IS CLIMATE CHANGE AFFECTING THE MONSOON MONTHS IN PUNE, INDIA?

ⓘD **Shalaka Shah[1]**
ⓘD **Shreenivas Londhe[2+]**

[1,2]*Vishwakarma Institute of Information Technology, Pune, Maharashtra, India.*
[1]*Email:* shalaka18@hotmail.com
[2]*Email:* shreenivas.londhe@viit.ac.in

(+ Corresponding author)

ABSTRACT

It is the need of the hour to study the changes occurring in the future rainfall trends for the Pune metropolitan region. General Circulation Models (GCMs) are used to study the variations in the climatic conditions occurring in the near and far future. GCMs yield the values of the required parameters over specific grid points which are not location specific and need to be downscaled. The current study suggests Artificial Neural Networks (ANN) as a tool for statistical downscaling. 5 precipitation causative parameters were mined from 5 GCMs each and then used as input to the neural networks, with the observed historical monthly cumulative precipitation as output. The best model was then used estimate the rainfall that would occur in the near future for Pune. The future rainfall analysis showed that the intensity of rainfall seems to be increasing for the non-monsoon months and decreasing for the monsoon months.

Contribution/Originality: Innovative methodology of using the parameters obtained by GCMs at 4 surrounding grid points as input (for training the ANNs) to downscale the parameter at the desired location. Average Mutual Information (AMI) is used for causative parameter selection. This is the first detailed rainfall analysis done for Pune, India.

1. INTRODUCTION

The recent years have shown that the effects of climate change on the environment are a cause for global concern. These effects bring along with them subsequent implications that the society faces, such as, drinking water and energy requirements, irrigation schemes, and hazard preventions. The impacts of the changing climate can be seen from the shifting weather patterns to the rise in sea levels, the magnitudes of which are unmatched in scale. A major impact of climate change can be seen in the hydrological cycle. As a result, the precipitation patterns are changing rapidly which again is a cause for sudden flood or drought ridden areas. Hence it is vital to investigate the changing trends of rainfall for future water management and life sustenance. The Indian climate is dominated by the southwest monsoon precipitation (June to September). India receives nearly 80% of its total rainfall during the southwest monsoon season. However, since the past few years, a rough observation can be made that the time window for this southwest monsoon rainfall is experiencing many changes. These changes in the southwest monsoons need to be studied in detail, region-wise. Only then would we be able to get an accurate idea of the impacts the changing climate would have over that particular region.

For the purpose of incorporating the effects of climate change in the future precipitation trends, General Circulation Models (GCMs) are globally used. GCMs are advanced tools used for simulating the variations in future climate conditions all over the world. These models mimic the physical processes that are occurring in the

15

atmosphere, ocean and land surface. By applying the Navier-Stokes equations over the rotating earth along with thermodynamic terms for various energy sources, GCMs form the basis of computer programs that simulate the Earth's atmospheric conditions. The GCMs are run for various warming 'scenarios'. A set of scenarios known as Representative Concentration Pathways (RCPs) have been adopted globally to provide a range of possible futures for the evolution of atmospheric composition.

GCMs exemplify the climate, using a 3D global model with a resolution covering 250 and 600 km horizontally and vertically it has 10 to 20 layers in the atmosphere. It may also sometimes comprise of around 30 layers in the oceans. Nonetheless, some sub-grid features at smaller scale cannot be accurately modelled by GCMs due to their higher spatial resolutions. For this purpose, the values obtained at the various grid points for any of the physical processes need to be downscaled to match the actual environmental conditions at regional level.

Several studies were conducted to judge the impact of changing climate on rainfall in India since the last century. The works done by Kripalani, et al. [1]; Ueda, et al. [2]; Stowasser, et al. [3]; Sabade, et al. [4]; Krishnan, et al. [5] concluded that climate change resulted in the southwest monsoons showing a weakening trend. Works of Lal, et al. [6]; Meehl and Arblaster [7]; Rupakumar, et al. [8] suggest that due to the global warming there is a probability of increase in magnitude of the monsoon rainfall over South Asia. Rupakumar, et al. [8] evaluated the effect of the changing climate in India by assessing the present-day simulation (1961–1990) of the PRECIS climate model. Their work reported an increase in extreme precipitation along the western regions of central India.

Rana, et al. [9] worked on the effects of climate change on the precipitation occurring in Mumbai. Their research indicated an increase in intensity of future rainfall (2010-2099). As per their findings, there is about ~15–20% average increase in the maximum rainfall for a 30-year time period and ~30–45% increase in the 90-year period. They also reported a shift in the southwest monsoon season and a delayed start of monsoon period. In their work, they adopted a statistical downscaling approach termed Distribution-based Scaling (DBS) technique which was tested and applied to scale the GCM data. It was concluded that, the results of this scaling technique, to downscale GCM projections to the local scale, could be potentially useful for impact studies.

