

# A Sustainable Climate Forecast System for Post-processing of Precipitation With Application of Machine Learning Computations

Adel Ghazikhani (✉ [a\\_ghazikhani@yahoo.com](mailto:a_ghazikhani@yahoo.com))

Imam Reza International University

Iman Babaeian

Climatological Research Institute, Mashhad, Iran

Mohammad Gheibi

Tecnologico de Monterrey, Escuela de Ingeniería y Ciencias, Puebla, Mexico

Mostafa Hajiaghaei-Keshteli

Tecnologico de Monterrey, Escuela de Ingeniería y Ciencias, Puebla, Mexico

Amir M. Fathollahi-Fard

Department of Electrical Engineering, École de Technologie Supérieure, University of Québec, Montréal, Canada

---

## Research Article

**Keywords:** CFSV2, Post-processing, Regression, Random Forest, Decision Support System, Sustainable Development

**Posted Date:** April 22nd, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1552614/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Although many meteorological prediction models have been developed recently, but their predictions are still unreliable. Post-processing is a task for improving meteorological predictions. This study proposes a post-processing method for the Climate Forecast System Version2 (CFSV2) model. The applicability of the proposed method is shown in Iran for an observation data from 1982 to 2017. A software has been implemented which could be used to automatically perform post-processing in meteorological organizations. With application of the present study, Decision Support System (DSS) is implemented for controlling precipitation based natural side effects such as flood disaster or drought phenomenon. Likewise, it is worth noting that the mentioned DSS meets Sustainable Development Goals (SDGs) through grantee of human health and environmental protection issues. Finally, the most important section of DSS is related to prediction and in the present study it is performed by Random Forest algorithm with more than 0.87 correlation coefficient.

## 1- Introduction

Weather predictions have a great impact on everyday life. Many political, economic, environmental and social programs are linking with an accurate weather prediction. Most of people take weather predictions serious in many of their schedules from their personal and business setting. Weather predictions are usually done by hydrological numerical models. These models predict different hydrological variables outputs such as precipitation, temperature and so on. Many researchers have studied the problems related to precipitation (Bodri & Čermák, 2000; Li, Chau, Cheng, & Li, 2006; Stojanovic, Milivojevic, Ivanovic, Milivojevic, & Divac, 2013).

One of the important concepts in weather prediction models is the post-processing. Post-processing, is a task in which the prediction model is updated to eliminate the errors in the model, lack of data for building the model and large-scale limitations. Most hydrological models have a limited scale. Therefore, they don't have an exact prediction for every point on earth surface. Post-processing could help to overcome this issue for having an accurate prediction.

The research on the post-processing models and prediction algorithms, is active and there are many post-processing methods for different hydrological variables. We can classify these papers into two main categories. The first category is a group of studies which developed new prediction models for the post-processing. The second category is utilized by different post-processing algorithms for post-processing using one or multiple variables.

To study most recent papers in the first category, Monache et al. (2011) proposed two post-processing methods based on Kalman filter and weighted average on analog data (Delle Monache, Nipen, Liu, Roux, & Stull, 2011). Robertson et al. (2013) proposed a post-processing method for rain forecasts (D. E. Robertson, D. L. Shrestha, & Wang, 2013). In their method, Bayesian joint probability modeling was used to produce rain probability distributions in different locations. In this regard, ensemble forecasts are

generated by combining these probabilities using Schaake shuffle. In another study, Scheuerer et al. (2014) proposed a statistic method for post-processing temperature ensemble forecasts in COSMO-DE (Scheuerer & Büermann, 2014). Madadgar et al. (2014) proposed a novel method based on copula functions for post-processing ensemble forecasts (Madadgar, Moradkhani, & Garen, 2014). As such, Chen et al. (2014) proposed a statistical post-processing method for ensemble forecasts using a stochastic weather generator (Chen, Brissette, & Li, 2014). Scheuerer et al. (2015) proposed a post-processing method that transforms raw ensemble precipitation forecasts from the Global Ensemble Forecast System (GEFS) into probability distributions and after that a regression model is used to link the distributions (Scheuerer & Hamill, 2015). Stauffer et al. (2017) a post-processing method for daily precipitation on the standardized anomaly model output statistics is proposed (Stauffer, Umlauf, Messner, Mayr, & Zeileis, 2017). Shrestha et al. (2015) used the Bayesian joint probability and Schaake shuffle to create calibrated quantitative precipitation forecasts ensembles (Shrestha, Robertson, Bennett, & Wang, 2015). Dabernig et al. (2017) proposed a new post-processing method based on standardized anomalies (Dabernig, Mayr, Messner, & Zeileis, 2017). Rasp et al. (2018) proposed a neural network based post-processing method for temperature in Germany (Rasp & Lerch, 2018). Wutzler et al. (2018) developed a package in R language for post-processing measurements of eddy covariance flux data (Wutzler, et al., 2018). At last but not least, El Ayari et al. (2019) proposed doubly truncated Bayesian model averaging for post-processing. This method is evaluated on water level forecasts on river Rhine (Mehrez El Ayari, Stephan Hemri, & Baran, 2019).

