Hybrid Deep Learning Approach in Predictability of Monthly Rainfall Using Endogenous property and Global Climatic Indices

Submitted by

Mandala Venkata Surendra

17CE33003

Under the supervision of

Prof. Rajib Maity

Department of Civil Engineering
Indian Institute of Technology Kharagpur
Kharagpur-721302, India

2021

CERTIFICATE

This is to certify that the Dissertation Report entitled, "**Hybrid Deep Learning Approach in Predictability of Monthly Rainfall Using Endogenous property and Global Climatic Indices**" submitted by **Mr. Mandala Venkata Surendra** to the Indian Institute of Technology, Kharagpur, India, is a record of bonafide Project work carried out by him under my supervision and guidance and is worthy of consideration for the award of the degree of Bachelor of Technology in Civil Engineering of the Institute. The Dissertation Report has fulfilled all the requirements as per the regulations of the institute and in my opinion, reached the standard for submission.

Abstract:

Among various Spatio-temporal scales, seasonal or monthly prediction of rainfall over a subdivision prediction is one of the most important tasks for agricultural and water resources management for a region. In general, climate and rainfall are highly non-linear and complicated phenomena that require advanced modeling strategies and improved simulation for acceptable prediction accuracy. One of the other major issues is that climate-induced time-varying characteristics are decreasing the performance of the model. So considering the time-varying association between the Indian Summer Monsoon Rainfall and Large scale climatic Indices(e.g. Pacific Decadal Oscillation, El Niño Modoki Index, North Atlantic Oscillation, Equatorial Indian Ocean Oscillation), we are going to propose a Hybrid DL approach, a combination of one-dimensional Convolutional Neural Network (Conv1D) and Multi-Layer Perceptron (MLP) (hereinafter referred to as hybrid Conv1D-MLP model), to capture the endogenous properties of sub-divisional monthly rainfall series as well as using the association of large scale indices with Indian rainfall with a goal for an improved prediction performance.The hybrid models are developed for five homogeneous monsoon regions (HMRs) of India, applied on future data and future trends are observed. The model performance is assessed through mean absolute error, root mean square error (RMSE), Pearson correlation (r), and Nash Sutcliffe coefficient of efficiency (NSE). Overall, this study establishes the potential of the proposed hybrid Conv1D-MLP model in capturing the hidden complex endogenous characteristics of regional monthly rainfall and also an association with large-scale indices that is effective for a reliable prediction performance.

Keywords

Rainfall prediction; Deep learning; Convolutional Neural Network; Large Scale Climatic Indices; Spatial Variability; homogeneous monsoon regions (HMRs).

1. INTRODUCTION

In India, annual rainfall is mainly accounted from the monsoon season i.e during months of June, July, August, September also referred to as summer monsoon rainfall. The spatiotemporal variability in the Indian Summer Monsoon Rainfall (ISMR) is linked with the atmospheric circulation patterns through hydroclimatic teleconnection (Kahya and Dracup, 1993; Ashok et al., 2001, 2004; Maity and Nagesh, 2008). The important teleconnection patterns known to influence the variability of summer monsoon rainfall are El Niño-Southern Oscillation (ENSO), Equatorial Indian Ocean Oscillation (EQUINOO), North Atlantic Oscillation (NAO), Pacific Decadal Oscillation (PDO), El Niño Modoki Index (EMI) (Nair et al., 2018). These large-scale indices are associated with ISMR at lead times of months to seasons and hence, used for long-range prediction of ISMR. However, the nature of association between different large-scale climatic indices and rainfall varies with both space and time. Therefore, information on the temporal evolution of large-scale indices and their impact may provide a better understanding of the spatiotemporal variability in the monthly rainfall. This is the focus of this study.

Several observational and modelling studies established the teleconnection pattern between the large scale indices and the summer monsoon rainfall (Pant and Parthasarathy, 1981; Rasmusson and Carpenter, 1983; Ju and Slingo, 1995; Meehl and Arblaster, 1998; Kumar et al., 1999). ENSO mode, for instance, is identified as the third largest component of Asian summer monsoon variations.

Another driver of rainfall variability in this region is EQUINOO, which is the atmospheric component of Indian Ocean Dipole (IOD) mode (Kumar et al., 2007; Rajeevan et al., 2007; Francis and Gadgil, 2010; Charlotte and Mathew, 2012). Recent meteorological observations indicate a strong link between ISMR and EQUINOO due to the association of large-scale monsoon rainfall over the Indian region with the northward propagation of convective system generated over the Indian Ocean region (Gadgil et al., 2004; Gadgil and Gadgil, 2006).

Another climatic index associated with ISMR is NAO, which is a temporal fluctuation of the zonal wind strength across the Atlantic Ocean due to pressure variations in the subtropical anticyclone belt and in the subpolar low near Iceland. PDO is another large-scale, strongly associated climatic index that modulates the ISMR–ENSO relation.The mechanism by which the PDO could affect the monsoon was hypothesized by Krishnamurthy and Krishnamurthy (2014). PDO index exhibits significant negative correlation with ISMR, similar to the relation between ISMR and ENSO. Lastly, EMI, warming in the Central Pacific (~Nino4 region) flanked by colder SST anomalies to the west and east, is considered to modulate the variability of ISMR as well.

The impact of ENSO events on India is seen to be limited and confined to Eastern Central India. In comparison, the impact from El Niño Modoki is seen over a larger area in Southern India (Nair et al., 2018). Thereby, it is evident that the large-scale climatic indices show temporal variability with the space and time-varying nature of interaction among the large-scale indices, which vastly impacts the rainfall pattern in the Indian region. Another important aspect is that the summer monsoon rainfall at **regional scale** over the Indian responds to the above-mentioned large-scale climatic indices in complex ways. However, it is difficult to treat India as a single unit for interpreting the association with the large-scale indices, as the association has seasonal and regional differences (Maity and Nagesh, 2006; Vathsala and Koolagudi, 2017).However, recent findings clearly state that ISMR at regional scale is influenced by the combined effect of large number of climatic indices (Nair et al., 2018). Thereby, it is vital to identify the complex association of the large-scale climatic indices and summer monsoon rainfall. Addressing these issues related to (a) complex temporal association of the large-scale indices and summer monsoon rainfall and (b) spatial variation in association and predictability of summer monsoon rainfall forms the motivation of this study. The objective of this study is to analyse the spatial variation in long-lead predictability of summer monsoon rainfall by identifying the time-varying association between the large-scale climatic indices and rainfall over homogeneous monsoon regions (HMRs) of India. The prediction models are developed for each HMR of India in order to assess the region-wise variation in association with the large-scale climatic indices and to improve the prediction performance at regional scale. These regions are divided based on the similarity in rainfall characteristics and association of sub divisional monsoon rainfall with regional/global circulation parameters (Parthasarathy et al., 1993), as per the specifications of Indian Institute of Tropical Meteorology.

2. LITERATURE REVIEW

Prediction methods have come a long way, from relying on an individual's experience to simple numeric methods to complex atmospheric models. Although machine learning algorithms like Artificial Neural Network (ANN) have been utilized by researchers to forecast rainfall. But studies on deep learning shows that DL may have potential applicability in many different domains and forecasting hydrologic variables at various spatiotemporal scales. For instance, (Liu et al. 2014) presented a DL based deep MLP approach to process a huge weather dataset which was used to forecast the weather for the next 24 hour. It was the first DL based study performed for detecting climate extremes. (Zhang et al.) presented a DL based deep belief network algorithm for forecasting next day precipitation using seven environmental factors from the previous day. They found a better accuracy in the forecast as compared with various ML and statistical algorithms. However, there were several days for which the forecast was not reasonably good. (Aswin et al. 2018) presented DL based architectures, namely LSTM and CNN for prediction of rainfall magnitude. In this study, rainfall dataset for January month (1979-2018), from the Global Precipitation Climatology Project (GPCP) was used. Both the DL architectures were trained and optimized on this Global Average Monthly (GAM) dataset. The proposed architecture predicted the GAM rainfall value. However, both the architectures have a similar root mean squared error (RMSE) indicating a scope of improvement in both the architectures. (Haidar and Verma 2018) used a DL based one-dimensional deep CNN approach (Conv 1D) to forecast the monthly rainfall at Innisfail, Australia. In this study, eleven climate indices and sunspot values were used as the predictors. The obtained result was compared with MLP and the forecasting model of the Bureau of Meteorology, Australia. The analysis revealed that Conv 1D model performance was better for the months having higher annual mean whereas it was not good for months having lower annual mean of rainfall. These studies form the motivation of this study, i.e. to explore the potential of the DL approaches in hydrometeorological studies. Hydrometeorological prediction of monthly rainfall prediction is considered in this study. Objective of this study is to extract the hidden sequential information in the HMRs monthly rainfall series and association of these monthly rainfall with the large-scale climatic indices in order to develop a monthly rainfall prediction model. Following specific contributions are made in this study:

- This study proposes a hybrid DL approach, a combination of one-dimensional Convolutional Neural Network (Conv1D) and Multi-Layer Perceptron (MLP) (hereinafter referred to as hybrid Conv1D-MLP model), for monthly rainfall prediction of Homogenous Monsoon regions (HMRs) of India.

- The hybrid model is trained on a dataset of rainfall data of India's different subdivisions which are further converted into five monsoon regions and different combinations of Large scale Indices from year 1951-2000 in such a way that the model uses previous months' data and Indices data as input with different lag periods to obtain the consecutive month rainfall data.

- Next, the trained model is tested on a dataset containing data from 2001-2015 and predicted the monthly data in time series form for all HMRs.

3. PROPOSED METHODOLOGY

A. MODEL ARCHITECTURE

The proposed hybrid Conv1D-MLP model is developed in the Jupyter notebook using Keras, which is a powerful library for large scale DL algorithms. The developed model is a sequential type. A sample schematic diagram of the model is shown in Fig 1. The first part consists of a Conv1D network and the second part consists of an MLP network. Conv1D is a type of CNN used in various applications including sequence prediction problems like time series analysis and forecasting. It comprises the input layer, a fully connected output layer with activation function,and between them, there is an arbitrary number of hidden layer(s) along with the activation functions. The function of the input layer is to receive the signal (input data) and transfer it to the hidden layer. Hidden layers are the computational engine of the model. These may have one or more layers of the Conv1D layer, flatten layer and fully connected layers.

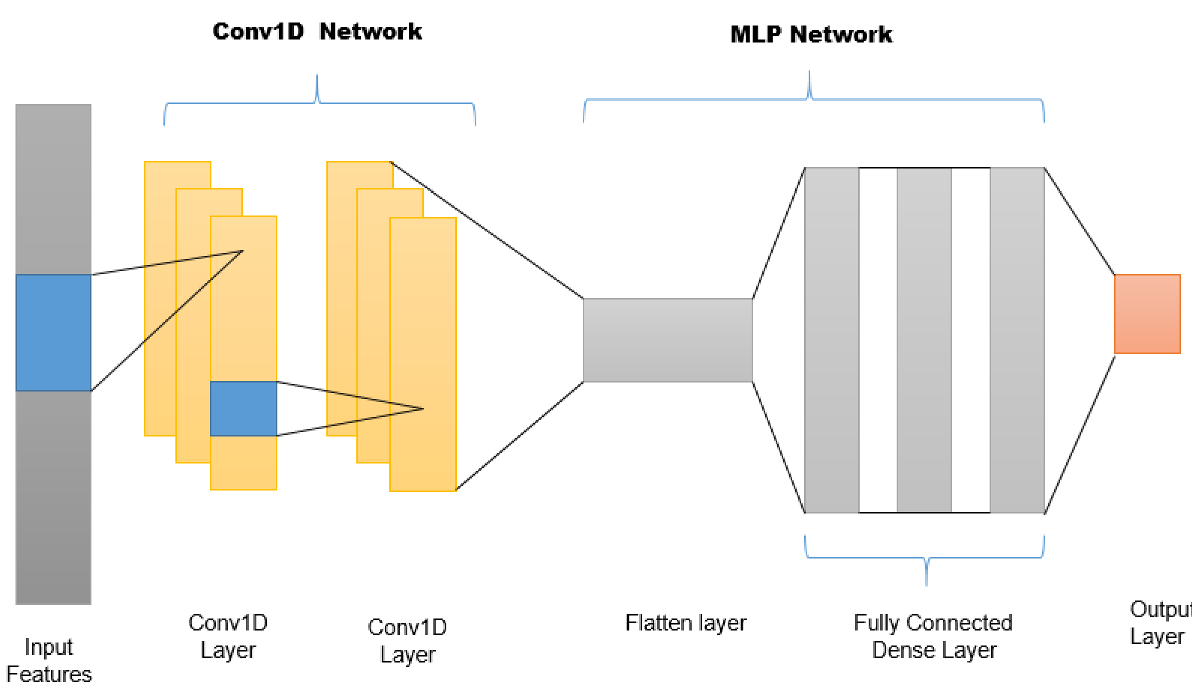


Figure 1. A schematic representation of the proposed hybrid Conv1D-MLP architecture.

Basic functioning and process of Convolution Neural networks is described below:
Basically for every model there are 2 sets i.e. Training dataset and Validation/test set which are given as input and obtain trained CNN model. Training starts with the initialization of network weights and bias. After initialization the process runs for the number of epochs required. In each epoch, the model processes the records of the training data cases, compares the actual values to the predicted values and calculates the loss function, all this process is known as forward propagation calculation. Now the model backpropagates the error through the layers and adjust the network weights.After updating the weights the model is validated on the validation dataset and check if better loss value is obtained and if so it saves the network weights and runs for epochs the model provided with. And finally after completion of all the iterations the trained model is obtained. Now we test this model on a test set.

The Conv1D layer is the main building block of CNN. It consists of filters to extract features from the input signal and kernels to specify the height of the filter. The model is trained on the defined dimension and extracts the hidden information of the sequence.MLP network receives the input from the Conv1D model (Fig. 1). It is also a fully connected ANN that receives the data in a one-dimensional vector form. Therefore, after the Conv1D network a flattened layer is added. Next, a fully connected dense layer along with the activation functions are added. The fully connected layers have a more number of neurons than output layers. In this way, neural networks are allowed to think wider before they converge to the output layer.There are several parameters to be specified to fix the model architecture. This is problem specific and details are provided in the results and discussion section of this report. After configuring the layers, the model is trained on a set of input and output data to learn the relationship between them. Training involves adjustments of weights and biases (parameters) to minimize the error. It is achieved through backpropagation, which adjusts the parameters considering the error (loss) in the predictions. There are several error based metrics, e.g., mean squared error (MSE) and log loss etc. Once the model is properly trained, it is ready for further use.

B. DATA DESCRIPTION

Rainfall prediction is clearly of great importance for India.India Meteorological Department (IMD) provides monthly rainfall data that is used in this study and the time period of the study is from 1951 to 2015 and the data for the aforementioned large-scale climate indices are obtained for the given time period.

