

Effect of Climate Change on the Rainfall Trends: A case study of Pune, India

Shalaka Shah

VIIT: Vishwakarma Institute of Information Technology

Shreenivas Londhe (✉ shreenivas.londhe@viit.ac.in)

VIIT: Vishwakarma Institute of Information Technology

Research Article

Keywords: Climate Change, Precipitation, Artificial Neural Networks

Posted Date: November 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1035885/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

It is the need of the hour to predict the impact of climate change, especially rainfall on the future environmental conditions on local as well as global scales. The present work aims at studying the impact of climate change on the rainfall occurring over Pune, the eighth largest city in India. The General Circulation Models (GCMs) are predominantly used to obtain the climate data all over the globe, at various grid points, for past and future years. Rainfall values obtained from these grid points need to be downscaled to make them location specific. This study proposes a soft computing tool, Artificial Neural Network (ANN) for the purpose of downscaling. The rainfall data at 4 grid points surrounding Pune, was extracted from 5 different GCMs and given as input to ANN with observed rainfall as output, thus forming 5 models. For comparison, a pre-existing downscaling technique, Distribution based scaling (DBS) was used. The coefficient of correlation (r) showed that ANN was working better than DBS. The value of r for ANN was 0.73 for its least accurate model whereas DBS managed to reach 0.73 for its most accurate model. The future rainfall estimated with the help of the trained ANN models show an increase in mean rainfall over the Pune region by $\sim 2 - 15\%$ and decrease in maximum rainfall by $\sim 40 - 65\%$. Peak prediction of rainfall simulated by ANN was not very accurate and hence there is still an opportunity for improvement which is the future scope of this study.

1. Introduction

The Earth's climate is constantly changing, due to the energy transfer cycle, from the sun to the earth's surface and back into the atmosphere. An observation can be drawn from the current scenario that the temperature of the earth is at a constant rise, leading to the increase in global warming. This phenomenon of the continual variations in the global climate is termed as climate change. The adverse effects if this changing climate can be observed over the global hydrological cycle. Certain areas, where the mean rainfall is forecasted to have a decreasing trend, are likely to face high intensity of extreme rainfall events (Semenov and Bengtsson 2002; Wilby and Wigley 2002). This increasing intensity of extreme rainfall events would majorly impact the intensity duration frequency relations (Kao and Ganguly 2011) and may result in flooding of urban areas (Mailhot et al. 2007). Thus, detailed risk analysis needs to be done to study these sudden trends which would prove helpful for design of infrastructures at urban localities.

This changing climate can be studied with the help General Circulation Models (GCMs). These models imitate physical processes in the atmosphere, ocean, cryosphere and land surface. They are one of the most advanced tools currently available for simulating the response of the global climate system to increase in greenhouse gas concentrations. GCMs give values of various climate variables along grid points. To obtain the accurate results, these values have to be downscaled so that they would represent a smaller and specific study area. Various statistical and dynamic downscaling techniques have been a major focus of research since the last decade.

Several studies are being conducted worldwide to gain a better understanding of how these changing climatic conditions are going to impact the rainfall trends. For the sub-continent of India, while some researchers have inferred that the southwest monsoon seems to be weakening due to impact of climate change (Kripalani et al. 2003; Ueda et al. 2006, Stowasser et al. 2009; Sabade et al. 2011 and Krishnan et al. 2013), others say that due to global warming the monsoon rainfall over a broad region encircling South Asia is likely to increase in intensity (Lal et al. 2000; Meehl and Arblaster 2003; and Rupakumar et al. 2006). Rupakumar et al., (2006), have reported an increase in extreme rainfall along the west coast and western parts of central India.

Rana et al., (2014) studied the effects of climate change on the precipitation taking place in Mumbai. They inferred that the intensity of rainfall is expected to increase in the future (2010 – 2099). As per their research the average increase in maximum rainfall is approximately 15–20% in each 30-year time period (2010-2040, 2041-2070, 2071-2099) and about 30–45% in the 90-year period. Along with a seasonal shift and delayed onset of monsoon, the study also showed an increase in the total accumulated annual rainfall, ranging from 300 to 500 mm in the ensemble. The statistical downscaling approach termed Distribution-based Scaling (DBS) technique was applied in their work, which was then tested and applied to scale GCM data. The results of this scaling technique seemed to be potentially useful for impact studies. The delay in onset of monsoon and fluctuations in intensity of rainfall in the Pune region of the state of Maharashtra, India, has given the authors' motivation to conduct this study to assess the impact of climate change on the rainfall intensity and pattern for this particular area.

Recently, numerous studies are being conducted in the field of downscaling of GCM data using soft computing tools like Artificial Neural Networks (ANN).

In the work done by Fistikoglu and Okkan (2011), the precipitation at three meteorological stations was investigated using ANN, which consisted of nine inputs selected as NCEP/NCAR parameters were used as input and the output was the monthly precipitation in the Tahtali basin, Turkey that needed to be downscaled. Levenberg-Marquardt (LM) algorithm was used for training the ANN with one hidden layer having 3 nodes. The results showed that ANN could successfully be used to downscale the grid-based NCEP/NCAR dataset. Chithra et al., (2014) worked with a similar methodology. In their work, a set of predictors chosen from the NCEP/NCAR reanalysis data were used as inputs to ANN models and maximum and minimum temperatures, the Chaliyar river basin in Kerala, India were the outputs. Their results too concluded that ANN is a feasible option for downscaling climate data. Swain et al., (2017) combined ANN-predicted daily runoff with ANN-predicted 30-day cumulative runoff to improve flow estimates. Their conclusions were that ANN performed well for representing 2007–2012 and the drier 1994–1997 periods and that ANN requires much less computational effort than a numerical model application, without compromising the accuracy of results. Abdullahi and Elkiran (2017) focused their work on forecasting the effect, climate change may have on reference evapotranspiration (ET₀) for Girne and Larnaca regions of Cyprus for the next 3 decades (2017 – 2050). A three-layer Feed Forward Back Propagation neural network (FFBP) trained by Levenberg-Marquardt (LM) optimization algorithm was used to predict evapotranspiration for the future. The study implemented two approaches; in the first, the input was kept constant (6 causative input parameters) while changing the number of hidden neurons and in the second approach, the inputs were varied from 2 to 6 parameters and the hidden neurons used were double of the inputs. Their conclusion was that the efficiency of ANN to predict future evapotranspiration in the regions could be significantly increased by adding more number of inputs. Tantelinaiina et al., (2021) chose 26 predictor variables with the help of coefficient of correlation between them and the desired output

(Temperature and pressure at Mangoky River, Madagascar). These 26 variables formed the input matrix for the neural networks with precipitation and temperature data as output. These trained models were then used to estimate the future rainfall and temperature.

In almost all the works mentioned above, standardization of the GCM data is suggested for the removal of systematic bias in the mean and standard deviation of the dataset. Also, it can be observed from many of the technical papers that methods such as bilinear interpolation (appendix 1) have been adopted for re-gridding the data as the grid points of the GCM differs from the grid points of the observed data. As the authors have considerable experience with soft computing techniques they feel that the above 2 steps can be reduced with the help of ANNs. This work proposes that the rainfall can be obtained from GCMs at the 4 nearest grid points surrounding the desired location and used as input variables for training the neural networks with observed rainfall as output. It can be seen in the results section that this novel approach is giving results a shed better than the traditional downscaling technique of Distribution Based Scaling (DBS) as suggested by Rana et al., (2014).

Another observation that was made in the earlier works was that the LM algorithm was used in a majority of papers for training the neural networks. Several other algorithms have proven efficient in solving non-linear problems and thus the accuracy of 9 different training algorithms would be checked and the best one would be adopted for future rainfall simulations. Detailed results for the same are given in the upcoming sections.

2. Artificial Neural Networks

As stated by Londhe and Shah (2019), Artificial Neural Networks (ANNs), possess excellent capabilities in modelling complex processes which have been applies to a widespread domain of problems in the field of hydrology. The structure of the biological neural network in the human brain is mimicked aptly by ANNs, where the input neurons are the causative variables and the phenomenon to be modelled forms the output node, which are finally connected by one or more hidden layers with neurons whose working resembles the biological neural network.

The training of ANNs is an iterative process which continues to minimize the error between the known observed and predicted variables (outputs), which can be done with the help of various training algorithms. Once the network is calibrated, it can be applied to evaluate the output for the unseen data. Londhe and Shah (2019), have shown that ANNs do understand the underlying physical nature of the problem under study and can't just be categorized as "Black Box" technique

The readers are referred to Bose and Liang (1998), Wassermann (1993), The ASCE Task Committee (2000), Maier and Dandy (2000), Dawson and Wilby (2001), Jain and Deo (2006) and Londhe and Panchang (2018) for understanding the preliminary concepts and working of ANN.

3. Study Area And Data Used

Pune, the eighth largest metropolitan city of India having population over 5 million is situated at approximately 18° 32' North latitude and 73° 51' East longitude. The city's total area is 729 square kilometres. Pune lies on the Western margin of the Deccan plateau, at an altitude of 560 m above sea level. As in the Indian subcontinent, Pune receives rainfall in the 'monsoon' spread over almost 4 months. The monsoon lasts from June to October, with moderate rainfall (214.85 mm as per the data acquired from India Meteorological Department, Pune) and temperatures ranging from 22 to 28°C. Figure 1 shows the location of Pune on the map of India.

The observed data was acquired from India Meteorological Department (IMD), Pune (<https://www.imdpune.gov.in/>). The data consists of cumulative monthly values of rainfall for 116 years, from 1901 to 2016, with no gaps.

Along with the observed values, GCM data from CM, obtained from the Copernicus Climate Change Service website (<https://cds.climate.copernicus.eu/>) was used in this study. Daily rainfall data (historical and future) from 5 GCMs was downloaded and extracted for historical (1901 – 2016) and future (2021-2050) time steps. The GCMs used in this study are IPSL_CMA5, HadGEM2_CC, NorEMS1_1, CNRM_CM5 and GFDL_CM3. Details of these are given in Table 1. These 5 GCMs are relatively more accurate for the Indian subcontinent and used by various researchers, like in the work done by Rana et al., (2014).