The second category is more active and many studies are done recently. For example, Lin et al. (2020) developed the post-processed precipitation forecasts in Canada at winter from GCM model (H. Lin, Brunet, & Derome, 2008). Rincon et al. (2010) applied three post-processing methods for short term irradiance (A. Rincon, O. Jorba, & Baldasano, 2010). Vashani et al. (2010) evaluated five different post-processing methods for temperature in WRF model in Iran (S. Vashani, M. Azadi, & Hajjam, 2010). Bentzien et al. (2010) post-processed precipitation from DOSMO-DE-EPS model in Germany using regression methods (Bentzien & Friederichs, 2012). Roulin et al. (2012) utilized extended logistic regression to post-process precipitation forecasts from ECMWF model in Belgium (Roulin & Vannitsem, 2012). Verkade et al. (2013) used a post-processing method based on regression for precipitation and temperature in ECMWF ensemble (Verkade, Brown, Reggiani, & Weerts, 2013). Sweeney et al. (2013) post-processed wind speed forecasts in COSMO model using seven different adaptive post-processing algorithms (Sweeney, Lynch, & Nolan, 2013). Williams et al. (2014) evaluated four different post-processing methods for post-processing extreme events in Lorenz 1996 model (Williams, Ferro, & Kwasniok, 2014). Bogner et al. (2016) evaluated different post-processing methods for updating flood forecasts in Switzerland (K. Bogner, K. Liechti, & Zappa, 2016). Vogel et al. (2017) used Bayesian Model Averaging (BMA) and Ensemble Model Output Statistics (EMOS) for post-processing precipitation forecasts in monsoon period in West Africa (Vogel, Gneiting, Knippertz, Fink, & Schlüter, 2017). Yang et al. (2017) investigated Bayesian model averaging and heteroscedastic censored logistic regression for post-processing precipitation forecasts in the U.S. Mid-Atlantic region (Yang, Sharma, Siddique, Greybush, & Mejia, 2017). Whan et al. (2017) utilized extended logistic regression, ensemble model output statistics

and quantile random forest for post-processing precipitation forecasts (Whan & Schmeits, 2017). Erickson et al. (2018) used bias correction for post-processing SREF system forecasts for fire weather days (Erickson, Colle, & Charney, 2018). Vogel et al. (2018) applied Bayesian model averaging and ensemble model output statistics for post-processing rainfall forecasts in north Africa (Vogel, Knippertz, Fink, Schlueter, & Gneiting, 2018). Wu et al. (2018) evaluated three different variants of Schaake shuffle for post-processing precipitation forecasts (Wu, et al., 2018). Taillardat et al. (2019) performed quantile regression forest and gradient regression for post-processing precipitation forecasts in France (Taillardat, Fougères, Naveau, & Mestre, 2019).

Building on previous studies on post-processing numerical weather predictions, a method for post-processing precipitation rate predictions of CFSV2 model is proposed. In the proposed method, CFSV2 model data from 1982–2017 and observation data from 274 weather stations in Iran is used. Methods based on regression are proposed for post-processing.

The rest of the paper is organized as follows. In Section 2 the regression methods that are used for post-processing are explained. In Section 3, the proposed method is detailed. In Section 4, the experimental results are reported and in section 5 the managerial insights and sustainability issues are discussed. Finally, the Section 6 evaluates the conclusion of present study and suggestion for future studies.

## **2- Material And Methods**

Post-processing is processing output of numerical weather forecast models in order to have more reliable predictions or having predictions in areas the model does not support. In this research, the goal is to have predictions in areas that the model does not support.

### **2-1- Material**

The material section is divided to CFSV2 model and case study which are presented in the following.

#### **2-1-1- CFSV2 model**

Climate Forecast System Version2 (CFSV2) is a numerical weather prediction model that predicts a great range of weather variables(Saha, et al., 2014). The variables are in different groups including: a) Surface and Radiative fluxes variables, b) 3-D Pressure level variables, c) 3-D Ocean data variables, d) 3-D Isentropic variables. CFSV2 is an ensemble prediction system executed 16 times every day. Four of the runs are for monthly prediction for the next nine months, three runs for season forecasts and nine runs for 45-day forecasts.

#### **2-1-2- Case study**

The research is done on CFSV2 model precipitation predictions in Iran. CFSV2 is a model with monthly forecasts. The CFSV2 data used here is from 1982 to 2017. The predictions used are in the surface and radiative fluxes variables. There are 107 variables in this group. Only 90 variables which had numerical

values were used in this research as input variables. The output variable is the precipitation observation from the weather station. The precipitation observation data are from 274 weather stations all over Iran. Figure 1 shows CFSV2 precipitation predictions in Iran for different regions compared to precipitation observations.

Each precipitation prediction in CFSV2 model is for a specific year and month. The model is executed several times each day and in different days. Therefore, it has multiple predictions for each month. These predictions with 90 variables for each and the observations for the same year and month are matched together so the dataset for post-processing is created. In Figure 1, the CFSV2 precipitation predictions in Iran has been compared with observations.

## **2-2- Methods**

In this section, the methods used in the research is reported.

### **2-2-1- Problem**

Post-processing is a task done on numerical weather predictions with different purposes. One of the purposes is that some models don't have predictions in some areas due to scalability limitations. Post-processing helps to have predictions everywhere. Another goal of post-processing in enhancing the predictions.

### **2-2-1- Post-processing**

Post-processing is a task done on numerical weather predictions with different purposes. One of the purposes is that some models don't have predictions in some areas due to scalability limitations. Post-processing helps to have predictions everywhere. Another goal of post-processing in enhancing the predictions.

### **2-2-2- Preprocessing methods**

In Machine learning, preprocessing are the tasks done on data before the learning task (Salvador García, Julián Luengo, & Herrera, 2015). Preprocessing makes the data ready for learning operations. The data was investigated and two main challenges were observed, which are imbalanced data and missing values. These concepts are detailed next.

#### **2-2-2-1- Imbalanced data**

Imbalanced data is an important challenge in machine learning (He & Garcia, 2009). This challenge usually occurs in classification tasks in which data in one class is much more than data in other class. Regression is another type of learning in which imbalance may occur (Torgo, Ribeiro, Pfahringer, & Branco, 2013). Imbalance in regression means that some output values occur much more than others.