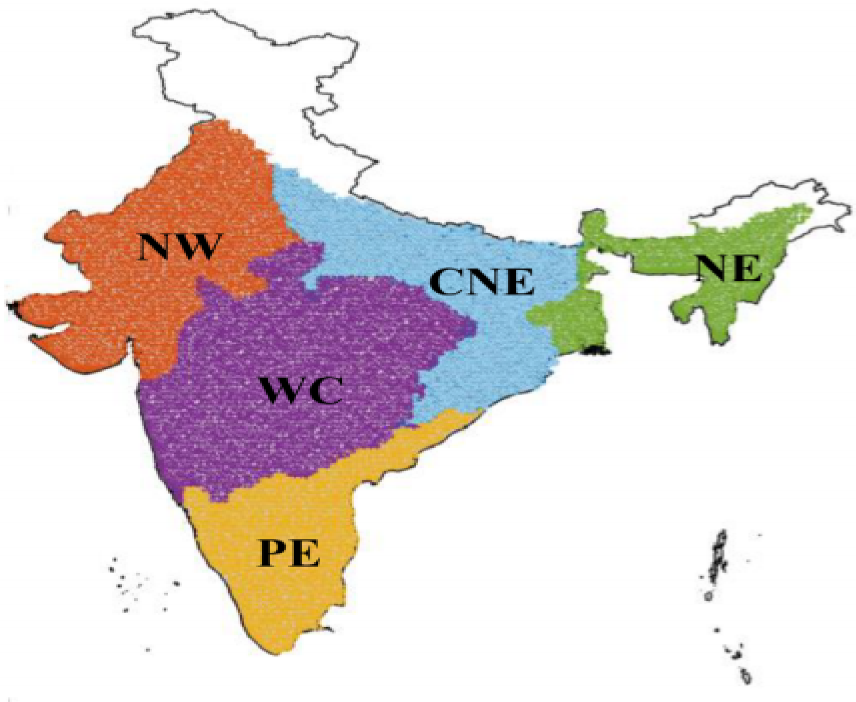


FIGURE 2: Map showing HMRs in India. These regions are the grouped contiguous sub-divisions based on the rainfall characteristics and association with global/regional circulation parameters to form the HMRs(Dutta R, Maity R.2020,p.5)

The dataset consists of the monthly rainfall for the period 1901-2015 for each state in India. The data for this study were obtained from IMD. (URL: https://data.gov.in/resources/subdivision-wise-Rainfall-and-its-departure-1901-2015 accessed in November 2020). The selected dataset contains 19 attributes (individual months, annual, and combinations of three consecutive months) for 36 sub divisions from 1901 to 2015. Now these different contiguous subdivisions are divided into 5

Homogeneous monsoon regions(HMRs) named Central North East(CNE), Peninsular(PE), Northwest(NW), West Central(WC), Northeast(NE) regions based on rainfall characteristics and association with global/ regional circulation parameters in order to assess the region-wise variation in association with the large-scale climatic indices and to improve the prediction performance at regional scale. Now based on the monthly rainfall data of subdivisions area weighted monthly rainfall series are prepared for the HMRs of India. The data for large scale Indices are prepared using different parameters. The Sea Level Pressure(SLP) and Zonal Wind data are utilized to derive the large-scale climatic indices. The ENSO index is evaluated as the SST anomaly over Niño3.4 region (120–170W, 5S–5N). EMI is derived from the difference in area average SST anomalies in the regions of 10S–10N and 165E–140W; 15S–5N and 110–70W; and 10S–20N and 125–145E (Ashok et al., 2007).PDO (Deser et al., 2016) is derived by evaluating the leading Empirical Orthogonal Function (EOF) of SST anomalies in the North Pacific basin (polewards of 20N). NAO (Hurrel et al., 2018) is derived by evaluating the leading EOF of SLP anomalies over the Atlantic sector, 20–80N, 90W–40E. Lastly, EQUINOO (Gadgil et al., 2004) index is computed as the negative of the zonal wind anomaly at surface in the Equatorial Indian Ocean region (60–90E, 2.5S–2.5N). Previous months rainfall data and different combinations of Indices values are used as input to predict the consecutive month rainfall, some of these monthly data had either zero or very few missing values that were handled during data preprocessing. Two different lags, i.e., 3 and 6 months are considered. Lag indicates the gap (number of months) between input i.e. past data and the starting month of prediction. For example, if we consider 3 months lag we use rainfall data of January, February and March of 2001 and any index or combination of these Indices corresponding to January, February and March of 2001 to predict the rainfall of April of 2001 and it goes on predicting the rainfall of all the consecutive months in the time series model until 2015. In the similar way 6 months lag is also considered in this study. The correlation coefficients between rainfall in different months and seasons and the same with annual rainfall are shown in heat maps(Figure 3 and 4).

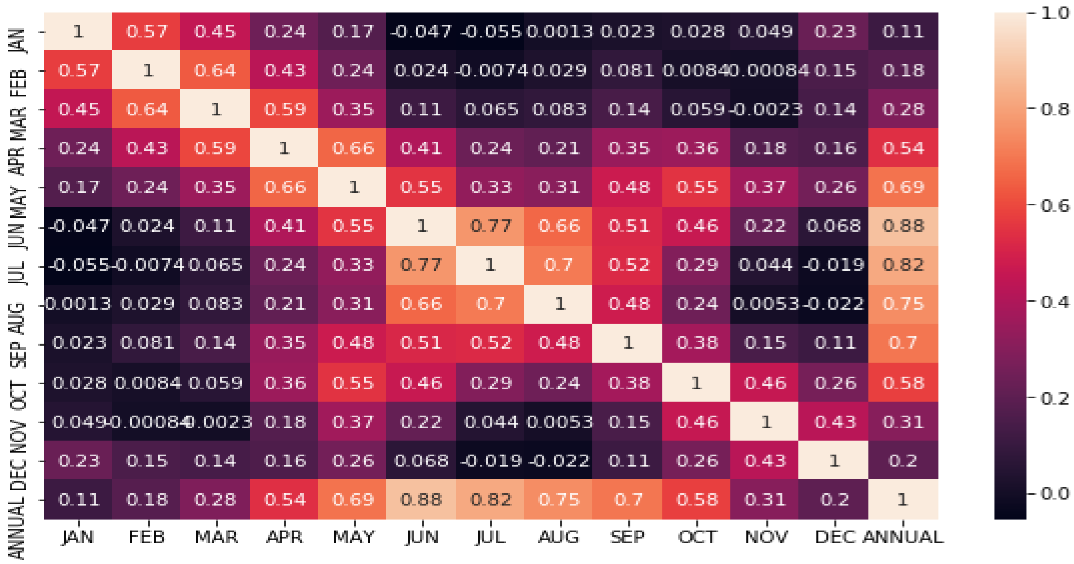


Figure 3. Heat map showing the correlation coefficients between rainfall in different months and the same with annual rainfall.

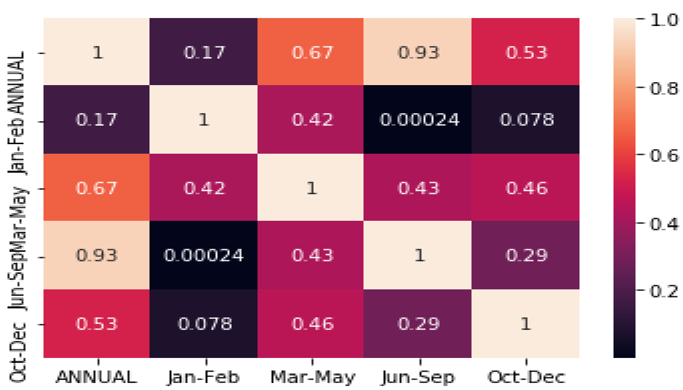


Figure 4. Heat map showing the correlation coefficients between rainfall in different seasons and the same with annual rainfall

C. PERFORMANCE EVALUATION CRITERIA

The performance of the hybrid Conv1D-MLP model is evaluated through three statistical measures viz. Root Mean Squared Error (RMSE), coefficient of correlation (r), Nash–Sutcliffe Efficiency (NSE). The coefficient of correlation is a measure of linear association between two variables. The value of r is computed as:

$$r = \frac{\sum_{t=1}^{n} (Y_t - \overline{Y}) (Y_t^{'} - \overline{Y}^{'})}{\sum_{t=1}^{n} (Y_t - \overline{Y}) \sum_{t=1}^{n} (Y_t^{'} - \overline{Y}^{'})}$$

The value of r ranges between [−1,1], where −1 represents a perfect negative linear association, 0 denotes no linear association, and 1 represents a perfect positive linear association. The higher the value of r, the better the model performs.

RMSE is used frequently to measure the difference between observed and predicted values. It is always positive and a lower RMSE indicates a better model performance. RMSE is expressed as:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (Y_t - Y_t^{'})^2}{n}}$$

NSE is used to measure the efficiency of the proposed model. The value NSE is computed by the equation:

$$NSE = 1 - \frac{\sum_{t=1}^{n} (Y_t - Y_t^{'})^2}{\sum_{t=1}^{n} (Y_t - \overline{Y})^2}$$

NSE ranges between (−∞, 1]. A value greater than 0 indicates a better efficiency of the model as compared to a value equal to 0 which signifies the predicted values are as good as the mean of the observed values. NSE values less than zero indicate an unacceptable model performance.

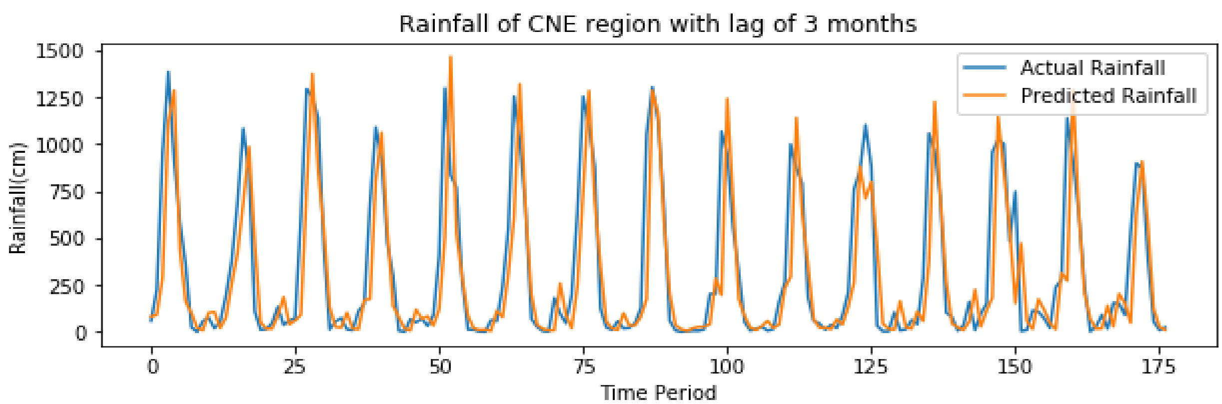
4. RESULTS AND DISCUSSION

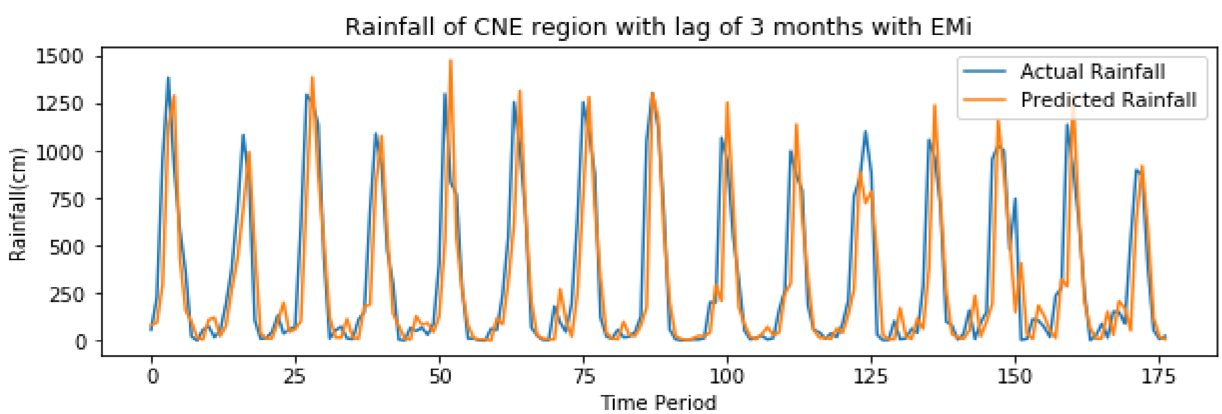
The models were created in python on the Jupyter notebook using Keras(https://github.com/surendra093/Rainfall_prediction/blob/main/Rainfall%20prediction.ipynb) deep learning. Several network architectures were evaluated by varying model parameters (viz. number of hidden layers, number of filters, kernel size) and optimizing several hyperparameters (viz. learning rate, batch size, number of epochs, loss functions, and activation functions) in order to ascertain the best possible architectural configuration. All the experiments were run for 10 epochs, but by using callbacks in Keras only the best weight for each test run was saved. The finalized architecture of the proposed hybrid model comprises seven layers. Details of the finalized configurations are shown in Table 1.

