$

$$Bel = \lambda(Tp)Eij / \sum \lambda(Tp)Eij) \tag{5}$$

where, N (L∩Eij) is the number of flood pixels in each class; N (L) is the total number of floods; N (Eij) is the number of pixels of each class; N (A) is the total pixels; N and D are the proportion of flood event areas and proportion of non-flood areas, respectively. Similarly, Dis values are also obtained by equations 7 and 8 as below:

$$\lambda(T\bar{p})Eij = K / H =](N(L) - N(L \quad Eij)) / N(L)] / [N(A) - N(L) - N(Eij) / (N(A) - N(L)] \tag{6}$$

$$Dis = \lambda(T\overline{p})Eij\,/(\sum \lambda(T\overline{p})Eij) \quad (7)$$

Where K is the proportion of flooding that does not occur, and H is the proportion of non-flooding occurrence areas in other attributes outside the class

Equations 5 and 6 were applied for all classes of factors influencing floods and equations 6 and 8 were used for Bel and Dis results. According to Dempster (1969) the main part of the theory is provided by Bel = Lower probability and Pls = upper probability; so, Pls is greater than or equal to Bel. Listed below are the three primary functions associated with the theory of Dempster-Shafer, according to Althuwaynee et al. (2012):

1- Basic probability assignment function (BPA);

2-Belief function (Bel);

3-Plausibility function (Pls)

Generally, the other parameters are calculated as follows:

Pls-Bel= Unc (8)

Dis=(1-Pls) or (1-Unc) (9)

Bel+Unc+Dis=1 (10)

4.4.2 Logistic Regression (LR) Model

Logistic regression is one of the Multivariate Statistical Analysis (MSA) models, which considers several parameters (such as slope, altitude, and plan curvature) that may affect the likelihood of flooding (Shahabi et al. 2014; Shirzadi et al. 2012). The purpose of using a LR model in this study was to determine the appropriate model for defining the relationship between the dependent variable and factors affecting floods, in order to

generate coefficients for each variable (Umar et al. 2014). Thus, in this study 70% of the flood training points and ten flood-conditioning factors were entered into the model as the dependent and independent variables, respectively (Hosmer and Lemeshow 2000). To run the LR model, Statistical Package for Social Science software (SPSS18) was used to carry out the Forward Conditional (stepwise) and Enter methods. Using the Forward method, variables were entered into the model individually and the results were evaluated based on the correlation between flood occurrence and individual variables; if there was no significant correlation, then they would be eliminated from the model (Papadopoulou-Vrynioti et al. 2013). The Forward Conditional calculations were based on the significant probability (Sig) in the LR model, with Sig <0.05 indicating that the flood conditioning factor or independent variable has a statistically significant impact on flooding (Tehrany et al. 2014a). Using the Enter method, all factors were entered into the model regardless of their correlation with flood occurrence. One of the advantages of LR models is that the data do not require a normal distribution and the conditioning factors can be continuous or discrete. In a quantitative way, the relationship between flood occurrence and its dependency on several independent variables can be computed as follows (Nampak et al. 2014):

$$P(event) = \frac{e^{Z}}{(1+e^{Z})} \tag{11}$$

where, *P* is the probability of an occurrence. As the value of Z changes from -∞ to + ∞, the probability of occurrence fluctuates between 0 and 1. Z is the linear combination function of affecting factors that indicate the linear relationship, and b_0 is the cutoff value of the model.

$$Z = b_0 + b_1 x_1 + b_2 x_2 + ... + b_n x_n \quad (12)$$

The general equation of a LR model is as follows:

$$Y = Logit(p) = \ln(\frac{p}{1-p}) = b_0 + b_1 x_1 + b_2 x_2 + ... + b_n x_n \quad (13)$$

Where Y is the probability of flooding, b_n (i = 0,1, ..., n) is estimated coefficients from sample data, n is the number of independent variables, and X_n (i = 0,1, ..., n) is independent variables. Using a LR model individually may be insufficient due to the fact that this model lacks the capacity to classify the weight for each factor (Tehrany et al. 2014a)); therefore, EBF and LR models were combined together to assess the probability of flooding events in the study area.

4.4.3 Validation of the Models

Validation of provided maps is an essential step in the development and identification of prone areas, as well as in the determination of map quality (Pourghasemi et al. 2012); without validated maps the models and their results will have no scientific significance (Chung and Fabbri 2003; Nampak et al. 2014). In this study, the receiver operating characteristics (ROC) curve was used for validation of the models. The area under the ROC curve (AUC) was used for the evaluation and quantitative comparisons.

4.5 Results

4.5.1 Multi-Collinearity Diagnosis

According to Table 3, the smallest tolerance and the largest VIF were 0.300 and 3.331, respectively; this indicates that there is no multi-collinearity between independent factors in the study area.

Table 16.Spatial relationship among conditioning factors and flooding occurrence extracted by EBF method

Factor	Class	No. of pixels in domain	Percentage of domain	No. of floods	Percentage of floods	Bel	Dis	Unc	Pls
Altitude(m)	328-350	112	0.02	0	0.00	0.000	0.108	0.892	0.892
	350-400	242	0.04	0	0.00	0.000	0.108	0.892	0.892
	400-450	466	0.08	2	1.32	0.255	0.107	0.638	0.893
	450-500	651	0.11	5	3.31	0.458	0.104	0.438	0.896
	500-1000	11230	1.98	46	30.46	0.243	0.077	0.680	0.923
	1000-2000	78535	13.82	43	28.48	0.032	0.090	0.878	0.910
	2000-3000	282516	49.70	55	36.42	0.012	0.136	0.852	0.864
	3000-4000	188954	33.24	0	0.00	0.000	0.162	0.838	0.838
	>4000	5722	1.01	0	0.00	0.000	0.109	0.891	0.891
Slope angle	0-5	21626	3.80	45	29.80	0.577	0.138	0.285	0.862
	5-10	43417	7.63	45	29.80	0.287	0.144	0.569	0.856
	10-15	73134	12.86	23	15.23	0.087	0.184	0.729	0.816
	15-25	203899	35.87	25	16.55	0.034	0.246	0.720	0.754
	>25	226352	39.82	13	8.60	0.016	0.288	0.696	0.712
Plan Curvature (100/m)	Convex	231228	40.67	23	15.23	0.116	0.469	0.414	0.531
	Flat	118098	20.77	44	29.13	0.436	0.294	0.271	0.706
	Concave	219102	38.54	84	55.62	0.448	0.237	0.315	0.763
TWI	1.8-2	72186	12.70	3	1.99	0.007	0.123	0.869	0.877
	2-3	143920	25.32	4	2.65	0.005	0.143	0.852	0.857
	3-4	135369	23.81	9	5.96	0.012	0.136	0.853	0.864
	4-5	95640	16.83	28	18.54	0.051	0.108	0.841	0.892
	5-6	60717	10.68	40	26.49	0.115	0.090	0.794	0.910
	6-7	36029	6.34	39	25.83	0.189	0.087	0.724	0.913
	7-8	17387	3.06	18	11.92	0.181	0.100	0.719	0.900
	8-10	4945	0.87	8	5.30	0.283	0.105	0.612	0.895
	10-12	2235	0.39	2	1.32	0.156	0.109	0.735	0.891
SPI	o-10	24	0.00	0	0.00	0.000	0.115	0.885	0.885
	1,50	84	0.01	0	0.00	0.000	0.115	0.885	0.885
	50-100	293	0.05	0	0.00	0.000	0.115	0.885	0.885
	100-500	3330	0.59	0	0.00	0.000	0.116	0.884	0.884
	500-1000	5494	0.97	3	1.99	0.462	0.114	0.424	0.886
	1000-2000	15970	2.81	1	0.66	0.053	0.118	0.829	0.882
	2000-5000	63050	11.09	8	5.30	0.107	0.123	0.770	0.877
	5000-10000	101088	17.78	11	7.28	0.092	0.130	0.778	0.870
	> 10000	379095	66.69	128	84.77	0.286	0.053	0.662	0.947
Distance from river (m)	0-500	53685	9.44	128	84.76	0.869	0.025	0.106	0.975
	500-1000	50645	8.90	9	5.96	0.065	0.156	0.779	0.844
	1000-1500	49097	8.63	5	3.31	0.037	0.160	0.803	0.840
	1500-2000	47719	8.39	2	1.32	0.015	0.163	0.822	0.837
	2000-2500	48645	8.55	1	0.66	0.007	0.164	0.828	0.836
	>2500	318637	56.05	6	3.97	0.007	0.331	0.662	0.669

Rainfall (mm)	183-267	2193	0.39	4	2.65	0.339	0.108	0.553	0.892
	267-329	9739	1.71	26	17.22	0.497	0.093	0.410	0.907
	329-375	91375	16.08	35	23.18	0.071	0.101	0.828	0.899
	375-409	190890	33.58	62	41.06	0.060	0.098	0.842	0.902
	409-434	170953	30.07	19	12.58	0.021	0.138	0.841	0.862
	434-468	72862	12.82	5	3.31	0.013	0.123	0.865	0.877
	468-514	11300	1.99	0	0.00	0.000	0.113	0.887	0.887
	514-576	12721	2.24	0	0.00	0.000	0.113	0.887	0.887
	>576	6395	1.13	0	0.00	0.000	0.112	0.888	0.888
Lithology	Cenozoic	218690	38.47	55	36.42	0.200	0.173	0.627	0.827
	Mesozoic	317901	55.93	87	57.62	0.218	0.161	0.621	0.839
	Paleozoic	26641	4.69	9	5.96	0.269	0.165	0.566	0.835
Landuse	Rangeland	52	2.40	132	87.42	0.016	0.210	0.773	0.790
	Bare land	526843	92.68	9	5.96	0.144	0.116	0.740	0.884
	forest	4043	0.71	5	3.31	0.010	0.126	0.865	0.874
	orchard	33421	5.88	0	0.00	0.138	0.122	0.739	0.878
	residential	759	0.13	1	0.66	0.049	0.122	0.829	0.878
	irrigation	1312	0.23	0	0.00	0.000	0.123	0.877	0.877
	water body	2004	0.35	4	2.65	0.664	0.119	0.216	0.881
NDVI	-0.69- -0.35	5057	0.89	5	3.31	0.199	0.096	0.705	0.904
	-0.34- -0.18	11625	2.05	7	4.64	0.121	0.096	0.783	0.904
	-0.17- -0.13	28452	5.01	17	11.26	0.120	0.092	0.788	0.908
	-0.12- -0.08	188356	33.14	15	9.93	0.016	0.133	0.851	0.867
	-0.07- -0.03	177441	31.22	32	21.19	0.036	0.113	0.851	0.887
	-0.02-0.05	27652	4.86	22	14.57	0.160	0.089	0.751	0.911
	0.06-0.18	26985	4.75	15	9.93	0.112	0.093	0.795	0.907
	0.19-0.33	28897	5.08	20	13.25	0.139	0.090	0.770	0.910
	0.34-0.47	38542	6.78	16	10.60	0.084	0.095	0.822	0.905
	0.48-0.73	35421	6.23	2	1.32	0.011	0.104	0.885	0.896

Table 17.The B coefficients and multi-collinearity diagnosis index for independent variables in the LR model

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	SE	Beta			Tolerance	VIF
Constant	0.130	0.191		0.679	0.489		
Altitude (m)	0.00	0.00	-0.300	-6.030	0.00	0.300	3.331
Slope angle	-0.005	0.002	-0.125	-2.731	0.007	0.353	2.833
Plan curvature (100/m)	-0.074	0.028	-0.082	-2.691	0.008	0.806	1.240
TWI	0.191	0.026	0.346	7.324	0.00	0.332	3.016
SPI	-5.612E-8	0.00	-0.025	-0.799	0.425	0.774	1.293
Distance from river (m)	0.00	0.00	-0.321	-7.228	0.00	0.377	2.656
Rainfall (mm)	0.00	0.00	0.023	0.716	0.474	0.714	1.401
Lithology	0.002	0.01	0.047	1.612	0.108	0.876	1.142
Landuse	-0.004	0.012	-0.011	-0.349	0.727	0.784	1.275
NDVI	0.105	0.086	0.038	1.218	0.224	0.779	1.284

4.5.2 Flood Susceptibility Mapping Using EBF Model

The EBF Model was used to determine the level of correlation between the flood occurrences and floods factors (Table 2). The results of this model show the impact of the class's weight of each factor on a flood event.

Analysis of elevation illustated that flooding occurs at lower elevations as in the classes of 3,000-4,000 m and > 4,000 m, no flood occurred. The first two classes on the slope angle map have the highest number of flood points (most likely to experience flooding), i.e. areas with low slope and altitude are more susceptible to flooding. Results of the plan curvature analysis showed that the greatest impact was represented by the concave curvature, followed by the flat curve, which represented high impacts, and finally that of the convex curvature which represented a small impact.

The TWI factor has a positive correlation with flood occurrence, which indicates a higher probability of flooding. The effect of SPI classes on the probability of flooding was also studied; with an increase in the SPI value, the number of recorded points increases and thus the possibility of flooding also increases. The highest flooding probability is related to the class 500 to 1,000. Distance from the river is one of the most important factors in mapping areas vulnerable to flooding (Tehrany et al. 2013). Results of distance from the river and flooding data showed that areas farther away from the river have lower risk of flooding; in fact, 85% of the flood events were placed in the first class or 0-500 m (Bel value equals to 0.869). According to the EBF method, the highest possibility of a flood occurrence is related to the class of 267-309 mm of rainfall. The geology of the study catchment consists of three classes of Mesozoic, Cenozoic, and Paleozoic Era. Almost 57%, 37% and 6% of floods were occurred in Mesozoic, Cenozoic and Paleozoic. With regard to the relationship

between the land-use map and the position of recorded flood points, it was obvious that 87.5% of flooding points were placed in rangeland areas and there is no recorded flood point in the orchard and irrigation land-use types. In terms of NDVI, the results indicate that there was no correlation observed between the EBF values and NDVI map classes. The highest and lowest probabilities of flood occurrence are related to NDVI ranging from -0.69 to -0.35 and 0.48 to 0.73 respectively.

Maps related to the four parameters of the EBF model are shown in Fig. 5. The Bel map (Fig. 5a) shows the distribution of flood occurrences. The Bel map was compared to the Dis map (Figure 5b), which showed that areas with a higher Bel value had a lower Dis and vice versa. This indicates that areas with higher susceptibility to flooding are characterized by a high degree of belief and a low degree of disbelief for each event. According to Pourghasemi and Beheshtirad (2014), detailed information cannot be extracted from the Unc map (Fig. 5c). Areas with a higher Bel and higher susceptibility to flooding were represented by the lowest values of Unc. The Pls map was almost the same as the Bel, the difference was that the contrast between the low and high degrees was less obvious than the Bel map (Fig. 5d).