Here, the output variable is precipitation observation in weather station. It was investigated that most of the observations are zero therefore data imbalance exists. In (Torgo, et al., 2013) a preprocessing algorithm based on SMOTE (Chawla, Bowyer, Hall, & Kegelmeyer, 2002) has been proposed to handle imbalance in regression. There is a R software package for this research (Branco, 2013).

### **2-2-2-2- Missed values**

Missed values is another challenge in machine learning, in which some features don't have values due to problems in data acquisition (W.-C. Lin & Tsai, 2020). Missed values could be handled by different methods. Here chained Equations are used to impute missed values (van Buuren & Groothuis-Oudshoorn, 2011). In (van Buuren & Groothuis-Oudshoorn, 2011), a R software package has been developed for imputing missed values using chained Equations.

### **2-2-2-3- Feature selection**

Feature selection is one of the most important preprocessing tasks in machine learning (Alpaydin, 2010). In feature selection it is aimed to reduce the dimensions of the learning problem. There are different methods for feature selection. Here, a filter method based on Pearson correlation was used to find the correlation between each variable and the observation. Variables that had low correlation were omitted.

As mentioned earlier, in the CFSV2 data there are 90 variables. After performing the feature selection method, the variables were reduced to 47. Therefore, the time for learning was reduced.

### **2-2-3- Regression methods**

Numerical weather predictions are usually continuous values and the post-processing method aims to change these forecasts to another continuous value. With this explanation, regression methods are a suitable mechanism for post-processing. In regression methods, the predicted variable is continuous (Alpaydin, 2010). In the next sections different regression methods used in this research are explained.

#### **2-2-3-1- General Regression Neural Network (GRNN)**

GRNN is a memory-based neural network suitable for linear and non-linear regression tasks (Specht, 1991). GRNN is built up of three layers, which are: pattern layer, summation layer and output layer. In the pattern layer, each neuron is a cluster center and the similarity of input to each cluster is computed. The summation layer sums up the result of the pattern layer and the output layer gives the final prediction.

#### **2-2-3-2- Extreme Learning Machine (ELM)**

ELM is a type of neural network in which the hidden layer weights are not trained and have random values (Huang, Zhu, & Siew, 2006). ELM can have multiple hidden layers. The output layer in ELM has weights and only these weights are trained. This enables ELM to estimate weights with an Equation and

no need to use backpropagation algorithm. ELM has faster training and doesn't fall into local minimums. ELM can be used for regression.

### **2-2-3-3- Neural Network (NN)**

Neural networks are a popular learning algorithm (Warren McCulloch & Pitts, 1943). Here a Multi-Layer Perceptron (MLP) is used for regression. The hidden layer has 50 neurons with tangent sigmoid activation. The output layer has one neuron with linear activation. The output layer neuron gives the final prediction of the network. Backpropagation is used for training the MLP.

### **2-2-3-4- Binary Regression Tree (BRT)**

Binary regression trees are a type of decision tree for regression (L. Breiman, J. Friedman, R. Olshen, & Stone, 1984). In this decision tree the nodes are divided based on limits on feature values. The features are selected based on GINI index. The learning function is recursive and the operation done on each leaf of the tree are the same. The training stops when there are no more leaves to extend and all leaves are labels not features.

### **2-2-3-5- Random Forest (RF)**

Random forest is an ensemble of decision trees which are combined based on Bagging approach (Breiman, 2001). In bagging, each learner gives a prediction or vote, and the result prediction is majority of votes (Kuncheva, 2004). In building each tree, random forest has a special strategy. It selects one of the attributes randomly. That's where the word random comes from.

### **2-2-3-6- Lasso Boosting (LB)**

Lasso Boosting is an ensemble of decision trees which are combined using boosting method (Zhao & Yu, 2004). It belongs to a big family of learners called "Gradient Boosting" methods. In boosting, the general idea is to start from a weak learner and try to enhance it iteratively based on the error in each iteration (Kuncheva, 2004). Lasso, generally is an iterative optimization method. In Lasso Boosting, Lasso is used in combination with boosting to optimize the training procedure.

## **3- Results And Discussions**

In this section experiments were conducted to evaluate the effectiveness of the proposed method. Finally, after evaluations of the outcomes with other researches, three different sections contain sustainability and climate change, Decision Support System (DSS) with focusing on managerial insights, and

### **3-1- Metrics**

Four different metrics were used to evaluate the results which are RMSE (Equation 1), Pearson Correlation (Equation 2), ROC analysis plot and Q-Q plot.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_t - y_p)^2} \quad (\text{Equation 1})$$

$$\text{Pearson Corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (\text{Equation 2})$$

K-fold cross validation with K=10, was used to compute the metrics. In Kfold the dataset is divided into  $K$  parts and  $(K-1)$  parts are used for training and one part for testing and this process is repeated  $K$  times. The final evaluation is the mean of  $K$  times of execution.

### 3-2- Results

In this section the results for the metrics are reported.

#### 3-2-1- RMSE and correlation metric

In Table 1, the RMSE and correlation results are shown. The results are the mean of 10 executions.

**Table 1.** The outcomes of RSME through the present study.

| Method | RSME         | Pear_Corr   |
|--------|--------------|-------------|
| GRNN   | 41.99        | 0.67        |
| NN     | 41.79        | 0.58        |
| ELM    | 51.19        | 0.15        |
| BRT    | 36.81        | 0.74        |
| RF     | <b>25.94</b> | <b>0.87</b> |
| LB     | 33.02        | 0.77        |

From the results of Table 1, it is concluded that in this data, tree based methods (BRT, RF, LB) have better results than neural network methods (GRNN, NN, ELM). Among the tree based methods Random Forest has the best results.