Table 1
Details of GCMs used for current study

Model	Institution	Grid Size (Lat * Long)
IPSL_CM5A_MR	Institut Pierre-Simon Laplace, France	1.2676*2.5
HadGEM2_CC	Met Office Hadley Centre, UK	1.25*1.875
CNRM_CM5	Centre National de Recherches Meteorologiques, France	1.4008*1.4063
NorESM1_I	Norwegian Climate Centre, Norway	1.8947*2.5
GFDL_CM3	Geophysical Fluid Dynamic Laboratory, USA	2*2.5

The monthly data for the entire globe was initially downloaded, in NetCDF format, which was then converted into a Comma Separated Value (.csv) file using MATLAB. Four (4) grid points surrounding the Pune region were then found and the rainfall data at all those 4 grid points were extracted from the

global data.

4. Methodology

As mentioned in the previous section, the climate data for rainfall was obtained from the 5 GCMs, at 4 grid points surrounding Pune. The technique of bilinear interpolation was used to obtain the value of the rainfall at the required grid point, which gave us the raw GCM data. This raw data was then used to obtain the downscaled data for Pune region using the Distribution based Scaling (DBS) technique, as suggested by Rana et al., (2014).

The DBS approach includes two steps. In the first step, the wet fraction (i.e. proportion of time steps with a non-zero rainfall) is adjusted to match the actual observations. Thus interpolated and observed precipitation data were sorted in descending order and a cut-off value was defined as the threshold that reduced the percentage of wet months in the GCM data to that of the observations. Months having rainfall larger than this threshold value were considered as wet months and the others as dry months. For detailed methodology, the readers are referred to Rana et al., (2014).

In the second step of DBS, the remaining non-zero rainfall were transformed to match the observed cumulative probability distribution in the reference data by fitting gamma distributions (equation 1) to both observed and interpolated monthly rainfall. A shape and scale parameter (α and β respectively) are needed to fit the gamma distribution function. As the distribution of rainfall values ($f(x)$) is highly tilted towards lower intensities, distribution parameters (shape and scale parameters) estimated by maximum likelihood could be dominated by the most frequently occurring values and fail to accurately describe extremes (Rana et al., 2014). To capture the characteristics of normal rainfall as well as extremes, in DBS the rainfall distribution was divided into two partitions separated by the 95th percentile. Two sets of parameters representing non-extreme values and above 95 percentile, representing extreme values – were estimated from observations and the GCM output for the historical data. These parameter sets were in turn used to bias-correct rainfall data from GCM outputs for the entire projection period up to 2050 (P_{DBS} and $P_{DBS,95}$) using inverse gamma function (equations 2 and 3). The authors had developed their own code using MATLAB 2013a, for the same.

$$f(x) = \frac{((x/\beta)^{\alpha-1})(\exp(-x/\beta))}{\beta\Gamma(x)} \quad (1)$$

$$P_{DBS} = f^{-1}(\alpha_{Obs}, \beta_{Obs}, f(P, \alpha_{Raw}, \beta_{Raw})) \quad (2)$$

$$P_{DBS,95} = f^{-1}(\alpha_{Obs,95}, \beta_{Obs,95}, f(P, \alpha_{Raw,95}, \beta_{Raw,95})) \quad (3)$$

Furthermore, using the 4 grid values obtained from each of the 5 GCM datasets as input and observed monthly rainfall as output, 5 separate ANN models were trained and tested. As per the authors' knowledge, ANN would work with reasonable accuracy as a statistical downscaling technique to correct the magnitude of data and make it applicable to a specific geographical area. Thus, the 116 years of monthly GCM data for the years 1901 to 2016 (1392 values) were used along with the observed data to form a 4*1 matrix to train the feedforward backpropagation type of neural network, using MATLAB 2013a. The data division for all models was 70% data (974 values) for training purposes and remaining 30% (418 values) data was used for testing the models.

During the training of the GCM models using it was observed that a 3 layer Feedforward Backpropagation (FFBP) network, having a single hidden layer was not sufficient to train the neural network, perhaps due to the highly non-linear nature of the problem. Figure 2 shows the time series plot for IPSL_MR_CMA5 using LM algorithm for training. It can be seen that the model is not able to finish training due to the complexity of the problem. Hence, the authors added another hidden layer, which proved to be effective in improving the results to a satisfactory extent.

As suggested by Londhe and Deo (2003) network performance as well as prediction accuracy could be improved by appropriate choice of training. The methods involved in their study were: (i) gradient descent with momentum (GDM), (ii) resilient backpropagation (RP), (iii) conjugate gradient Fletcher-Reeves update (CGF), (iv) conjugate gradient Polak-Ribiere update (CGP), (v) Powell Beale restarts (CGB), (vi) scaled conjugate gradient (SCG), (vii) Broyden, Fletcher, Goldfarb and Shanno update (BFG), (viii) one step secant algorithm (OSS) and (ix) Levenberg-Marquardt algorithm (LM). The details of these algorithms can be found in Demuth et al., (1998). Thus while training the GCM models all 9 algorithms were used. Comparison of the results of these 9 algorithms shows that LM algorithm was dominating the other algorithms in majority of the cases. For 4 out of the 5 GCMs, LM seems to be having a comparatively better value of coefficient of correlation(r). 3 out of the 5 models show that LM seems to be dominating the other algorithm when the values of RMSE are considered. Even though the difference in the r and RMSE values is not major, considering the computation time required and overall efficiency of working of all the algorithms LM was a clear winner. Thus it was adopted for training the neural networks in this study. The consolidated results of all the algorithms are shown in table 2.

All the other parameters of the models were fixed by trial and error method. Lastly, 5 separate models were finalised as the best performing models of each GCM. The network details for each of the final 5 models are given in table 2. These 5 trained networks were then used to estimate the rainfall that would be occurring over Pune for the future 30 years, that is, from 2021 to 2050. For this purpose the future rainfall (2021 to 2050) for the 4 grid points surrounding Pune, extracted from the 5 GCMs mentioned in table 1, were used as unseen inputs.

Table 2. Comparison of different ANN algorithms

GCM	IPSL -CMA5_MR		HadGEM2_CC		NorESM1_I		CNRM_CM5		GFDL_CM3	
Algorithm	r	RMSE	r	RMSE	r	RMSE	r	RMSE	r	RMSE
GDM	0.56	99.34	0.59	99.95	0.73	100.99	0.65	96.24	0.73	81.62
RP	0.62	93.66	0.74	80.08	0.76	78.67	0.73	81.30	0.74	81.04
CGF	0.63	91.93	0.71	87.14	0.70	85.83	0.74	79.85	0.74	81.05
CGP	0.65	90.25	0.64	91.76	0.76	78.38	0.74	79.98	0.74	81.05
CGB	0.65	90.28	0.64	91.79	0.76	78.48	0.74	79.87	0.74	81.69
SCG	0.65	90.06	0.74	80.93	0.76	78.45	0.74	79.88	0.74	81.02
BFG	0.66	89.30	0.75	79.58	0.76	78.33	0.74	79.88	0.74	81.04
OSS	0.64	91.44	0.73	81.87	0.75	79.56	0.74	80.19	0.74	81.04
LM	0.73	81.13	0.74	84.06	0.76	78.20	0.75	77.97	0.74	81.04

Table 3. Details of ANN models developed

GCM	No of Hidden neurons	No of Epochs	Transfer and Activation Functions	Normalization Range
IPSL_CM5A_MR	Hidden layer 1 & 2 – 3 neurons	115	logsig logsig purelin	-1 to 1
HadGEM2_CC	Hidden layer 1 & 2 – 5 & 4 neurons resp.	120	logsig logsig purelin	0 to 1
CNRM_CM5	Hidden layer 1 & 2 – 2 neurons	120	logsig logsig purelin	-1 to 1
NorESM1_I	Hidden layer 1 & 2 – 1 neurons	120	logsig logsig purelin	0 to 1
GFDL_CM3	Hidden layer 1 & 2 – 1 neurons	120	logsig logsig purelin	0 to 1

5. Results And Discussion

To judge the accuracy of both the

downscaling techniques, the authors' used statistical parameters as well as error measures. As mentioned by Dawson et al., (2007), error measures can be classified broadly into three categories, that is, Absolute Parameters, Relative Parameters and Dimensionless Coefficients. Hence, for the present work two absolute parameters (namely, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE)) and two dimensionless parameters (Coefficient of Correlation (r) and Coefficient of Error (CE)) were used. Since the precision in prediction of peaks plays an important role in this study, another relative error measure, Percentage Error in Peak (PEP) was also used. Readers are referred to Dawson et al., (2007) for detailed description of all the error measures. As for the statistical parameters, the mean, standard deviation and maximum value of time series were used. Statistical parameters help in judging how precise the mean, standard deviation and maximum rainfall values from DBS and ANN models, can get to the parameters calculated from observed rainfall series.

Table 4 shows comparison between the raw GCM data and the monthly precipitation values obtained from DBS and ANN. It can be inferred from the table that ANN is consistently working a shed better than the other two methods. Also, an extensive result analysis, of observed, DBS and ANN models, was done for monthly as well as annual rainfall for the past 116 years. The trained monthly models were used to compute the annual values. No separate annual models were trained or tested. The annual rainfall analysis was further bifurcated as annual accumulated rainfall and annual average rainfall. The rainfall for all 12 months of each year was summed up to obtain the annual accumulated rainfall and their mean gives the annual average rainfall. This detailed analysis would throw light on the distribution of rainfall over the entire year. For example, if the average rainfall shows an increasing trend but the maximum rainfall does not, that shows a more distributed rainfall pattern throughout the year. If it is the other way around, then the total rainfall occurring over the region might be less, with extreme rainfall on certain days and no rainfall on the rest. Tables 5, 6 and 7 show the comparison of both the techniques for monthly, annual accumulated and annual average rainfall respectively. Alongside, to keep a check on the performance of the ANN

models, decade wise RMSE analysis was carried out for all the 116 years. No sudden spike or fall in the RMSE value indicates that the ANN models are reliable and would also help in comparison of the accuracy of all GCMs with each other. Table 8 gives an insight on the decade wise RMSE analysis for all GCMs. Figures 4 to 13 give a pictorial representation of the results, in the form of time series plot and scatter plots, of all 5 GCMs. All the result analysis have been done for both training and testing datasets. Result analysis of only testing datasets was not feasible as comparison with DBS would have been difficult since DBS requires entire dataset for its working methodology.