The Flood Susceptibility Map (FSM) was implemented using the EBF model (Fig. 6-a) and the following equation:

$$FSM = (\text{altitude}_{\text{Bel}}) + (\text{slope}_{Bel}) + (curvature_{Bel}) + (\text{TWI}_{Bel}) + (\text{SPI}_{Bel}) + (\text{distance from river}_{Bel}) + (\text{rainfall}_{Bel}) + (\text{lithology}_{Bel}) + (\text{landuse}_{Bel}) + (\text{NDVI}_{Bel}) \quad (14)$$

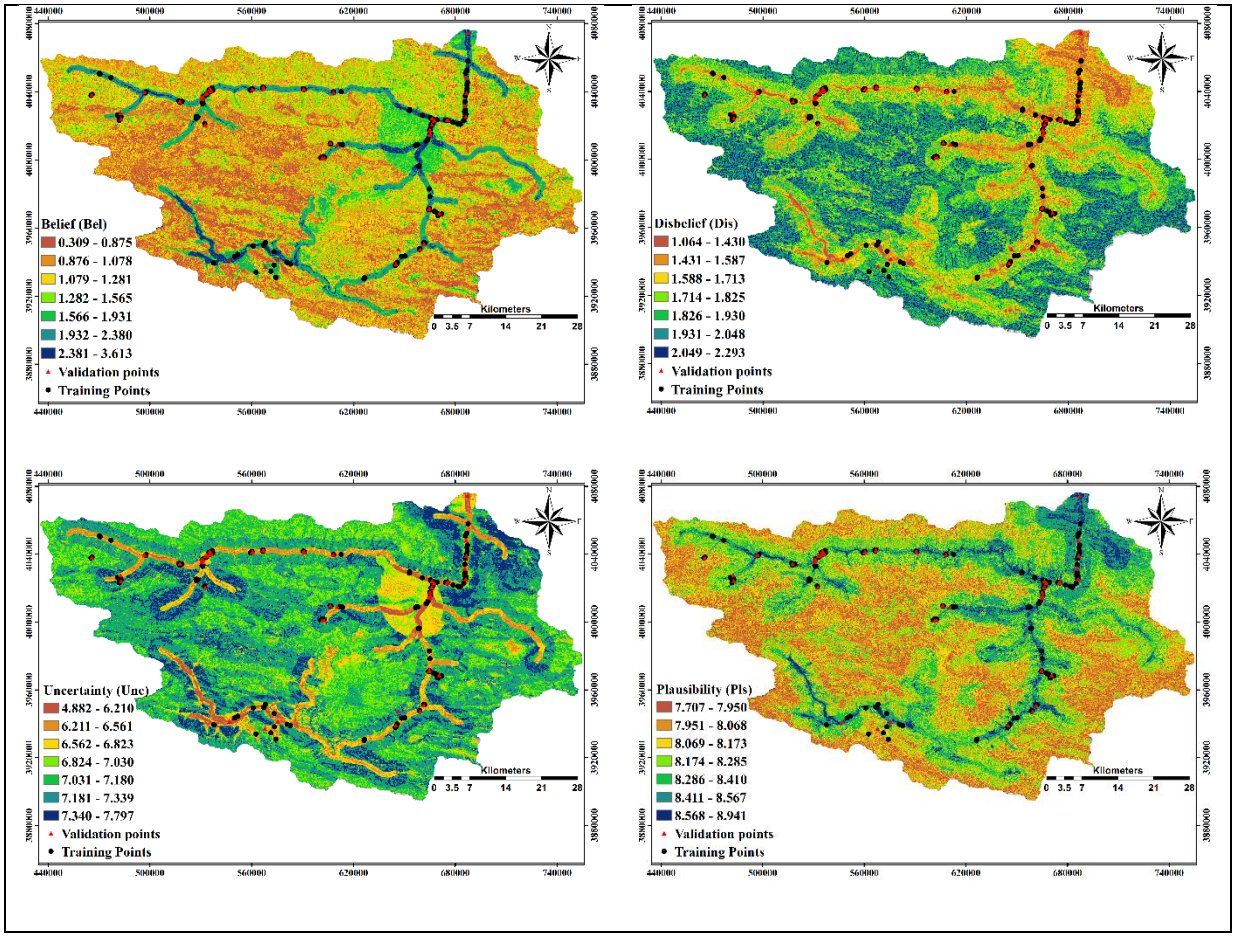


Figure 30. Integrated results of EBF model: (a) belief, (b) disbelief, (c) uncertainty, (d) plausibility

4.5.3 Flood Susceptibility Mapping Using LR Model

In this study, two methods of "enter" and "stepwise" have been used in a LR model. In the enter method, ten conditioning factors of altitude, slope angle, plan curvature, TWI, SPI, distance from river, rainfall, lithology, land-use, and NDVI were assumed with coefficients of -0.004, -0.162, -0.638, 3.32, 0.00, 0.002, -0.003, 0.10, -0.362 and -0.453, respectively. In the stepwise method, the conditioning factors that have a significant impact on flooding

were selected; these factors are: slope angle, distance from river, altitude, and TWI with coefficients of -0.141, -0.002, -0.004 and 3.402 and degrees of significance of 0.021, 0.00, 0.00 and 0.00, respectively (Table 4). Negative weight values for LR coefficients indicate that the flooding events are negatively related to the independent variables (Tehrany et al. 2014a). After running the LR model and obtaining the coefficients, each coefficient was multiplied by the corresponding conditioning factor in order to find the z-value, as shown below:

$$Z(enter) = (-0.004\text{altitude}) + (-0.162\text{slope}) + (-0.638\text{curvature}) + (3.32\text{TWI}) + (0\text{SPI}) + (0.002\text{distance from river}) + (-0.003\text{rainfall}) + (0.1\text{lithology}) + (-0.362\text{landuse}) + (-0.453\text{NDVI}) + 1.02 \quad (15)$$

$$Z(stepwise) = (-0.004\text{altitude}) + (-0.141\text{slope}) + (3.402\text{TWI}) + (-0.002\text{distance from river}) - 2.19 \quad (16)$$

Table 18. Conditioning factors coefficients of LR method

Method	Factors	B
Enter	Altitude (m)	-0.004
	Slope angle	-0.162
	Plan curvature (100/m)	-0.638
	TWI	3.320
	SPI	0.00
	Distance from river(m)	0.002
	Rainfall (mm)	-0.003
	Lithology	0.10
	Landuse	-0.362
	NDVI	-0.453
	Constant	1.025
Forward (Stepwise)	Altitude (m)	-0.004
	Slope	-0.141
	TWI	3.402
	Distance from river (m)	-0.002
	Constant	-2.196

4.5.4 Flood Susceptibility Mapping Using EBF-LR And EBF From LR Model

The results of the LR model (stepwise) showed that slope, distance from the river, altitude and TWI (with significance values equal to 0.021, 0.00, 0.00 and 0.00, respectively) are the most important factors Influencing flood occurrence. The difference in -2log likelihood (-2LL) was considered as the indicator of effectively improving on the null model (Table 5). Nagelkerke's *and* Cox/Snell's R-squared tests were used to measure the model's efficacy. A better model has a higher Nagelkerke's and Cox / Snell's R-square (Althuwaynee et al. 2014). Finally, equations 21 and 22 for the EBF-LR (enter and stepwise) hybrid model were entered into the raster calculator of ArcGIS10.1 and implemented as follows (Figs. 6 (b-c)):

$$Z(EBF-LR.Enter) = (-0.004\text{altitude}_{\text{EBF}}) + (-0.162\text{slope}_{EBFl}) + (-0.63curvature_{EBFl}) + (3.32\text{TWI}_{EBFl}) + (0\text{SPI}_{EBFl}) + (0.002\text{distance from river}_{EBFl}) + (-0.003\text{rainfall}_{EBFl}) + (0.1\text{lithology}_{EBFl}) + (-0.362\text{landuse}_{EBF}) + (-0.45\text{NDVI}_{EBFl}) + 1.025 \quad (17)$$

$$Z(EBF-LR.Stepwise) = (-0.004\text{altitude}_{\text{EBF}}) + (-0.141\text{slope}_{EBFl}) + (3.402\text{TWI}_{EBFl}) + (-0.002\text{distance from river}_{EBFl}) - 2.196 \quad (18)$$

$$FSM(EBF-LR.enter) = P = \frac{e^{z\,(\text{EBF-LR}.enter)}}{1+e^{z\,(\text{EBF-LR}.enter)}} \quad (19)$$

$$FSM(EBF-LR.stepwise) = P = \frac{e^{Z\,(\text{EBF-LR.stepwise})}}{1+e^{Z\,(\text{EBF-LR.stepwise})}} \quad (20)$$

Based on four factors affecting flooding obtained by the LR model, the EBF model was re-run and the final map was produced which is shown in Fig 6d. Parameters with no significant values were deleted from the EBF model.

$$FSM(\text{EBF from LR}) = (\text{altitude}_{\text{Bel}}) + (\text{slope}_{\text{Bel}}) + (\text{distance from river}_{\text{Bel}}) + (\text{TWI}_{\text{Bel}}) \quad (21)$$

4.5.5 Validation of The Flood Susceptibility Maps and Their Comparison

In this study, 70% of the chosen flood locations were used for model training and the remaining 30% were used for model validation. Using success-rate and prediction-rate curves, the results of flood susceptibility maps were compared with existent flood positions (Chung and Fabbri 2003) and the accuracy of four FSMs was evaluated. Fig. 7(a-b) shows the success-rate and prediction rate curves for four FSMs. As it can be seen, the EBF model has the largest AUC (94.61%) and the EBF-LR (stepwise) model has the lowest AUC (56.31%). Due to the fact that the training points of the model were used for the success-rate, this approach may not be an appropriate method for assessing the predictive ability of the model (Bui et al. 2012a; Nampak et al. 2014; Pradhan 2013). AUC for prediction rate indicates how well the model predicts the flood (Brenning 2005; Bui et al. 2011; Chung and Fabbri 2003). The prediction rate was evaluated using the 30% of points that were not used in the training model, i.e. the 30% used for model validation. This method shows the predictive capability of the model (Maier and Dandy 2000). The AUC of the prediction rate for models of EBF, EBF from LR, EBF-LR (Enter), and EBF-LR (Stepwise) were 94.55, 66.41, 83.19, and 52.98%, respectively. For example, the AUC of 0.9455, which is for the EBF model, shows a predicting accuracy of 94.55%; hence, the most accurate prediction of the flood susceptibility map was the EBF model. Conversely, the lowest predictive accuracy was related to the EBF-LR (Stepwise). Ultimately, the EBF, EBF from LR, and EBF-LR (Enter) models displayed acceptable accuracy in predicting the final map.

Table 19. LR model summary

Method	Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
Enter	1	42.604[a]	0.712	0.949
Forward (Stepwise)	1	198.454[a]	0.518	0.690
	2	83.501[b]	0.670	0.894
	3	53.481[c]	0.702	0.935
	4	47.401[c]	0.708	0.943

4.6. Comparison o of Different Models

Based on the results from the four implemented models, the final flood susceptibility maps were prepared and categorized into very-low, low, moderate, high, and very-high susceptibility classes (Fig. 8). The EBF model was chosen as a basis for model comparison because it showed the highest accuracy among each of the study models. According to the EBF model, the study areas' percentages of each susceptibility class were 23.7% (very-low), 38.7% (low), 23.04% (moderate), 9.88% (high), and 4.67% (very-high). The susceptibility class percentages according to the EBF-LR(Enter) were 43.67%, 30.68%, 18.17%, 7.28%, and 0.21%, for EBF-LR(Stepwise) were 38.93%, 22.85%, 26.68%, 10.38%, 1.17%, and for EBF from LR were 72.37%, 10.68%, 5%, 9.95% and 2%. The results show that the implemented models of EBF-LR (Enter), EBF-LR (Stepwise), and EBF from LR overestimated the area of the very-low class compared to results from the EBF model, but they *under*estimated the area of the low susceptibility class compared to the outputs from the EBF model. All four models estimated roughly equal area percentages for the high susceptibility class category, and the EBF-LR (Enter), EBF-LR (Stepwise), and EBF from LR models displayed lower values for the very-high susceptibility class when compared with the EBF model, although all four models exhibited comparatively small values for the very-high susceptibility class.