Artificial Neural Networks (ANN) are being actively used for various hydrological modelling purposes Londhe and Charhate [10]; Londhe, et al. [11]; Londhe and Shah [12]. Londhe and Shah [13] estimated the evaporation values at 3 stations for the city of Pune, using 5 different causative), variables with the help of ANNs. Their work proved that ANN understands the physics of the parameter being modelled. ANNs are also currently used by numerous researchers worldwide for downscaling of GCM data.

Fistikoglu and Okkan [14] worked on NCEP/NCAR reanalysis data to downscale the monthly rainfall in the Tahtali basin, Turkey. The rainfall at three meteorological stations was analysed using ANN consisting of nine input parameters and a single hidden layer with three nodes. In this study, the suitable predictors were examined through three regression methods. As per the authors, this was one of the finest ways to choose a subset of causative parameters in a statistical model where there was a wide range of potential causative parameters. This regression analysis was carried out using three criteria, namely maximum determination coefficient (R^2), the maximum adjusted determination coefficient ($AdjR^2$) and the minimum of Mallow's Cp. This regression analysis generated linear equations that forecast the dependent variable as a function of several independent variables. Training of ANNs was done using the Levenberg-Marquardt (LM) algorithm. Their results showed that ANN can be used to downscale the grid-based NCEP/NCAR data set to station scale. Chithra, et al. [15] worked with a similar methodology to predict the maximum and minimum temperature at the Chaliyar river basin in Kerala, India. A set of predictors chosen from the NCEP/NCAR reanalysis data were used as inputs to ANN models with maximum and minimum temperatures are output. Selection of the most suitable predictors was done based on the correlation between the NCEP predictors and the predictand (Tmax and Tmin). Results obtained indicated that the Artificial Neural Network (ANN) approach is a feasible option for downscaling climate data. Swain, et al. [16] developed a method to combine daily flows (ANN predicted) with 30-day cumulative flows (ANN predicted) to improve runoff estimations. Their results showed that

ANN had a satisfactory performance when representing the 2007–2012 and the comparatively drier 1994–1997 periods. They also concluded that ANN requires far less computational effort as compared to a numerical model and still give results with sufficient accuracy to evaluate hydrological scenarios. Sarzaeim, et al. [17] worked on runoff prediction under the influence of climate change. They used 3 statistical downscaling tools, namely, genetic programming (GP), artificial neural network (ANN), and support vector machine (SVM) to forecast the regional rainfall and temperature with the Hadley Centre Coupled Atmosphere-Ocean General Circulation Model version 3 (HadCM3) followed by runoff prediction with GP, ANN, and SVM in the Aidoghmoush Basin, Iran. The correlation criterion was applied to choose the input parameters. Abdullahi and Elkiran [18] worked on predicting the consequences of climate change on evapotranspiration (ETo) for Girne and Larnaca regions of Cyprus for the future 3 decades (2017 – 2050). ANN was used to forecast evapotranspiration. three-layer Feed Forward Back Propagation neural network (FFBP) trained by Levenberg-Marquardt (LM) optimization algorithm was used for the same. In their first approach, 6 predictor input parameters were kept constant while varying the number of hidden neurons and in the second approach, varying inputs (2 to 6 parameters) were tried and the hidden neurons were fixed to double the number of inputs. They concluded that neural networks could efficiently forecast future evapotranspiration even with the constraint of having limited climate parameters as inputs, Additionally, they are commented that the accuracy could be significantly improved by addition of a greater number of inputs. Rabezanahary, et al. [19] also focused their work on predicting the future precipitation and temperature at Mangoky River, Madagascar. 26 causative variables were chosen with the help of a linear measure (coefficient of correlation) calculated between the predictors and the desired output (Temperature and pressure). These variables were then given as input to ANN and were trained with the help of known rainfall and temperature as output. The future rainfall and temperature were then estimated using the previously trained ANN models.

From all of the studies mentioned earlier it could be observed that out of a large set of predictor variables, the major influencing parameters were chosen with the help of liner methods such as coefficient of correlation, determination coefficient and adjusted determination coefficient. As these are all linear tools, they might not be able to catch the nonlinear relationship between the independent and dependent variables. The current study suggests the use of Average Mutual Information (AMI) in place of correlation analysis for deciding the input variables which would quantify the non-linear relationship between the predictor and predictand variables. AMI is a non-linear tool which tells us what information can be extracted from the independent variable pertaining to specifically the dependent variable. It was suggested by Londhe and Shah [13] for their work on evaporation modelling using various causative parameters as input. As per the authors' knowledge no studies have been done in the field of climate change using AMI as an input selection tool.