Table 4
Coefficient of correlation for all 3 methods

GCM	Coefficient of correlation (r)		
	Raw GCM	DBS	ANN
IPSL_CM5A_MR	0.56	0.54	0.73
HadGEM2_CC	0.60	0.62	0.74
CNRM_CM5	0.64	0.67	0.75
NorESMLI	0.72	0.73	0.76
GFDL_CM3	0.73	0.73	0.74

Table 5
Result analysis of monthly rainfall data (mm)

	Obs	IPSL_CMA5		HadGEM2_CC		CNRM_CM5		NorESM1_1		GFDL_CM3	
		DBS	ANN	DBS	ANN	DBS	ANN	DBS	ANN	DBS	ANN
Mean	84.5	90.44	87.48	79.33	88.9	103.51	88.05	101.05	92.85	92.22	94.14
Max	703.1	869.81	454.51	1014.19	395.08	926.85	355.44	763.45	335.00	551.25	389.65
Standard Deviation	118.71	141.01	89.65	164.89	112.91	133.64	92.97	141.53	102.82	113.03	102.55
RMSE		125.23	81.13	129.97	84.06	105.06	77.97	99.98	78.20	85.93	81.03
MAE		74.10	50.93	69.32	49.70	63.11	49.63	59.67	51.04	52.00	51.34
CE		-0.11	0.53	-0.20	0.50	0.21	0.57	0.29	0.57	0.47	0.53
PEP		-23.71	35.35	-44.24	43.80	-31.82	49.45	-8.58	52.35	21.59	44.58

Table 6
Result analysis of annual accumulated rainfall data (mm)

	Obs	IPSL_CMA5		HadGEM2_CC		CNRM_CM5		NorESM1_1		GFDL_CM3	
		DBS	ANN	DBS	ANN	DBS	ANN	DBS	ANN	DBS	ANN
Mean	1013.80	1085.3	1049.8	952.1	1067.3	1242.1	1056.7	1212.6	1114.1	1106.7	1129.7
Max	1753.6	2110.2	1570.00	2185.9	1426.2	1891.2	1891.2	1787.8	1504.2	1472.00	1494.8
Standard Deviation	260.53	313.55	169.28	401.90	191.68	243.47	243.47	215.59	135.07	160.18	142.36
RMSE		403.50	303.62	446.35	302.01	401.34	276.47	393.37	326.05	335.19	328.77
MAE		334.81	241.32	381.31	230.93	320.63	216.26	316.39	256.46	265.64	254.92
CE		-1.42	-0.37	-1.96	-0.35	-1.39	-0.13	-1.3	-0.58	-0.66	-0.61
PEP		-20.33	10.47	-24.65	18.67	-36.32	93.89	-1.95	14.22	16.06	14.76

Table 7
Result analysis of annual average rainfall data (mm)

	Obs	IPSL_CMA5		HadGEM2_CC		CNRM_CM5		NorESM1_1		GFDL_CM3	
		DBS	ANN	DBS	ANN	DBS	ANN	DBS	ANN	DBS	ANN
Mean	84.5	90.4	87.5	79.3	88.9	103.5	88.1	101	92.8	92.9	94.1
Max	146.1	175.8	130.8	182.2	118.9	157.6	115.6	149	125.3	122.7	124.6
Standard Deviation	21.71	26.13	14.10	33.49	15.97	20.29	11.42	17.96	11.26	13.34	11.86
RMSE		33.63	25.30	37.19	25.16	33.44	23.04	32.78	27.17	27.93	27.39
MAE		27.90	20.11	31.77	19.24	26.80	18.02	26.36	21.37	22.13	21.24
CE		-1.42	-0.37	-1.96	-0.36	-1.39	-0.13	-1.3	-0.58	-0.66	-0.61
PEP		-20.33	10.47	-24.65	18.67	-36.32	93.89	-1.95	14.22	16.06	14.76

Table 8
RMSE analysis per decade (mm)

GCM	IPSL_CMA5	HadGEM2_CC	CNRM_CM5	NorESM1_1	GFDL_CM3
RMSE per 10 years					
1901-1910	72.62	64.26	62.25	68.28	72.77
1911-1920	69.22	73.07	76.71	73.68	78.38
1921-1930	71.40	58.79	63.91	65.73	70.10
1931-1940	77.89	81.44	78.92	70.78	69.21
1941-1950	91.66	66.01	93.51	78.76	77.13
1951-1960	87.04	68.70	81.97	72.74	82.89
1961-1970	84.16	61.03	87.83	71.06	70.92
1971-1980	88.57	68.10	100.47	82.03	85.76
1981-1990	70.36	71.67	71.98	70.67	75.59
1991-2000	79.91	74.05	85.42	79.47	74.97
2001-2010	67.75	89.35	74.99	80.67	90.00
2011-2016	76.22	81.00	72.08	87.82	86.42
Average RMSE	78.07	71.46	79.17	75.14	77.85

The result analysis for monthly models (Table 5) throws light on the superior performance of ANN over DBS. The superiority of ANN models can be observed from the RMSE values of all 5 models that lie in the range of 77.97 mm (CNRM_CM5) to 84.06 mm (HadGEM2_CC) in case of ANN models and in between 85.93 mm (GFDL_CM3) and 129.97 mm (HadGEM2_CC) in case of DBS models. Similarly, the least accurate model in case of MAE is of GFDL_CM3 (51.34mm) for ANN which is still better than the best DBS model as per MAE, that is, GFDL_CM3 (52.00 mm). The values of CE also show a similar trend. Where ANN ranges between 0.50 (HadGEM2_CC) to 0.57 (CNRM_CM5 and NorESM1_1), DBS shows negative results in case of IPSL_CMA5 (-0.11) and HadGEM2_CC (-0.20) models. Likewise, the result analysis for annual accumulated (Table 6) and annual average rainfall (Table 7) values also show that ANN seems to be working a shed better than DBS as a downscaling technique.

However, ANN seems to be lagging in the aspect of modelling the peak rainfall values, both monthly and annually. The maximum rainfall and PEP values are an indication of the accuracy by which the peaks are being simulated. A positive PEP value indicates underestimation and a negative value indicates overestimation of the peaks. It can be observed for both monthly (Table 5) and annual (Table 6 and Table7) results that ANN has consistently under predicted the peak rainfall values, as compared to DBS, where they are being consistently over predicted.

A similar pattern can be observed in the time series plots of all the 5 GCMs as well (Figures 4, 6, 8, 10 and 12). In each of these figures, ANN appears to be under predicting the higher order values and DBS is showing a over predicting trend.

A brief glance at the scatter plots for all 5 GCMs (Figures 5, 7, 9, 11, 13) show that as compared to DBS, ANN values are aligned better along the best fit line. As discussed earlier, from these figures as well it can be noted that ANN is under predicting results and not giving satisfactory results for higher order values. It can be observed from figure 7 that there is a saturation of ANN values at approximately 400mm. A similar trend is observed in the scatter plots of the other GCMs. This is consistent with the PEP values hence proving that the models are not up to the mark when it comes to higher order values. The network is not able to predict values beyond a certain limit. This might be occurring due to insufficient information being provided to the neural networks.

Thus, the authors feel that this drawback could be overcome by using rainfall causative parameters, along with previous year's rainfall, as input to the ANN models, so that they get more information about the physical process being modelled.

Table 8, gives the detailed RMSE analysis for all the monthly climate models, indicates that ANN models have been trained with sufficient accuracy as there is no sudden increase or decrease in the error value. They lie in a similar range for all models. This inference can be backed up by figure 14, which depicts a plot of the decade wise RMSE analysis for all climate models. This study also shows that all 5 GCMs are working almost at par to model the rainfall at Pune. Out of all 5, HADGEM2_CC seems to be having an upper hand, followed by NorESM1_1. However, comparison of these models could be carried out only after the peak prediction of the models can be improved.

Further, the 5 trained ANN models were used to estimate real time rainfall for future 30 years, that is, 2021-2050. These models were run by giving the rainfall at 4 surrounding grid points as input to the pre-trained ANN models having the parameters mentioned in Table 2. For the future rainfall series obtained using ANN the mean and maximum value was calculated for both monthly as well as annual data series. These future parameters can be studied from Table 9. The table also shows us the percentage increase in the mean and maximum rainfall, as compared to the previous years. It can be seen in this table that the mean rainfall seems to be increasing in the future years. The percentage increase varies from 2.09% (HadGEM2_CC) to 15.76% (NorESM1_1). Whereas, the monthly maximum rainfall seems to be decreasing by a range of -39.19% (IPSL_CMA5) to -64.90% (HadGEM2_CC). An increase in mean rainfall and decreasing maximum rainfall would point to an equally distributed rainfall throughout the year. No particular month(s) might have extremely high rainfall and the monsoon season might be extended beyond the months of June to September. The authors' being residents of Pune, have been observing a similar rainfall pattern for the region. The monsoon season seems to continue even after the month of September. But, as per just raw observation, there seems to be an increase in the average as well as peak rainfall for the region of Pune. Rana et al., (2014) also have shown in their work that for the city of Mumbai, which lies fairly close to Pune, the average increase in maximum rainfall is about ~15–20% in each 30-year time slice (2010-2040, 2041-2070, 2071-2099) and ~30–45% in the 90-year transient period. Even though these two cities have a fair amount of variation in their rainfall trends, so much difference in the future maximum rainfall does not seem plausible. It has already been discussed earlier in the paper that ANN is constantly under predicting the peak values, as a result of which, the same error is being carried forward to the future rainfall as well. Hence, the dependability of these results could be increased after the peak estimation of the trained models is more accurate.