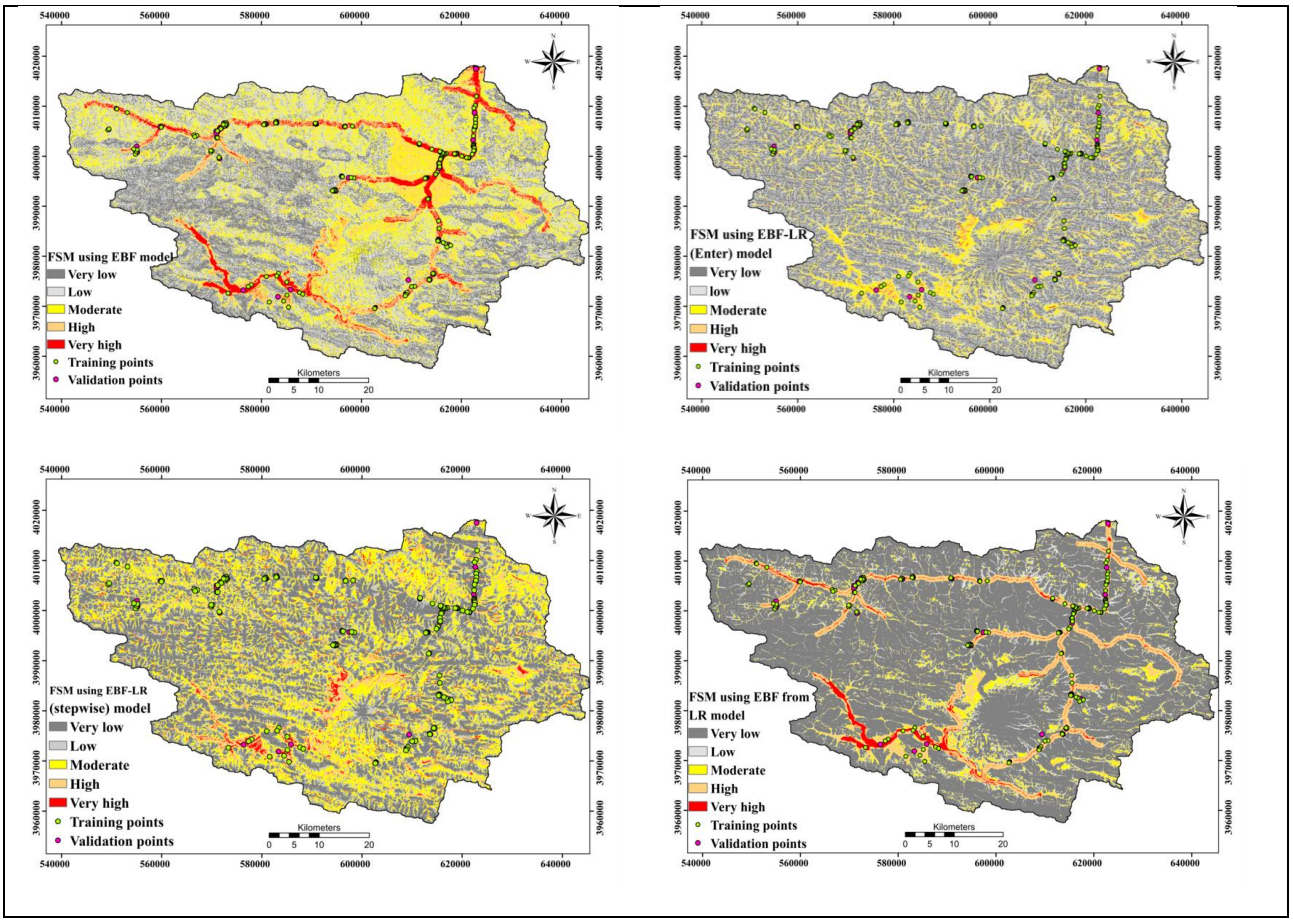


Figure 31. Flood Susceptibility Index using (a) EBF, (b) EBF-LR (Enter), (C) EBF-LR (Stepwise), (d) EBF from LR

4.6 Discussion

Analysis of elevation maps as well as the position of flood points revealed that flooding usually occurs at lower elevations. It also indicates that the frequency of recorded flooding events decreased from lower elevation classes towards higher classes as floods occurs in a flat, lower elevation and lower slope areas where water can get together. Also, as TWI shows the wetness, thus the areas with high TWI have a saturated soil, and therefore the potential of flood is higher for these areas. The river and its surrounding areas maintain higher flood susceptibility than any other region, and the river flood risk would be theoretically reduced by moving away from the river. Rainfall volume and EBF values

showed a decrease trend; this may be due to the fact that rainfall usually increases at higher altitudes although the risk of flooding significantly decreases. This results are according to the result of Khosravi et al (2016a,b; 2018), Tien Bui (2018a,b). In terms of land use villages close to Haraz River, agricultural areas, such as citrus orchards, and areas with a low topographic gradient are more susceptible to flooding. The results show that most points recorded in the rangeland are close to the registered residential area and gardens, therefore, a single flood event may turn to a fatal natural hazard, causing catastrophic financial damage as well as claiming human lives. Such devastation has been one of the leading causes of death and economic distress in Haraz Catchment, according to annual reports by authorities.

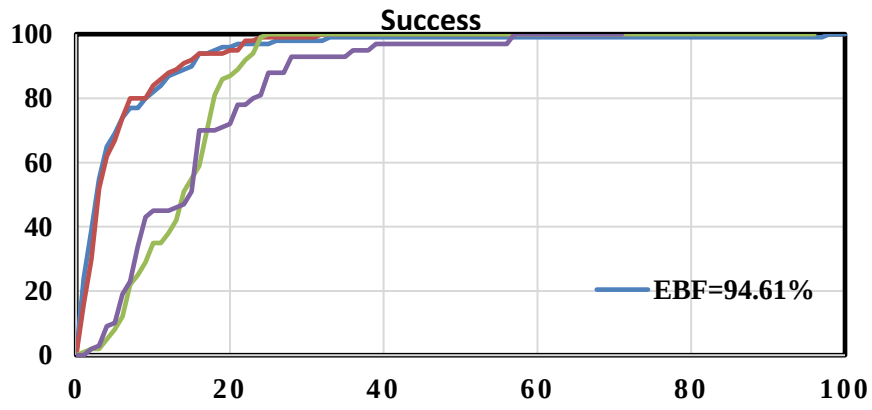
According to Park (2010), the main limitation of the EBF method is that if a flood event did not happen in a represented class, the Bel results would be equal to zero; in this case, the Dis would also be zero, and the case Unc or uncertainty values would be equal to 1. Wally (1987) stated that if the observations had complete information about the study area, we would expect that Pls-Bel is equal to zero (in this case Bel is called Bayesian belief function) (Walley 1987), confirming that the two maps of Bel and Pls should have the same results (Althuwaynee et al. 2012). The results of this study are consistent with Carranza and Hale (2002); Althuwaynee et al. (2014); Bui et al. (2012b); Carranza et al. (2008); Nampak et al. (2014). According to the results of Nampak et al. (2014), the main advantage of Dempster-Shafer theory is that the application of an EBF model not only provides the predictive maps of desired areas, but also provides the predictive degree of uncertainty.

Result showed that the greater number of inputs parameter can enhance the result of modeling, which is in according to result of Donati and Turrini (2002) explained that a

greater number of parameters would likely result in improved model accuracy, and in fact the high accuracy of the EBF model resulting from this study was in line with those of Nampak et al. (2014); Pourghasemi and Beheshtirad (2014) and rahmati and Pourghasemi (2017).

Results from the achieved maps showed that the areas nearest to the Haraz River where the slope is low, curvature is flat, altitude is low, and TWI is high, are completely susceptible to flooding, so, our findings are in agreement with Tehrany et al. (2013), Khosravi et al. (2016a,b;2018). Results from the current study could be eminently useful in the pursuit to prevent and protect from flooding hazards and damage. By avoiding the construction of homes, villas, or industries in susceptible zones, and by employing both structural and non-structural approaches for future flood mitigation, the enduring damage and devastation caused by floods can be greatly reduced.

Recommendations for future research include a comprehensive study involving an assessment of the accuracy and simplicity of assorted bivariate, multivariate, machine learning, data mining, and multi criteria decision making models and also their coupled for improved flood predictive power.



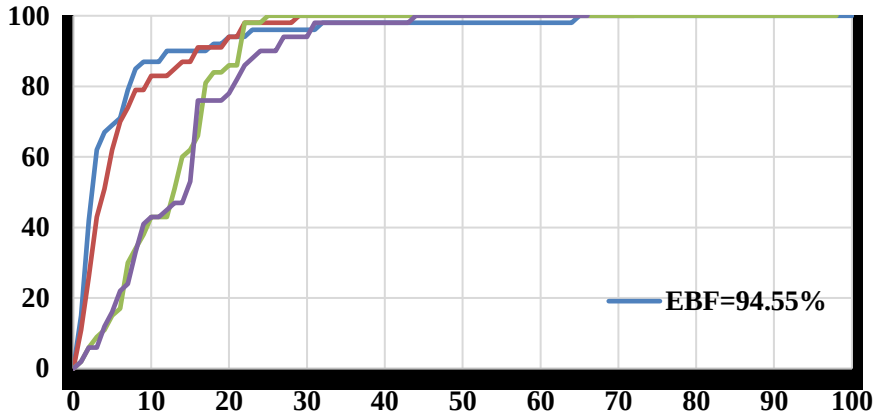


Figure 32.The success and prediction rate curves for flooding map; (a) success rate and (b) prediction rate.

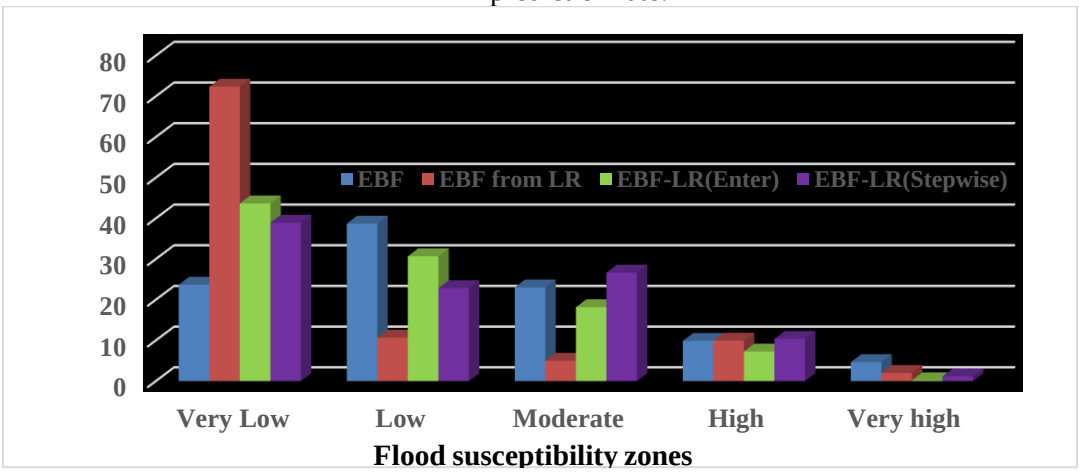


Figure 33. A histogram showing the percentage of flood zones that fall into the various classes of four models

4.7 Conclusion

The main goal of this study was to assess the performance of an EBF model that is rarely used for the development of flood susceptibility maps. In order to determine the accuracy of the EBF model individually *and* in combination with a LR model, the two models were combined. The LR model was implemented using independent variables that were weighted and reclassified by the EBF model. Based on the coefficients obtained from the

LR model, the TWI parameter had the highest weight and impact on flooding. The relationship between factors affecting flooding and the final maps generated by the models indicated that most flooding events occur in areas where the topography is more flattened. The flooding maps resulting from the current study illustrate that most flooding in Haraz Catchment occurs in the areas directly adjacent to the river, which are most often characterized by low slope, concave curvature, or flat conformation. Because of the steep, mountainous areas in Haraz Catchment, frequent rapid runoffs occur and water flows down toward the Haraz River causing flash flooding in areas where the topography permits. Most residential and agricultural territories within this catchment are located in areas with low slope and flat constitutions, which are susceptible to flooding.

According to the success rate and prediction rate curves, the highest accuracy belonged to the EBF model, which exhibited 94.61% for success rate and 94.5% for prediction rate. The developed EBF model had an accuracy of 66.41% based on significant parameters from the LR model (EBF from LR), and the difference between the two models was 28.5%. Based on the expressed curves, the accuracy of the EBF-LR Enter and Stepwise models were 83.19% and 52.98%, respectively; the difference between these two models was 30.21%. The lowest accuracy belonged to the EBF-LR (Stepwise) model, which represented the weakness or failure of this method to determine the flood prone areas. These results suggest that considering significant parameters such as altitude, slope, TWI, and distance from the river alone in flooding occurrence, may not be enough to adequately validate the model. The occurrence of a natural hazard, like flooding, is very complex and cannot be predicted with high accuracy when limited parameters are being considered in model development. According to the flood susceptibility map resulting from the EBF

model, 14% of the total area was located in the high and very-high susceptibility classes while 62% of the area was located within the low and very low susceptibility classes. It's recommended that researchers and stakeholders identify flood-prone areas in additional catchments using the EBF model, which is simple with high accuracy .

The results of this research also indicate that the impact of factors' classes is more important and more effective on the natural hazards assessment and mapping than weights of layers. In general, the developed model can be used as a tool for decision making by management agencies such as the Department of Water and Natural Resources. The results and insights gained from this modeling effort can provide an improved understanding of flood susceptibility and will improve the process by which researchers identify areas prone to flood hazards. In addition, the results of such a model are beneficial and necessary for flood warning and preventing damage in future floods, specifically in the northern areas of Iran. For future study, it's recommended that researchers focus on the flood prone areas in order to identify best management practices and to implement structural and non-structural methods for potential damage reduction.

5 Comprehensive Conclusion

Hydrological models frequently comprise uncertainties with negative effects on the estimated results, thereby on the model accuracy. To attain more accurate model estimates, it is definitely better to evaluate and improve models using different approaches like parameter optimization, operative management, and design space exploration. Developing a model that is compatible with the nonadditive and nonlinear character of the hydrological process is very essential in hydrology. The most recent models do not recognize the spatiotemporal heterogeneity of hydrological phenomena's and they have lots of assumptions which consider hydrological process as a linear system.

In this context, the unique characteristic of the ensemble machine learning, and data mining algorithm can overcome this downside. Because these methods can perform:

1- Missing data recovery and data cleaning

2- Feature importance analysis, Sensitivity analysis and uncertainty analysis

3- Highly nonlinear prediction

4- High accurate forecasting

All proposed machine learning algorithm in this study are low bias and low variance as they have highly accurate predictive performance and they have been evaluated on different preprocessed data set. These models can consider all possible interactions between the input parameters which account for nonlinear behavior of any hydrological phenomena like streamflow and flood. More importantly, the robustness of these model can contribute to bias estimates of water availability and uncertainty in forecasting of potential future climate changes.

In this research, two new ensemble machines learning and data mining algorithms have been proposed: NESN-MP and RNESN coupled with twelve different ensemble data mining algorithms including: PCA, TF, TF-K-means clustering for data cleaning and KNN, Bootstrap aggregating, SVM and RandF for feature selection have been investigated. These models were tested in three different regions across US: Berkshire County (MA), Tuolumne County (CA), and Wasco County (OR). The method was tested with data provided by the United States Geological Survey (USGS), Natural Resource Conservation Service (NRCS), National Weather Service Climate Prediction Center (NOAA) and Daymet Data Set from NASA through the Earth Science Data and Information System (ESDIS). Each data set includes the daily records of the local observed hydrological and large-scale weather/climate variability parameters.

The propose RENESN and NESN-MP streamflow modeling approach can incorporate a clustered large data entry as input. This characteristic will decrease the prediction uncertainty related to input parameters and it will increase the model reliability as it is capable of applying all contributing factor in governing streamflow. So generally potential advantage of the model can be recognized as fallows;

1- The accurate forecasting of meteorological anomalies like precipitation which has large uncertainty in the measurement process.
2- The quantification of existing natural reservoir of snowpack; as the result agreed in the western United States, it is no surprise that precipitation is winter dominant and a large portion (by some accounts more than 70%) of streamflow there originates from melting snow and thus streamflow is largest during April–June. This will

make western water resources managers highly dependent on wintertime snow surveys to venture water availability in the coming spring and summer.