Another observation that can be made from the above works is that in majority of the studies, standardization is carried out for the minimization of bias in the average and standard deviation of the dataset. Also, for regridding the dataset, interpolation methods such as bilinear interpolation have been adopted by many researchers. Using the authors' considerable knowledge in soft computing techniques, the above 2 steps (standardization and re-gridding) can be minimized by using ANNs. Hence, the current work suggests that the rainfall causative parameters (at the 4 closest points around the desired grid location) extracted from GCMs can be added as input for training and testing the ANNs with actual precipitation as output. The results conclude that that this innovative approach is performing a shed better than the traditional downscaling tool (Distribution Based Scaling (DBS)), as suggested by Rana, et al. [9].

Also, all of the studies mentioned above have been carried out at a regional level. Currently, an erratic rainfall pattern and shift in monsoon season can be observed even for the city of Pune and again as per the authors' knowledge, no detailed analysis for the future rainfall trend for Pune has been done. The monsoon season here, seems to be increasing beyond the months of June and September. This is causing havoc in all aspects. From farmers not being able to get hundred percent benefit from their agricultural produce to the water management committee not being

17

able to implement an efficient water resources management, this change in monsoon period is affecting everyone. Thus, being natives of the Pune city (the eighth largest city in India having population over 5 million), the authors felt the need to evaluate the rainfall conditions that would be encompassing the city for the future years.

For the current work, five causative parameters, which would impact the magnitude and frequency of rainfall were chosen with help of the study done by Gowariker, et al. [20] namely, daily minimum and maximum temperature, pressure at sea level, 10m wind speed and humidity. These 5 parameters were extracted from 5 different GCMs for 4 grid points surrounding the Pune region. These were then fed to ANN as input with the observed rainfall as output. Since ANN understands the physics behind the parameters being modelled, they were able to fairly overcome the steps of interpolating the grid point data to get rainfall at a particular point and also downscale the data to match the rainfall at that particular region. Various models were formed for each GCM by varying the inputs given to the neural networks. Details of the model formulation are given in the methodology section below.

2. STUDY AREA AND DATA

As mentioned above, the present study was conducted for the city of Pune, in Maharashtra, India. Pune is the second largest city of the state and usually is a recipient of moderate precipitation (214.85 mm as per the data acquired from India Meteorological Department, Pune) majorly occurring during the southwest monsoon season (June – September). It is located at approximately 18° 32' North latitude and 73° 51' East longitude. It lies on the Western margin of the Deccan plateau, at an altitude of 560 m above sea level. Figure 1 shows the location of Pune on the map of India.

The actual rainfall data was obtained from India Meteorological Department (IMD), Pune (https://www.imdpune.gov.in/). It consisted of 116 years (1901-2016) of rainfall data in the form of monthly cumulative with no gaps.

Along with the observed values, GCM data for rainfall, minimum and maximum temperature, surface humidity, 10m wind speed and pressure at mean sea level obtained from the Copernicus Climate Change Service website (https://cds.climate.copernicus.eu/) was used in this study. Data for all parameters from 5 GCMs was downloaded and extracted for past (1901 − 2016) and upcoming (2021-2050) years. The GCMs used in this study summarized in Table 1. These 5 GCMs are comparatively more precise for the Indian subcontinent and used by various researchers, like in the work done by Rana, et al. [9].

Table 1. Details of GCMs used.

Model	Institution	Grid Intervals (in degrees)
IPSL_CM5A_MR	Institut Pierre-Simon Laplace, France	Latitude - 1.2676 Longitude - 2.5
HadGEM2_CC	Met Office Hadley Centre, UK	Latitude -1.25 Longitude - 1.875
CNRM_CM5	Centre National de Recherches Meteorologiques, France	Latitude -1.4008 Longitude - 1.4063
NorESMI_I	Norwegian Climate Centre, Norway	Latitude -1.8947 Longitude - 2.5
GFDL_CM3	Geophysical Fluid Dynamic Laboratory, USA	Latitude -2 Longitude - 2.5

Figure 1. Study area (Pune) shown on the map of India.

3. ARTIFICIAL NEURAL NETWORKS

As stated by Londhe and Shah [13] Artificial Neural Networks (ANNs), have excellent skills that have been applied in modelling of complex processes, spread over a wide range of issues faced in the sector of hydrologic modelling. ANNs aptly mimic the framework of the human neural network, where the input neurons are the predictor (causative) variables and the variable to be modelled is the output node. Like the biological neural network, these nodes are interconnected with the help of one or more hidden layers with neurons.

ANNs use an iterative process for their training stage which helps in continuous minimization of the error between the observed and predicted variables. A variety of training algorithms are available for this purpose. The trained network can be used to assess the output over unseen data. Londhe and Shah [13] have concluded that ANNs understand the physics of the process under study and are no longer to be categorized as "Black Box" technique.

The readers are referred to Bose and Liang [21]; Wasserman [22]; The ASCE Task Committee [23]; Maier and Dandy [24]; Dawson and Wilby [25]; Jain and Deo [26]; Londhe and Panchang [27] for understanding the basic concepts and working of neural networks.