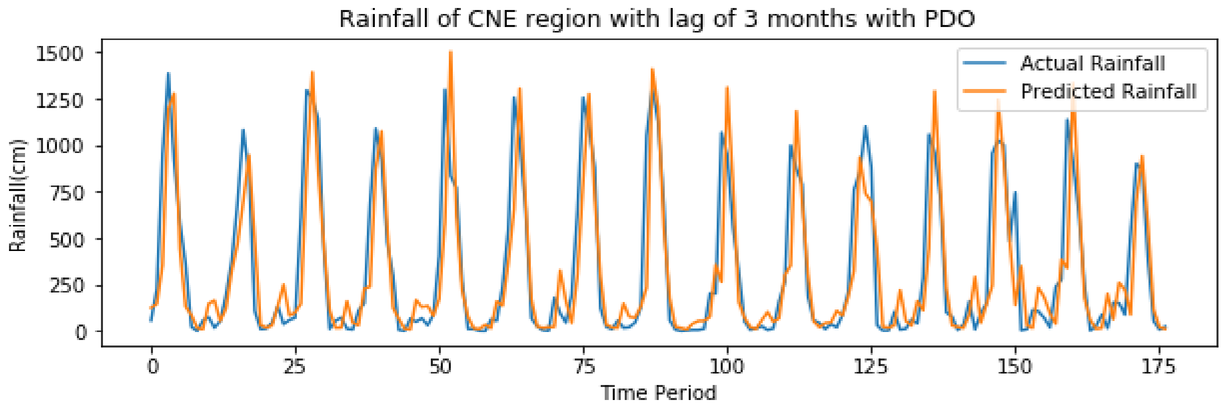
TABLE 1. Configurations of the proposed hybrid Conv1D-MLP model

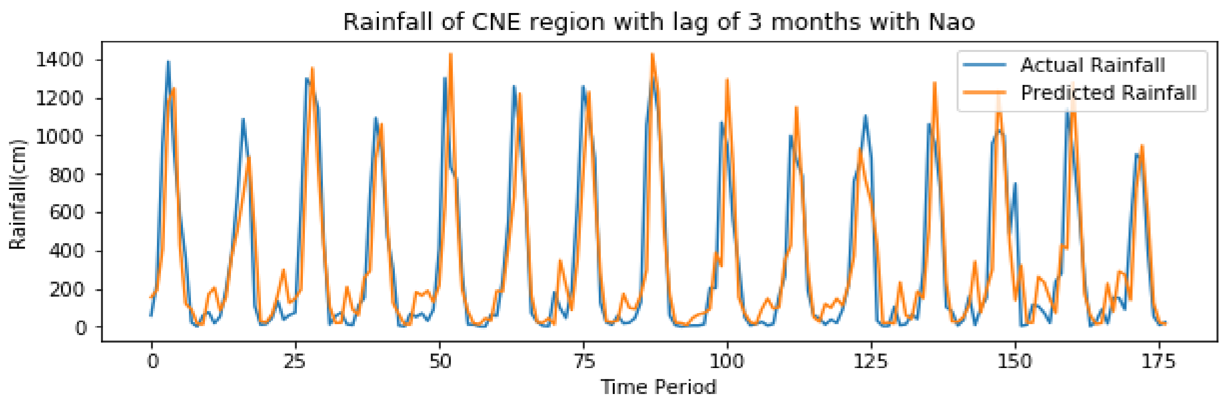
Layer no.	Layer	Type	Parameters of layers			
			Activation func.	Kernel Size	No.of filters	Neurons
1	Conv1D	Convolution Layer	ReLU	1	64	-
2	Conv1D	Convolution Layer	ReLU	2	128	-
3	Flatten	Flatten Layer	-	-	-	-
4	dense	Fully connected Layer	ReLU	-	-	128
5	dense	Fully connected Layer	ReLU	-	-	64
6	dense	Fully connected Layer	ReLU	-	-	32
7	dense	Fully connected Layer (output layer)	Linear	-	-	1

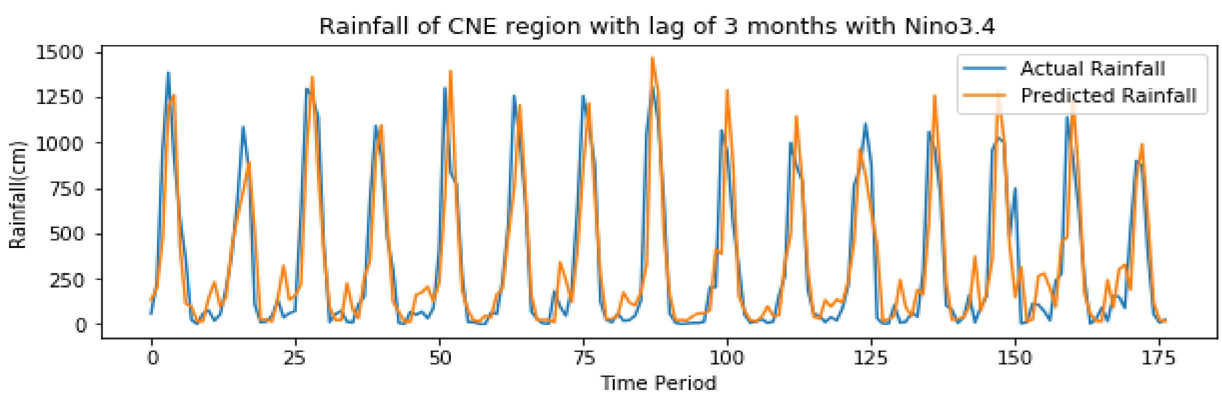
The above mentioned architecture is finalized based on multiple observations of the results obtained when trained on different data sets i.e trained on data of 5 different HMRs of India. We tried to analyze different models for different HMRs for region specific analysis and finalized the above mentioned architecture for each of the models will be having their own parameters based on the data they trained during their training period. Now, different trained models with the combination of different Indices (i.e compared results obtained only using previous months rainfall data, combination of previous month's rainfall data along with combination of Indices and previous month's rainfall data along with all the Indices) with aforementioned architecture are tested on CNE, PE, NW, WC, NE regions with their respective models and observed the predicted rainfall for all the months from 2001-2015 in time series fashion and the model performance is evaluated using RMSE, coefficient of correlation and NSE values.

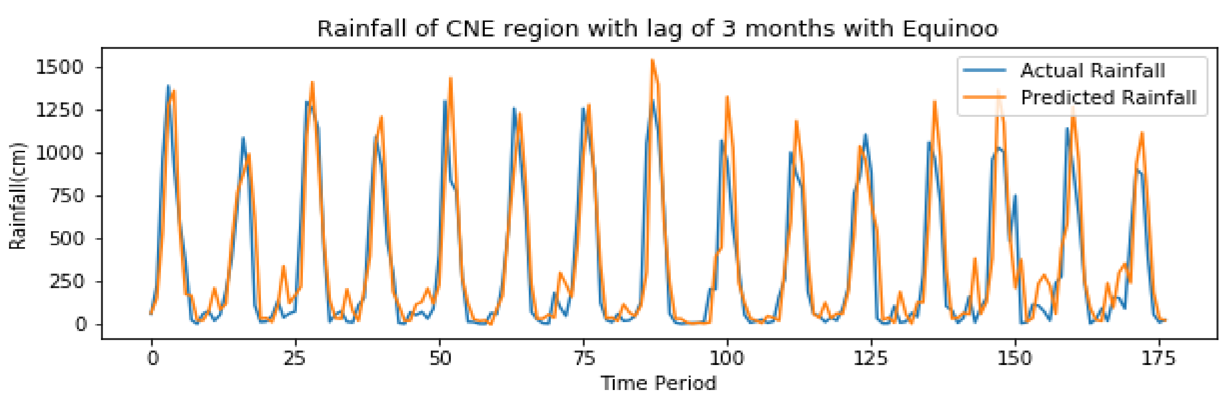





Rainfall of CNE region with lag of 3 months with PDO
Actual Rainfall
Predicted Rainfall
Rainfall(cm)
Time Period


Rainfall of CNE region with lag of 3 months with Nao
Actual Rainfall
Predicted Rainfall
Rainfall(cm)
Time Period


Rainfall of CNE region with lag of 3 months with Nino3.4
Actual Rainfall
Predicted Rainfall
Rainfall(cm)
Time Period


Rainfall of CNE region with lag of 3 months with Equinoo
Actual Rainfall
Predicted Rainfall
Rainfall(cm)
Time Period

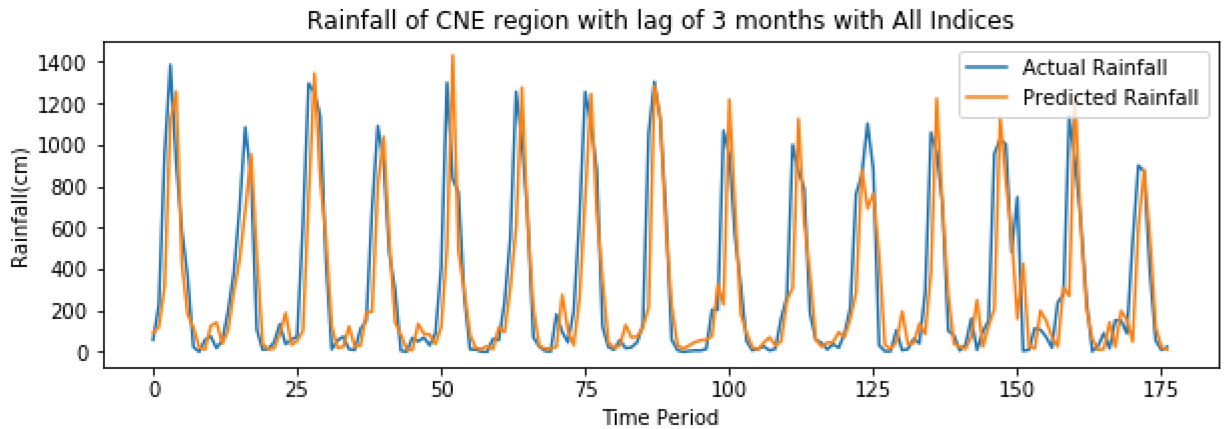
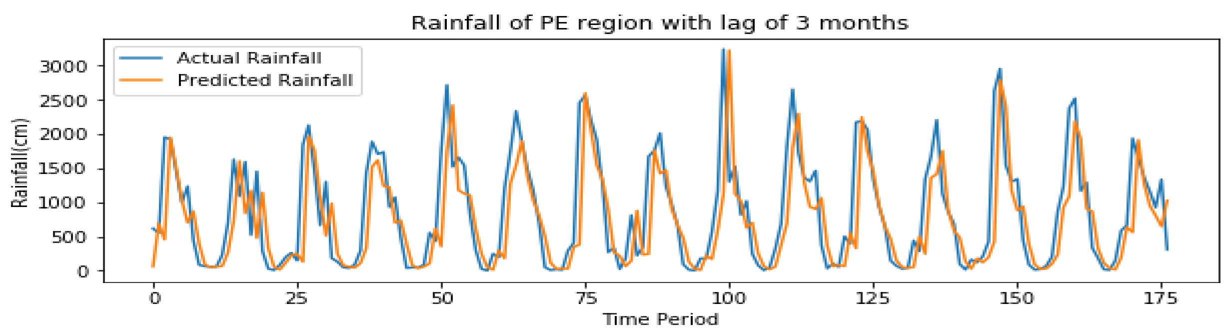
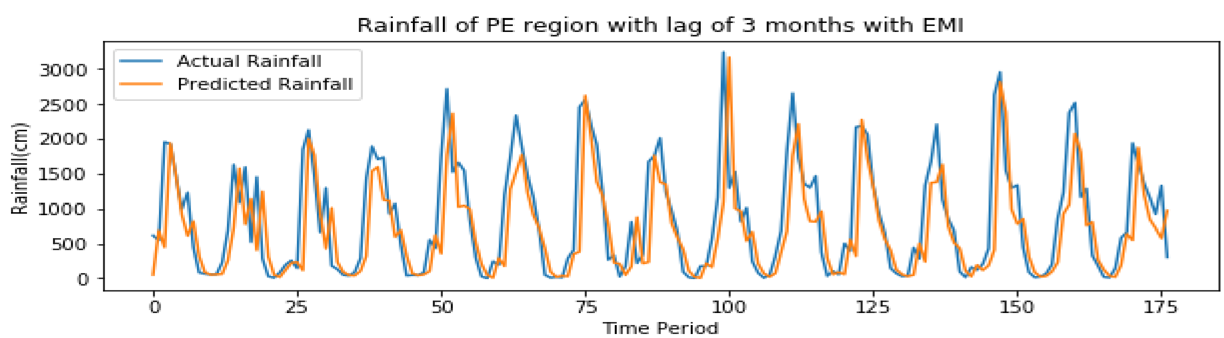
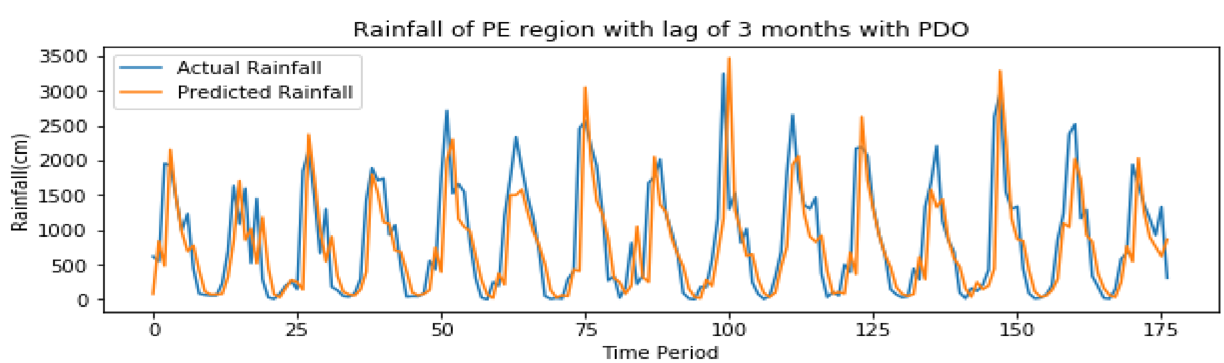


Figure 5. Comparison plots between actual and predicted monthly using inputs from 3 previous months and different combinations of Indices for CNE region. X-axis shows the months from the year 2001 to 2015.