3- A third possible contributor of streamflow forecast is information about current state of soil moisture. We tested our result in different months of the year in the western US to indicate relative importance of soil moisture initialization when other seasons are considered. In warmer season, even large fraction of the streamflow occurs during the other three seasons, the streamflow variability is noticeable, suggesting that the dominant impact of soil moisture initialization in these seasons is important. The possible reason might be dryness of the top soil which makes incident water at the surface, either snowmelt or precipitation, infiltrate and then evaporate back rather than run off into streams; a wet soil, on the other hand, may boost greater streamflow and a more efficient filling of reservoirs during the season.

4- A very interesting observation from the results of data mining algorithms is that TF-k-means clustering subsets clearly outperform the other created subsets, and also show acceptable classification accuracy of 96.58 % in combination with Random Forest classification algorithm. Among the machine learning algorithms investigated, the SVM accuracy was surprisingly low on the PCA subsets about 80.49%. Even though SVMs perform very well on TF-k-means clustering subsets of the original features and it has been commonly reported as an accurate feature importance methods, they achieve only the lowest accuracy for the PCA subsets. A possible reason might be because of the sensitivity of feature importance strategy on the type of performed dataset. In many cases, where the search for the best feature set is still an active research topic, the classification accuracy achieved with

cleaned data sets is often significantly better than with the full data set. In application contexts where feature sets are already well established, the differences between diverse dimensionality reduction strategies are much smaller.

5- The feature importance methods indicated that the most effective factor in governing streamflow is Precipitation fallowed by SWE, soil moisture , NAO, AO, soil temperature and so on. There is slightly different between feature importance percentage among three case studies and the only difference lies is climate variability indices. In case study 1 the most important climate variability indices are NAO, AO and PNA respectively, because this region locates in North East US and is Cleary affected by AO and NAO much more than the other case studies. There is not much difference in in case study 2 and 3, as they both are under effect of PNA and not much influenced by AO and NAO.

6- For the best feature subset selection the factors including: vapor pressure, solar radiation, wind amplitude and humidity has not been utilized for the forecasting phase. Because their contributing percentage is low and can be neglected as it does not have any significant impact on the forecasting accuracy.

7- The utilized deep learning algorithm (RNESN) for streamflow forecasting is also a high nonlinear approach in prediction of time series data as it considers all possible interactions between the internal states of inputs through performing different orders of multivariate polynomial. In the case of E_{NS}, the RNESN gives the respective value of 0.96, which is well below the E_{NS} of 0.84 and 0.47 for NESN-MP and ANFIS, respectively.

8- The performed training approach in RNESN, minimizes the direct impact of each input data on the output, otherwise stated, due to the structure of the RNESN, the sensitivity of each input parameter is almost the same as the others because the interactions of the input parameters and the internal states are used to train the output matrix and the input parameters are not directly involved in the process of training.

9- The proposed calibration process (RLS) enhance the learning capability through reducing variance. Besides, to improve the forecasting ability and reducing the computational load, the number of the internal states has been significantly decreased through applying modified time series analysis in readout.

10- The recommended method has a simple structure, far less computation, and does not need parameter tuning, optimization task, and complex training. Simulation results show the dominance of RNESN over NESN and ANFIS.

In this research, two new ensemble machine learning algorithms for flood susceptibility mapping has been introduced : EBF-LR a new intelligent hybrid model (BaggTree -Cubic KNN) for modeling flood. These models were tested at the Haraz watershed, the Mazandaran province, Iran. Ten flood conditioning factors, including slope angle, elevation, curvature, stream power index (SPI), topographic wetness index (TWI), land use, rainfall, drainage density and distance to the river were selected for flood modelling. The information gain ratio (IGR) was used to optimize the most important conditioning factors for the flood model.

The potential advantage of the model can be recognized as follows;

1- The result of IGR showed that all the factors were significant in the model training; however, distance to a river was the most important factor, followed by slope, curvature.

2- The most of floods in the study area taken place during the heavy rainfall and due to overbanking the rivers. This process leads to inundation the areas adjacent the rivers, flood plains, during the heavy rainfall where are distinguished in the prepared flood susceptibility maps.

3- There are several agricultural fields near the rivers whose their tillage practices is causing heavy sedimentation and making river less storage capacity for holding streamflow which cause easy over banking. Therefore, the water manger can use the result of the current study to modify the agricultural practices.

4- Because the distance to the river was the most effective factor on flooding in the studied area, the decision makers have to be more concerned about cropping practices and widening streams. Because when the discharge becomes too high, the stream widens its channel by overtopping its banks and flooding the low-lying areas surrounding the stream. This process are a major part of the erosional process, working in conjunction with weathering and mass wasting.

5- The Receiver Operating Characteristic curve was drawn and the Area Under the Curve (AUC) was calculated to obtain accuracy of the flood susceptibility maps prepared by success rates (the training data) and prediction rates (validation data). The AUC results indicated that the EBF, EBF from LR, EBF-LR (enter), and EBF-LR (stepwise) success rates were 94.61%, 67.94%, 86.45%, and 56.31%, respectively, and the prediction rates were 94.55%, 66.41%, 83.19%, and 52.98%.

The results showed that the EBF model had the highest accuracy in predicting the flood susceptibility map, in which 14% of the total areas were located in high and very high susceptibility classes and 62% were located in low and very low susceptibility classes.

The KNN as one of the most utilized neighborhood classifiers belongs to the weak classifiers in which they are very simple and highly efficient in some fields of studies (Hassanat, 2014). It is remarkable that memory requirement and time complexity are the main limitations of performance ability of KNN classifier and they are completely dependent upon every example in the training set (Hassanat et al., 2014).

In the context, potential advantage of this model can be recognized as fallows;

1- To enhance the performance of KNN, we used of Bootstrap aggregating Meta classifier. The combination of the Bootstrap aggregating ensemble technique and the KNN classifier can provide a solution to build a flood model. It has been shown that, the area under the receiver operating characteristic curve (AUC=0.800) value of the proposed BaggTree -CubicKNN model has the best performance.
2- The proposed model could significantly enhance the prediction accuracy of CubicKNN classifier as a base classifier. General speaking, the ensemble model can outperform the basic classifier because the ensemble model integrates the advantage of each classifier (Tien Bui et al. 2016; Shirzadi et al. 2018; Dou et al. 2019).
3- Additionally, this model (AUC=0.800) could greatly decrease the over-fitting and variance problems between the training datasets and eventually enhance the prediction accuracy of cubicKNN model (AUC=0.660).

4- The results of these research also indicate that the impact of factors' classes is more important and more effective on the natural hazards assessment and mapping than weights of layers.

5- The proposed ML model can enable decision-makers for a less-expensive early field survey of the district they are meant to cope with high reliability. Thus, the achievement of this research could assist the managers to recognize the flood-prone zones in watershed more accurately. After the assessment of the frequent susceptible areas, decision-managers could prepare more accurate and more useful data related to these regions (such as rainfall and river data) using proposed models to produce accurate flood maps for mitigating further damage.

In general, despite the alluring prospect of Machine learning forecasting methods, they are often criticized by hydrologists for the lack of physical hydrologic meanings and poor robustness. In this context, all three proposed modeling approach will work only in case of mainly operational purposes not with regards to the theoretical revelation of the hydrological process. Therefore, one must define the exact objective of the modeling task to get the right answer for the right reason. For instance, one might just need methods that predict stream flows, groundwater levels, flood susceptibility or water quality with enough accuracy for the task at hand. At this point, the main objective is developing an optimal accurate model with the correct input parameters. However, if a physical interpretation of the process is needed, then the parsimonious conceptual/ANN models would be preferred as a Neural Network black box nature will not explore interpretation of the results. This makes them a valued complement to physical models, mainly in data-scarce regions for

model parameterization and warrants additional research into their development and application.

Therefore, in any case our first aim was to examine the accurate predictability of daily streamflow and flood susceptibility and then improving the performance of operational streamflow and flood forecast model, for which the optimization of bias and model robustness were a concern.

The developed modelling approaches used in our study do not rely on complex calibration to observed anomalies or complex, expensive and time-consuming data preprocessing; it relies instead exclusively on the integration of antecedent meteorological and observed cleaned data at the large scale. This study thus provides the sound basis that such large-scale, cleaned data entry models may prove useful predictions at a basin-scale, in coincidence with existing operational approaches that rely on calibrated, statistics-based approaches.

For the limitation of the present study, as the cross-validation techniques may not diagnose the serious model deficiencies, in follow-up research, one could also enhance the introduced algorithms' model performance by incorporating data analysis to find the relationship of interest from both the observational data and the simulation data, and then compare them against one another. The other limitation belongs to neglecting the time variability in stream routing. In the current modeling approach, we did not consider the length of time needed to deliver a grid cell's stream water to a gauge station.

5.1 Future Work

This work won't be the end of a journey but counts as a new beginning to start of the further research including the following.

- Integrating numerical analysis into coastal zone flood modeling.
- Data analytics improvements
- Improving Urban flood management in middle eastern countries by incorporating adjusted streamflow from dams with higher-resolution terrain [].
- Combining flood forecasts with transportation models by integrating location information, and merging them with other boundary conditions, like transportation. This helps to priority check and recovery after disaster while will expand evacuation planning and first respond placement.

Therefore, it is highly recommended to apply a hybrid modeling approach (e.g. data-driven models with physically-based/conceptual models and/or empirical relationships between high flows and influencing factors) for more accurate forecasting of high flow events.

For the future work we aim to apply the proposed modeling approaches to a different flood event in various catchments. As we evaluate the performance of machine Learning in the small watershed in Northern part of Iran, one might investigate the application of the proposed approaches to a large flood event in a large catchment to validate the model applicability in terms of infrequent flood events in multiple regions.

As a future recommendation, for a better modeling task in the hydrological process, data restrictions especially for soil standard should be lightened somewhat by new measurement

technologies, or by new hydrologic observatory networks or even through the mathematically incorporation of signals or satellite retrievals into the land modeling environment. However, these newly investments in hydrologic measurement infrastructure are likely to be expensive, and hydrologists will need to make an undoubted case for them. Surprisingly, as the new proposed method indicated a good performance in data scarce region, there might be a tradeoff between installing necessitate gauging station and removing useless gauging station where their data will not make a big change in the future forecasting

With all these developments we are on the edge of a new era in flood plain mapping. Imagine by these advancements, there might be a phone routing apps that track possibly inundated roads and can inform users of their daily routs from home to office and helps for risk mitigation plans.

6 References

Abrahart, R. J., Heppenstall, A. J., & See, L. M. (2007). Timing error correction procedure applied to neural network rainfall-runoff modelling. *Hydrological Sciences Journal*, *52*(3), 414–431. https://doi.org/10.1623/hysj.52.3.414

Adamowski, J., & Sun, K. (2010). Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology*, *390*(1–2), 85–91. https://doi.org/10.1016/j.jhydrol.2010.06.033

Alfieri, L., Salamon, P., Pappenberger, F., Wetterhall, F., & Thielen, J. (2012). Operational early warning systems for water-related hazards in Europe. *Environmental Science and Policy*, *21*, 35–49. https://doi.org/10.1016/j.envsci.2012.01.008

Alfieri, L., Thielen, J., & Pappenberger, F. (2012). Ensemble hydro-meteorological simulation for flash flood early detection in southern Switzerland. *Journal of Hydrology*, *424–425*, 143–153. https://doi.org/10.1016/j.jhydrol.2011.12.038

Alvisi, S., & Franchini, M. (2011). Fuzzy neural networks for water level and discharge forecasting with uncertainty. *Environmental Modelling & Software*. Elsevier Ltd.

S. Bahrami, P. E. Wigand, "Bad Data Analysis on Streamflow Forecasting Using Nonlinear Echo State Network", International Journal of Advanced Research in Science, Engineering and Technology, 2018, pg no: 7054-7060.

S. Bahrami, P. E. Wigand, "Sensitivity Analysis on Daily Streamflow Forecasting", International Journal of Advanced Research in Science, Engineering and Technology, 2018, pg no: 7312-7317

S. Bahrami, P. E. Wigand, "Daily Streamflow Forecasting Using Nonlinear Echo State Network", International Journal of Advanced Research in Science, Engineering and Technology, 2018, pg no: 6720-6727.

Chen, P. A., Chang, L. C., & Chang, F. J. (2013). Reinforced recurrent neural networks for multi-step-ahead flood forecasts. *Journal of Hydrology*, *497*, 71–79. https://doi.org/10.1016/j.jhydrol.2013.05.038

Guha-sapir, D., Hoyois, P., & Below, R. (2011). Annual Disaster Statistical Review 2016. *Review Literature And Arts Of The Americas*, 1–50. https://doi.org/10.1093/rof/rfs003

Kasiviswanathan, K. S., Cibin, R., Sudheer, K. P., & Chaubey, I. (2013). Constructing prediction interval for artificial neural network rainfall runoff models based on ensemble simulations. *Journal of Hydrology*, *499*, 275–288. https://doi.org/10.1016/j.jhydrol.2013.06.043

Kasiviswanathan, K. S., He, J., Sudheer, K. P., & Tay, J. H. (2016). Potential application of wavelet neural network ensemble to forecast streamflow for flood management. *Journal of Hydrology*, *536*, 161–173. https://doi.org/10.1016/j.jhydrol.2016.02.044

Liu, Y., Brown, J., Demargne, J., & Seo, D. J. (2011). A wavelet-based approach to assessing timing errors in hydrologic predictions. *Journal of Hydrology*, *397*(3–4), 210–224.

https://doi.org/10.1016/j.jhydrol.2010.11.040

Nayak, P. C., Sudheer, K. P., Rangan, D. M., & Ramasastri, K. S. (2005). Short-term flood forecasting with a neurofuzzy model. *Water Resources Research*, *41*(4), 1–16. https://doi.org/10.1029/2004WR003562

Nourani Vahid A4 - Komasi, Mehdi A4 - Mano, Akira, V. A.-N. (2009). A Multivariate ANN-Wavelet Approach for Rainfall-Runoff Modeling. *Water Resources Management, v. 23*(14), 2877-2894–2009 v.23 no.14. https://doi.org/10.1007/s11269-009-9414-5

Pappenberger, F., Cloke, H. L., Parker, D. J., Wetterhall, F., Richardson, D. S., & Thielen, J. (2015). The monetary benefit of early flood warnings in Europe. *Environmental Science and Policy*, *51*, 278–291. https://doi.org/10.1016/j.envsci.2015.04.016

Prakash, O., Sudheer, K. P., & Srinivasan, K. (2014). Improved higher lead time river flow forecasts using sequential neural network with error updating. *Journal of Hydrology and Hydromechanics*, *62*(1), 60–74. https://doi.org/10.2478/johh-2014-0010