4. METHODOLOGY

As discussed earlier 5 GCMs were chosen for the current study. The global data was downloaded in NetCDF format and then, using MATLAB, converted into a Comma Separated Value (.csv) file. The nearest 4 grid points encircling Pune were then extracted from the global data. This procedure was followed for extracting the data for all the 6 physical parameters (for rainfall, minimum and maximum temperature, surface humidity, 10m wind speed and

pressure at mean sea level). It should be noted that this entire study is conducted using monthly cumulative rainfall values. Observed rainfall was available from the year 1901 to 2016 (116 years). For each year 12 values (1 for each month) of rainfall were available, hence the dataset consisted of a total of 1392 values.

The next step after data extraction was to calculate the Average Mutual Information (AMI) between each of the causative parameters and the observed rainfall. AMI measures how closely two variables are related and how much information that any two variables share. It gives an idea of the reduction in uncertainty of the dependent variable, given that we have sufficient information of the independent variable. AMI is zero for two statistically independent random variables. In case of highly correlated random variables, the AMI score would be on the higher side. AMI between the 5 causative parameters and observed rainfall was calculated for each GCM. Along with minimum and maximum temperature as inputs, the average temperature was also used as an alternative causative parameter. Table 2 shows the AMI values for each of the 5 GCMs. With the help of the AMI values in Table 2, various input-output models were then formed. The parameter having highest AMI was the sole input for the first model (Model 1) for each GCM. The parameter having the 2nd highest AMI was added as an input for the second model (Model 2) of each GCM. Thus, 5 models for each GCM were prepared with varying inputs as per their AMI values. The 6th model of every GCM consisted of average temperature as input in place of maximum and minimum temperature and the for 7th model, along with the initial 5 causative parameters, rainfall values received from GCMs was also added as an input so as to add more degrees of freedom to the models. Thus, 7 models were prepared for each of the 5 GCMs, except for NORESMI_1 in which case 6 models were formulated as the causative parameter wind speed was not available. Each causative input parameter consisted of 4 values (4 surrounding grid points' values). The details of model formulation are given in Table 3.

Table 2. AMI values for model formulation.

Causative Parameter	AMI values				
	IPSL_CMA5	NorESMI_1	CNRM_CM5	GFDL_CM3	HADGEM2_CC
Max Temp	0.12	0.16	0.1422	0.1	0.13
Min Temp	0.35	0.32	0.3319	0.27	0.32
PSL	0.29	0.39	0.2796	0.3	0.33
Humidity	0.5	0.49	0.5196	0.51	0.51
Wind Speed	0.34	xx	0.2817	0.34	0.4
Avg Temp	0.21	0.25	0.1478	0.24	0.15

Note: xx denotes that wind speed values are not available for NorESMI_1 and hence there is no AMI value for it.

Table 3. Model formulation.

Model Name		Input	Output
IPSL_CMA5	1A	Humidity	Obs. Precipitation
	2A	Humidity, Min Temp	
	3A	Humidity, Min Temp, Wind Speed	
	4A	Humidity, Min Temp, Wind Speed, PSL	
	5A	Humidity, Min Temp, Wind Speed, PSL, Max Temp	
	6A	Humidity, Wind Speed, PSL, Average Temp	
	7A	Precipitation, Humidity, Min Temp, Wind Speed, PSL, Max Temp	
NorESMI_1	1B	Humidity	Obs. Precipitation
	2B	Humidity, PSL	
	3B	Humidity, PSL, Min Temp	
	4B	Humidity, PSL, Min Temp, Max Temp	
	5B	Humidity, PSL, Average Temp	
	6B	Precipitation, Humidity, PSL, Min Temp, Max Temp	
CNRM_CM5	1C	Humidity	Obs. Precipitation
	2C	Humidity, Min Temp	
	3C	Humidity, Min Temp, Wind Speed	
	4C	Humidity, Min Temp, Wind Speed, PSL	
	5C	Humidity, Min Temp, Wind Speed, PSL, Max Temp	

Model Name		Input	Output
	6C	Humidity, Wind Speed, PSL, Average Temp	
	7C	Precipitation, Humidity, Min Temp, Wind Speed, PSL, Max Temp	
GFDL_CM3	1D	Humidity	Obs. Precipitation
	2D	Humidity, Wind Speed	
	3D	Humidity, Wind Speed, PSL	
	4D	Humidity, Wind Speed, PSL, Min Temp	
	5D	Humidity, Wind Speed, PSL, Min Temp, Max Temp	
	6D	Humidity, Wind Speed, PSL, Avg Temp	
	7D	Precipitation, Humidity, Wind Speed, PSL, Min Temp, Max Temp	
HADGEM2_CC	1E	Humidity	Obs. Precipitation
	2E	Humidity, Wind Speed	
	3E	Humidity, Wind Speed, PSL	
	4E	Humidity, Wind Speed, PSL, Min Temp	
	5E	Humidity, Wind Speed, PSL, Min Temp, Max Temp	
	6E	Humidity, Wind Speed, PSL, Avg Temp	
	7E	Precipitation, Humidity, Wind Speed, PSL, Min Temp, Max Temp	