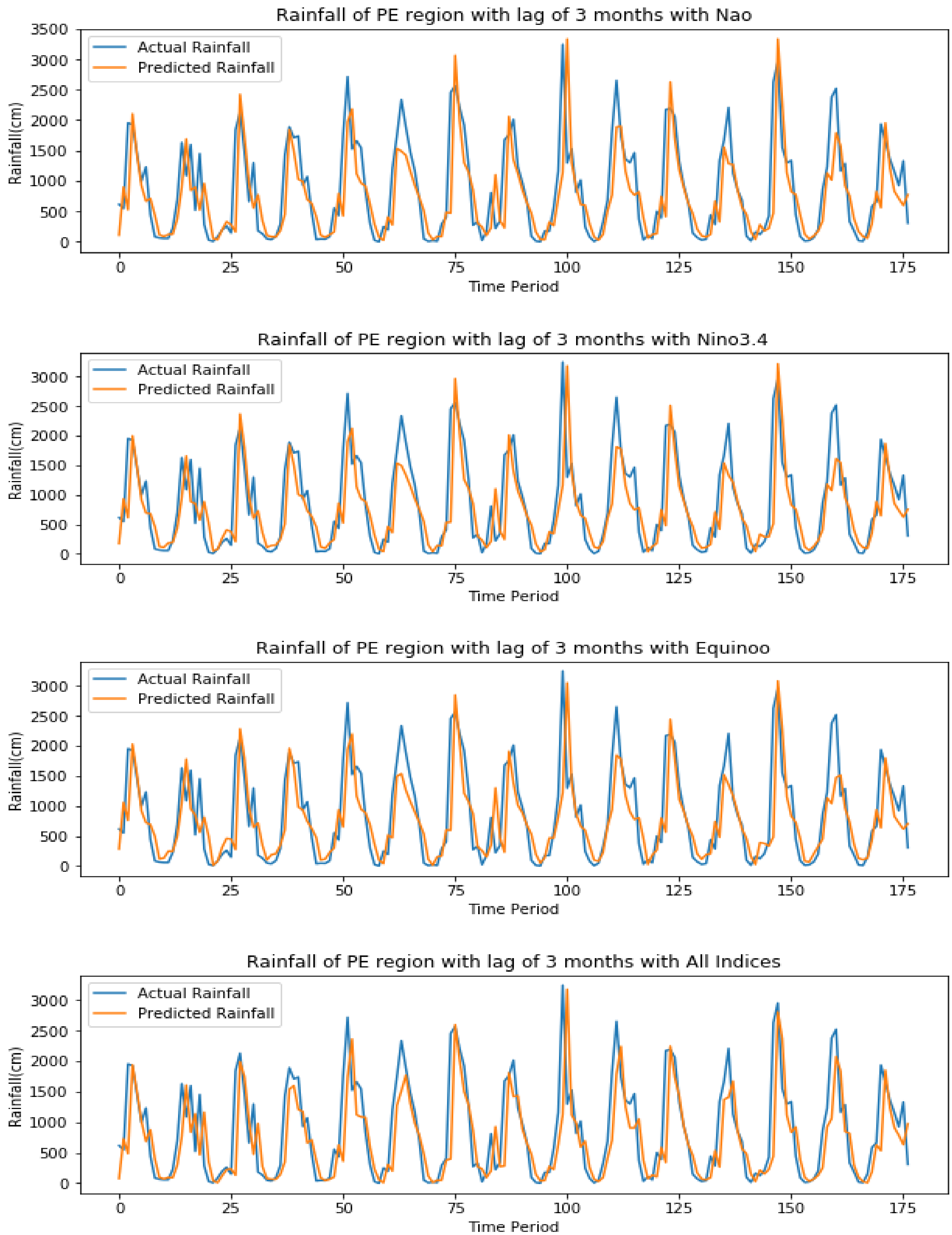
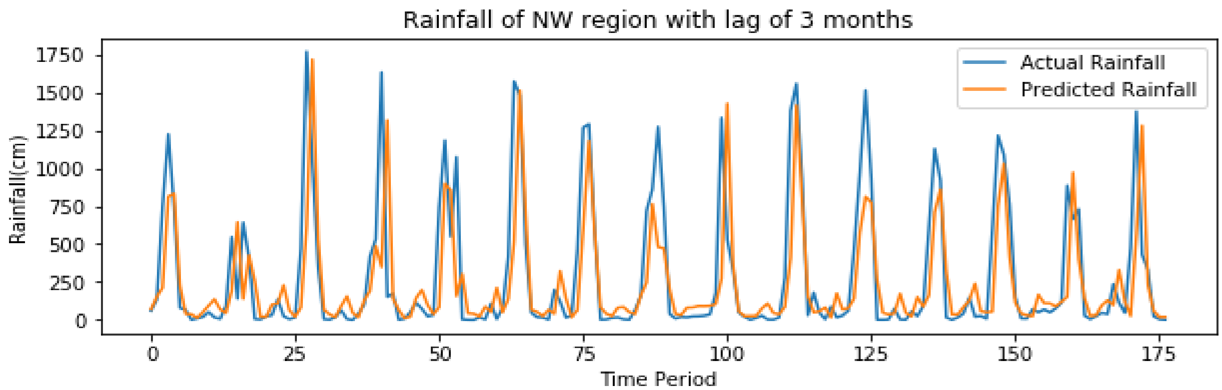
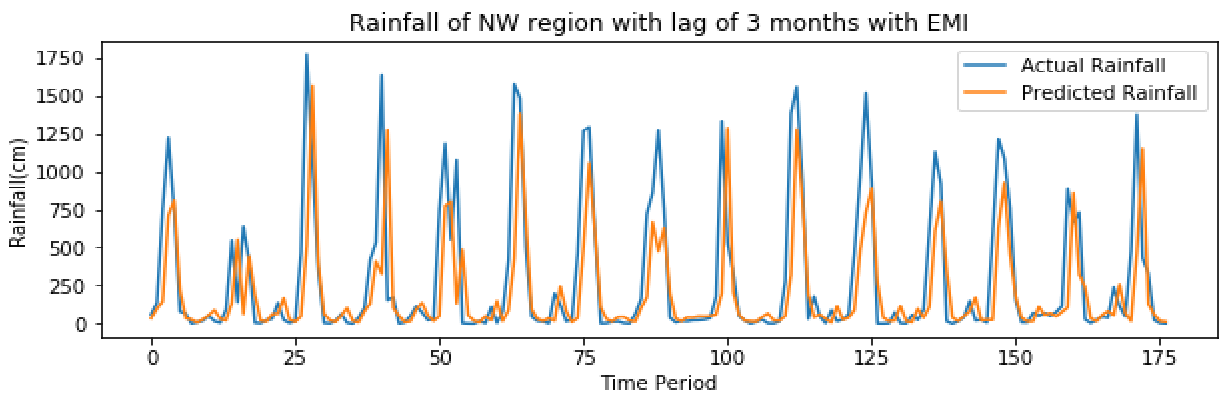
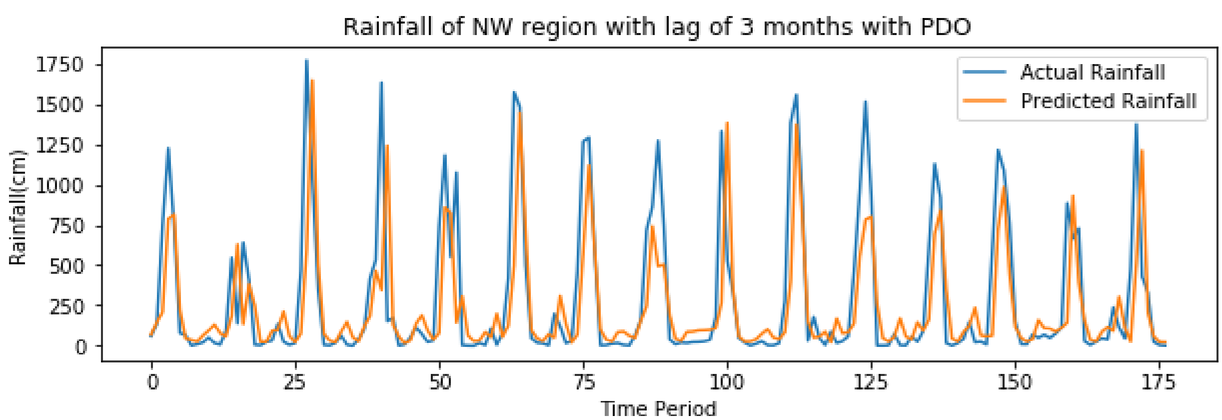
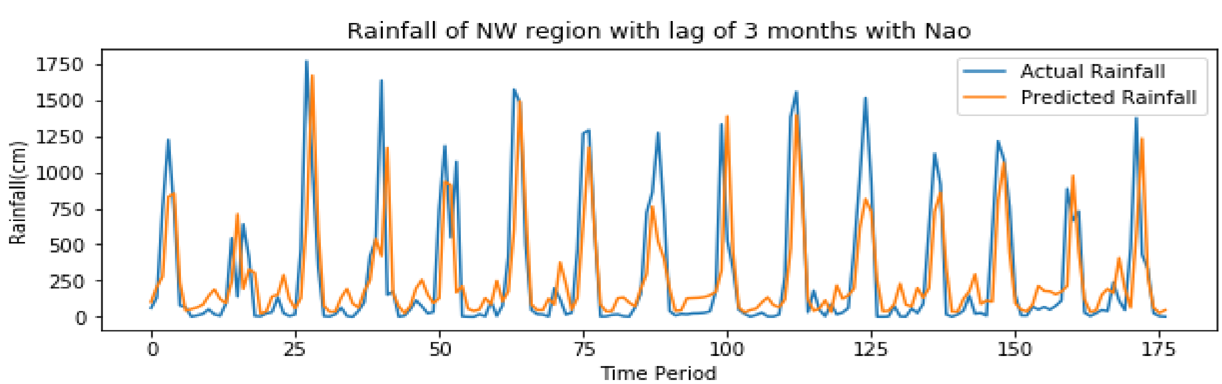


Figure 6. Comparison plots between actual and predicted monthly using inputs from 3 previous months and different combinations of Indices for PE region. X-axis shows the months from the year 2001 to 2015


Rainfall of NW region with lag of 3 months
Rainfall(cm)
1750
1500
1250
1000
750
500
250
0
0
25
50
75
100
125
150
175
Time Period
Actual Rainfall
Predicted Rainfall


Rainfall of NW region with lag of 3 months with EMI
Rainfall(cm)
1750
1500
1250
1000
750
500
250
0
0
25
50
75
100
125
150
175
Time Period
Actual Rainfall
Predicted Rainfall


Rainfall of NW region with lag of 3 months with PDO
Rainfall(cm)
1750
1500
1250
1000
750
500
250
0
0
25
50
75
100
125
150
175
Time Period
Actual Rainfall
Predicted Rainfall


Rainfall of NW region with lag of 3 months with Nao
Rainfall(cm)
1750
1500
1250
1000
750
500
250
0
0
25
50
75
100
125
150
175
Time Period
Actual Rainfall
Predicted Rainfall

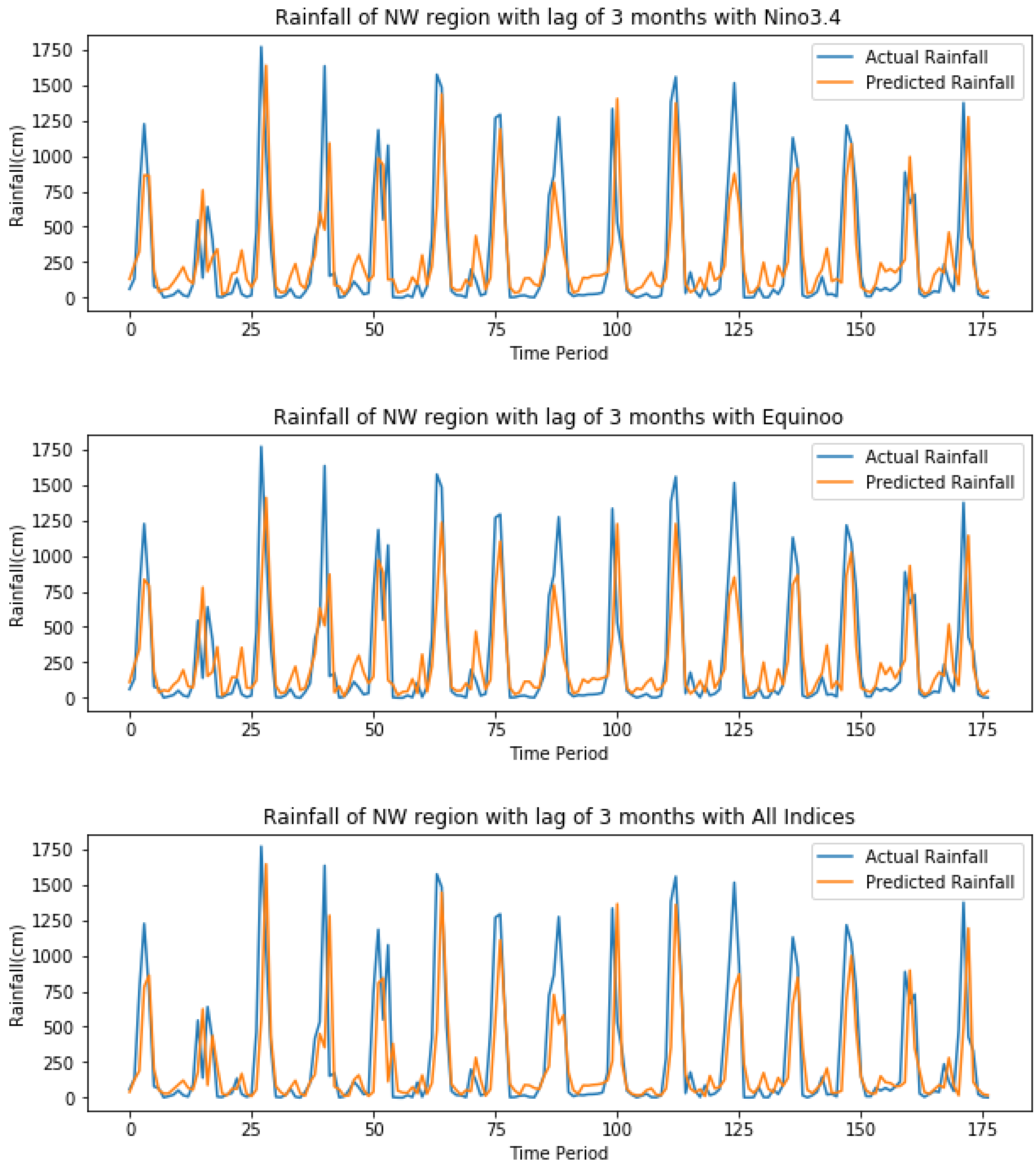
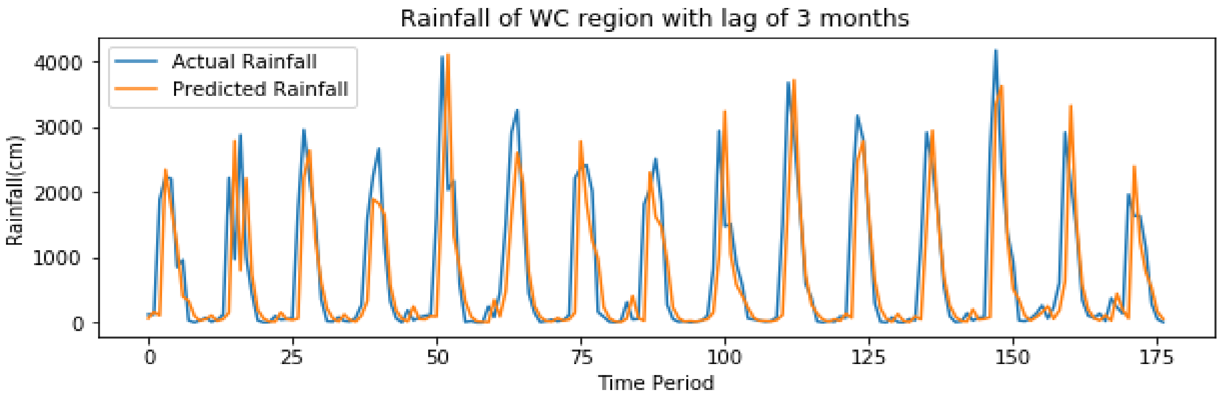
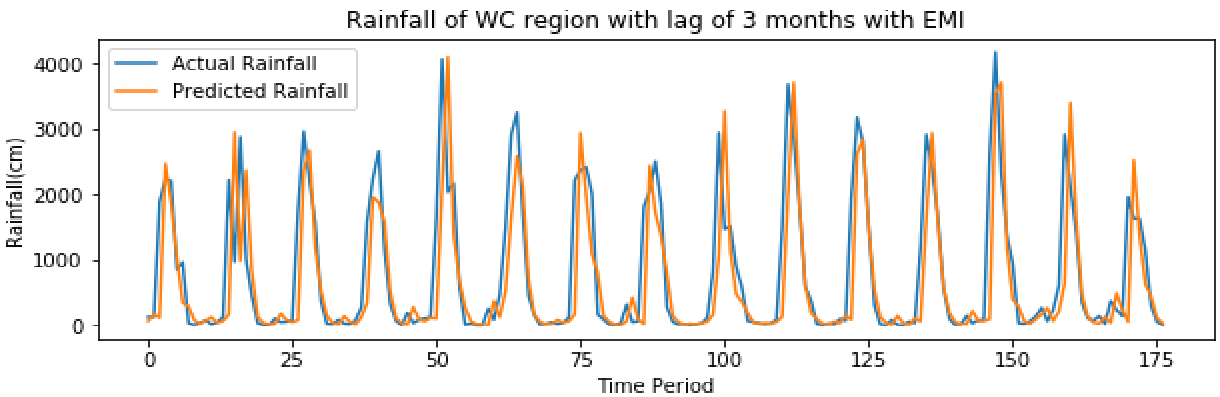
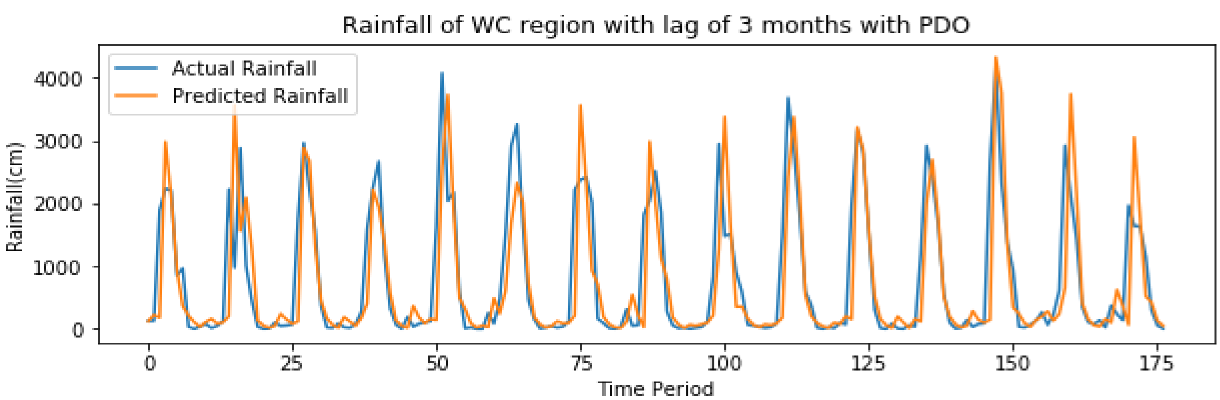
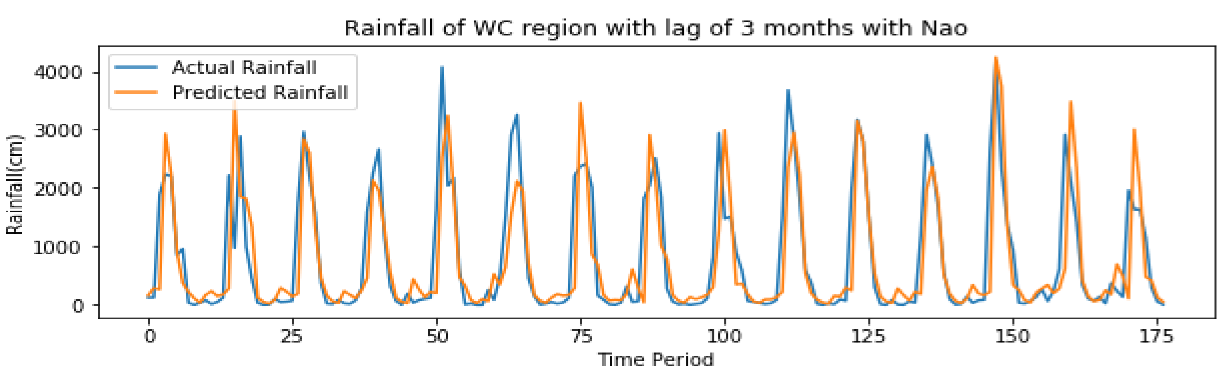