Table 9
Statistics of rainfall predicted to occur from 2021 – 2050.

GCM		Future 30 year values										
Historical Data		IPSL_CMA5		HadGEM2_CC		CNRM_CM5		NorESM1_1		GFDL_CM3		
		Past (mm)	Actual (mm)	% Increase								
Monthly	Mean	84.5	86.97	2.92	93.61	10.78	86.27	2.09	97.82	15.76	95.62	13.16
	Max	703.1	427.57	-39.19	246.76	-64.90	297.70	-57.66	327.46	-53.43	396.88	-43.55
Annual Accumulated	Mean	1013.8	1043.64	2.94	1123.34	10.80	1035.19	2.11	1173.85	15.79	1147.41	13.18
	Max	1753.6	1468.94	-16.23	1455.20	-17.02	1241.49	-29.20	1488.90	-15.09	1431.16	-18.39
Annual Average	Mean	84.5	86.97	2.92	93.61	10.78	86.27	2.09	97.82	15.76	95.62	13.16
	Max	146.1	122.41	-16.21	121.27	-17.00	103.46	-29.19	124.08	-15.08	119.26	-18.37

Table 10 and figure 15 show how the future rainfall trend would be per decade, that is, from 2021-2030, 2031-2040 and 2041-2050. Except HADGEM2_CC, all other models show a slight increase in total rainfall up to 2040 and then it seems to be decreasing slightly. HADGEM2_CC shows a sudden rise in the rainfall for the last decade. If compared to decade wise RMSE analysis (figure 14), it can be seen that the RMSE value for this model increases for the last 2 decades and thus, the operational model also might have an increased error percentage. Here again, the rainfall trend over the next 3 decades could be depicted more accurately if the peak prediction by ANN models could be improved upon.

Table 10
Total Future Rainfall (Decade Wise in mm).

GCM	IPSL_CMA5	HadGEM2_CC	CNRM_CM5	NorESM1_1	GFDL_CM3
Total Rainfall in 10 year window (mm)					
2021-2030	10127.23	10841.52	10244.47	11527.54	11628.59
2031-2040	10912.64	10595.55	10534.77	11902.56	11499.98
2041-2050	10269.19	12263.02	10276.48	11785.49	11293.77

From the above results, it can be clearly concluded that ANN is working better than DBS, as a downscaling technique. Not only is it more accurate but also a less tedious process as the grid points can be directly used to train the models. The drawback that is seen in the results of ANN is that the peaks are not estimated with satisfactory accuracy. The authors' feel that this can be improved by using rainfall causative parameters, such as temperature, air

pressure, wind and humidity, as input parameters along with the rainfall. This would give ANN more data to learn the rainfall trend from. This is the future scope of the current study.

6. Conclusion And Future Scope

The present environmental scenario clearly dictates the need for an estimate of how the changing climate is going to affect the various hydrological parameters such as rainfall. Global Climate Models are being used all over the globe for this purpose. These GCMs give climatic parameters over large grid points for historic and future time periods. The parameter values obtained need to be downscaled so that they can be made location specific. The current work suggests the use of a soft computing tool, Artificial Neural Networks, as a statistical downscaling technique, to study the rainfall 30 years into the future. The study has been carried out for the city of Pune.

The rainfall values obtained from 5 GCMs, IPSL_CMA5, HADGEM2_CC, NorEMS1_1, CNRM_CM5 and GFDL_CM3, for the 4 grid points surrounding Pune were used as input for each model. The output was the actual measured rainfall obtained from the Indian Meteorological Department, from years 1901 to 2016. A technique, namely, Distribution Based Scaling (DBS) was also used for comparison with ANN, as suggested by Rana et al., (2014). Statistical parameters and error measures were used to judge the accuracy of the models, along with time series plots and scatter plots. The results clearly indicated that ANN is performing better than DBS in all aspects.

The future rainfall estimated with the help of the trained ANN models show an increase in mean rainfall over the Pune region by $\sim 2 - 15\%$ and decrease in maximum rainfall by $\sim 40 - 65\%$. There is a scope of improvement for the percentage decrease in the future rainfall as it was observed that ANN was not able to catch the peaks with the desired accuracy. For this, the authors suggest using rainfall causative parameters as input to ANN models along with rainfall. Study for the same is being carried out by the authors currently. Furthermore, analysis of the climate models showed that all 5 GCMs are working at par, but HADGEM2_CC seems to be working a tad better as compared to the others apart from the slight increase in error for the last 2 decades.

The inherent characteristic of ANN to learn from the input data fed to it thus, proves to be an advantage for the current problem and ANN could be explored further as a statistical downscaling tool for modelling various hydrological parameters using climate models.

7. Appendix I - Bilinear Interpolation

Bilinear interpolation is an extension of linear interpolation for interpolating functions of two variables (e.g. x and y) on a regular grid. In this technique linear interpolation is first performed in one direction, and then again in the other direction. If values of four points, (x_1, y_1) , (x_1, y_2) , (x_2, y_1) and (x_2, y_2) , are known and it is required to find value at point (x, y) , a linear interpolation is done between the points (x_1, y_1) and (x_2, y_1) to obtain value at (x, y_1) . The same is done between point (x_1, y_2) and (x_2, y_2) to obtain value at (x, y_2) . Then a linear interpolation is done between the newly interpolated values to obtain value at the required point (x, y) .

Declarations

Data Availability Statement

All data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Financial Declarations

The authors have no relevant financial or non-financial interests to disclose.

Authors' Contribution

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by both the authors.

Code Availability

All models or code that support the findings of this study are available from the corresponding author upon reasonable request.

Ethics Approval

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Consent to participate

Informed consent was obtained from all individual authors included in the study.

Consent for publication

Both the authors have consented to the submission of the technical paper to the journal.

Acknowledgements

The authors sincerely thank Indian Meteorological Department for providing the data as per the request.

References

1. Abdullahi J, Elkiran G (2017) Prediction of the future impact of climate change on reference evapotranspiration in Cyprus using artificial neural network. *Procedia Computer Science* 120:276–283
2. Bose NK, Liang P (1998) *Neural Network Fundamentals with Graphs, Algorithms, and Applications*. Tata McGraw-Hill
3. Chithra NR, Thampi SG, Surapaneni S, Nannapaneni R, Reddy AAK, Kumar J, D (2014) Prediction of the likely impact of climate change on monthly mean maximum and minimum temperature in the Chaliyar river basin, India, using ANN-based models. *Theor Appl Climatol* 121:581–590
4. Dawson CW, Abrahart RJ, See IM (2007) HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environmental Modelling & Software* 22:1034–1052
5. Dawson CW, Wilby RL (2001) Hydrological Modelling using Artificial Neural Networks. *Prog Phys Geogr* 25(1):80–108
6. Demuth H, Beale M, Hagen M (1998) *Neural network toolbox user's guide*. The Mathworks Inc.
7. Fistikoglu O, Okkan U (2011) Statistical Downscaling of Monthly Precipitation Using NCEP/NCAR Reanalysis Data for Tahtali River Basin in Turkey. *ASCE J Hydrol Eng* 16(2):157–164
8. Jain P, Deo MC (2006) Neural Networks in Ocean Engineering. *Ships and Offshore Structures* 1(1):25–35
9. Kao S-C, Ganguly AR (2011) Intensity, duration, and frequency of precipitation extremes under 21st-century warming scenarios. *J Geophys Res* 116(D16):D16119
10. Kripalani RH, Kulkarni A, Sabade SS, Khandekar ML (2003) Indian monsoon variability in a global warming scenario. *Nat Hazards* 29(2):189–206
11. Krishnan R, Sabin TP, Ayantika DC, Kitoh A, Sugi M, Murakami H, Turner AG, Slingo JM, Rajendran K (2013) Will the South Asian monsoon overturning circulation stabilize any further? *Clim Dyn* 40(1–2):187–211
12. Lal M, Meehl GA, Arblaster JM (2000) Simulation of Indian summer monsoon rainfall and its intraseasonal variability in the NCAR climate system model. *Reg. Environ. Change* 1. (3–4),163–179
13. Londhe SN, Deo MC (2003) Wave tranquility studies using neural networks. *Mar Struct* 16(6):419–436
14. Londhe SN, Panchang V (2018) ANN Techniques: A Survey of Coastal Applications. *Advances in Coastal Hydraulics. World Scientific*,119–234
15. Londhe SN, Shah SS (2019) A Novel Approach for Knowledge Extraction from Artificial Neural Networks. *ISH Journal of Hydraulic Engineering* 25(3):269–281
16. Maier HR, Dandy GC (2000) Neural networks for Prediction and Forecasting of Water resources variables: a review of modelling issues and applications. *Elsevier Environ Model Soft* 15:101–124
17. Mailhot A, Duchesne S, Caya D, Talbot G (2007) Assessment of future change in intensity-duration-frequency (IDF) curves for Southern Quebec using the Canadian Regional Climate Model (CRCM). *J Hydrol* 347(1–2):197–210
18. Meehl GA, Arblaster JM (2003) Mechanisms for projected future changes in south Asian monsoon precipitation. *Clim Dyn* 21(7–8):659–675
19. Rabezanahary Tanteliniaina MF, Rahaman MH, Zhai J, Water (2021) 2021. 13, 1239, <https://doi.org/10.3390/w13091239>
20. Rana A, Foster K, Bosshard T, Olsson J, Bengtsson L (2014) Impact of climate change on rainfall over Mumbai using Distribution-based Scaling of Global Climate Model projections. *Journal of Hydrology: Regional Studies* 1:1107–1128
21. Rupakumar K, Sahai AK, Kumar KK, Patwardhan SK, Mishra PK, Revadekar JV, Kamala K, Pant GB (2006) High-resolution climate change scenarios for India for the 21st century. *Curr Sci* 90:334–345
22. Sabade SS, Kulkarni A, Kripalani RH (2011) Projected changes in South Asian summer monsoon by multi-model global warming experiments. *Theor Appl Climatol* 103(3–4):543–565
23. Semenov VS, Bengtsson LB (2002) Secular trends in daily precipitation characteristics: greenhouse gas simulation with a coupled AOGCM. *Clim Dyn* 19(2):123–140
24. Stowasser M, Annamalai H, Hafner J (2009) Response of the South Asian summer monsoon to global warming: mean and synoptic systems. *J Clim* 22(4):1014–1036
25. Swain ED, Gomez-Fragoso J, Torres-Gonzalez S (2017) Impacts of Climate Change on Water Availability Using Artificial Neural Network Techniques. *J Water Resour Plann Manage* 143(12):04017068
26. The ASCE task committee (2000) Artificial neural networks in hydrology, II: hydrologic applications. *J Hydrol Eng* 5(2):124–137
27. Ueda H, Iwai A, Kuwako K, Hori ME (2006) Impact of anthropogenic forcing on the Asian summer monsoon as simulated by eight GCMs. *Geophys Res Lett* 33(6):L06703
28. Wasserman PD (1993) *Advanced methods in neural computing*. Van Nostrand Reinhold, New York, NY.264
29. Wilby RL, Wigley TML (2002) Future changes in the distribution of daily precipitation totals across North America. *Geophys Res Lett* 29(7):1135