#### 3-2-2- ROC curve

The ROC plot is a metric used for classification problems. Here the problem is a regression problem, therefore it needs to be converted. To achieve this, the predictions and observations are categorized into three groups Below Normal (BN), Normal (NN), Above Normal (AN). Below normal means the precipitation is less than 80 percent of average long-term reforecast precipitation. Normal means precipitation is between 80 percent and 120 percent average and above normal means precipitation is higher than 120 percent average. In Figure 2, the result without post-processing is shown.

From Figure 2 it could be concluded that CFSV2 doesn't have suitable predictions on observations above normal. In Figure 3, the ROC plots for six post-processing algorithms is shown.

From Figure 3 it could be concluded that the post-processing algorithms have improved the CFSV2 predictions in BN and NN categories. For observations above normal all methods except ELM have results similar to CFSV2. This issue is related to low precipitation in Iran. Between the six post-processing algorithms, RF has the best ROC plot result.

### **3-2-3- Q-Q plot**

The Q-Q plot is used to investigate if two sample data are from the same distribution. Here, the two sample data are the observations and predictions. If the observation and predictions come from the same distribution the result is a linear plot. In Figure 4, the Q-Q plot is shown for predictions before post-processing.

From Figure 4, it could be concluded that CFSV2 prediction don't have similar distribution to the observations. In Figure 5, the Q-Q plots after post-processing is shown.

From Figure 5 it could be concluded that the post-processing algorithms have improved the CFSV2 predictions. Between the six algorithms, GRNN and BRT have the best Q-Q plot result.

## **3-3- Sensitivity analysis**

In this section, the sensitivity of the learned post-processing algorithms was analyzed. In this analysis, CFSV2 precipitation predictions and observation data from Iran weather stations in 2018 were collected and used.

### **3-3-1- ROC plot**

The ROC plot for CFSV2 precipitation predictions before post-processing is shown in Figure 6.

From Figure 6 it could be concluded that CFSV2 has better predictions for the BN category.

Figure 7 shows that post-processing algorithms have improved CFSV2 predictions in BN category. RF and LB have better results compared to other algorithms.

The sensitivity analysis of learned post-processing algorithms with ROC metric on CFSV2 data in 2018 has similar results to the main results in 1982-2017.

### **3-3-2- Q-Q plot**

The Q-Q plot before post-processing is shown in Figure 8.

From Figure 8, it could be concluded that CFSV2 predictions have similar distribution to the observations approximately. The Q-Q plot for post-processing algorithms are shown in Figure 9.

From Figure 9 it could be concluded that the post-processing algorithms have improved the CFSV2 predictions. Between the six algorithms, RF and LB have the best Q-Q plot result.

The sensitivity analysis of learned post-processing algorithms with Q-Q plot on CFSV2 data in 2018 doesn't have similar results to the main results in 1982-2017.

## 4- Implemented Software

A software was developed for Iran Meteorological Organization (IMO) to use this research for post-processing in practice (Figure 10). The software was designed and implemented using MATLAB 2017 and Mysql database version 8. The functions are implemented in MATLAB 2017 and the data are stored in Mysql database. This software could be used to perform the post-processing tasks in IMO.

The main part of this software is the automatic post-processing (Figure 11). Automatic post-processing means that by pressing the start button, the software starts to download CFSV2 model predictions from the site and saves them in the specified path and then the post-processing function is called and it is performed for the specified regions. The result of post-processing is saved in the specified path as excel files and maps.

Whenever this process stops, in result of software or hardware reasons. The process could be continued after restarting the software.

The last part of the software is Maps (Figure 12). In this tab the post-processing outputs could be viewed as maps. There are four types of maps.

Through the research of Khan et al. (2021), applied NASA's Goddard Earth Observing System (GEOS) for prediction of precipitations in different climates with concentration on El Niño and the best correlation coefficient of the mentioned study was around 0.6 which is less than outcomes of the present research [45]. Therefore, this result shows the validity of this study's outputs. Plus, Andrade et al., (2021) presented a novel method for prediction of rainfall through the El Niño phenomenon in Africa with application of three method contain European Centre for Medium-Range Weather Forecasts (ECMWF), National Centers for Environmental Prediction (NCEP), and Met Office (UKMO) models [46]. The results of their study demonstrated that with the declared models, the unpredictable El Niño event can predict with more than 0.9 correlation coefficient. But, the mentioned computation systems are black-box models and it is not flexible for benchmarking in the undefined situations. But, this study focused on application of programming for estimation of precipitation and it is so appropriate from economic and execution aspects. Likewise, Peng et al., (2021) evaluated combination of metaheuristic algorithm and machine learning computations for prediction of precipitation. The outcomes illustrated that efficiency of participation forecasting system by Deep neural networks (DNN) has acceptable precision [47]. But, because of integration machine learning and optimization methods, volume of computations is so huge and in the following, run time is extended. Besides, in the present investigation, the speed of computations is so high and it can be utilized as a real time soft-sensor.

## **5- A Discussion On Sustainability Issues**

Through this section meeting the Sustainable Development Goals (SDGs) is appraised. Then, the application of Decision Support System (DSS), and research position are argued.