Figure 7. Comparison plots between actual and predicted monthly using inputs from 3 previous months and different combinations of Indices for NW region. X-axis shows the months from the year 2001 to 2015


Rainfall of WC region with lag of 3 months
Actual Rainfall
Predicted Rainfall
Rainfall(cm)
Time Period


Rainfall of WC region with lag of 3 months with EMI
Actual Rainfall
Predicted Rainfall
Rainfall(cm)
Time Period


Rainfall of WC region with lag of 3 months with PDO
Actual Rainfall
Predicted Rainfall
Rainfall(cm)
Time Period


Rainfall of WC region with lag of 3 months with Nao
Actual Rainfall
Predicted Rainfall
Rainfall(cm)
Time Period

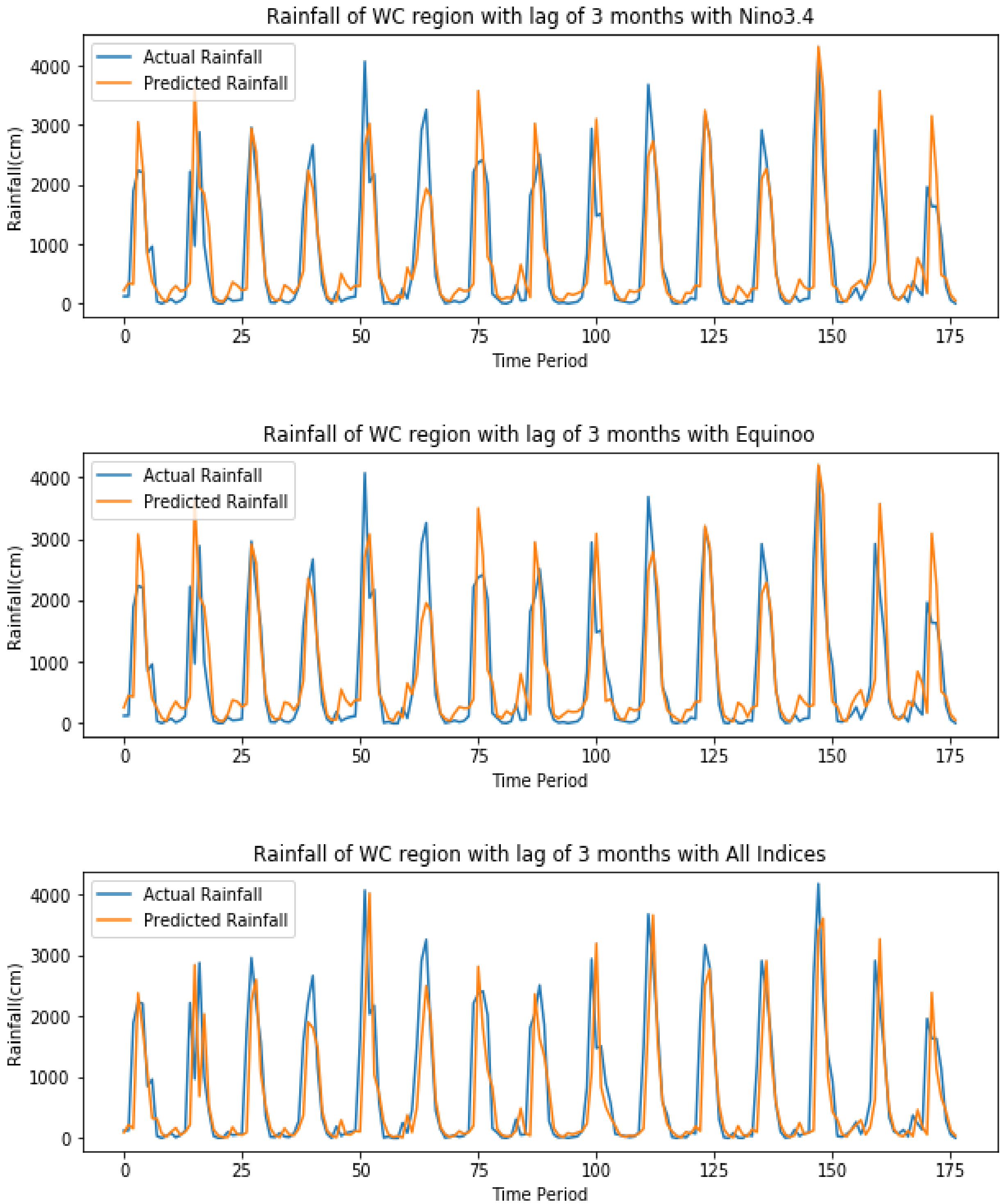
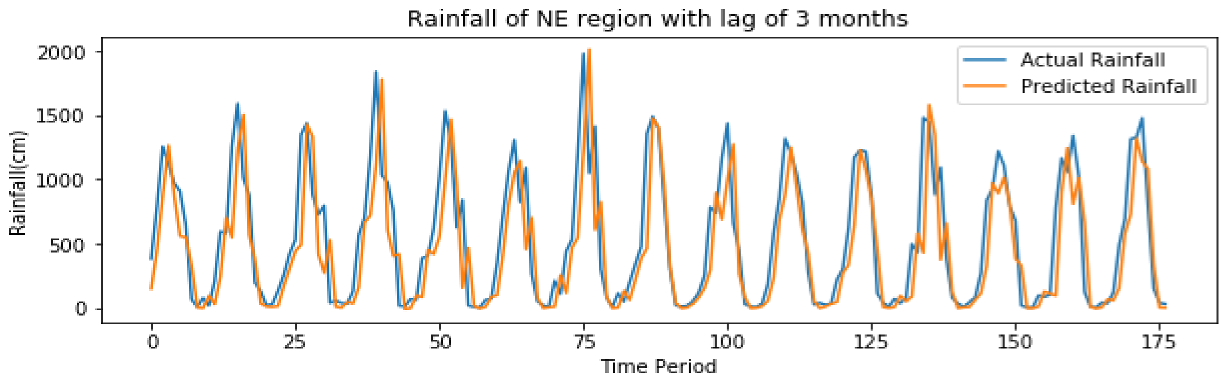
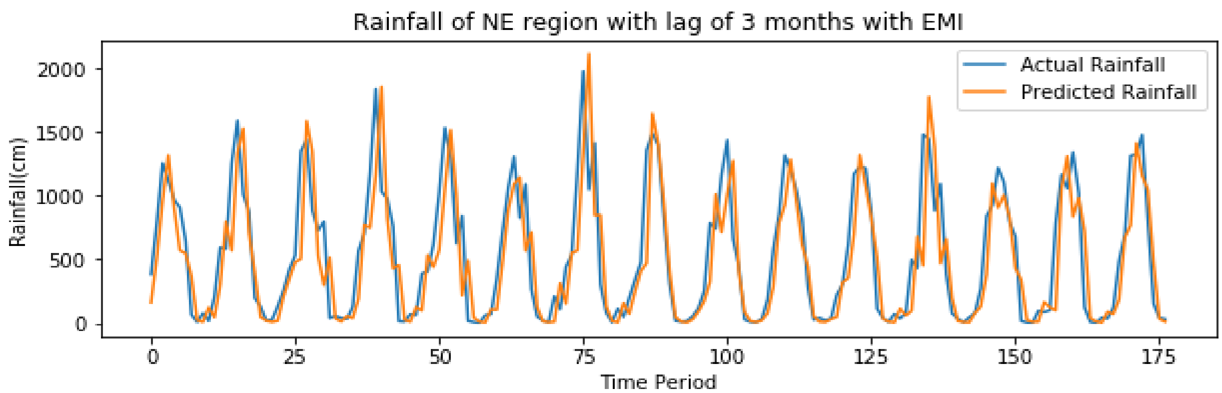
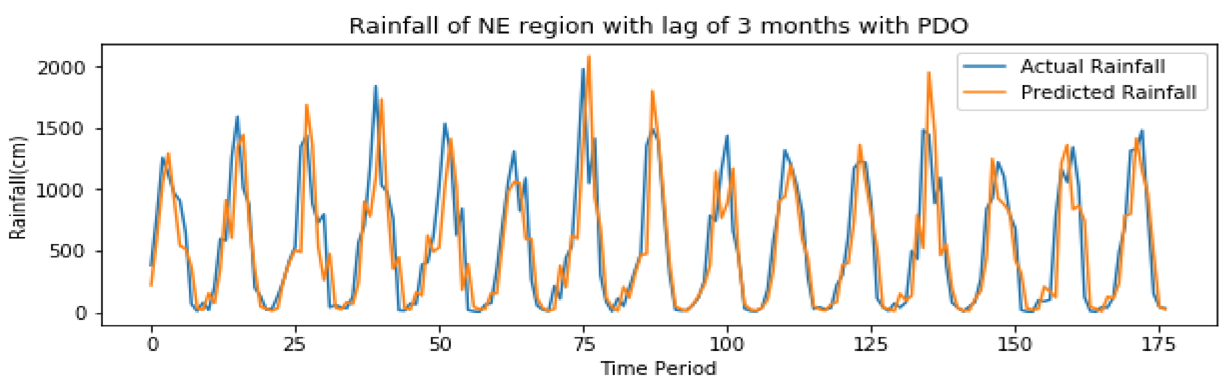
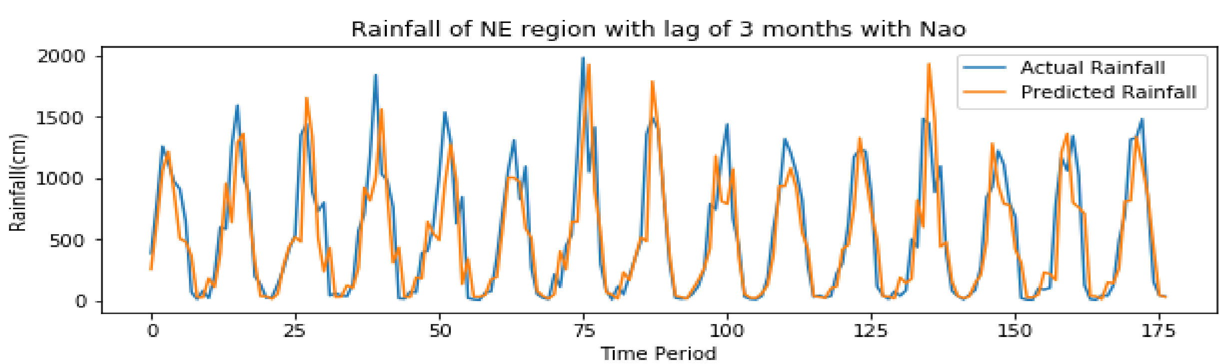


Figure 8. Comparison plots between actual and predicted monthly using inputs from 3 previous months and different combinations of Indices for WC region. X-axis shows the months from the year 2001 to 2015


Rainfall of NE region with lag of 3 months
Actual Rainfall
Predicted Rainfall
Rainfall(cm)
Time Period


Rainfall of NE region with lag of 3 months with EMI
Actual Rainfall
Predicted Rainfall
Rainfall(cm)
Time Period


Rainfall of NE region with lag of 3 months with PDO
Actual Rainfall
Predicted Rainfall
Rainfall(cm)
Time Period


Rainfall of NE region with lag of 3 months with Nao
Actual Rainfall
Predicted Rainfall
Rainfall(cm)
Time Period