Figures

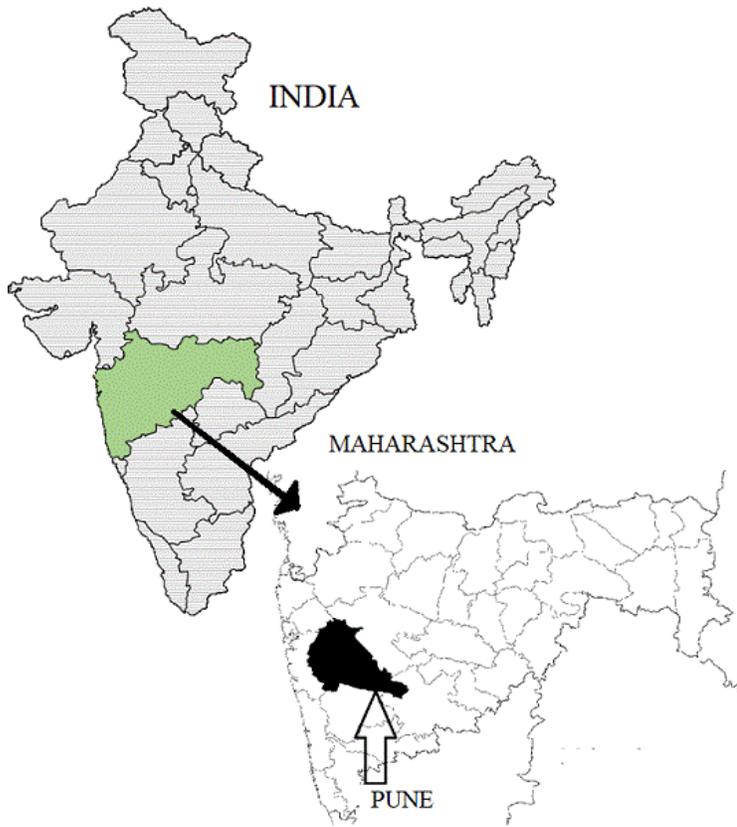


Figure 1

Location of Pune on the map of Maharashtra

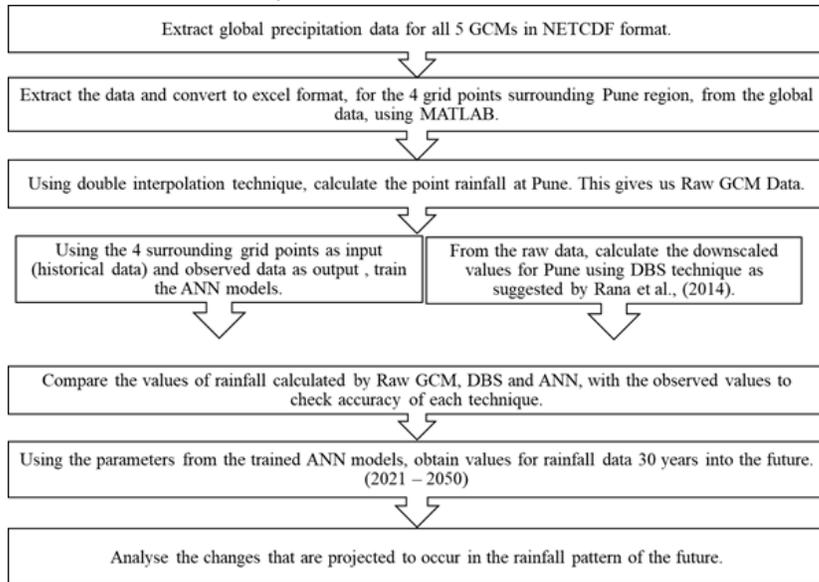


Figure 2

Flowchart explaining methodology of the present work

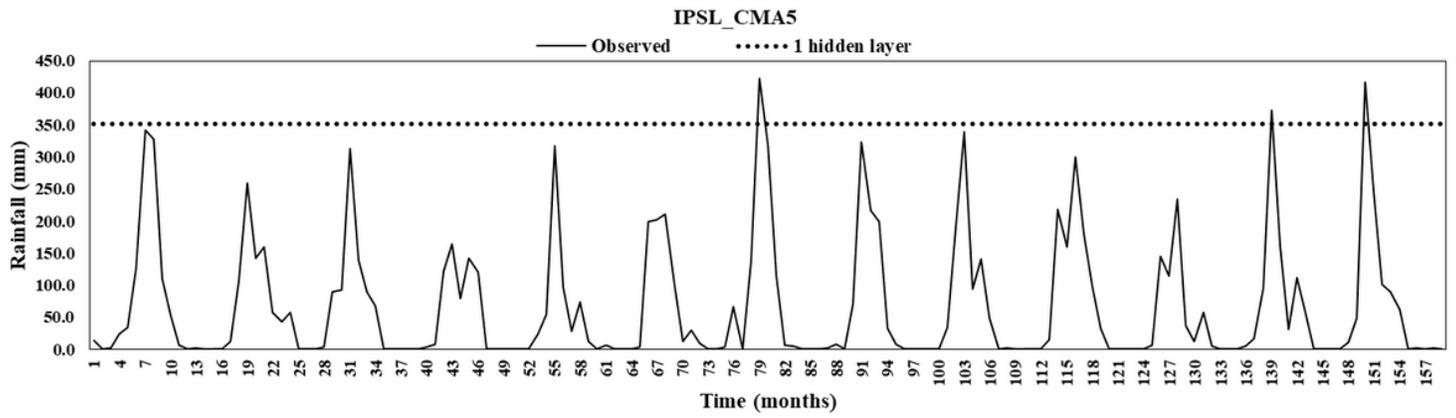


Figure 3

Time Series plot for IPSL_CMA5 (1 hidden layer)

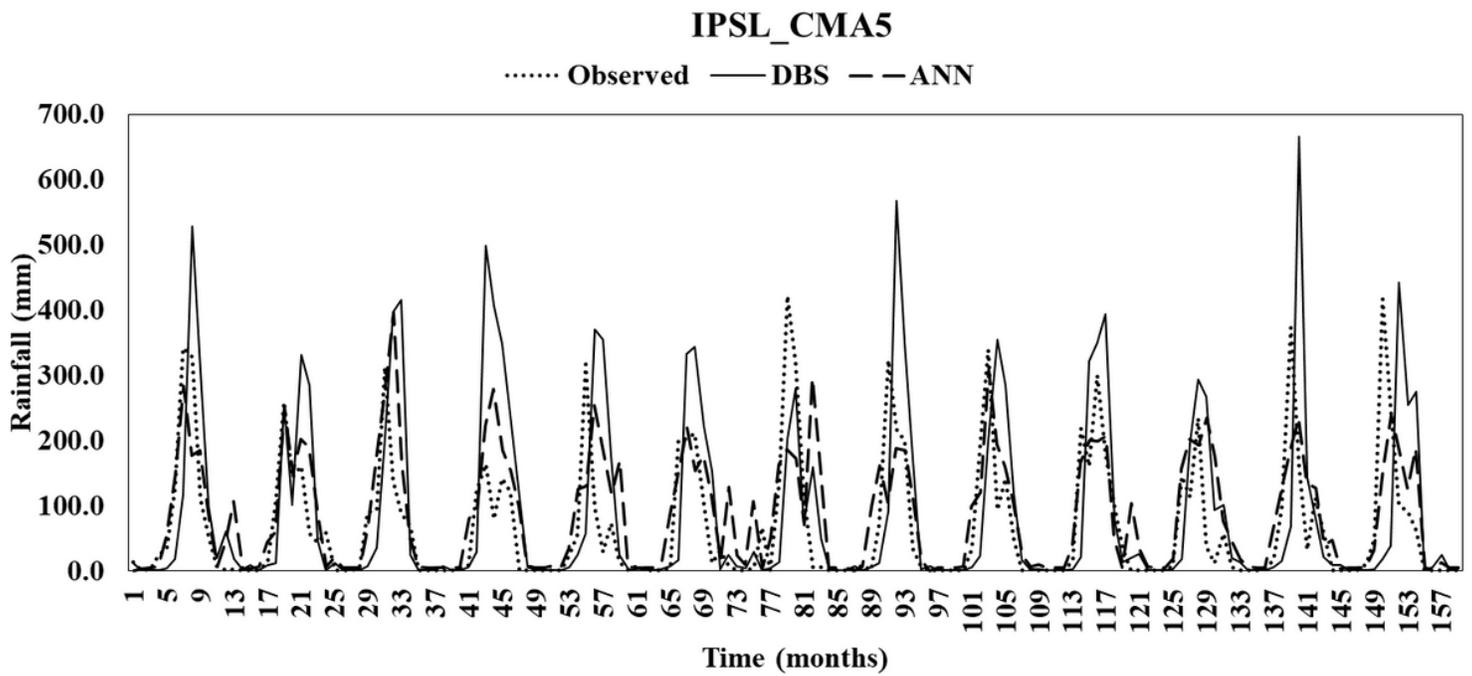


Figure 4

Time Series plot (IPSL_CMA5)