### **5-1- Sustainability**

According to Figure 13, through implementation of the present study's outcomes, precipitation can be predicted with high efficiency and then in the following, drought and flood disasters can be controlled. By the outputs, human risks, environmental and infrastructure damages are reduced. Finally, after execution of the early reaction systems, social satisfactions are increased and they trust to local government. Therefore, as well, technical aspects, the results of this investigation help with social and economic aspects. Finally, two aspects of SDGs include Sustainable Cities and Communities (Ghadami, et al., 2021; Shahsavar, et al., 2021) and Good Health and Well Being (Amini, Arab, Faramarz, Ghazikhani, & Gheibi, 2021; Mohammadi, et al., 2021) are met.

### **5-2- DSS concept**

One of the main goals through this study is related to implementation of DSS for monitoring, prediction and controlling side-effects of increased and decreased precipitation events such as drought and flood. The conceptual model of the DSS is illustrated as per Figure 14. Through the mentioned DSS, monitoring section is organized by online/offline achieved data. Then, the efficiency of GRNN, NN, ELM, BRT, RF, and LB are assessed by rainfall data and the best algorithm is selected for future estimation. Finally, by predicted precipitation amounts, alarm management are done based on comparison with thresholds. While, the thresholds are determined in specific values which are variances in different regions.

According to World Bank database, Iran is divided into 6 main watersheds (Figure 15) and the combination of temperature and precipitation diagrams of the mentioned zones from 1991 to 2020 are presented according to Figure 16 (a-f). The Figures express that Iran has lots of fluctuations throughout the whole period. Therefore, prediction of rainfall in Iran is so complex and this post-processing system operates multifaceted problem through climatology issues.

### **5-3- Importance of viewpoints**

For determination of study position in scientific communications, the library evaluation is operated by the application of VOSviewer software and Scopus database. Whereas, for the declared goals, the precipitations and machine learning keywords are documented with simple searching and then the outputs are filtered by authors (Figure 17-a), country (Figure 17-b), and keywords occurrences (Figure 17-c). Based on Figure 17-a, Y. Liu, X. Zhang, and Y. Zhang contribute more than more studies about the usage of the machine learning in precipitation estimation research area. Also, as per Figure 17-b, United State and China published the most documents in the declared field. While, rainfall estimation is hot issue in Iran which is illustrated in Figure 17-b. Finally, according to Figure 17-c integration of machine

learning and precipitation issues are combined by climate changes as a novel subject which is suggested by this investigation.

Based on Figure 18 which is provided from World Bank datacenter, the distribution of rainfall in the world map demonstrated that Iran is located in the arid area through the time. Thus, the exact prediction of precipitation is so necessary in the case study and based on flood/drought phenomenon occurrences in Iran, establishment of high efficient forecasting system is assumed as crucial implementation. Considering outcomes of this study, it is clear that the declared gap can be filled.

Although this study provided a strong prediction model in comparison with majority of previous literature, there are many limitations which can help us for future work. First, the proposed model may be extended by other hydrological numerical models (Lawrence & Hosein, 2019). Finally, our prediction models can be combined by recent advances in swarm intelligence and computational methods (Fathollahi-Fard, Hajiaghaei-Keshteli, & Tavakkoli-Moghaddam, 2018, 2020) to improve the accuracy and robustness of our model.

## **6- Conclusion**

Weather predictions are an important issue in everyday life. Hydrological numerical predictions have errors and are sometimes unreliable. Post-processing methods could be used to manage this issue. Increasing the scale of predictions is another goal for post-processing. In this study regression methods are used for improving CFSV2 precipitations in Iran. CFSV2 predictions and weather station observations in 1982-2017 build up the data. Generalized regression neural network, neural network, extreme learning machine, binary regression tree, random forest and lasso boosting are the methods used for post-processing. The results show improvements in predictions with different metrics. Random forest shows better results in RMSE and correlation and ROC plot. Generalized regression neural network and Binary regression tree show better results in Q-Q plot.

Finally, the sensitivity of learned models was analyzed. The analysis is done for CFSV2 predictions and weather station observations in 2018. The results are approximately similar to the results in 1982-2017 with some minor differences.

For future studies, this research suggests to application of metaheuristic algorithms for optimization of machine learning errors through the precipitation process. Also, after forecasting precipitation, Multi Criteria Decision Making (MCDM) techniques can be coupled with machine learning computations for online decision making through flood or drought controlling systems.

## **Declarations**

### **Ethical Approval**

Author declare that they don't have conflict of interest.

## Consent to Publish

In this paper all authors agree with publish this paper in title “A SUSTAINABLE CLIMATE FORECAST SYSTEM FOR POST-PROCESSING OF PRECIPITATION WITH APPLICATION OF MACHINE LEARNING COMPUTATIONS” in the Theoretical and Applied Climatology journal.

## Authors Contributions

In this paper Adel Ghazikhani and Iman Babaeian devised the project, the main conceptual ideas and proof outline. Adel, Iman, and Mohammad Gheibi worked out almost all of the technical details, mathematical formulation and performed the prediction systems. Mostafa Hajiaghaei-Keshteli contributed to do comparison of proposed algorithms and revised the manuscript from technical and grammatical aspects. Also, Amir M. Fathollahi-Fard supports the scientific and grammatical issues of manuscripts. All authors discussed the results and contributed to the final manuscript.

## Funding

In this paper our financial issues and requirements are supplied by Imam Reza International University.

## Competing Interests

Author declare that they don't have conflict of interest.

## Availability of data and materials

In this paper the data of CFSV2 predictions are collected from the official website and the observation data are gathered from weather stations in Iran. Finally, all requirement data is available for declaration.

## Acknowledgements

No Acknowledgements.