These models were then trained and tested using the neural network toolbox in MATLAB. It was observed that a three-layered feedforward backpropagation network was unable to catch the high level of nonlinearity of the process being modelled, and hence a four-layered network (2 hidden layers, 1 input and 1 output layer) was adopted for the current study. The transfer and activation functions used were log sigmoidal between first three layers and purely linear between the second hidden and output layer. After testing 9 different training algorithms, it was concluded that the Levenberg–marquardt algorithm (LM) was giving best results, hence LM was adopted as the training algorithm for all models.

As mentioned in the section, the pre-existing downscaling technique, Distribution based scaling (DBS), as suggested by Rana, et al. [9] was also used for the current work for the purpose of comparison.

The Distribution-based Scaling (DBS) Method can be used for downscaling and bias-correction of historic and future GCM data. This method assumes that GCM simulations fully cover a wide range of processes that occur in the present climatic conditions, thus representing the present climatic conditions with the error of a stationary and systematic bias. DBS works in two steps. First is to adjust the wet fraction of rainfall to match the observed values. For this purpose, the raw interpolated and observed rainfall data were arranged in descending order and a threshold value was finalized. The months having precipitation value greater than the threshold value were deemed as wet months and the rest as dry months. After this, the non-zero precipitation months were altered to match the cumulative probability distribution in the actual rainfall data by fitting gamma distribution function Equation 1 to both observed and interpolated monthly precipitation. α and β (shape and scale parameter respectively) are required to fit the gamma distribution function. As the distribution of precipitation values (f(x)) has a major inclination towards lower intensities, α and β which are estimated by maximum likelihood could be dominated by the frequently occurring values and might be unable to precisely describe extreme values [9]. To capture the trend of normal as well as extreme rainfall, in DBS the rainfall distribution was divided into two parts separated by the 95th percentile. Two sets of parameters representing normal values and above 95 percentiles, were estimated from actual and the GCM output for the historical data. These parameter sets were then used to bias-correct GCM rainfall data for the entire forecast period up to 2050 (PDBS and PDBS,95) using inverse gamma function Equations 2 and 3. For detailed methodology, the readers are referred to Rana, et al. [9]. The authors have their own codes for this developed using MATLAB (version - 2013a, win64).

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

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

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

To judge the accuracy of the models formed using DBS as well as neural networks various error measures were used. Statistical parameters like mean, maximum and standard deviation obtained from the simulated and actual datasets were compared. Four error measures were used in the current study. Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were the absolute parameters used and Coefficient of Error (CE) was the dimensionless parameter used. To keep a track on the efficiency of peak prediction, Percentage Error in Peak (PEP) was also used. Readers are referred to Dawson, et al. [28] for detailed description of all the error measures. Table 4 show the detailed results of all the 5 GCMs.

Table 4. Result analysis.