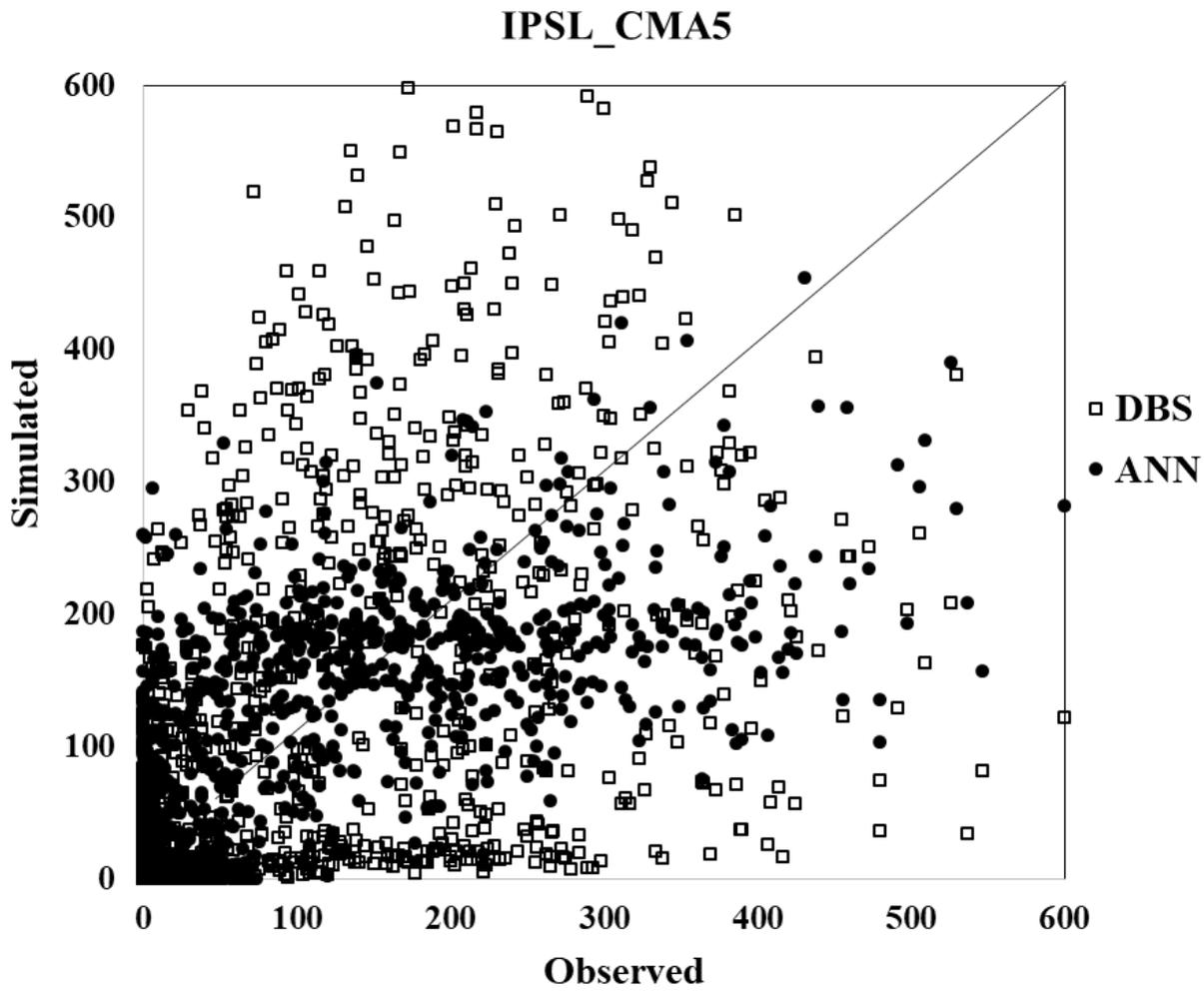


Figure 5

Scatter plot (IPSL_CMA5)

Figure 6

Time Series plot (HadGEM2_CC)

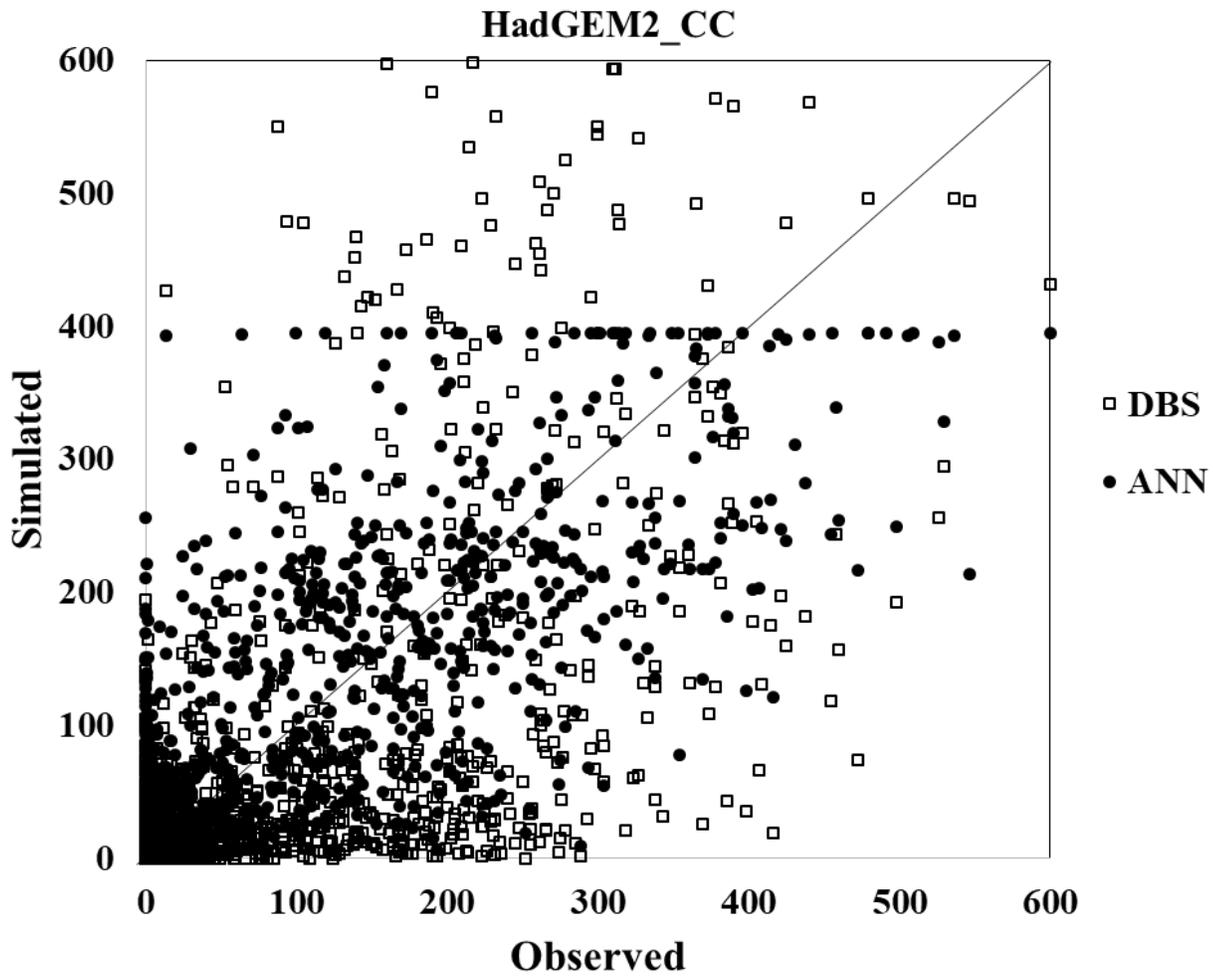


Figure 7
Scatter plot (HadGEM2_CC)