Proud, S. R., Fensholt, R., Rasmussen, L. V., & Sandholt, I. (2011). Rapid response flood detection using the MSG geostationary satellite. *International Journal of Applied Earth Observation and Geoinformation*, *13*(4), 536–544. https://doi.org/10.1016/j.jag.2011.02.002

Shamseldin, A. Y., & O'Connor, K. M. (2010). A non-linear neural network technique for updating of river flow forecasts. *Hydrology and Earth System Sciences*, *5*(4), 577–598. https://doi.org/10.5194/hess-5-577-2001

Sperna Weiland, F. C., Van Beek, L. P. H., Kwadijk, J. C. J., & Bierkens, M. F. P. (2010). The ability of a GCM-forced hydrological model to reproduce global discharge variability. *Hydrology and Earth System Sciences*, *14*(8), 1595–1621. https://doi.org/10.5194/hess-14-1595-2010

Wang, J., Hong, Y., Li, L., Gourley, J. J., Khan, S. I., Yilmaz, K. K., … Okello, L. (2011). The coupled routing and excess storage (CREST) distributed hydrological model. *Hydrological Sciences Journal*, *56*(1), 84–98. https://doi.org/10.1080/02626667.2010.543087

Abrahart, R. J., Heppenstall, A. J., & See, L. M. (2007). Timing error correction procedure applied to neural network rainfall-runoff modelling. *Hydrological Sciences Journal*, *52*(3), 414–431. https://doi.org/10.1623/hysj.52.3.414

Adamowski, J., & Sun, K. (2010). Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology*, *390*(1–2), 85–91. https://doi.org/10.1016/j.jhydrol.2010.06.033

Alfieri, L., Salamon, P., Pappenberger, F., Wetterhall, F., & Thielen, J. (2012). Operational early warning systems for water-related hazards in Europe. *Environmental Science and Policy*, *21*, 35–49. https://doi.org/10.1016/j.envsci.2012.01.008

Alfieri, L., Thielen, J., & Pappenberger, F. (2012). Ensemble hydro-meteorological simulation for flash flood early detection in southern Switzerland. *Journal of Hydrology*, *424–425*, 143–153. https://doi.org/10.1016/j.jhydrol.2011.12.038

Alvisi, S., & Franchini, M. (2011). Fuzzy neural networks for water level and discharge forecasting with uncertainty. *Environmental Modelling & Software*. Elsevier Ltd.

Chen, P. A., Chang, L. C., & Chang, F. J. (2013). Reinforced recurrent neural networks for multi-step-ahead flood forecasts. *Journal of Hydrology*, *497*, 71–79. https://doi.org/10.1016/j.jhydrol.2013.05.038

Guha-sapir, D., Hoyois, P., & Below, R. (2011). Annual Disaster Statistical Review 2016. *Review Literature And Arts Of The Americas*, 1–50. https://doi.org/10.1093/rof/rfs003

Kasiviswanathan, K. S., Cibin, R., Sudheer, K. P., & Chaubey, I. (2013). Constructing prediction interval for artificial neural network rainfall runoff models based on ensemble simulations. *Journal of Hydrology*, *499*, 275–288. https://doi.org/10.1016/j.jhydrol.2013.06.043

Kasiviswanathan, K. S., He, J., Sudheer, K. P., & Tay, J. H. (2016). Potential application of wavelet neural network ensemble to forecast streamflow for flood management. *Journal of Hydrology*, *536*, 161–173. https://doi.org/10.1016/j.jhydrol.2016.02.044

Liu, Y., Brown, J., Demargne, J., & Seo, D. J. (2011). A wavelet-based approach to assessing timing errors in hydrologic predictions. *Journal of Hydrology*, *397*(3–4), 210–224. https://doi.org/10.1016/j.jhydrol.2010.11.040

Nayak, P. C., Sudheer, K. P., Rangan, D. M., & Ramasastri, K. S. (2005). Short-term flood forecasting with a neurofuzzy model. *Water Resources Research*, *41*(4), 1–16. https://doi.org/10.1029/2004WR003562

Nourani Vahid A4 - Komasi, Mehdi A4 - Mano, Akira, V. A.-N. (2009). A Multivariate ANN-Wavelet Approach for Rainfall-Runoff Modeling. *Water Resources Management*, *v. 23*(14), 2877-2894–2009 v.23 no.14. https://doi.org/10.1007/s11269-009-9414-5

Pappenberger, F., Cloke, H. L., Parker, D. J., Wetterhall, F., Richardson, D. S., & Thielen, J. (2015). The monetary benefit of early flood warnings in Europe. *Environmental Science and Policy*, *51*, 278–291. https://doi.org/10.1016/j.envsci.2015.04.016

Prakash, O., Sudheer, K. P., & Srinivasan, K. (2014). Improved higher lead time river flow forecasts using sequential neural network with error updating. *Journal of Hydrology and Hydromechanics*, *62*(1), 60–74. https://doi.org/10.2478/johh-2014-0010

Proud, S. R., Fensholt, R., Rasmussen, L. V., & Sandholt, I. (2011). Rapid response flood detection using the MSG geostationary satellite. *International Journal of Applied Earth Observation and Geoinformation*, *13*(4), 536–544. https://doi.org/10.1016/j.jag.2011.02.002

Shamseldin, A. Y., & O'Connor, K. M. (2010). A non-linear neural network technique for updating of river flow forecasts. *Hydrology and Earth System Sciences*, *5*(4), 577–598. https://doi.org/10.5194/hess-5-577-2001

Sperna Weiland, F. C., Van Beek, L. P. H., Kwadijk, J. C. J., & Bierkens, M. F. P. (2010). The ability of a GCM-forced hydrological model to reproduce global discharge variability.

Hydrology and Earth System Sciences, *14*(8), 1595–1621. https://doi.org/10.5194/hess-14-1595-2010

Tien Bui, D.; Khosravi, K.; Shahabi, H.; Daggupati, P.; Adamowski, J.F.; Melesse, A.M.; Thai Pham, B.; Pourghasemi, H.R.; Mahmoudi, M.; Bahrami, S. Flood spatial modeling in northern Iran using remote sensing and gis: A comparison between evidential belief functions and its ensemble with a multivariate logistic regression model. Remote Sens. 2019, 11, 1589.

Van den Broeck J, Cunningham SA, Eeckels R, Herbst K. Data cleaning: detecting, diagnosing, and editing data abnormalities. PLoS Med. 2005;2(10):e267. doi:10.1371/journal.pmed.0020267

Wang, J., Hong, Y., Li, L., Gourley, J. J., Khan, S. I., Yilmaz, K. K., … Okello, L. (2011). The coupled routing and excess storage (CREST) distributed hydrological model. *Hydrological Sciences Journal*, *56*(1), 84–98. https://doi.org/10.1080/02626667.2010.543087

Addor, N., Newman, A.J., Mizukami, N., Clark, M.P., 2017. The CAMELS data set: catchment attributes and meteorology for large-sample studies. Hydrol. Earth Syst. Sci. 21, 5293–5313. https://doi.org/10.5194/hess-21-5293-2017

Ahmed, J.A., Sarma, A.K., 2007. Artificial neural network model for synthetic streamflow generation. Water Resour. Manag. 21, 1015–1029. https://doi.org/10.1007/s11269-006-9070-y

Arnal, L., Wood, A.W., Stephens, E., Cloke, H.L., Pappenberger, F., Arnal, L., Wood, A.W., Stephens, E., Cloke, H.L., Pappenberger, F., 2017. An Efficient Approach for Estimating Streamflow Forecast Skill Elasticity. J. Hydrometeorol. 18, 1715–1729. https://doi.org/10.1175/JHM-D-16-0259.1

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. LARGE AREA HYDROLOGIC MODELING AND ASSESSMENT PART I: MODEL DEVELOPMENT. J. Am. Water Resour. Assoc. 34, 73–89. https://doi.org/10.1111/j.1752-1688.1998.tb05961.x

Asadi, H., Shahedi, K., Jarihani, B., Sidle, R., Asadi, H., Shahedi, K., Jarihani, B., Sidle, R.C., 2019. Rainfall-Runoff Modelling Using Hydrological Connectivity Index and Artificial Neural Network Approach. Water 11, 212. https://doi.org/10.3390/w11020212

Barge, J., Sharif, H., Barge, J.T., Sharif, H.O., 2016. An Ensemble Empirical Mode Decomposition, Self-Organizing Map, and Linear Genetic Programming Approach for Forecasting River Streamflow. Water 8, 247. https://doi.org/10.3390/w8060247

Bayazit, M., 2015. Nonstationarity of Hydrological Records and Recent Trends in Trend Analysis: A State-of-the-art Review. Environ. Process. 2, 527–542. https://doi.org/10.1007/s40710-015-0081-7

Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Networks 5, 157–166.

https://doi.org/10.1109/72.279181

Beven, K., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrol. Process. 6, 279–298. https://doi.org/10.1002/hyp.3360060305

Bhuiyan, M.A.E., Anagnostou, E.N., Kirstetter, P.-E., 2017. A Nonparametric Statistical Technique for Modeling Overland TMI (2A12) Rainfall Retrieval Error. IEEE Geosci. Remote Sens. Lett. 14, 1898–1902. https://doi.org/10.1109/LGRS.2017.2728658

Brown, R.D., Goodison, B.E., Brown, R.D., Goodison, B.E., 1996. Interannual Variability in Reconstructed Canadian Snow Cover, 1915–1992. J. Clim. 9, 1299–1318. https://doi.org/10.1175/1520-0442(1996)009<1299:IVIRCS>2.0.CO;2

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

Carquex, C., Rosenberg, C., Bhattacharya, K., 2018. State Estimation in Power Distribution Systems Based on Ensemble Kalman Filtering. IEEE Trans. Power Syst. 33, 6600–6610. https://doi.org/10.1109/TPWRS.2018.2847289

Chen, C.-S., Chou, F.N.-F., Chen, B.P.-T., 2010. Spatial Information-Based Back-Propagation Neural Network Modeling for Outflow Estimation of Ungauged Catchment. Water Resour. Manag. 24, 4175–4197. https://doi.org/10.1007/s11269-010-9652-6

Chitsazan, M.A., Sami Fadali, M., Trzynadlowski, A.M., 2019. Wind speed and wind direction forecasting using echo state network with nonlinear functions. Renew. Energy 131, 879–889. https://doi.org/10.1016/j.renene.2018.07.060

Chua, L.H.C., Wong, T.S.W., 2011. Runoff forecasting for an asphalt plane by Artificial Neural Networks and comparisons with kinematic wave and autoregressive moving average models. J. Hydrol. 397, 191–201. https://doi.org/10.1016/J.JHYDROL.2010.11.030

Coulibaly, P., Burn, D.H., Coulibaly, P., Burn, D.H., 2005. Spatial and Temporal Variability of Canadian Seasonal Streamflows. J. Clim. 18, 191–210. https://doi.org/10.1175/JCLI-3258.1

Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. Math. Control. Signals, Syst. 2, 303–314. https://doi.org/10.1007/BF02551274

Dehghani, M., Saghafian, B., Nasiri Saleh, F., Farokhnia, A., Noori, R., 2014. Uncertainty analysis of streamflow drought forecast using artificial neural networks and Monte-Carlo simulation. Int. J. Climatol. 34, 1169–1180. https://doi.org/10.1002/joc.3754

Dettinger, M.D., Cayan, D.R., Meyer, M.K., Jeton, A.E., 2004. Simulated Hydrologic Responses to Climate Variations and Change in the Merced, Carson, and American River Basins, Sierra Nevada, California, 1900–2099. Clim. Change 62, 283–317. https://doi.org/10.1023/B:CLIM.0000013683.13346.4f

Doherty, J., Johnston, J.M., 2003. METHODOLOGIES FOR CALIBRATION AND PREDICTIVE ANALYSIS OF A WATERSHED MODEL. J. Am. Water Resour. Assoc.

39, 251–265. https://doi.org/10.1111/j.1752-1688.2003.tb04381.x

Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour. Res. 28, 1015–1031. https://doi.org/10.1029/91WR02985

Farmer, W.H., Levin, S., 2018. Characterizing Uncertainty in Daily Streamflow Estimates at Ungauged Locations for the Massachusetts Sustainable Yield Estimator. JAWRA J. Am. Water Resour. Assoc. 54, 198–210. https://doi.org/10.1111/1752-1688.12603

Flato, G.M., 2011. Earth system models: an overview. Wiley Interdiscip. Rev. Clim. Chang. 2, 783–800. https://doi.org/10.1002/wcc.148

Gansterer, W.N., Ecker, G.F., n.d. On the Relationship Between Feature Selection and Classification Accuracy 90–105.

Guo, X., Hu, T., Wu, C., Zhang, T., Lv, Y., 2013. Multi-Objective Optimization of the Proposed Multi-Reservoir Operating Policy Using Improved NSPSO. Water Resour. Manag. 27, 2137–2153. https://doi.org/10.1007/s11269-013-0280-9

Gupta, A., Raman, K., 2014. BOARD DIVERSITY AND CEO SELECTION. J. Financ. Res. 37, 495–518. https://doi.org/10.1111/jfir.12044

Hochreiter, S., Hochreiter, S., Bengio, Y., Frasconi, P., Schmidhuber, J., 2001. Gradient Flow in Recurrent Nets: the Difficulty of Learning Long-Term Dependencies.

Jang, J.-S.R., 1993. ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans. Syst. Man. Cybern. 23, 665–685. https://doi.org/10.1109/21.256541

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

Kennedy, A.M., Garen, D.C., Koch, R.W., 2009. The association between climate teleconnection indices and Upper Klamath seasonal streamflow: Trans-Niño Index. Hydrol. Process. 23, 973–984. https://doi.org/10.1002/hyp.7200

King, G., Zeng, L., 2001. Logistic Regression in Rare Events Data. Polit. Anal. 9, 137–163. https://doi.org/10.1093/oxfordjournals.pan.a004868

Kirchner, J.W., 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resour. Res. 42. https://doi.org/10.1029/2005WR004362

Kisi, O., 2015. Streamflow Forecasting and Estimation Using Least Square Support Vector Regression and Adaptive Neuro-Fuzzy Embedded Fuzzy c-means Clustering. Water Resour. Manag. 29, 5109–5127. https://doi.org/10.1007/s11269-015-1107-7

Makkeasorn, A., Chang, N.-B., Li, J., 2009. Seasonal change detection of riparian zones with

remote sensing images and genetic programming in a semi-arid watershed. J. Environ. Manage. 90, 1069–1080. https://doi.org/10.1016/J.JENVMAN.2008.04.004

Newman, A.J., Clark, M.P., Sampson, K., Wood, A., Hay, L.E., Bock, A., Viger, R.J., Blodgett, D., Brekke, L., Arnold, J.R., Hopson, T., Duan, Q., 2015. Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. Hydrol. Earth Syst. Sci. 19, 209–223. https://doi.org/10.5194/hess-19-209-2015

Pagano, T.C., Garen, D.C., Perkins, T.R., Pasteris, P.A., 2009. Daily Updating of Operational Statistical Seasonal Water Supply Forecasts for the western U.S. JAWRA J. Am. Water Resour. Assoc. 45, 767–778. https://doi.org/10.1111/j.1752-1688.2009.00321.x

Partington, D., Brunner, P., Simmons, C.T., Werner, A.D., Therrien, R., Maier, H.R., Dandy, G.C., 2012. Evaluation of outputs from automated baseflow separation methods against simulated baseflow from a physically based, surface water-groundwater flow model. J. Hydrol. 458–459, 28–39. https://doi.org/10.1016/J.JHYDROL.2012.06.029

Peng, T., Zhou, J., Zhang, C., Fu, W., 2017. Streamflow forecasting using empirical wavelet transform and artificial neural networks. Water (Switzerland). https://doi.org/10.3390/w9060406

Rasouli, K., Hsieh, W.W., Cannon, A.J., 2011. Daily streamflow forecasting by machine 1 learning methods with weather and climate 2 inputs.