Parameter		Mean (mm)	SD (mm)	Max (mm)	PEP (%)	RMSE (mm)	MAE (mm)	CE
Actual		84.5	118.711	703.1	--	--	--	--
IPSL_CMA5	DBS	90.44	869.81	141.01	15.09	125.23	74.10	-0.11
	1A	87.68	100.56	331.43	-52.86	70.41	42.71	0.64
	2A	87.29	103.79	382.35	-45.62	65.99	40.1	0.69
	3A	87.76	108.29	525.24	-25.3	63.95	38.61	0.71
	4A	88.12	106.96	584.09	-16.92	64.34	39.54	0.71
	5A	89.16	113.08	720.06	2.41	61.3	36.79	0.73
	6A	88.55	110.03	461.74	2.36	64.38	38.88	0.71
	7A	86.88	108.96	677.69	-3.61	60.63	36.93	0.74
NorESMI_1	DBS	101.05	763.45	141.53	8.58	99.98	59.67	0.29
	1B	85.27	94.77	257.61	-63.36	73.41	43.85	0.62
	2B	87.65	103.84	377.66	-46.28	67.35	40.87	0.68
	3B	87.31	103.71	369.41	-47.45	67.28	40.77	0.68
	4B	87.42	106.4	477.53	-32.08	66.68	41.1	0.68
	5B	87.69	105.39	396.98	-43.54	66.36	40.08	0.68
	6B	87.7	106.21	577.96	-17.79	65.89	39.93	0.69
CNRM_CM5	DBS	103.51	926.85	133.64	31.82	105.06	63.11	0.21
	1C	86.51	95.59	241.12	-65.71	72.93	43.7	0.62
	2C	87.27	102.17	444.39	-36.79	66.86	40.21	0.68
	3C	87.42	104.01	412.55	-41.32	66.56	39.46	0.69
	4C	88.35	103.08	413.56	-41.18	66.86	40.59	0.69
	5C	86.99	102.51	415.81	-40.86	65.99	39.87	0.69
	6C	88.97	108.83	538.32	-23.44	66.54	39.92	0.69
	7C	86.08	104.37	593.57	-15.57	64.22	38.83	0.71
GFDL_CM3	DBS	92.22	551.25	113.03	-21.59	85.93	52.00	0.47
	1D	86.49	103.41	401.3	-42.92	68.65	41.6	0.67
	2D	86.15	104.62	482.27	-31.41	67	40.74	0.68
	3D	87.75	105.02	476.25	-32.26	66.5	40.46	0.69
	4D	91.16	110.17	466.38	-33.67	66.39	39.79	0.69
	5D	86.44	104.32	515.16	-26.73	64.3	38.44	0.71
	6D	90.37	109.33	574.84	-18.24	65.05	39.38	0.70
	7D	85.78	108.16	612.71	-12.86	61.83	37.75	0.73
HADGEM2_CC	DBS	79.33	1014.19	164.89	44.24	129.97	69.32	-0.20
	1E	86.75	97.73	289.02	-58.89	71.57	44.48	0.64
	2E	87.6	103.43	361.73	-48.55	67.48	40.85	0.68
	3E	87.93	105.21	435.11	-38.12	68.45	42.10	0.67
	4E	89.52	108.74	460.45	-34.51	65.87	39.96	0.70
	5E	86.53	108.13	522.1	-25.74	65.50	38.98	0.70
	6E	87.57	108.26	543.62	-22.68	66.76	40.64	0.68
	7E	86.22	105.63	567.18	-19.33	65	39.35	0.7

5. RESULTS

Models 1A to 7A are the ones formulated using the GCM IPSL_CMA5. It can be seen from the result table that both the statistical parameters as well as error measures show that neural networks is more accurate as compared to DBS as a downscaling technique. The RMSE and MAE value for DBS (125.23 mm and 74.10mm respectively) are considerably on the higher end as compared to the ANN models. Studying the ANN models with varying inputs, an inference can be drawn that there is a consistent improvement in results from model 1A to model 7A. The RMSE values decreases from 70.41 mm to 60.63 mm, MAE from 42.71 mm to 36.93 mm and CE improves from 0.64 to 0.74 as the causative parameters are added as inputs to the training models. The error in peak prediction (PEP) values show a vast improvement from -52.86 % to 2.36 % (Model 6A) and -3.61 % (Model 7A). Here the negative values mean that the peak values are underestimated. Thus, in case of IPSL_CMA5 model 7A (having inputs maximum and minimum temperature, pressure at sea level, humidity, wind speed and precipitation) seems to be the best working model as compared to the rest.

Models 1B to 6B represent the GCM NorESMI_1. Here again similar conclusions can be drawn upon. The performance of ANN models, triumphs over the DBS model performances as can be reflected by the high RMSE and MAE values and extremely low CE value. Comparing the 6 ANN models, there is an improvement in the models as the number of input parameters increase. The best model in this case is also model 6B with maximum and minimum temperature, pressure at sea level, humidity and precipitation as its input parameters.

CNRM_CM5 results are reflected by the models 1C to 7C. Following the previous trends, in this case also ANN is performing a shed better as compared to DBS. Model 7C (having inputs maximum and minimum temperature, pressure at sea level, humidity, wind speed and precipitation) seems to be having the best results as compared to the other 6 ANN models. A considerable progress in the prediction of peaks can be observed form the values of PEP which have improved from -65.71% (model 1C) to -15.57% (model 7C).

Models 1D to 7D show us the performance of ANN using the GCM GFDL_CM3. Comparing DBS and all the 7 ANN models, it can be witnessed that ANN model 7D seems to be working better than the rest. Having the lowest PEP, RMSE and MAE values (-12.86 %, 61.83 mm and 37.75 mm respectively) and highest CE value (0.70), the performance of this model seems to be a shed better than the rest of the models. The inputs for this model were maximum and minimum temperature, pressure at sea level, humidity, wind speed and precipitation.

In case of the GCM HADGEM2_CC, models 1E to 7E show that the performance of ANN models' is much advanced in comparison with the DBS model performance. This can be inferred from the values of all the error measures and statistical parameters. Out of the 7 ANN models, 5E (Humidity, Wind Speed, PSL, Average Temp) seems to be working the best, closely followed by model 7E (Precipitation, Humidity, Wind Speed, PSL, Min Temp, Max Temp).