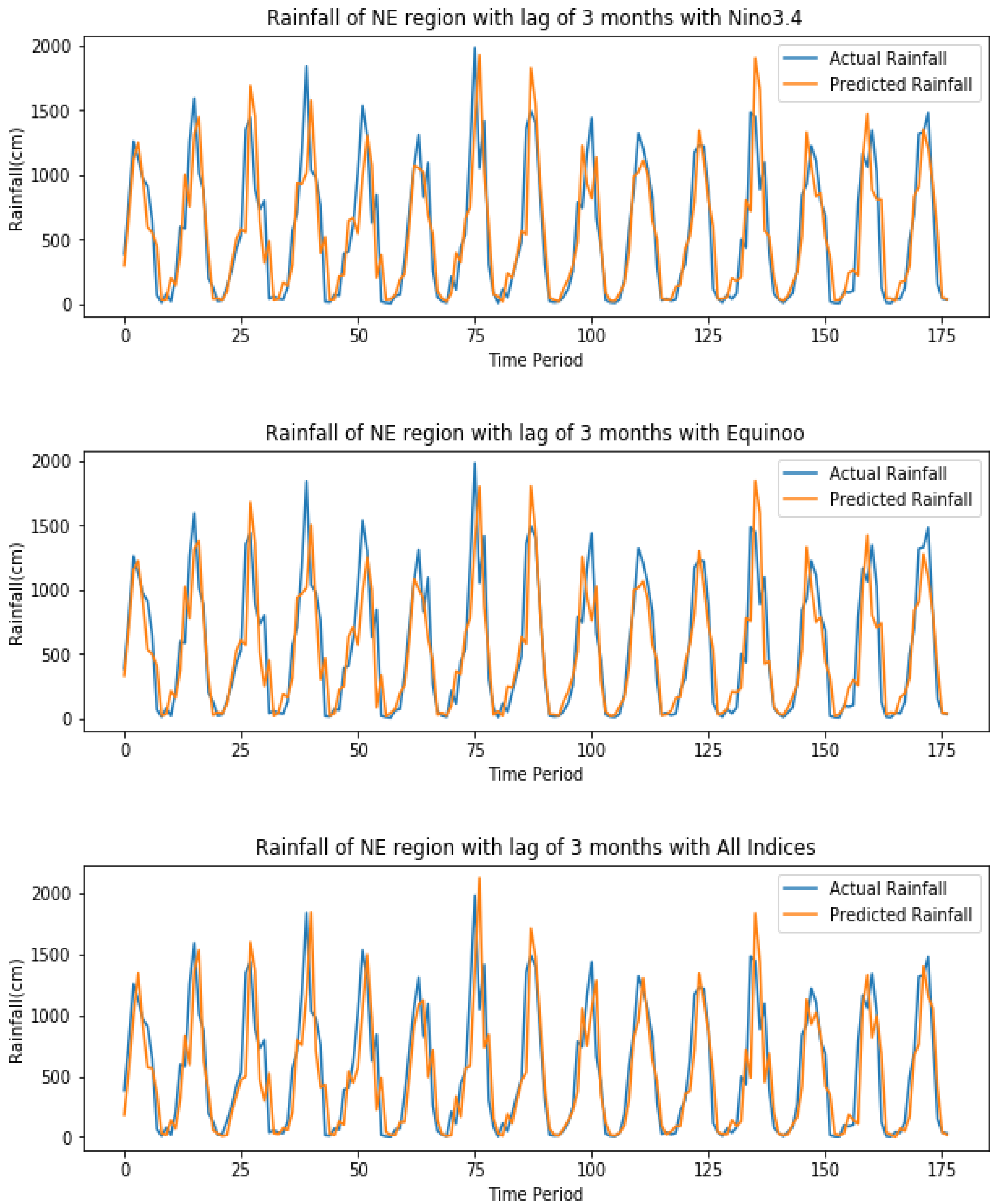


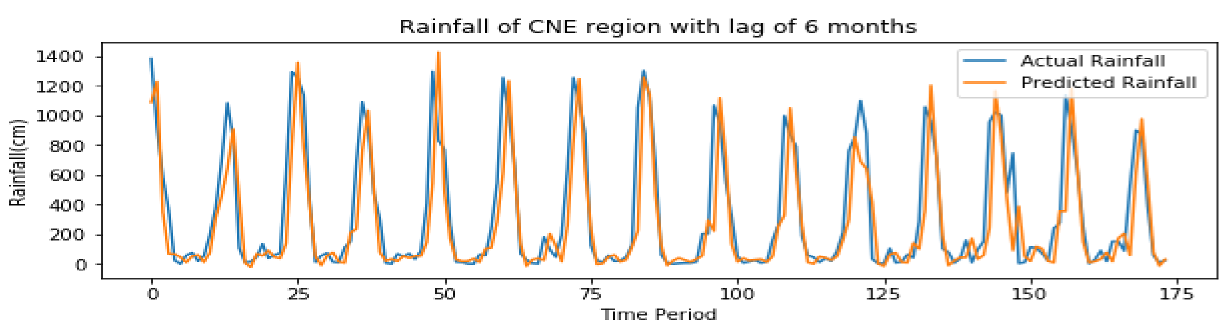
Figure 9. Comparison plots between actual and predicted monthly using inputs from 3 previous months and different combinations of Indices for NE region. X-axis shows the months from the year 2001 to 2015.

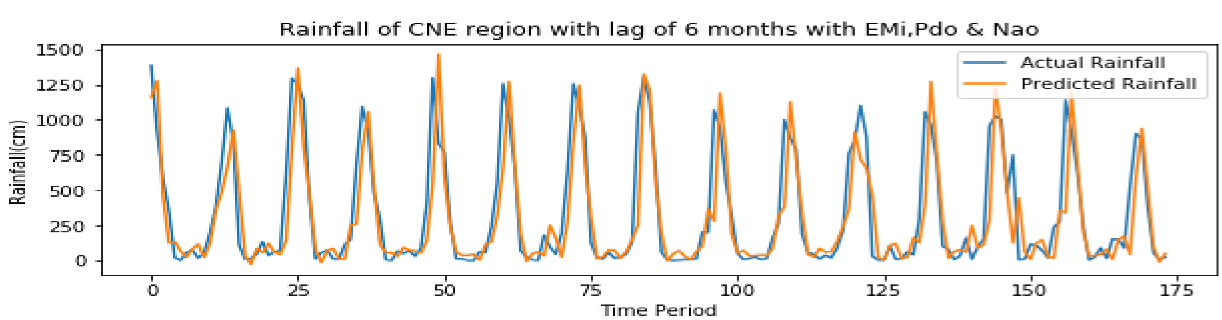
TABLE 2: Performance statistics viz. r, RMSE and NSE for different combinations of Indices for different HMRs considering 3 months lag period.

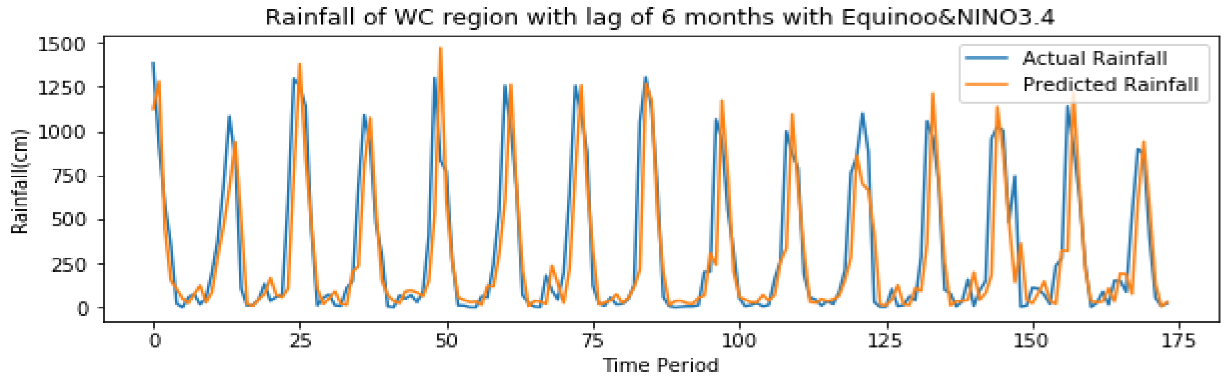
HMR	Index Combination	Performance Statistics: RMSE	Correlation coefficient(r)	NSE coefficient
CNE region	No Index	249.4	0.807	0.539
	EMI	246.19	0.813	0.56
	PDO	235.33	0.825	0.604
	NAO	223.90	0.838	0.62
	NINO3.4	211.26	0.857	0.679
	EQUINOO	208.83	0.872	0.744
	All Indices	244.03	0.812	0.528
PE region	No Index	588.03	0.707	0.250
	EMI	593.46	0.709	0.202
	PDO	581.08	0.713	0.294
	NAO	571.34	0.721	0.264
	NINO3.4	553.25	0.733	0.245
	EQUINOO	541.85	0.740	0.235
	All Indices	583.32	0.710	0.223
NW region	No Index	322.66	0.692	0.05
	EMI	336.83	0.671	-0.167
	PDO	321.84	0.692	-0.018
	NAO	309.86	0.710	0.068
	NINO3.4	300.96	0.728	0.135
	EQUINOO	289.61	0.751	0.095
	All Indices	328.57	0.682	-0.043

WC region	No Index	724.77	0.744	0.392
	EMI	717.02	0.755	0.431
	PDO	703.04	0.771	0.536
	NAO	677.34	0.778	0.528
	NINO3.4	670.85	0.781	0.534
	EQUINOO	663.64	0.785	0.545
	All Indices	717.27	0.750	0.377
NE region	No Index	320.4	0.805	0.521
	EMI	310.03	0.812	0.585
	PDO	302.61	0.818	0.603
	NAO	296.05	0.823	0.582
	NINO3.4	286.04	0.832	0.632
	EQUINOO	282.47	0.836	0.613
	All Indices	309.32	0.811	0.588

Now rainfall is predicted use other different combination of large scale Indices with 6 months lag period:







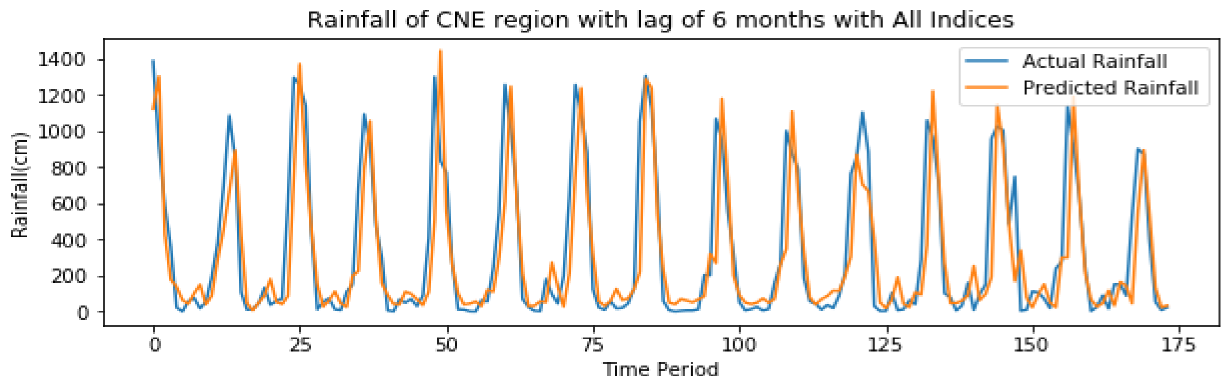
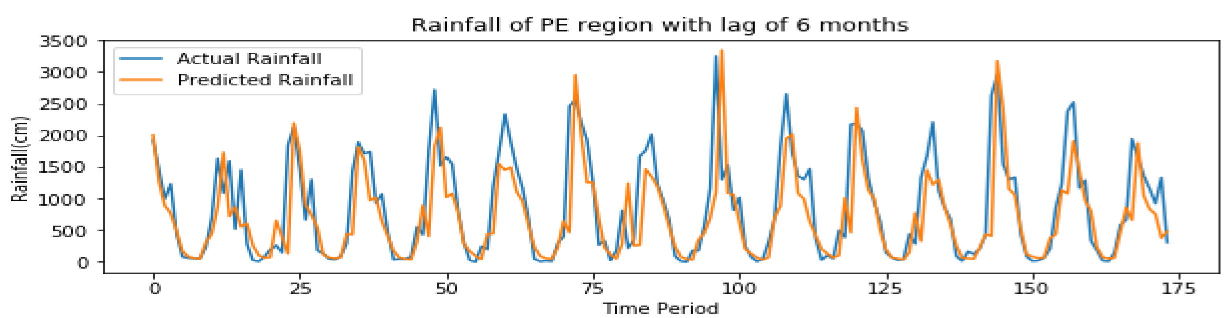
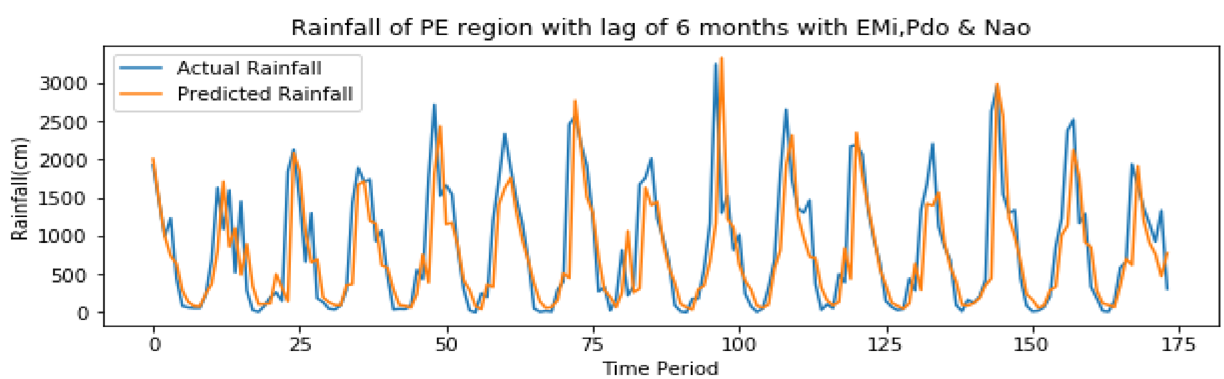


Figure 10. Comparison plots between actual and predicted monthly using inputs from 6 previous months and different combinations of Indices for CNE region. X-axis shows the months from the year 2001 to 2015.