## References

1. A. Rincon, O. Jorba, & Baldasano, J. M. (2010). Development of a short-term irradiance prediction system using post-processing tools on WRF-ARW meteorological forecasts in Spain. In *European Conf. on Applied Meteorology* (Vol. 7). Zurich, Switzerland, European Meteorological Society.
2. Alpaydin, E. (2010). *Introduction to Machine Learning Second Edition*: The MIT Press.
3. Amini, M. H., Arab, M., Faramarz, M. G., Ghazikhani, A., & Gheibi, M. (2021). Presenting a soft sensor for monitoring and controlling well health and pump performance using machine learning, statistical analysis, and Petri net modeling. *Environ Sci Pollut Res Int*.
4. Bentzien, S., & Friederichs, P. (2012). Ensemble postprocessing for probabilistic quantitative precipitation forecasts. In *AGU Fall Meeting Abstracts* (Vol. 2012).

5. Bodri, L., & Čermák, V. (2000). Prediction of extreme precipitation using a neural network: application to summer flood occurrence in Moravia. *Advances in Engineering Software*, *31*, 311–321.
6. Branco, P. (2013). Smote for Regression R Package. In.
7. Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 27.
8. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *16*, 321–357.
9. Chen, J., Brissette, F. P., & Li, Z. (2014). Postprocessing of Ensemble Weather Forecasts Using a Stochastic Weather Generator. *Monthly Weather Review*, *142*, 1106–1124.
10. D. E. Robertson, D. L. Shrestha, & Wang, Q. J. (2013). Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting. *Hydrology and Earth System Sciences*, *17*, 17.
11. Dabernig, M., Mayr, G. J., Messner, J. W., & Zeileis, A. (2017). Spatial ensemble post-processing with standardized anomalies. *Quarterly Journal of the Royal Meteorological Society*, *143*, 909–916.
12. Delle Monache, L., Nipen, T., Liu, Y., Roux, G., & Stull, R. (2011). Kalman Filter and Analog Schemes to Postprocess Numerical Weather Predictions. *Monthly Weather Review*, *139*, 3554–3570.
13. Erickson, M. J., Colle, B. A., & Charney, J. J. (2018). Evaluation and Postprocessing of Ensemble Fire Weather Predictions over the Northeast United States. *Journal Of Applied Meteorology and Climatology*, *57*, 1135–1153.
14. Fathollahi-Fard, A. M., Hajiaghaei-Keshteli, M., & Tavakkoli-Moghaddam, R. (2018). The Social Engineering Optimizer (SEO). *Engineering Applications of Artificial Intelligence*, *72*, 267–293.
15. Fathollahi-Fard, A. M., Hajiaghaei-Keshteli, M., & Tavakkoli-Moghaddam, R. (2020). Red deer algorithm (RDA): a new nature-inspired meta-heuristic. *Soft Computing*, *24*, 14637–14665.
16. Ghadami, N., Gheibi, M., Kian, Z., Faramarz, M. G., Naghedi, R., Eftekhari, M., Fathollahi-Fard, A. M., Dulebenets, M. A., & Tian, G. (2021). Implementation of solar energy in smart cities using an integration of artificial neural network, photovoltaic system and classical Delphi methods. *Sustainable Cities and Society*, *74*, 103149.
17. He, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, *21*, 1263–1284.
18. Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, *70*, 489–501.
19. K. Bogner, K. Liechti, & Zappa, M. (2016). Post-Processing of Stream Flows in Switzerland with an Emphasis on Low Flows and Floods. *Water*, *8*.
20. Kuncheva, L. I. (2004). *Combining Pattern Classifiers Methods and Algorithms*: JOHN WILEY & SONS.
21. L. Breiman, J. Friedman, R. Olshen, & Stone, C. (1984). *Classification and Regression Trees*. Boca Raton, FL: CRC Press.

22. Lawrence, T., & Hosein, P. (2019). Stochastic dynamic programming heuristics for influence maximization–revenue optimization. *International Journal of Data Science and Analytics*, *8*, 1–14.
23. Li, X.-Y., Chau, K. W., Cheng, C.-T., & Li, Y. S. (2006). A Web-based flood forecasting system for Shuangpai region. *Advances in Engineering Software*, *37*, 146–158.
24. Lin, H., Brunet, G., & Derome, J. (2008). Seasonal Forecasts of Canadian Winter Precipitation by Postprocessing GCM Integrations. *Monthly Weather Review*, *136*, 769–783.
25. Lin, W.-C., & Tsai, C.-F. (2020). Missing value imputation: a review and analysis of the literature (2006–2017). *Artificial Intelligence Review*, *53*, 1487–1509.
26. Madadgar, S., Moradkhani, H., & Garen, D. (2014). Towards improved post-processing of hydrologic forecast ensembles. *Hydrological Processes*, *28*, 104–122.
27. Mehrez El Ayari, Stephan Hemri, & Baran, S. (2019). Statistical post-processing of hydrological forecasts using Bayesian model averaging. *Geophysical Research Abstracts*, *21*.
28. Mohammadi, M., Gheibi, M., Fathollahi-Fard, A. M., Eftekhari, M., Kian, Z., & Tian, G. (2021). A hybrid computational intelligence approach for bioremediation of amoxicillin based on fungus activities from soil resources and aflatoxin B1 controls. *J Environ Manage*, *299*, 113594.
29. Rasp, S., & Lerch, S. (2018). Neural Networks for Postprocessing Ensemble Weather Forecasts. *Monthly Weather Review*, *146*, 3885–3900.
30. Roulin, E., & Vannitsem, S. (2012). Postprocessing of Ensemble Precipitation Predictions with Extended Logistic Regression Based on Hindcasts. *Monthly Weather Review*, *140*, 874–888.
31. S. Vashani, M. Azadi, & Hajjam, S. (2010). Comparative Evaluation of Different Post Processing Methods for Numerical Weather Prediction of Temperature Forecasts over Iran. *Research Journal of Environmental Sciences*, *4*, 305–316.
32. Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-y., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., van den Dool, H., Zhang, Q., Wang, W., Chen, M., & Becker, E. (2014). The NCEP Climate Forecast System Version 2. *Journal of Climate*, *27*, 2185–2208.
33. Salvador García, Julián Luengo, & Herrera, F. (2015). *Data Preprocessing in Data Mining*. Springer.
34. Scheuerer, M., & Büermann, L. (2014). Spatially adaptive post-processing of ensemble forecasts for temperature. *Journal of the Royal Statistical Society Series C Applied Statistics*, *63*, 405–422.
35. Scheuerer, M., & Hamill, T. M. (2015). Statistical Postprocessing of Ensemble Precipitation Forecasts by Fitting Censored, Shifted Gamma Distributions. *Monthly Weather Review*, *143*, 4578–4596.
36. Shahsavari, M. M., Akrami, M., Gheibi, M., Kavianpour, B., Fathollahi-Fard, A. M., & Behzadian, K. (2021). Constructing a smart framework for supplying the biogas energy in green buildings using an integration of response surface methodology, artificial intelligence and petri net modelling. *Energy Conversion and Management*, *248*, 114794.
37. Shrestha, D. L., Robertson, D. E., Bennett, J. C., & Wang, Q. J. (2015). Improving Precipitation Forecasts by Generating Ensembles through Postprocessing. *Monthly Weather Review*, *143*, 3642–3663.