Therefore, studying all the models discussed above, it could be seen that for almost all GCMs the 7th model of each GCM tends to be giving the best results. The reason behind this might be that the neural network is getting sufficient information about the physical process being modelled through all the causative input parameters. An important observation that can be made here is that for each GCM, all the other models (having lesser number of inputs), also are performing considerably well in comparison with DBS. The inherent property of ANNs to learn the physical relationship between the input and output gives it an advantage over other downscaling techniques and is able to model the precipitation with substantial accuracy even when all input parameters aren't available.

Apart from this, a detailed decade wise RMSE analysis was done for performed for each GCM. The best model for each GCM was chosen and the RMSE value for each decade was calculated. This exercise was done to make sure that the ANN model was not oscillating towards any extreme value. From Table 5 and Figure 2 it can be observed that the RMSE value for each decade in all the models was confined to a particular range thus concluding that no model oscillation was occurring. This exercise was also useful to compare the performance of the GCMs against each

other. From table – it can be seen that IPSL_CMA5 seems to be working the best, followed by CNRM_CM5 and NorESMI_1 ranked 3ʳᵈ in their performance.

Hence these 3 best performing GCMs (IPSL_CMA5, CNRM_CM5 and NorESMI_1) were chosen to predict and study the rainfall patters 30 years into the near future (2021-2050). Using the best ANN models from each of the 3 GCMs (Model 7A, 7C and 6B), the rainfall that would be occurring from 2021-2050 was determined.

Table 5. Decade wise RMSE analysis (mm).

| GCM	IPSL_CMA5	HadGEM2_CC	CNRM_CM5	NorESM1_1	GFDL_CM3
RMSE per 10 years					
1901-1910	62.9	64.26	52.17	56.62	71.47
1911-1920	73.41	73.07	69.84	64.64	79.19
1921-1930	65.15	58.79	53.21	56.86	71.93
1931-1940	57.75	81.44	59.01	59.91	67.08
1941-1950	53.4	66.01	58.51	68.27	79.85
1951-1960	46.7	68.7	57.28	60.92	77.92
1961-1970	46.21	61.03	60.67	55.25	70.71
1971-1980	60.44	68.1	74.34	72.19	88.62
1981-1990	55.84	71.67	55.88	53.07	71.4
1991-2000	56.04	74.05	60.34	65.9	74.31
2001-2010	57.77	133.27	63.92	63.52	91.89
2011-2016	56.24	81.00	70.37	70.37	86.42
Average	57.65	75.12	61.29	62.29	77.57
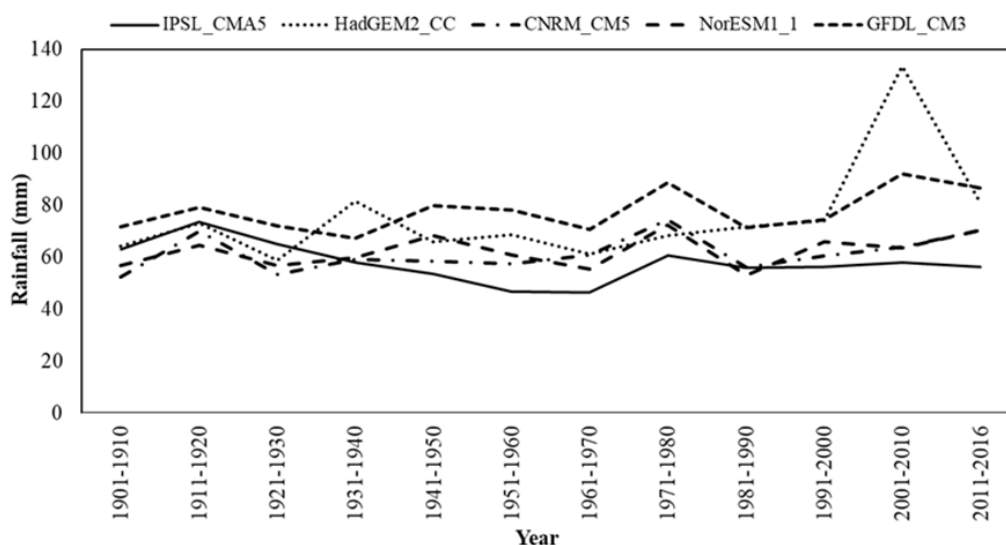
Figure 2. Decade wise RMSE analysis.