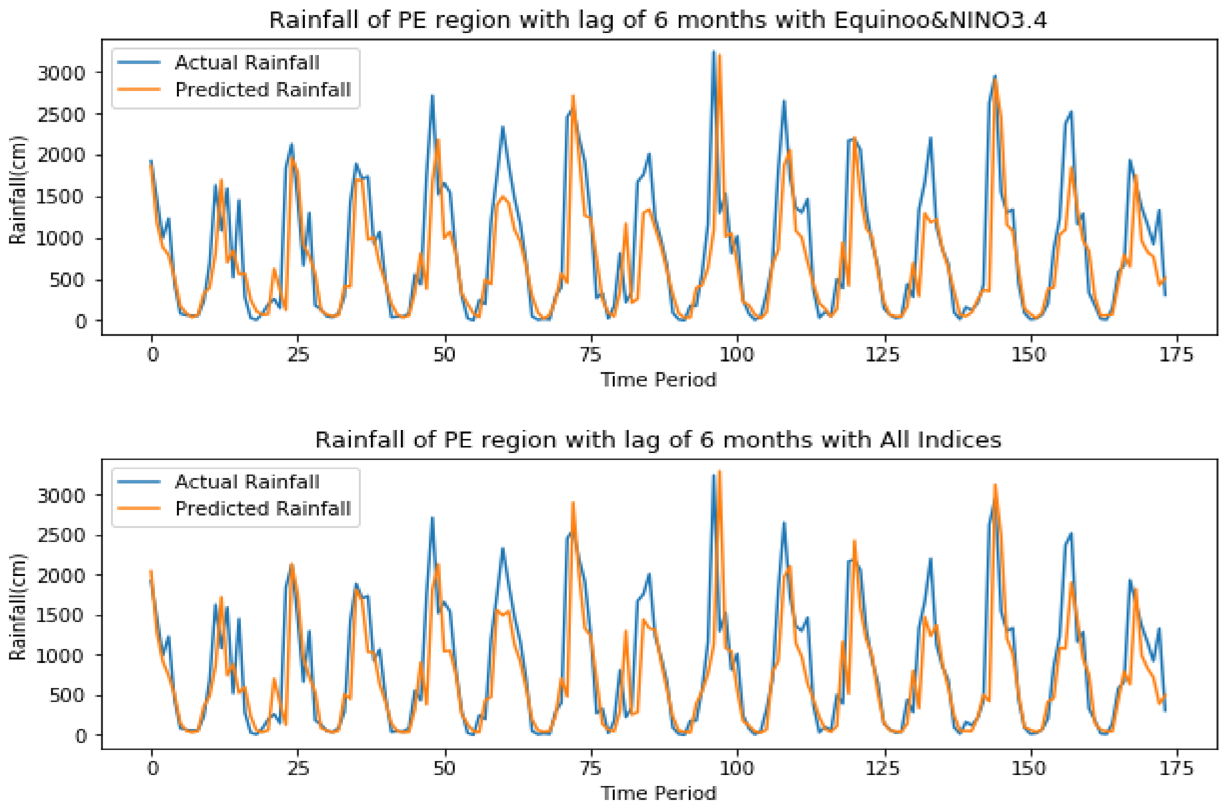
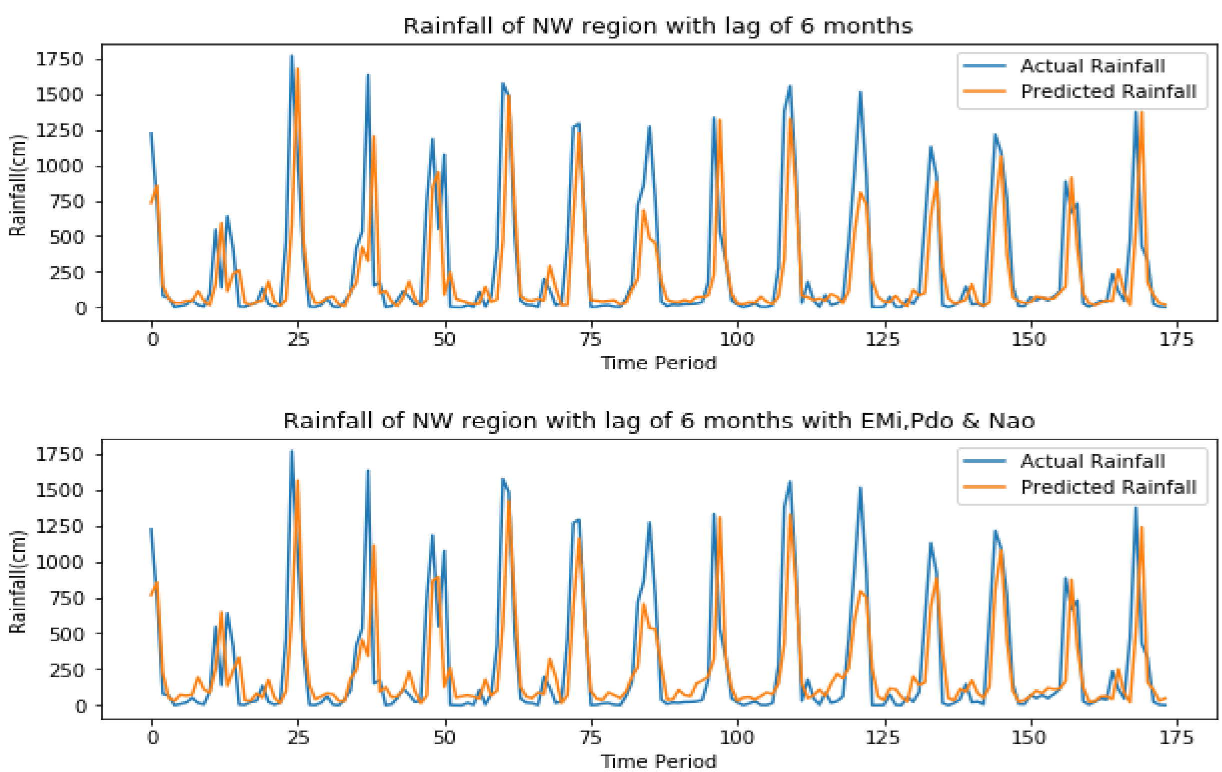


Figure 11. Comparison plots between actual and predicted monthly using inputs from 6 previous months and different combinations of Indices for PE region. X-axis shows the months from the year 2001 to 2015.



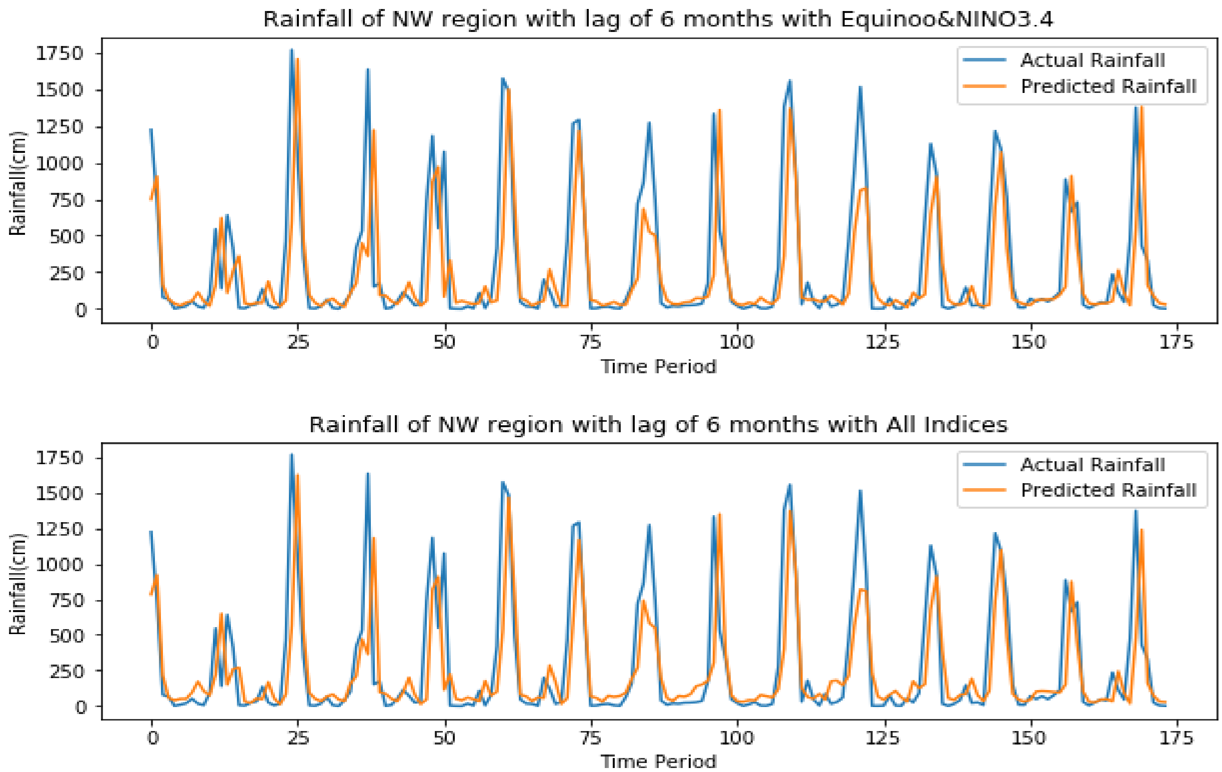
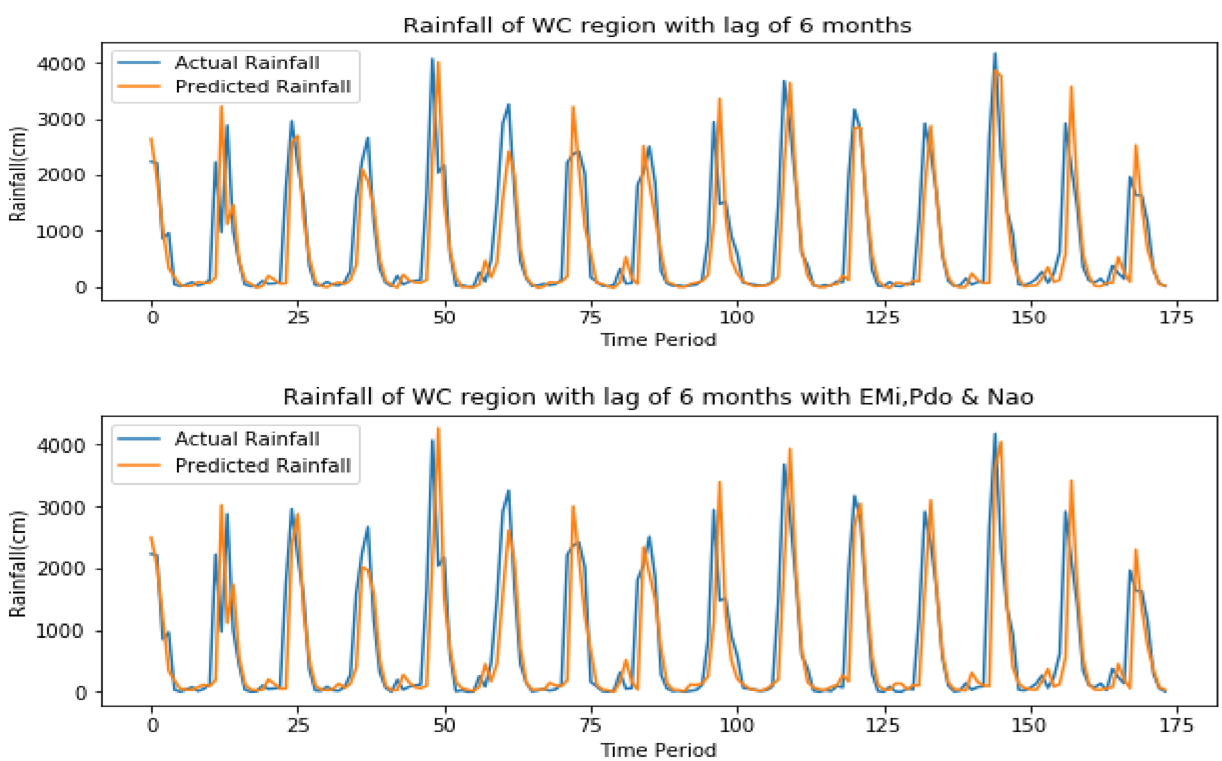


Figure 12. Comparison plots between actual and predicted monthly using inputs from 6 previous months and different combinations of Indices for NW region. X-axis shows the months from the year 2001 to 2015.