38. Specht, D. F. (1991). A general regression neural network. *IEEE Transaction Neural Networks*, *2*, 568–576.
39. Stauffer, R., Umlauf, N., Messner, J. W., Mayr, G. J., & Zeileis, A. (2017). Ensemble Postprocessing of Daily Precipitation Sums over Complex Terrain Using Censored High-Resolution Standardized Anomalies. *Monthly Weather Review*, *145*, 955–969.
40. Stojanovic, B., Milivojevic, M., Ivanovic, M., Milivojevic, N., & Divac, D. (2013). Adaptive system for dam behavior modeling based on linear regression and genetic algorithms. *Advances in Engineering Software*, *65*, 182–190.
41. Sweeney, C. P., Lynch, P., & Nolan, P. (2013). Reducing errors of wind speed forecasts by an optimal combination of post-processing methods. *Meteorological Applications*, *20*, 32–40.
42. Taillardat, M., Fougères, A.-L., Naveau, P., & Mestre, O. (2019). Forest-Based and Semiparametric Methods for the Postprocessing of Rainfall Ensemble Forecasting. *Weather and Forecasting*, *34*, 617–634.
43. Torgo, L., Ribeiro, R. P., Pfahringer, B., & Branco, P. (2013). SMOTE for Regression. In (pp. 378–389). Berlin, Heidelberg: Springer Berlin Heidelberg.
44. van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *2011, 45, 67%* *Journal of Statistical Software*.
45. Verkade, J. S., Brown, J. D., Reggiani, P., & Weerts, A. H. (2013). Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *Journal of Hydrology*, *501*, 73–91.
46. Vogel, P., Gneiting, T., Knippertz, P., Fink, A. H., & Schlüter, A. (2017). Statistical ensemble postprocessing for precipitation forecasting during the West African Monsoon. In *EGU General Assembly Conference Abstracts* (pp. 14208).
47. Vogel, P., Knippertz, P., Fink, A. H., Schlueter, A., & Gneiting, T. (2018). Skill of Global Raw and Postprocessed Ensemble Predictions of Rainfall over Northern Tropical Africa. *Weather and Forecasting*, *33*, 369–388.
48. Warren McCulloch, & Pitts, W. (1943). A Logical Calculus of Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, *5*, 18.
49. Whan, K., & Schmeits, M. (2017). Probabilistic forecasts of extreme local precipitation using HARMONIE predictors and comparing 3 different post-processing methods. In *EGU General Assembly Conference Abstracts* (pp. 5596).
50. Williams, R. M., Ferro, C. A. T., & Kwasniok, F. (2014). A comparison of ensemble post-processing methods for extreme events. *Quarterly Journal of the Royal Meteorological Society*, *140*, 1112–1120.
51. Wu, L., Zhang, Y., Adams, T., Lee, H., Liu, Y., & Schaake, J. (2018). Comparative Evaluation of Three Schaake Shuffle Schemes in Postprocessing GEFS Precipitation Ensemble Forecasts. *Journal Of Hydrometeorology*, *19*, 575–598.
52. Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Šigut, L., Menzer, O., & Reichstein, M. (2018). Basic and extensible post-processing of eddy covariance flux data with REdDyProc.

*Biogeosciences*, 15, 5015–5030.

53. Yang, X., Sharma, S., Siddique, R., Greybush, S. J., & Mejia, A. (2017). Postprocessing of GEFS Precipitation Ensemble Reforecasts over the U.S. Mid-Atlantic Region. *Monthly Weather Review*, 145, 1641–1658.
54. Zhao, P., & Yu, B. (2004). *Boosted Lasso*: california univ berkeley Department of statistics.

## Figures

Figure 1

Left: CFSV2 predictions in January 2017, Right: Weather station observations in January 2017.