Saltelli, A. (Andrea), 2008. Global sensitivity analysis : the primer. John Wiley.

Samuel, J., Coulibaly, P., Metcalfe, R.A., 2011. Estimation of Continuous Streamflow in Ontario Ungauged Basins: Comparison of Regionalization Methods. J. Hydrol. Eng. 16, 447–459. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000338

Schnier, S., Cai, X., 2014. Prediction of regional streamflow frequency using model tree ensembles. J. Hydrol. 517, 298–309. https://doi.org/10.1016/j.jhydrol.2014.05.029

Sene, K., Tych, W., Beven, K., 2018. Exploratory studies into seasonal flow forecasting potential for large lakes. Hydrol. Earth Syst. Sci. 22, 127–141. https://doi.org/10.5194/hess-22-127-2018

Senent-Aparicio, J., Jimeno-Sáez, P., Bueno-Crespo, A., Pérez-Sánchez, J., Pulido-Velázquez, D., 2019. Coupling machine-learning techniques with SWAT model for instantaneous peak flow prediction. Biosyst. Eng. 177, 67–77. https://doi.org/10.1016/J.BIOSYSTEMSENG.2018.04.022

Shortridge, J.E., Guikema, S.D., Zaitchik, B.F., 2015. Empirical streamflow simulation for water resource management in data-scarce seasonal watersheds 11083–11127. https://doi.org/10.5194/hessd-12-11083-2015

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

Singh, V., Goyal, M.K., 2017. Curve number modifications and parameterization sensitivity analysis for reducing model uncertainty in simulated and projected streamflows in a Himalayan catchment. Ecol. Eng. 108, 17–29. https://doi.org/10.1016/J.ECOLENG.2017.08.002

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., Xu, C., 2015. Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. J. Hydrol. 523, 739–757. https://doi.org/10.1016/J.JHYDROL.2015.02.013

Thornton, P.H., Ocasio, W., Lounsbury, M., 2012. The institutional logics perspective : a new approach to culture, structure, and process. Oxford University Press.

Tree-Ring Society., Y., University of Arizona. Laboratory of Tree-Ring Research., X., Wei, W., Yu, S., Gong, Y., Trouet, V.M., 2001. Tree-ring research., Tree-Ring Research. [Tree-Ring Society with the cooperation of the Laboratory of Tree-Ring Research].

USGS.gov | Science for a changing world [WWW Document], n.d. URL https://www.usgs.gov/ (accessed 3.17.19).

Valizadeh, N., Mirzaei, M., Allawi, M.F., Afan, H.A., Mohd, N.S., Hussain, A., El-Shafie, A., 2017. Artificial intelligence and geo-statistical models for stream-flow forecasting in ungauged stations: state of the art. Nat. Hazards 86, 1377–1392. https://doi.org/10.1007/s11069-017-2740-7

Viel, C., Beaulant, A.-L., Soubeyroux, J.-M., Céron, J.-P., 2016. How seasonal forecast could help a decision maker: an example of climate service for water resource management. Adv. Sci. Res. 13, 51–55. https://doi.org/10.5194/asr-13-51-2016

Xu, X., Zhou, J.T., Tsang, I., Qin, Z., Goh, R.S.M., Liu, Y., 2016. Simple and Efficient Learning using Privileged Information.

Yaseen, Z.M., Fu, M., Wang, C., Mohtar, W.H.M.W., Deo, R.C., El-shafie, A., 2018. Application of the Hybrid Artificial Neural Network Coupled with Rolling Mechanism and Grey Model Algorithms for Streamflow Forecasting Over Multiple Time Horizons. Water Resour. Manag. 32, 1883–1899. https://doi.org/10.1007/s11269-018-1909-5

Ahmadlou, M., Karimi, M., Alizadeh, S., Shirzadi, A., Parvinnejhad, D., Shahabi, H., Panahi, M., 2018. Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA). Geocarto International, 1-21.

Al-Juaidi, A.E., Nassar, A.M., Al-Juaidi, O.E., 2018. Evaluation of flood susceptibility mapping using logistic regression and GIS conditioning factors. Arabian Journal of Geosciences. 11, 765.

Alfieri, L., Bisselink, B., Dottori, F., Naumann, G., de Roo, A., Salamon, P., Wyser, K., Feyen, L., 2017. Global projections of river flood risk in a warmer world. Earth's Future. 5, 171-182.

Bahrami, S., Wigand, E., Sensitivity Analysis on Daily Streamflow Forecasting.

Bahrami, S., Wigand, E., 2018. Sensitivity Analysis on Daily Streamflow Forecasting. International Journal of Advanced Research in Science ,engineering and Technology. 5, 1-6.

Bauer, E., Kohavi, R., 1999. An empirical comparison of voting classification algorithms: bagging, boosting, and variants. Machine learning. 36, 105-139.

Bui, D.T., Panahi, M., Shahabi, H., Singh, V.P., Shirzadi, A., Chapi, K., Khosravi, K., Chen, W., Panahi, S., Li, S., 2018. Novel hybrid evolutionary algorithms for spatial prediction of floods. Scientific reports. 8, 15364.

Chapi, K., Singh, V.P., Shirzadi, A., Shahabi, H., Bui, D.T., Pham, B.T., Khosravi, K., 2017. A novel hybrid artificial intelligence approach for flood susceptibility assessment. Environmental modelling & software. 95, 229-245.

Dash, M., Liu, H., 1997. Feature selection for classification. Intelligent data analysis. 1, 131-156.

De Brito, M., Evers, M., 2015. Multi-criteria decision making for flood risk management: a survey of the current state-of-the-art. Natural Hazards and Earth System Sciences Discussions. 3, 6689-6726.

de Brito, M.M., Evers, M., Almoradie, A.D.S., 2018. Participatory flood vulnerability assessment: a multi-criteria approach. Hydrology & Earth System Sciences. 22.

Dietterich, T.G., 2000. Ensemble methods in machine learning, International workshop on multiple classifier systems. Springer, pp. 1-15.

Duda, R.O., Hart, P.E., Stork, D.G., 2012. Pattern classification. John Wiley & Sons.

Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do we need hundreds of classifiers to solve real world classification problems? The Journal of Machine Learning Research. 15, 3133-3181.

Giacinto, G., Roli, F., 2001. Design of effective neural network ensembles for image classification purposes. Image and Vision Computing. 19, 699-707.

Guo, G., Wang, H., Bell, D., Bi, Y., Greer, K., 2003. KNN model-based approach in classification, OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". Springer, pp. 986-996.

Hall, M.A., Holmes, G., 2002. Benchmarking attribute selection techniques for discrete class data mining.

Hassanat, A.B., 2014. Visual passwords using automatic lip reading. arXiv preprint arXiv:1409.0924.

Hassanat, A.B., Abbadi, M.A., Altarawneh, G.A., Alhasanat, A.A., 2014. Solving the problem of the K parameter in the KNN classifier using an ensemble learning approach. arXiv preprint arXiv:1409.0919.

He, Q.P., Wang, J., 2007. Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes. IEEE transactions on semiconductor manufacturing. 20, 345-354.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., Kanae, S., 2013. Global flood risk under climate change. Nature Climate Change. 3, 816.

Hong, H., Panahi, M., Shirzadi, A., Ma, T., Liu, J., Zhu, A.-X., Chen, W., Kougias, I., Kazakis, N., 2018. Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. Science of The Total Environment. 621, 1124-1141.

Hu, L.-Y., Huang, M.-W., Ke, S.-W., Tsai, C.-F., 2016. The distance function effect on k-nearest neighbor classification for medical datasets. SpringerPlus. 5, 1304.

Kamali, T., Boostani, R., Parsaei, H., 2014. A multi-classifier approach to MUAP classification for diagnosis of neuromuscular disorders. IEEE transactions on neural systems and rehabilitation engineering. 22, 191-200.

Kantardzic, M., 2011. Data mining: concepts, models, methods, and algorithms. John Wiley & Sons.

Kazakis, N., Kougias, I., Patsialis, T., 2015. Assessment of flood hazard areas at a regional scale using an index-based approach and Analytical Hierarchy Process: Application in Rhodope–Evros region, Greece. Science of the Total Environment. 538, 555-563.

Khosravi, K., Nohani, E., Maroufinia, E., Pourghasemi, H.R., 2016a. A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. Natural Hazards. 83, 947-987.

Khosravi, K., Panahi, M., Tien Bui, D., 2018a. Spatial Prediction of Groundwater Spring Potential Mapping Based on Adaptive Neuro-Fuzzy Inference System and Metaheuristic Optimization.

Khosravi, K., Pham, B.T., Chapi, K., Shirzadi, A., Shahabi, H., Revhaug, I., Prakash, I., Bui, D.T., 2018b. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. Science of the Total Environment. 627, 744-755.

Khosravi, K., Pourghasemi, H.R., Chapi, K., Bahri, M., 2016b. Flash flood susceptibility analysis and its mapping using different bivariate models in Iran: a comparison between Shannon's entropy, statistical index, and weighting factor models. Environmental monitoring and assessment. 188, 656.

Khosravi, K., Sartaj, M., Tsai, F.T.-C., Singh, V.P., Kazakis, N., Melesse, A.M., Prakash, I., Bui, D.T., Pham, B.T., 2018c. A comparison study of DRASTIC methods with various objective methods for groundwater vulnerability assessment. Science of the total environment. 642, 1032-1049.

Khosravi, K., Shahabi, H., Pham, B.T., Adamawoski, J., Shirzadi, A., Pradhan, B., Dou, J., Ly, H.-B., Gróf, G., Ho, H.L., 2019. A Comparative Assessment of Flood Susceptibility Modeling Using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. Journal of Hydrology. 573, 311-323.

Kira, K., Rendell, L.A., 1992. A practical approach to feature selection, Machine Learning Proceedings 1992. Elsevier, pp. 249-256.

Kononenko, I., 1994. Estimating attributes: analysis and extensions of RELIEF, European conference on machine learning. Springer, pp. 171-182.

Kron, W., 2002. Keynote lecture: Flood risk= hazard× exposure× vulnerability. Flood defence, 82-97.

Liu, C.-L., Lee, C.-H., Lin, P.-M., 2010. A fall detection system using k-nearest neighbor classifier. Expert systems with applications. 37, 7174-7181.

Liu, X.-s., Xiao, H., Wang, T.-l., 2014. Rapid assessment of flood loss based on neural network ensemble. Transactions of Nonferrous Metals Society of China. 24, 2636-2641.

Maclin, R., Opitz, D., 1997. An empirical evaluation of bagging and boosting. In Proceedings of the Fourteenth National Conference on Artificial Intelligence. 1997, 546-551.

Messner, F., Meyer, V., 2006. Flood damage, vulnerability and risk perception–challenges for flood damage research, Flood risk management: hazards, vulnerability and mitigation measures. Springer, pp. 149-167.

Pradhan, B., 2010. Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. Journal of Spatial Hydrology. 9.

Rahmati, O., Pourghasemi, H.R., 2017. Identification of critical flood prone areas in data-scarce and ungauged regions: A comparison of three data mining models. Water resources management. 31, 1473-1487.

Rahmati, O., Pourghasemi, H.R., Zeinivand, H., 2016a. Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. Geocarto International. 31, 42-70.

Rahmati, O., Zeinivand, H., Besharat, M., 2016b. Flood hazard zoning in Yasooj region, Iran, using GIS and multi-criteria decision analysis. Geomatics, Natural Hazards and Risk. 7, 1000-1017.

Ramaswami, M., Bhaskaran, R., 2009. A study on feature selection techniques in educational data mining. arXiv preprint arXiv:0912.3924.

Robnik-Šikonja, M., Kononenko, I., 2003. Theoretical and empirical analysis of ReliefF and RReliefF. Machine learning. 53, 23-69.

Sarhadi, A., Soltani, S., Modarres, R., 2012. Probabilistic flood inundation mapping of ungauged rivers: Linking GIS techniques and frequency analysis. Journal of Hydrology. 458, 68-86.

Shafizadeh-Moghadam, H., Valavi, R., Shahabi, H., Chapi, K., Shirzadi, A., 2018. Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping. Journal of environmental management. 217, 1-11.

Shahabi, H., Hashim, M., 2015. Landslide susceptibility mapping using GIS-based statistical models and Remote sensing data in tropical environment. Scientific reports. 5, 9899.

Shirzadi, A., Solaimani, K., Roshan, M.H., Kavian, A., Chapi, K., Shahabi, H., Keesstra, S., Ahmad, B.B., Bui, D.T., 2019. Uncertainties of prediction accuracy in shallow landslide modeling: Sample size and raster resolution. Catena. 178, 172-188.

Shirzadi, A., Soliamani, K., Habibnejhad, M., Kavian, A., Chapi, K., Shahabi, H., Chen, W., Khosravi, K., Thai Pham, B., Pradhan, B., 2018. Novel GIS based machine learning algorithms for shallow landslide susceptibility mapping. Sensors. 18, 3777.

Tehrany, M.S., Pradhan, B., Jebur, M.N., 2014. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. Journal of hydrology. 512, 332-343.

Tien Bui, D., Khosravi, K., Li, S., Shahabi, H., Panahi, M., Singh, V., Chapi, K., Shirzadi, A., Panahi, S., Chen, W., 2018. New hybrids of anfis with several optimization algorithms for flood susceptibility modeling. Water. 10, 1210.