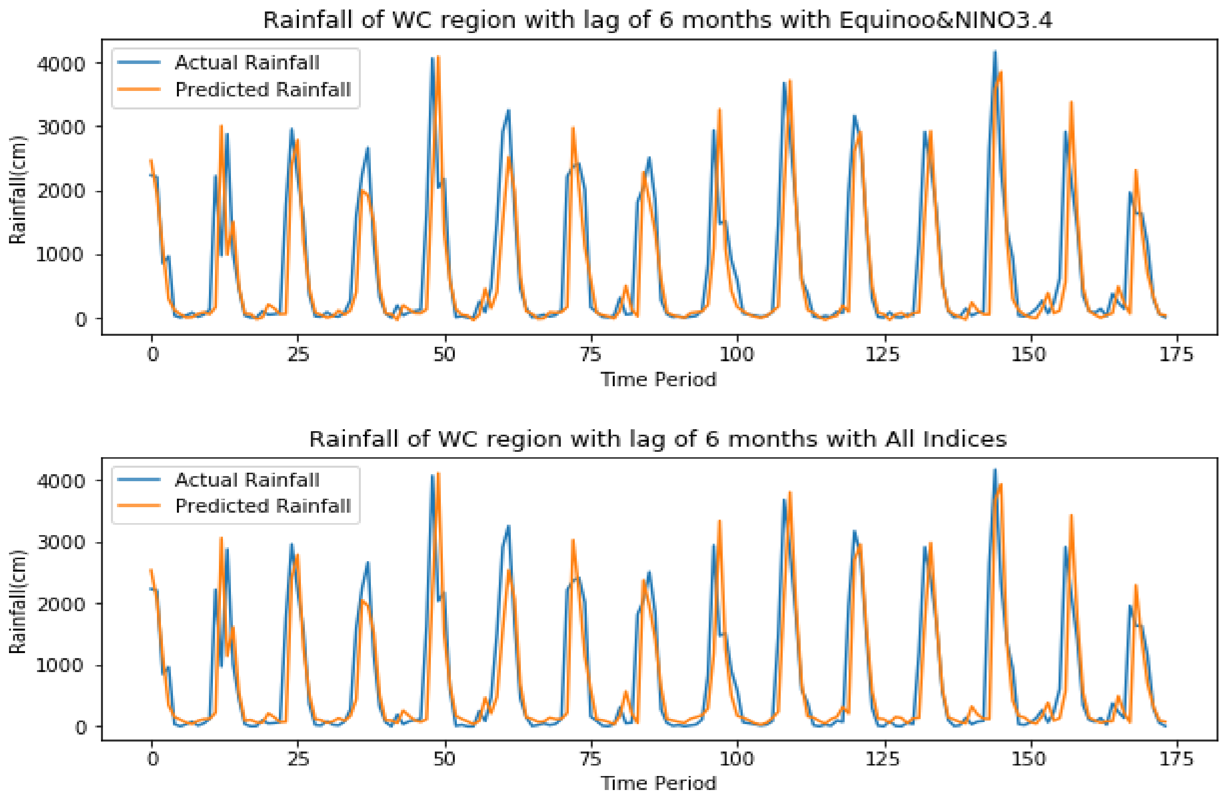
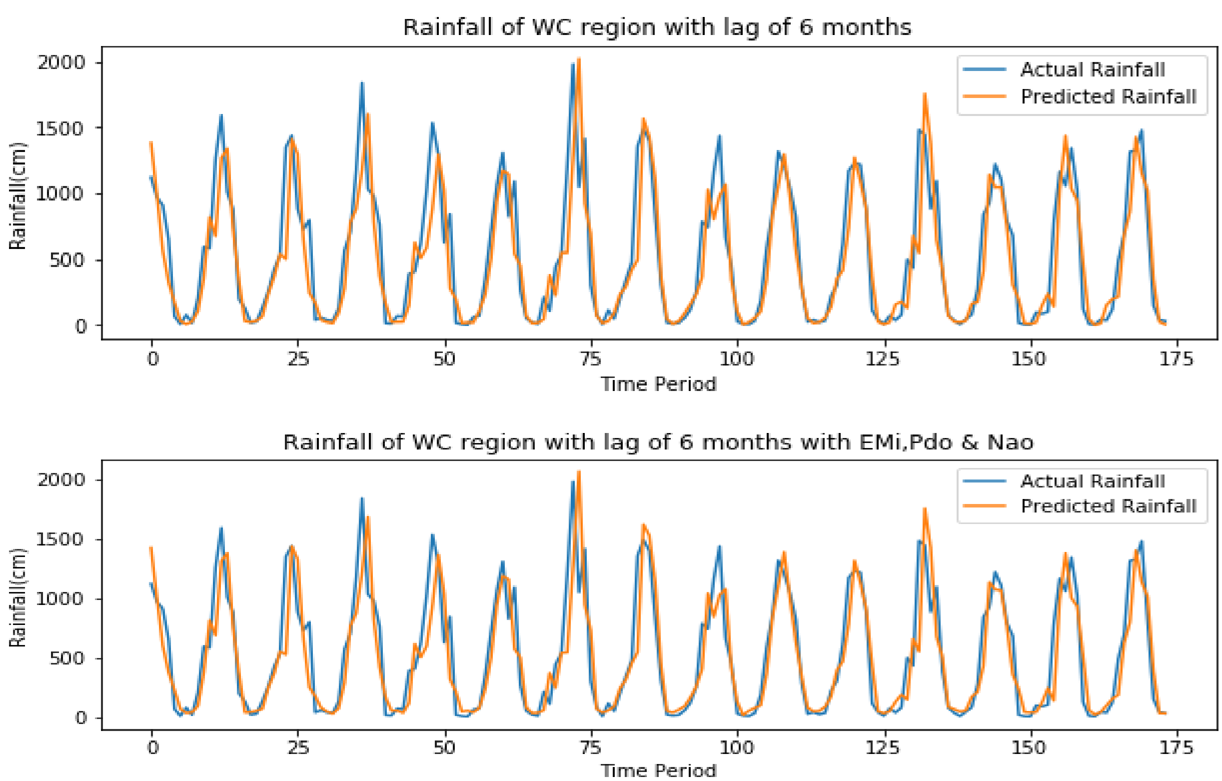


Figure 13. Comparison plots between actual and predicted monthly using inputs from 6 previous months and different combinations of Indices for WC region. X-axis shows the months from the year 2001 to 2015.



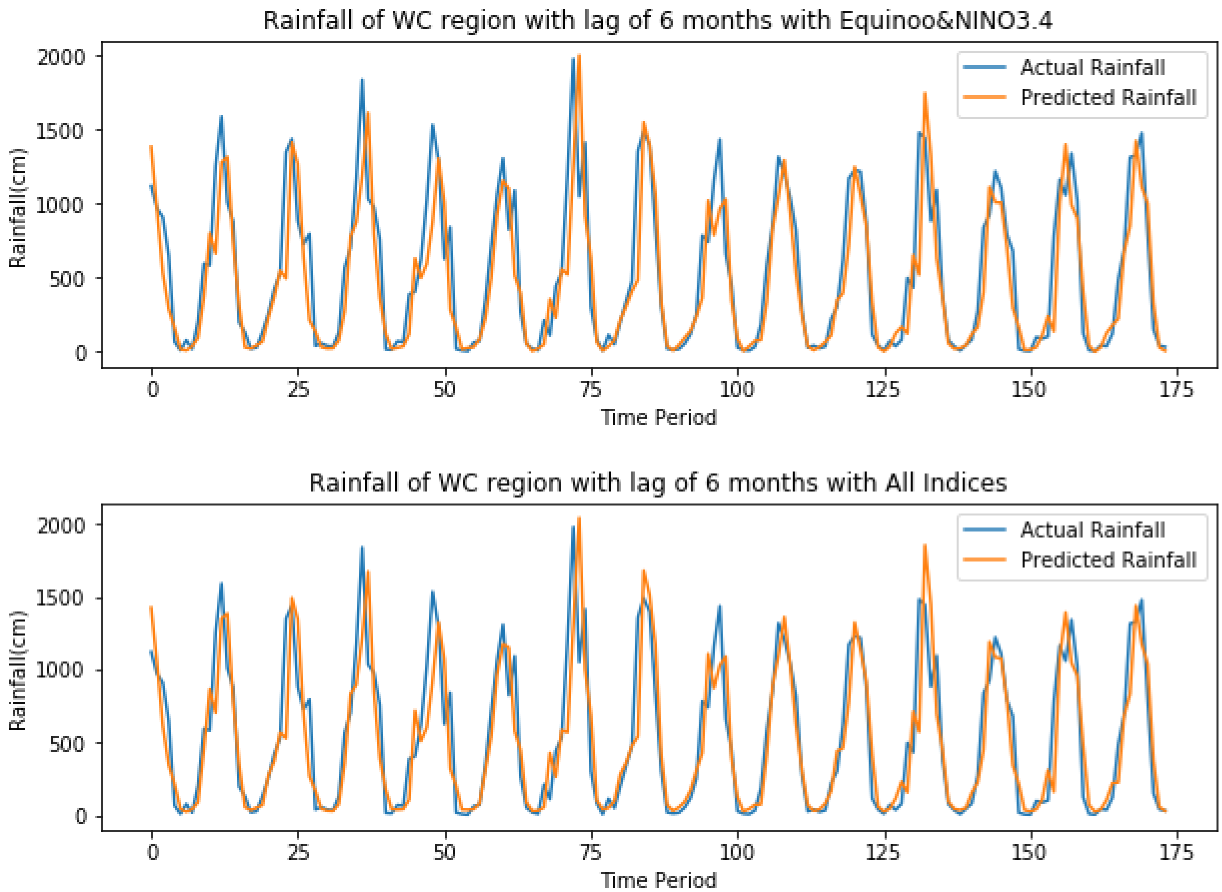


Figure 13. Comparison plots between actual and predicted monthly using inputs from 6 previous months and different combinations of Indices for NE region. X-axis shows the months from the year 2001 to 2015.

TABLE 3: Performance statistics viz. r, RMSE and NSE for different combinations of Indices for different HMRs considering 6 months lag period.

HMR	Index Combination	Performance Statistics: RMSE	Correlation coefficient(r)	NSE coefficient
CNE region	No Index	234.44	0.837	0.575
	(EMI+PDO +NAO)	229.56	0.835	0.609
	(NINO3.4+ EQINOO)	235.97	0.827	0.576
	All Indices	237.94	0.820	0.546

PE region	No Index	551.03	0.750	0.315
	(EMI+PDO +NAO)	552.51	0.741	0.338
	(NINO3.4+ EQINOO)	555.58	0.746	0.228
	All Indices	549.28	0.753	0.336
NW region	No Index	329.35	0.687	0.012
	(EMI+PDO +NAO)	312.67	0.712	0.026
	(NINO3.4+ EQINOO)	328.37	0.688	0.061
	All Indices	317.57	0.704	0.065
WC region	No Index	695.89	0.776	0.507
	(EMI+PDO +NAO)	703.67	0.767	0.496
	(NINO3.4+ EQINOO)	703.47	0.769	0.468
	All Indices	690.82	0.771	0.493
NE region	No Index	275.28	0.852	0.658
	(EMI+PDO +NAO)	273.97	0.851	0.666
	(NINO3.4+ EQINOO)	275.67	0.855	0.651
	All Indices	270.73	0.854	0.685

So from the observations of performance statistics shown in Table 2 and 3, for the described architecture of hybrid model Northwest region(NW) seems to produce poor results which can be observed from NSE coefficient values for both 3 months lag period and 6 months lag period.so the obtained model with mentioned architecture can be discarded from applying to NW region of India. Now coming to the West central(WC) region of India we can clearly see from both the tables the r value and NSE coefficient are under acceptable limits but giving very high RMSE values which are not under acceptable limits. The Central Northeast(CNE) region seems to produce pretty good results with the mentioned model architecture and model with Input of previous year rainfall data and EQUINOO index seems to produce better results and also most of the cases 6 month lag period input produced better results compared to 3 months lag period and also combination of EMI,PDO and NAO indices produces better results compared to the combination of EQUINOO,NINO3.4 and combination of all Indices for CNE region.

The Northeast(NE) region also produced better results, input with 6 months lag period produced much better results compared to 3 month lag period input and also with 6 months lag period NE region had much better correlation coefficient and NSE coefficient. The Peninsular(PE) region of India also produced some high RMSE values but have somewhat better r coefficient values, here also input taken as combination of all indices, EQUINOO index gave better results compared to other combinations but did not produce as much better results as CNE and NE regions. From all these observations we can also say that results produce only based previous months data i.e hidden endogenous properties are not producing better results compared to input taken as both combination of previous months rainfall data and large-scale indices. So for some of the monsoon regions we tried to obtain the variation in association with the large-scale climatic indices and also from all the above plots we can observe that for most of the regions models are able to capture the peak rainfall value which is one of the most important things to predict.

5. CONCLUSIONS

The problem of spatial variation in predictability of the summer monsoon rainfall for different HMRs in India has been addressed, through a Hybrid deep learning model, which uses lagged scale previous rainfall data and lagged large-scale climatic indices as the input variables. The conditional independence structure is employed to identify the complex association among the different lags of climatic indices and summer monsoon rainfall. The index of EQUINOO used for this study captured better performance from CNE, PE and NE regions and also with a 6 months lag period captured much better results. The NW and WC regions are not that effective the proposed model and may try to improve with updating the parameters of the proposed architecture. From the performance of the proposed model, it is concluded that deep learning has the potential to capture the non-linear relationship between the past data,association of large scale Indices with monthly rainfall variability. Therefore it can be effectively used for its prediction with a couple of months ahead. And also with the introduction of Global climatic Indices to the model, the DL model is giving us better and good results of the rainfall predictability when compared to the model only with input of previous rainfall data. The proposed Conv1D-MLP model is also able to capture the higher rainfall values in a better way which was observed during the work. Thus, the hybrid Conv1D-MLP model may be a better choice, considering the extreme rainfall events. So in general, the predictions obtained from the proposed hybrid DL model can be helpful in agriculture, irrigation scheduling, and even flooding due to heavy rainfall.

REFERENCES

Aswin.S, Geetha.P, and Vinayakumar.R, "Deep learning models for the prediction of rainfall," in Proc. Int. Conf. Commun. Signal Process. (ICCSP), Apr. 2018, pp. 657–661.

Ashok, K., Guan, Z. and Yamagata, T. (2001) Impact of the Indian Ocean dipole on the relationship between the Indian Monsoon Rainfall and ENSO. Geophysical Research Letters, 28(23), 4499– 4502. https://doi.org/10.1029/2001GL013294.

Dutta R, Maity R. Spatial variation in long-lead predictability of summer monsoon rainfall using a time-varying model and global climatic indices. Int J Climatol. 2020;1–16. https://doi.org/10.1002/joc.6556

Guhathakurta, P., Rajeevan, M., Sikka, D.R. and Tyagi, A. (2015) Observed changes in southwest monsoon rainfall over India during 1901–2011. International Journal of Climatology, 35(8), 1881–1898. https://doi.org/10.1002/joc.4095.

Haidar.A and Verma. B, "Monthly rainfall forecasting using one dimensional deep convolutional neural network," IEEE Access, vol. 6, pp. 69053–69063, 2018.

Ju, J. and Slingo, J. (1995) The Asian summer monsoon and ENSO. Quarterly Journal of the Royal Meteorological Society, 121(525), 1133–1168. https://doi.org/10.1002/qj.49712152509.

Kahya E, Dracup J a. 1993. U.S. streamflow patterns in relation to the E1 Niño/Southern Oscillation, 29(8): 2491–2503.

Khan.M.I and Maity. R, "Hybrid Deep Learning Approach for Multi-Step-Ahead Daily Rainfall Prediction Using GCM Simulations" IEEE Access March 16 ,2020.

Kumar, D.N., Reddy, M.J. and Maity, R. (2007) Regional rainfall forecasting using large scale climate teleconnections and artificial intelligence techniques. Journal of Intelligent Systems, 16 (4), 307–322. https://doi.org/10.1515/JISYS.2007.16.4.307.

Liu.J. N. K., Y. Hu, J. J. You, and P. W. Chan, “Deep neural network based feature representation for weather forecasting,” in Proc. Int. Conf. Artif. Intell. (ICAI), 2014, p.1.

Meehl, G.A. and Arblaster, J.M. (1998) The Asian-Australian monsoon and El Nino-Southern Oscillation in the NCAR climate system model. Journal of Climate, 11(6), 1356–1385. https://doi. org/10.1175/1520-0442(1998)011<1356:TAAMAE>2.0.CO;2.

Nair, P.J., Chakraborty, A., Varikoden, H., Francis, P.A. and Kuttippurath, J. (2018) The local and global climate forcings induced inhomogeneity of Indian rainfall. Scientific Reports, 8 (1), 1–12. https://doi.org/10.1038/s41598-018-24021-x.

Pant, G.B. and Parthasarathy, S.B. (1981) Some aspects of an association between the southern oscillation and Indian summer monsoon. Archives for Meteorology, Geophysics, and Bioclimatology Series B, 29(3), 245–252. https://doi.org/10.1007/BF02263246.

Rasmusson, E.M. and Carpenter, T.H. (1983) The relationship between eastern equatorial Pacific Sea surface temperatures and rainfall over India and Sri Lanka. Monthly Weather Review, 111(3), 517–528. https://doi.org/10.1175/1520-0493(1983)111<0517:TRBEEP>2.0.CO;2

Vathsala, H. and Koolagudi, S.G. (2017) Long-range prediction of Indian summer monsoon rainfall using data mining and statistical approaches. Theoretical and Applied Climatology, 130 (1–2), 19–33. https://doi.org/10.1007/s00704-016-1862-2.

Zhang.P, Zhang.L, Leung.H, and Wang.J, “A deep-learning based precipitation forecasting approach using multiple environmental factors,” in Proc. IEEE Int. Congr. Big Data (BigData Congress), Jun. 2017, pp. 193–200.