CNRM_CM5

..... Observed — DBS - - ANN

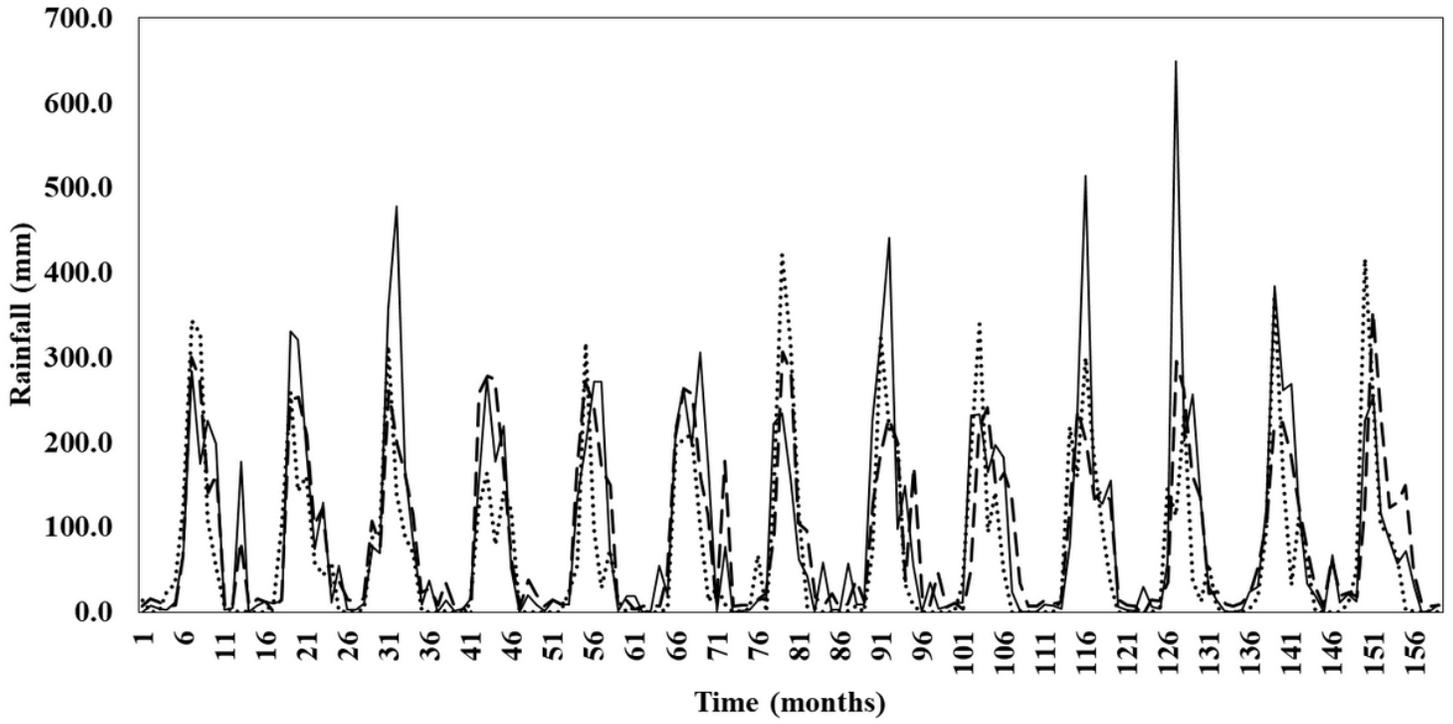


Figure 8

Time Series plot (CNRM_CM5)

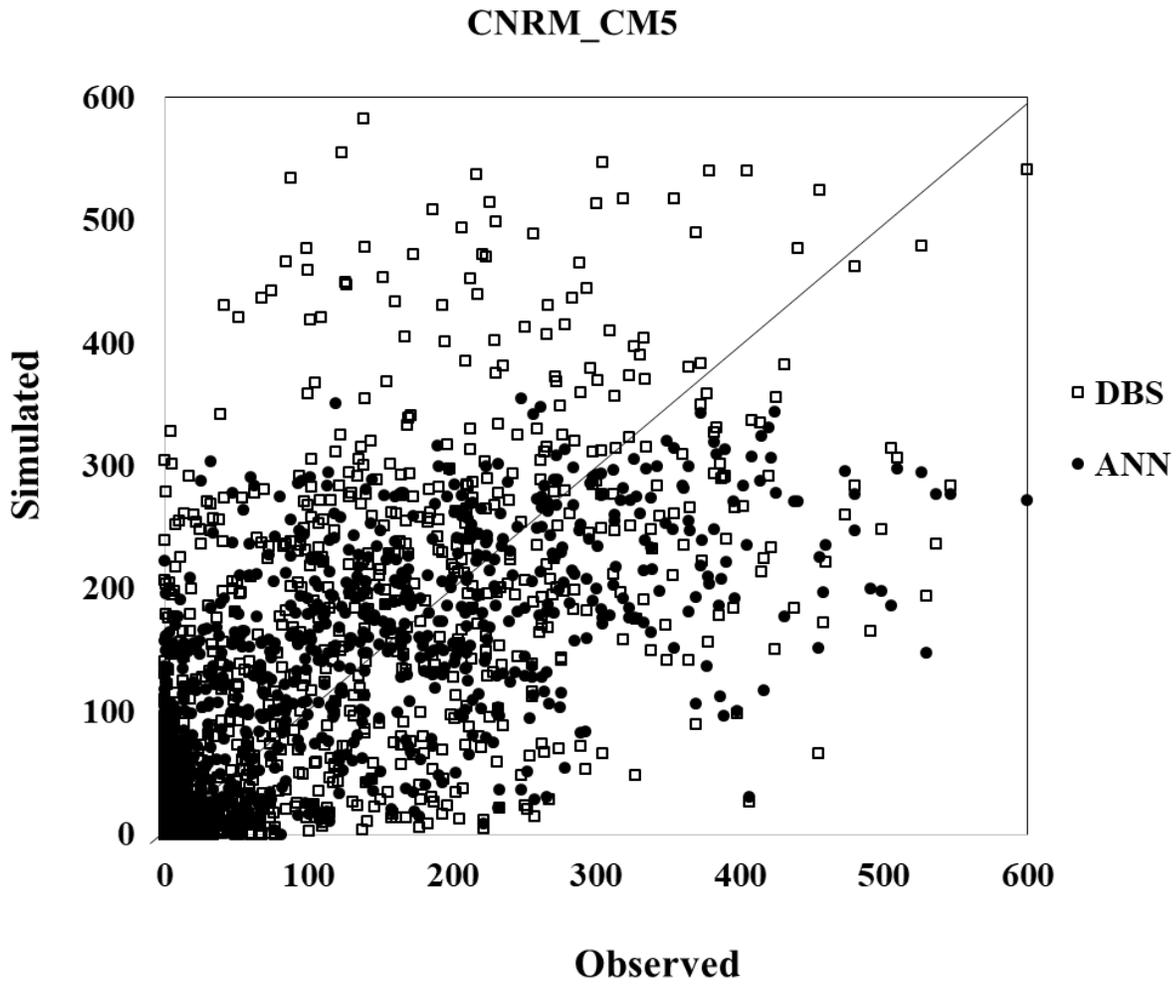


Figure 9
Scatter plot (CNRM_CM5)

Figure 10
Time Series plot (NorESM1_1)

NorESM1_1

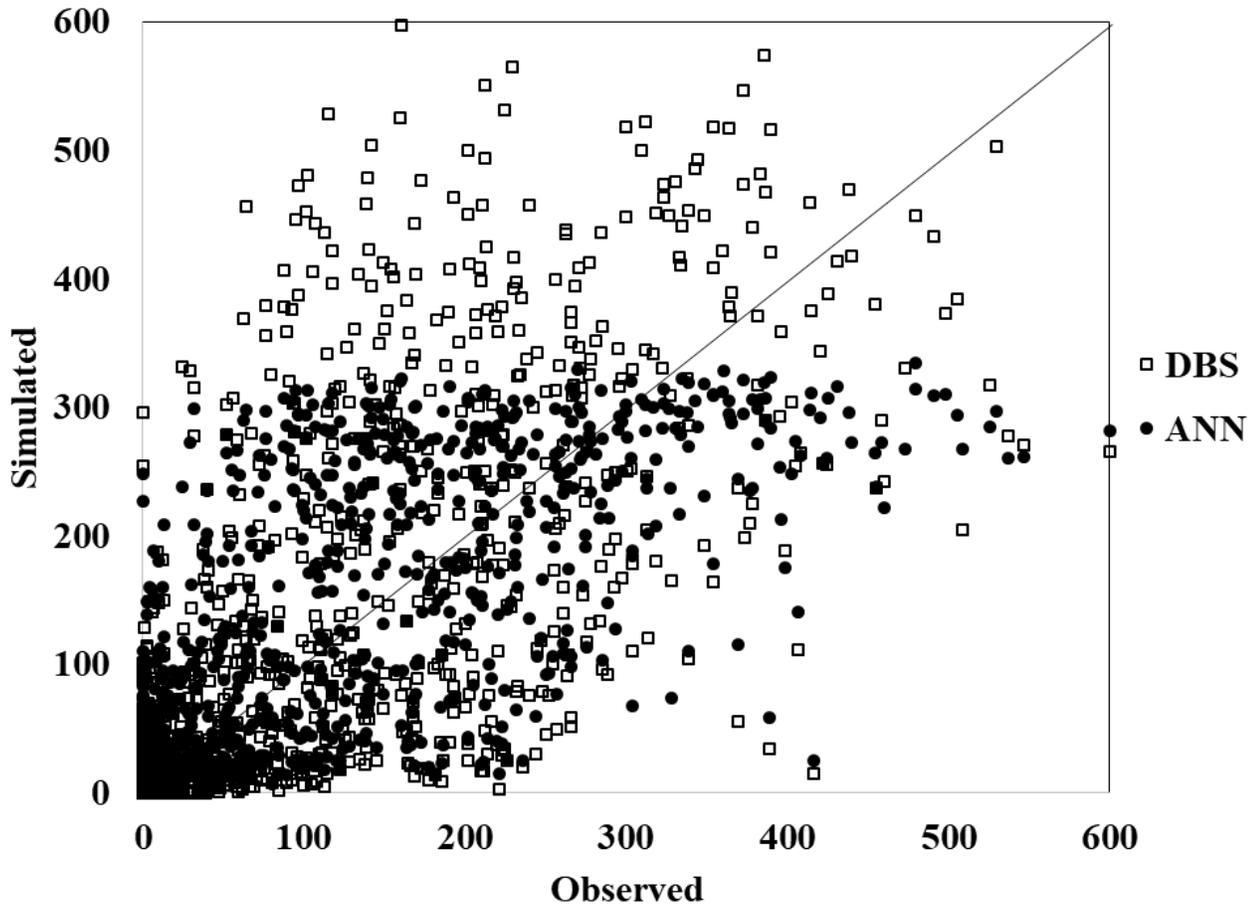


Figure 11

Scatter plot (NorESM1_1)