Waske, B., van der Linden, S., Benediktsson, J.A., Rabe, A., Hostert, P., 2010. Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing. 48, 2880-2889.

Wettschereck, D., Aha, D.W., Mohri, T., 1997. A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. Artificial Intelligence Review. 11, 273-314.

Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Philip, S.Y., 2008. Top 10 algorithms in data mining. Knowledge and information systems. 14, 1-37.

Yu, J., Qin, X., Larsen, O., 2013. Joint Monte Carlo and possibilistic simulation for flood damage assessment. Stochastic environmental research and risk assessment. 27, 725-735.

Ahmadlou M, Karimi M, Alizadeh S, Shirzadi A, Parvinnejhad D, Shahabi H, Panahi M (2018) Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA) Geocarto International:1-21

Althuwaynee OF, Pradhan B, Lee S (2012) Application of an evidential belief function model in landslide susceptibility mapping Computers & Geosciences 44:120-135

Althuwaynee OF, Pradhan B, Park H-J, Lee JH (2014) A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping Catena 114:21-36

Bednarik M, Magulová B, Matys M, Marschalko M (2010) Landslide susceptibility assessment of the Kraľovany–Liptovský Mikuláš railway case study Physics and Chemistry of the Earth, Parts A/B/C 35:162-171

Brenning A (2005) Spatial prediction models for landslide hazards: review, comparison and evaluation Natural Hazards and Earth System Science 5:853-862

Bubeck P, Botzen WJ, Aerts JC (2012) A review of risk perceptions and other factors that influence flood mitigation behavior Risk Analysis: An International Journal 32:1481-1495

Bui DT, Lofman O, Revhaug I, Dick O (2011) Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression Natural hazards 59:1413

Bui DT et al. (2018) Novel hybrid evolutionary algorithms for spatial prediction of floods Scientific reports 8:15364

Bui DT, Pradhan B, Lofman O, Revhaug I, Dick OB (2012a) Landslide susceptibility mapping at Hoa Binh province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS Computers & Geosciences 45:199-211

Bui DT, Pradhan B, Lofman O, Revhaug I, Dick OB (2012b) Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): a comparative assessment of the efficacy of evidential belief functions and fuzzy logic models Catena 96:28-40

Carranza EJM, Hale M (2003) Evidential belief functions for data-driven geologically constrained mapping of gold potential, Baguio district, Philippines Ore Geology Reviews 22:117-132

Carranza EJM, Van Ruitenbeek F, Hecker C, van der Meijde M, van der Meer FD (2008) Knowledge-guided data-driven evidential belief modeling of mineral prospectivity in Cabo de Gata, SE Spain International Journal of Applied Earth Observation and Geoinformation 10:374-387

Çelik HE, Coskun G, Cigizoglu HK, Ağıralioğlu N, Aydın A, Esin AI (2012) The analysis of 2004 flood on Kozdere Stream in Istanbul Natural hazards 63:461-477

Chapi K, Singh VP, Shirzadi A, Shahabi H, Bui DT, Pham BT, Khosravi K (2017) A novel hybrid artificial intelligence approach for flood susceptibility assessment Environmental modelling & software 95:229-245

Chen W, Pradhan B, Li S, Shahabi H, Rizeei HM, Hou E, Wang S (2019a) Novel Hybrid Integration Approach of Bagging-Based Fisher's Linear Discriminant Function for Groundwater Potential Analysis Natural Resources Research:1-20

Chen W et al. (2019b) Spatial prediction of landslide susceptibility by combining evidential belief function, logistic regression and logistic model tree Geocarto International:1-25

Chung C-JF, Fabbri AG (2003) Validation of spatial prediction models for landslide hazard mapping Natural Hazards 30:451-472

Dempster AP (1969) Upper and lower probability inferences for families of hypotheses with monotone density ratios The Annals of Mathematical Statistics 40:953-969

Donati L, Turrini M (2002) An objective method to rank the importance of the factors predisposing to landslides with the GIS methodology: application to an area of the Apennines (Valnerina; Perugia, Italy) Engineering Geology 63:277-289

Glenn EP, Morino K, Nagler PL, Murray RS, Pearlstein S, Hultine KR (2012) Roles of saltcedar (Tamarix spp.) and capillary rise in salinizing a non-flooding terrace on a flow-regulated desert river Journal of arid environments 79:56-65

Gokceoglu C, Sonmez H, Nefeslioglu HA, Duman TY, Can T (2005) The 17 March 2005 Kuzulu landslide (Sivas, Turkey) and landslide-susceptibility map of its near vicinity Engineering geology 81:65-83

Grahn T, Nyberg L (2017) Assessment of pluvial flood exposure and vulnerability of residential areas International Journal of Disaster Risk Reduction 21:367-375

Haghizadeh A, Siahkamari S, Haghiabi AH, Rahmati O (2017) Forecasting flood-prone areas using Shannon's entropy model Journal of Earth System Science 126:39

Hosmer DW, Lemeshow S (2000) Applied logistic regression.(Sl). Wiley,

Jaafari A, Panahi M, Pham BT, Shahabi H, Bui DT, Rezaie F, Lee S (2019) Meta optimization of an adaptive neuro-fuzzy inference system with grey wolf optimizer and biogeography-based optimization algorithms for spatial prediction of landslide susceptibility Catena 175:430-445

Khosravi K, Nohani E, Maroufinia E, Pourghasemi HR (2016a) A GIS-based flood susceptibility assessment and its mapping in Iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique Natural Hazards 83:947-987

Khosravi, K., Pourghasemi, H.R., Chapi, K., Bahri, M. 2016b. Flash flood susceptibility analysis and its mapping using different bivariate models in Iran: a comparison between Shannon's entropy, statistical index, and weighting factor models. Environmental monitoring and assessment 188 (12), 656

Khosravi K, Panahi M, Tien Bui D (2018a) Spatial Prediction of Groundwater Spring Potential Mapping Based on Adaptive Neuro-Fuzzy Inference System and Metaheuristic Optimization. Hydrology and earth system science. 22, 4771-4792.

Khosravi K et al. (2018b) A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran Science of the Total Environment 627:744-755

Kia MB, Pirasteh S, Pradhan B, Mahmud AR, Sulaiman WNA, Moradi A (2012) An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia Environmental Earth Sciences 67:251-264

Lee J-m, Hyun K-h, Choi J-s, Yoon Y-j, Geronimo FKF (2012) Flood reduction analysis on watershed of LID design demonstration district using SWMM5 Desalination and Water Treatment 38:255-261

Levy JK, Hartmann J, Li KW, An Y, Asgary A (2007) Multi-criteria decision support systems for flood hazard mitigation and emergency response in urban watersheds 1 JAWRA Journal of the American Water Resources Association 43:346-358

Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications Environmental modelling & software 15:101-124

Manandhar B, Balla MK, Awal R, Pradhan BM Floodplain analysis and risk assessment of lothar khola (stream). In: 11th ESRI India User Conference, 2010.

Merz B, Thieken A, Gocht M (2007) Flood risk mapping at the local scale: concepts and challenges. In: Flood risk management in Europe. Springer, pp 231-251

Moore ID, Grayson R, Ladson A (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications Hydrological processes 5:3-30

Nampak H, Pradhan B, Manap MA (2014) Application of GIS based data driven evidential belief function model to predict groundwater potential zonation Journal of Hydrology 513:283-300

O'brien RM (2007) A caution regarding rules of thumb for variance inflation factors Quality & quantity 41:673-690

Oh H-J, Pradhan B (2011) Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area Computers & Geosciences 37:1264-1276

Opolot E (2013) Application of remote sensing and geographical information systems in flood management: a review Research journal of applied sciences engineering and technology 6:1884-1894

Papadopoulou-Vrynioti K, Bathrellos GD, Skilodimou HD, Kaviris G, Makropoulos K (2013) Karst collapse susceptibility mapping considering peak ground acceleration in a rapidly growing urban area Engineering Geology 158:77-88

Park N-W (2011) Application of Dempster-Shafer theory of evidence to GIS-based landslide susceptibility analysis Environmental Earth Sciences 62:367-376

Pham BT, Prakash I, Singh SK, Shirzadi A, Shahabi H, Bui DT (2019) Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches CATENA 175:203-218

Pourghasemi HR, Beheshtirad M (2015) Assessment of a data-driven evidential belief function model and GIS for groundwater potential mapping in the Koohrang Watershed, Iran Geocarto International 30:662-685

Pourghasemi HR, Mohammady M, Pradhan B (2012) Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran Catena 97:71-84

Pradhan B (2010) Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing Journal of Spatial Hydrology 9

Pradhan B (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS Computers & Geosciences 51:350-365

Pradhan B, Abokharima MH, Jebur MN, Tehrany MS (2014) Land subsidence susceptibility mapping at Kinta Valley (Malaysia) using the evidential belief function model in GIS Natural hazards 73:1019-1042

Rahmati O, Pourghasemi HR (2017) Identification of critical flood prone areas in data-scarce and ungauged regions: A comparison of three data mining models Water resources management 31:1473-1487

Rahmati O, Pourghasemi HR, Zeinivand H (2016a) Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran Geocarto International 31:42-70

Rahmati O, Zeinivand H, Besharat M (2016b) Flood hazard zoning in Yasooj region, Iran, using GIS and multi-criteria decision analysis Geomatics, Natural Hazards and Risk 7:1000-1017

Sadeghi-Pouya A, Nouri J, Mansouri N, Kia-Lashaki A (2017) An indexing approach to assess flood vulnerability in the western coastal cities of Mazandaran, Iran International journal of disaster risk reduction 22:304-316

Shafer G (1976) A mathematical theory of evidence vol 42. Princeton university press,

Shafizadeh-Moghadam H, Valavi R, Shahabi H, Chapi K, Shirzadi A (2018) Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping Journal of environmental management 217:1-11

Shahabi H, Khezri S, Ahmad BB, Hashim M (2014) Landslide susceptibility mapping at central Zab basin, Iran: a comparison between analytical hierarchy process, frequency ratio and logistic regression models Catena 115:55-70

Shirzadi A, Saro L, Joo OH, Chapi K (2012) A GIS-based logistic regression model in rock-fall susceptibility mapping along a mountainous road: Salavat Abad case study, Kurdistan, Iran Natural hazards 64:1639-1656

Srivastava OS, Denis D, Srivastava SK, Kumar M, Kumar N (2014) Morphometric analysis of a Semi Urban Watershed, trans Yamuna, draining at Allahabad using Cartosat (DEM) data and GIS International Journal of Engineering And Science (IJES) 3:71-79

Tehrany MS, Lee M-J, Pradhan B, Jebur MN, Lee S (2014a) Flood susceptibility mapping using integrated bivariate and multivariate statistical models Environmental earth sciences 72:4001-4015

Tehrany MS, Pradhan B, Jebur MN (2013) Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS Journal of Hydrology 504:69-79

Tehrany MS, Pradhan B, Jebur MN (2014b) Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS Journal of hydrology 512:332-343

Tehrany MS, Pradhan B, Jebur MN (2015a) Flood susceptibility analysis and its verification using a novel ensemble support vector machine and frequency ratio method Stochastic environmental research and risk assessment 29:1149-1165

Tehrany MS, Pradhan B, Mansor S, Ahmad N (2015b) Flood susceptibility assessment using GIS-based support vector machine model with different kernel types Catena 125:91-101

Termeh SVR, Kornejady A, Pourghasemi HR, Keesstra S (2018) Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms Science of the Total Environment 615:438-451

Tien Bui D et al. (2018a) New hybrids of anfis with several optimization algorithms for flood susceptibility modeling Water 10:1210

Tien Bui D et al. (2018b) Landslide detection and susceptibility mapping by airsar data using support vector machine and index of entropy models in cameron highlands, malaysia Remote Sensing 10:1527

Tien Bui D et al. (2018c) A novel integrated approach of relevance vector machine optimized by imperialist competitive algorithm for spatial modeling of shallow landslides Remote Sensing 10:1538

Tien Bui D et al. (2018d) Land subsidence susceptibility mapping in south korea using machine learning algorithms Sensors 18:2464

Umar Z, Pradhan B, Ahmad A, Jebur MN, Tehrany MS (2014) Earthquake induced landslide susceptibility mapping using an integrated ensemble frequency ratio and logistic regression models in West Sumatera Province, Indonesia Catena 118:124-135

Varoonchotikul P (2003) Flood forecasting using artificial neural networks. CRC Press,

Walley P (1987) Belief function representations of statistical evidence The annals of statistics 15:1439-1465

Yalcin A, Reis S, Aydinoglu A, Yomralioglu T (2011) A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey Catena 85:274-287

Youssef AM, Pradhan B, Hassan AM (2011) Flash flood risk estimation along the St. Katherine road, southern Sinai, Egypt using GIS based morphometry and satellite imagery Environmental Earth Sciences 62:611-623

7 Appendix

To indicate the model applicability and ensure its spatial consistency in different climatic regions, five different basins distributed across the US are chosen. In Massachusetts basin, Berkshire County as the snow-dominant watershed is selected (Case Study 1). In California, we chose Tuolumne County as a mixed pluvial-nival (Case Study 2). In Oregon basins, Wasco County is chosen as the rain dominate watershed (Case Study 3). In Texas, we use Denton County as a Rain dominated (Case Study 4) and in Missouri we chose St. Louis County (Case Study 5). Table 1-a shows a summary of the selected watersheds (USGS.gov).

Table a-1. the catchment characteristics

Region	Berkshire County, Massachusetts	Tuolumne County, California	Wasco County, Oregon	Denton County Texas	St. Louis County Missouri
Watershed	USGS 01333000 Green River at Williamstown, MA	USGS 11284400, Big C AB Whites Gulch NR Groveland CA	USGS 14096850 Beaver Creek below Quartz Creek, NR Simnasho, OR	USGS 08051135 Elm Fk Trinity Rv at Greenbelt nr Pilot Point, TX	USGS 06935965 Missouri River at St. Charles, MO
Hydrologic unit code	2020003	18040009	17070306	12030103	10300200
Coordinates	Latitude: 42°42'32", Longitude: 73°11'50" NAD27	Latitude: 37°50'31", Longitude: 120°11'02" NAD27	Latitude: 44°57'32", Longitude: 121°23'35" NAD27	Latitude:33°20'59" Longitude:97°02'08" NAD83	Latitude:38°47'19.9" Longitude:90°28'14.6" NAD83
Drainage area (Meter Sq.)	110.34	42.47	375,548,276	179,700,000	135,700,000
Gage datum (Meter)	186.85 meter above NAVD88	780.50 meter above NGVD29	688.85 meter above NGVD29	158.801 meters above NGVD29	125 meters above NAVD88

7.1 Simulation Results

The table a-2 provides the evaluation indices for the test stage covering 50 days for 5 different regions.