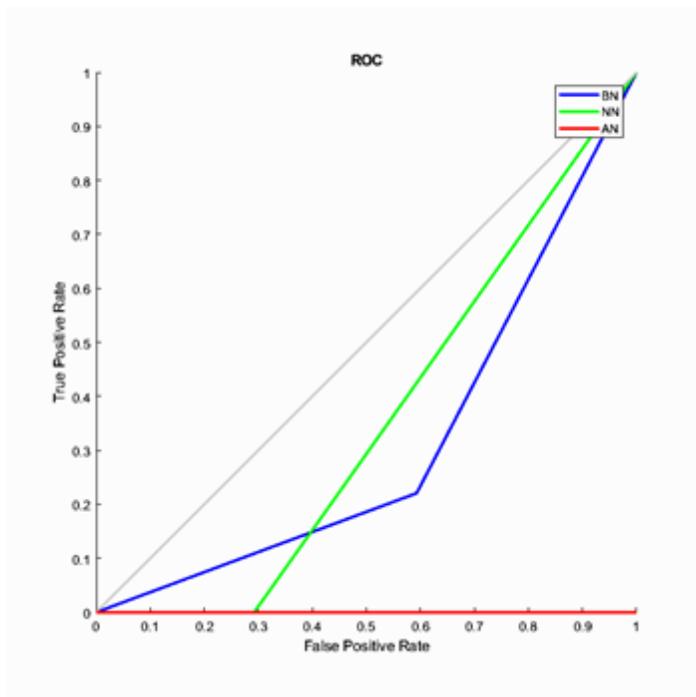
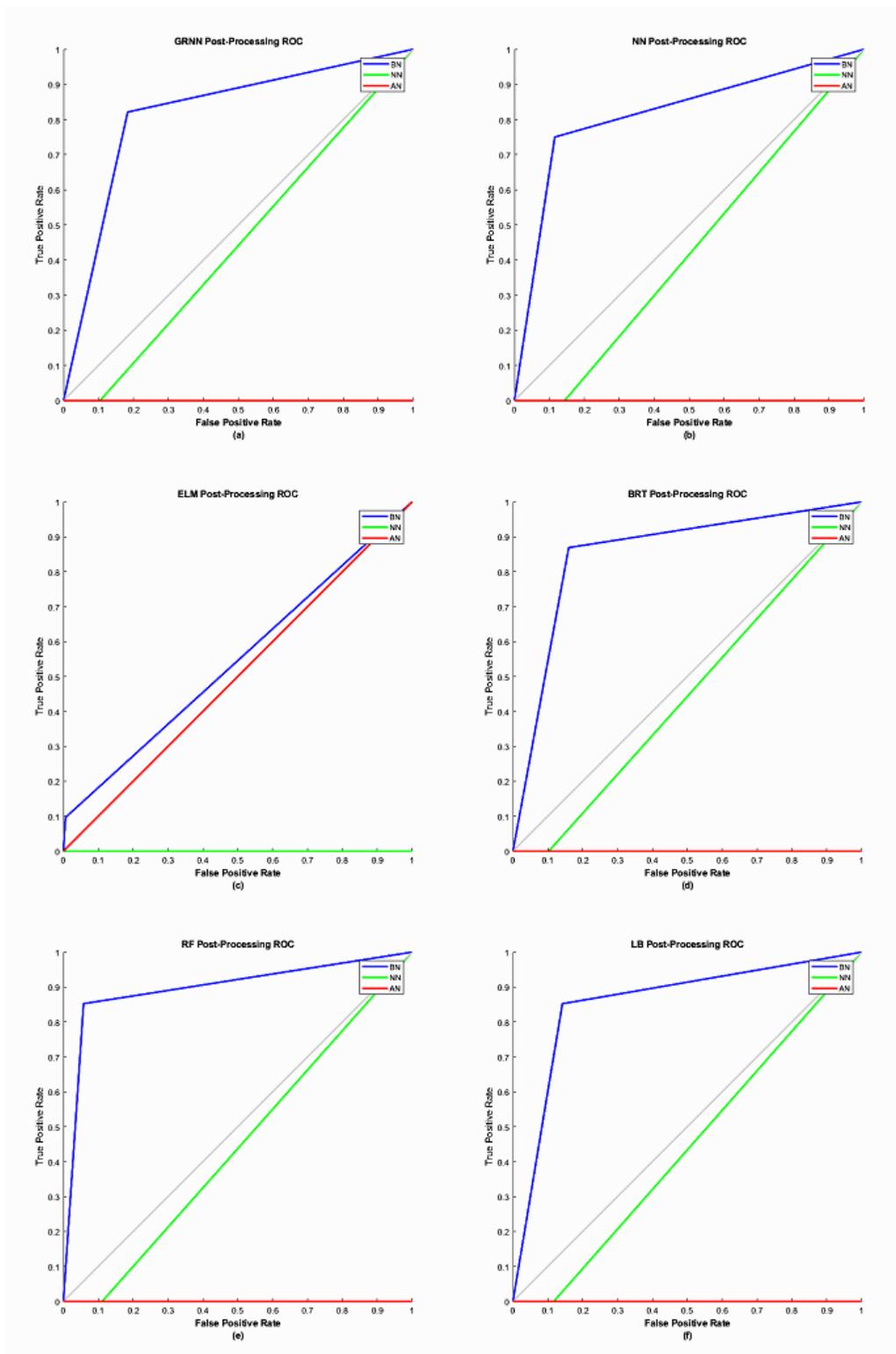


Figure 2

ROC plot for CFSV2 predictions compared to observations.



**Figure 3**

ROC plot for six Post-Processing algorithms. (a) GRNN, (b) NN, (c) ELM, (d) BRT, (e) RF and (f) LB.

## Figure 5

Q-Q plot for six Post-Processing algorithms. (a) GRNN, (b) NN, (c) ELM, (d) BRT, (e) RF and (f) LB.

## Figure 14

The Decision Support System in the present study.