GFDL_CM3

..... Observed — DBS — — ANN

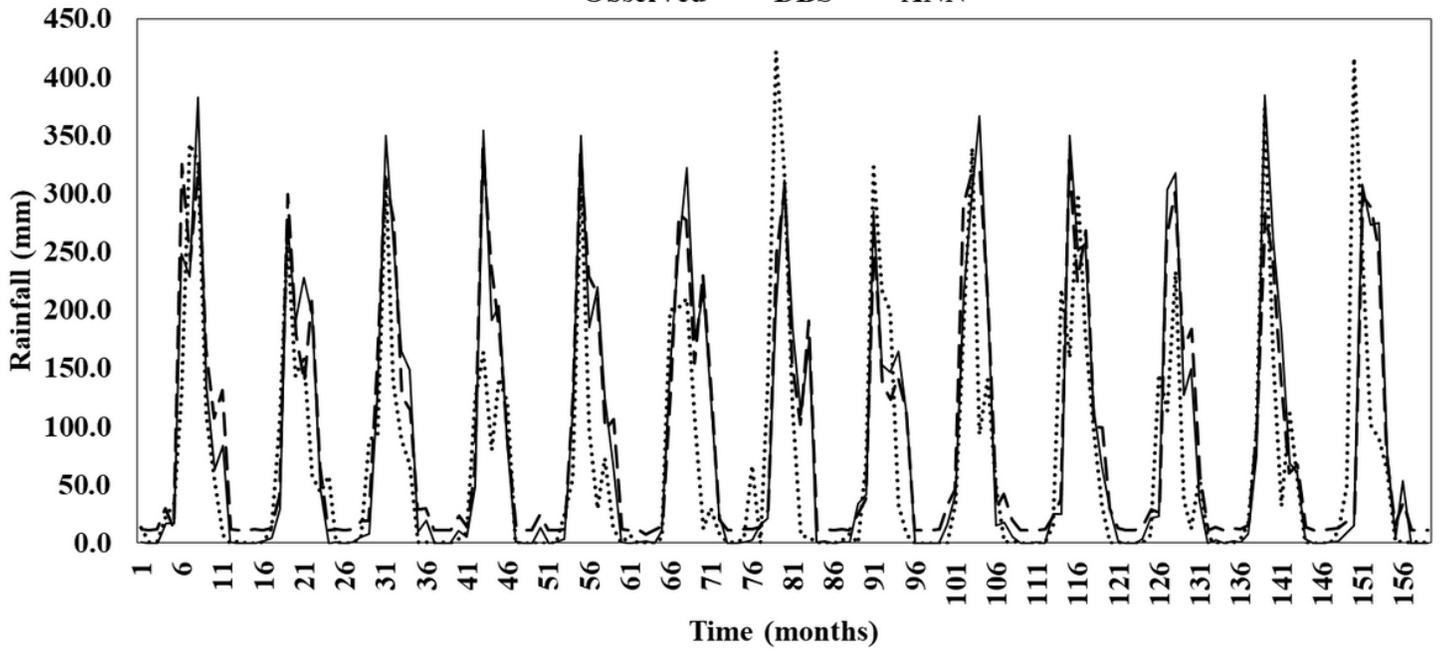


Figure 12

Time Series plot (GFDL_CM3)

GFDL_CM3

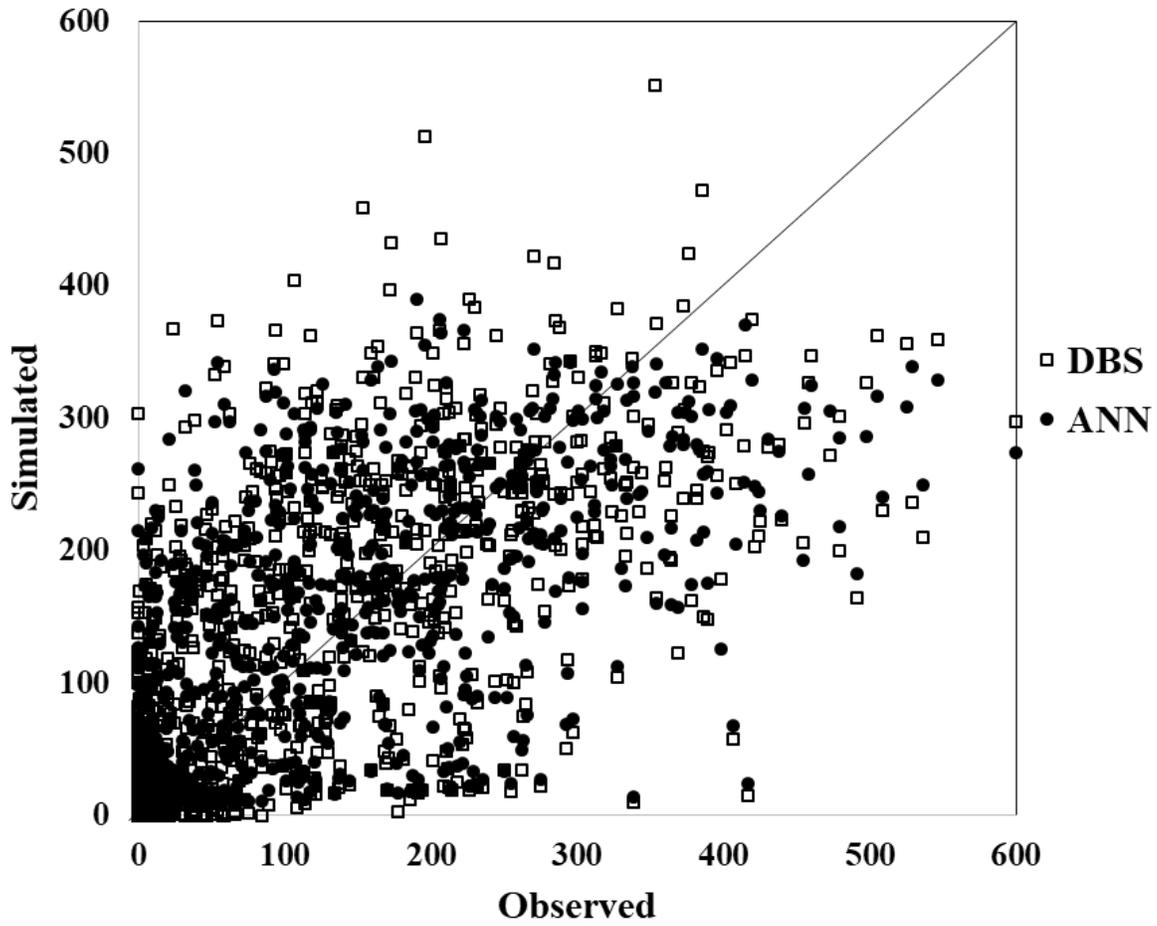


Figure 13

Scatter plot (GFDL_CM3)

Figure 14

RMSE trend for all GCMs used.

Annual Accumulated Rainfall Trend for Future 3 Decades

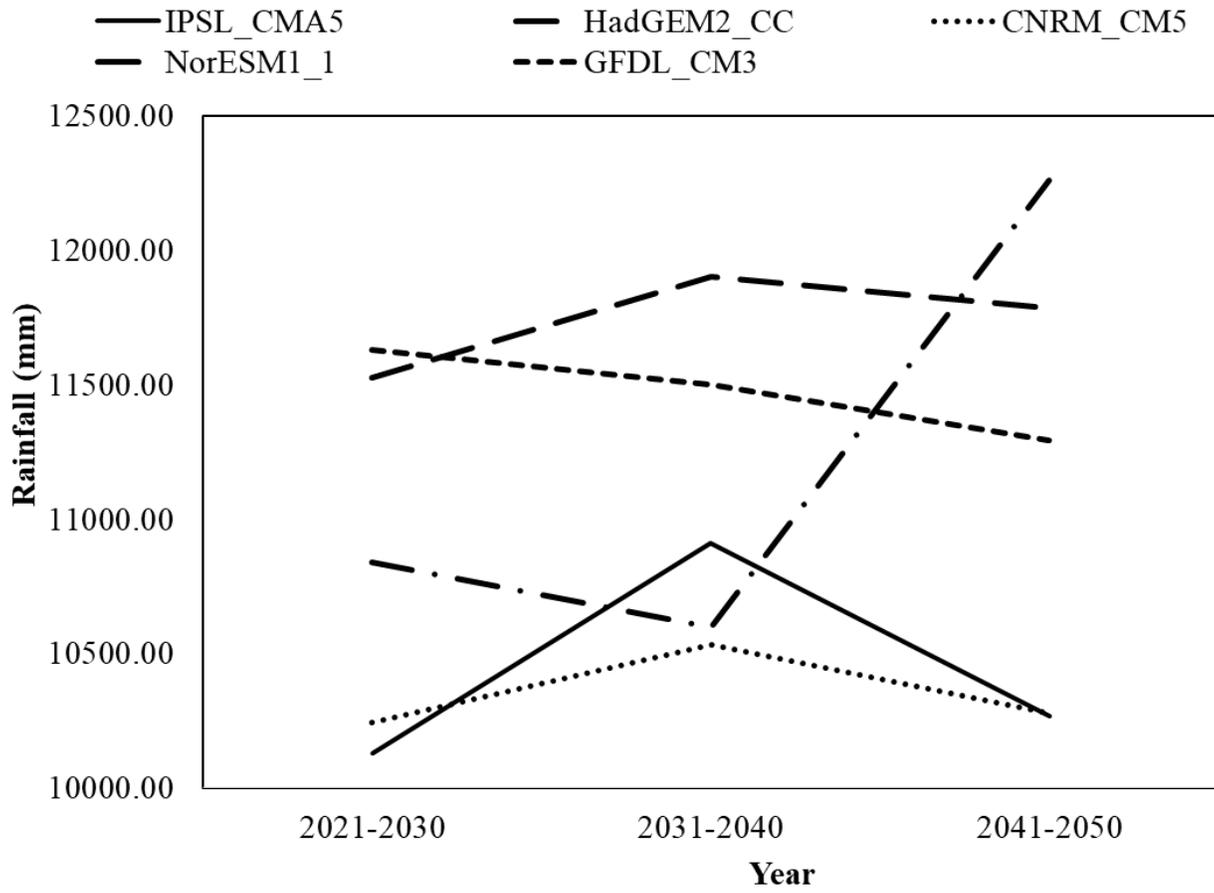


Figure 15

Comparison of Annual Accumulated Future Rainfall for all GCMs.