Table a-2. Forecasting results using evaluation indices for the three case studies: R, RMSE, MAE, WI, and E_{SN}

		Streamflow Forecast					**Bad Data**					**Limited Data**				
		R	WI	E_{SN}	MAE (cfs)	RMSE (cfs)	R	WI	E_{SN}	MAE (cfs)	RMSE (cfs)	R	WI	E_{SN}	MAE (cfs)	RMSE (cfs)
Case Study 1	RNESN	0.98	0.99	0.96	14.8	2.29	0.95	0.95	0.92	15.3	2.45	0.96	0.98	0.92	14.95	2.34
	NESN	0.93	0.96	0.84	31.1	4.65	0.86	0.86	0.81	31.5	5.3	0.88	0.93	0.78	31.9	5.0
	ANFIS	0.78	0.87	0.47	57.4	8.43	0.72	0.64	0.45	72.6	8.57	0.58	0.70	0.39	63.85	8.73
Case Study 2	RNESN	0.99	0.99	0.98	5.38	0.86	0.92	0.94	0.92	5.42	0.92	0.94	0.96	0.95	5.44	0.89
	NESN	0.97	0.98	0.93	12.33	1.88	0.96	0.96	0.83	12.4	2.22	0.89	0.97	0.92	13.25	2.01
	ANFIS	0.87	0.93	0.72	25.98	3.77	0.74	0.75	0.68	27.1	5.15	0.68	0.82	0.69	33.6	4.65

0.06	0.12	0.23	1.1	2.83	5.82	0.055	0.11	0.28
0.39	0.7	1.71	6.92	15.76	37.4	0.29	0.7	1.56
0.96	0.93	0.68	0.95	0.95	0.63	0.97	0.95	0.69
0.96	0.91	0.83	0.93	0.94	0.79	0.97	0.93	0.86
0.97	0.93	0.81	0.94	0.86	0.63	0.98	0.95	0.84
0.06	0.11	0.26	0.92	2.45	7.1	0.047	0.10	0.23
0.4	0.73	1.8	6.73	14.2	32.5	0.37	0.66	1.7
0.92	0.82	0.85	0.87	0.86	0.64	0.95	0.85	0.86
0.96	0.87	0.79	0.94	0.93	0.73	0.98	0.89	0.8
0.95	0.93	0.69	0.96	0.92	0.68	0.97	0.95	0.7
0.06	0.11	0.23	0.94	2.23	4.58	0.05	0.103	0.21
0.39	0.69	1.52	6.32	11.72	28.91	0.32	0.63	1.46
0.99	0.96	0.86	0.94	0.88	0.70	0.99	0.96	0.87
0.997	0.992	0.964	0.94	0.91	0.83	0.97	0.99	0.98
0.99	0.98	0.94	0.97	0.93	0.87	0.98	0.97	0.97
RNESN	NESN	ANFIS	RNESN	NESN	ANFIS	RNESN	NESN	ANFIS
Case Study 3			**Case Study 4**			**Case Study 5**		

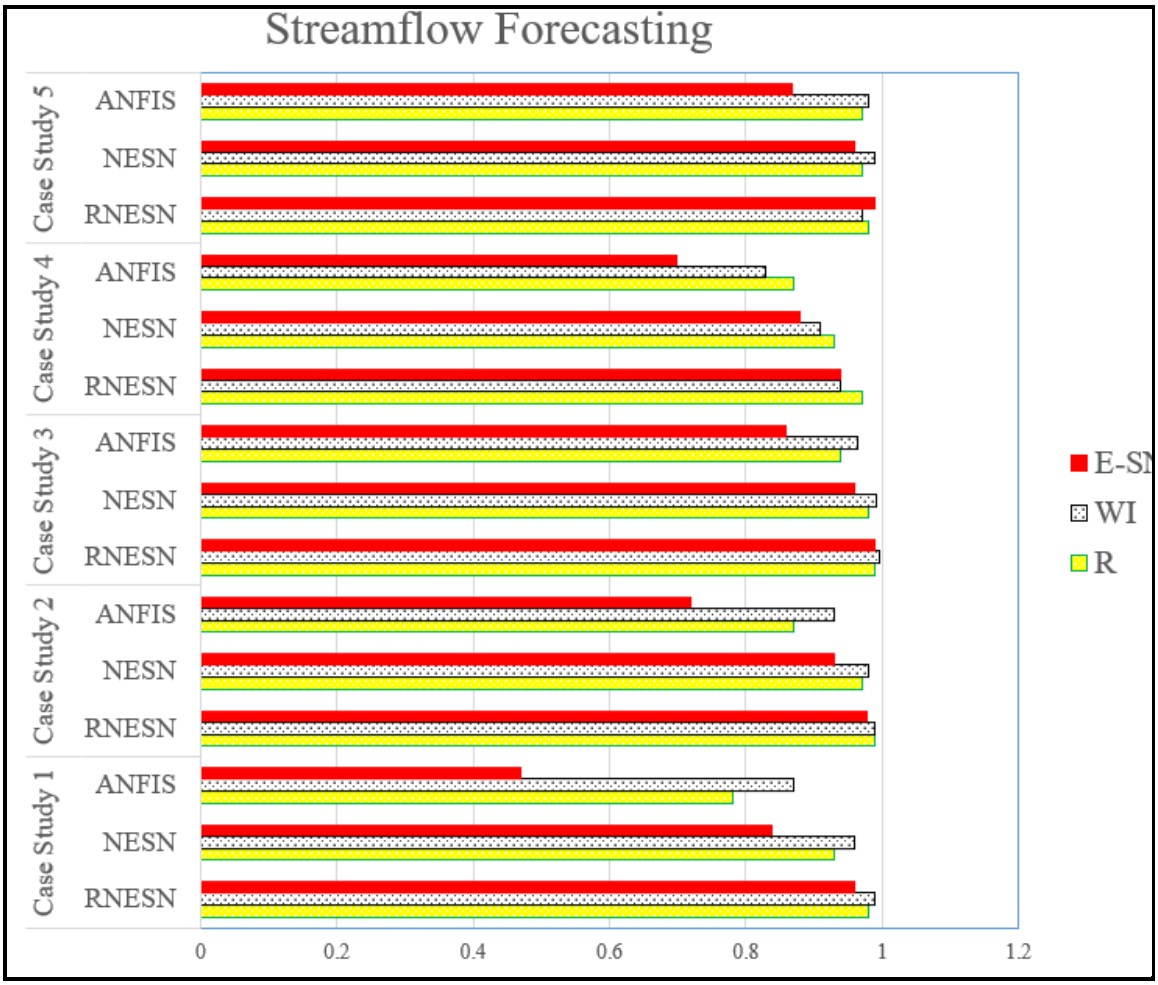


Figure a-1. statistic indices comparison for 5 different case studies (whole data set)

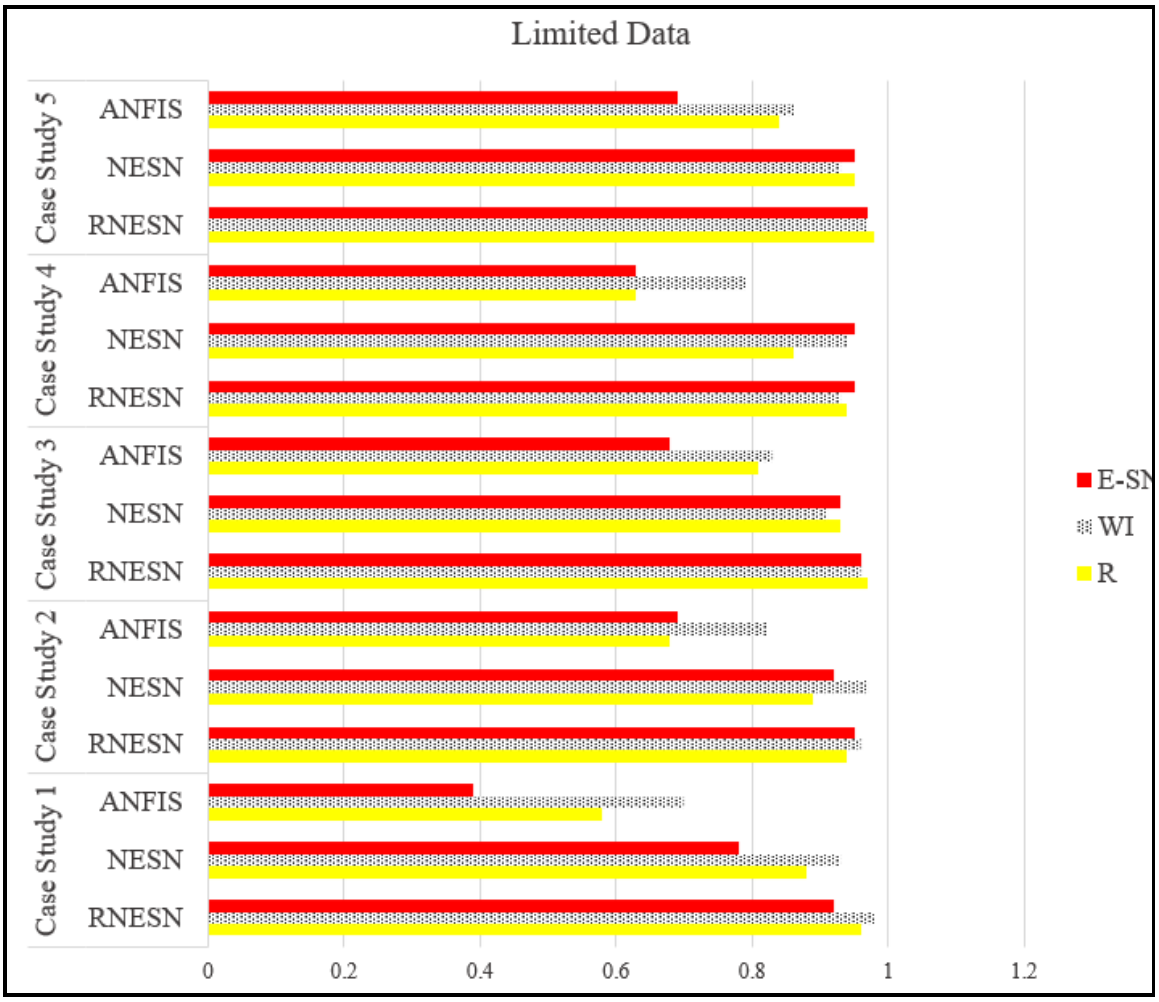


Figure a-2. statistic indices comparison for 5 different case studies (limited data)

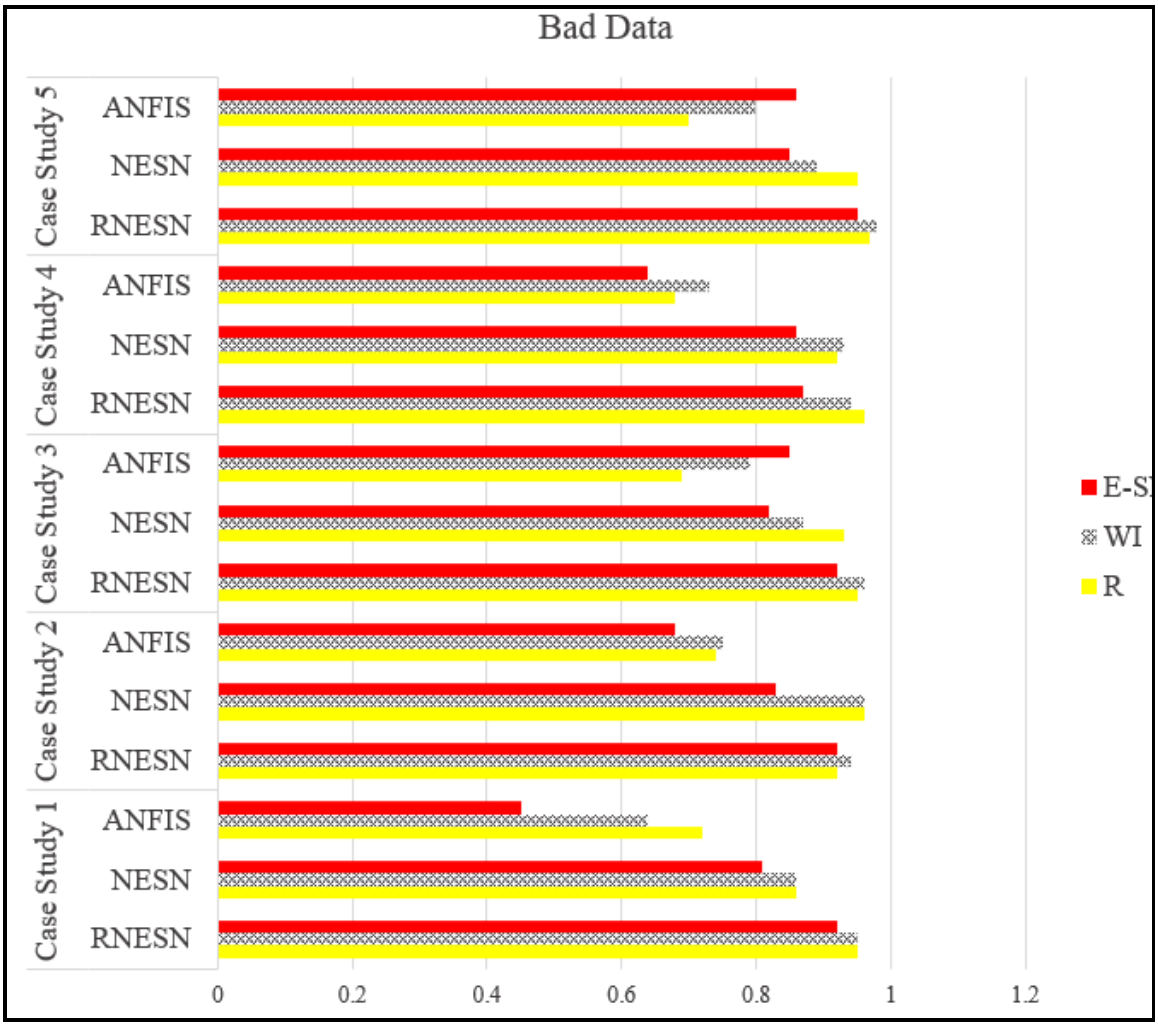


Figure a-3. statistic indices comparison for 5 different case studies (bad data)

As the Results show the model accuracy dose not deteriorate at the global scale and the proposed RNESN can be applied in different topographic region with various watershed characteristic.