6. FUTURE CLIMATE SCENARIOS

As mentioned in the previous section, using models 7A, 7C and 6B the future rainfall occurring at Pune was determined. For this purpose, the causative parameters used for training each of the best models were first extracted at the 4 grid points surrounding Pune for the time period of 2021 to 2050. These were then input to the pre-trained ANN models and the output received from the unseen models was the monthly rainfall at Pune for the above said time period. India usually receives rainfall in the southwest monsoon season, with maximum intensity in the months of May to October. But the recent climatic conditions have shown occurrences of heavy rainfall even during the non-monsoon months. Hence the current study shows a detailed comparison of the variation in rainfall pattern for the monsoon as well as non-monsoon months. To study the changes in trend that would occur in the future rainfall pattern, it needs to be compared with the historical rainfall trends. Hence the historical rainfall data was divided into 30-year time slices (1901-1930, 1931-1960, 1961-1990 and 1991 to 2016) and the average rainfall occurring each

month for each time slice was calculated. For comparison with past rainfall, the monthly average of the future rainfall computed from the 3 ANN models also was calculated. This comparison can be studied from Table 6 and Figures 3 and 4. Figure 3 shows the changes in average rainfall for non-monsoon months and Figure 4 shows the same for monsoon months. From the figures it can be seen that for the non-monsoon months there is a considerable increase in the intensity of rainfall that would occur in the near future. On the contrary, for the monsoon months, the future rainfall is either at par with or lesser than the historical rainfall intensity. This inference can be further backed up by Table 7 which shows the percentage increase or decrease in the future rainfall events as compared to the past records. This table shows us that there is an approximate 200-500% increase in rainfall that would occur in the month of January, 200-300% increase for February and March and 100-150% increase for April and December. Whereas, for the months of June-September the rainfall seems to be decreasing by approximately 20-30%.

These records show that for the coming years, the notion of a monsoon season might not hold true for the city of Pune as there seems to be a significant quantity of rainfall occurring throughout the year. This inference is of utmost importance to the agriculture sector of the country. The usual sowing and reaping practices followed by farmers currently would have to be modified as per the expected rainfall to occur in the coming years. The irrigation and water supply management schemes also would need modifications to suit the upcoming changes in the water availability. Apart from this, the current work would also benefit the drought and flood management projects and would dictate the steps to be followed as per the expected climatic conditions.

Table 6. Comparison of Monthly average rainfall for past and future (2021-2050) years (mm).

| Year/GCM Month	1901-1930	1931-1960	1961-1990	1991-2016	Model 7A	Model 7C	Model 6B
Jan	2.83	2.19	0.69	1.03	7.17	11.23	7.44
Feb	1.58	0.42	0.78	7.15	8.10	8.25	9.86
Mar	1.99	2.71	1.88	6.23	11.52	11.43	11.45
Apr	10.01	11.47	6.12	5.32	21.41	17.95	11.73
May	22.83	28.18	27.33	16.70	64.67	117.99	40.38
Jun	146.86	164.90	169.00	189.10	184.10	238.84	202.76
Jul	271.55	377.82	354.96	239.54	234.13	232.64	217.15
Aug	178.69	244.38	257.87	182.93	214.43	204.07	195.68
Sep	146.81	183.51	173.69	141.93	149.87	149.93	175.99
Oct	57.22	99.82	73.21	94.86	112.39	89.64	94.23
Nov	28.65	37.84	23.51	23.11	43.71	46.17	31.51
Dec	4.00	4.96	7.16	6.53	14.52	17.28	9.29
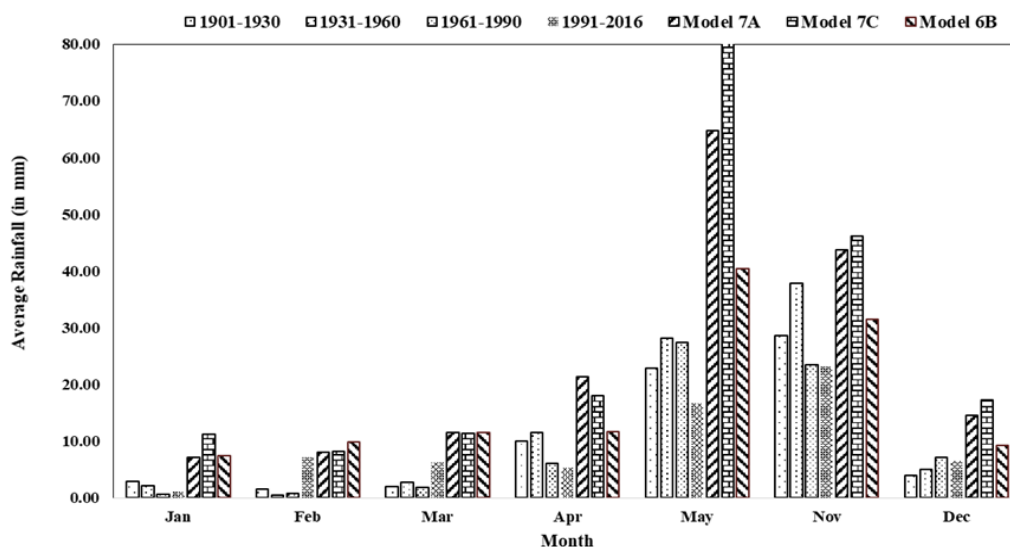
Figure 3. Changes in future average rainfall for non-monsoon months.
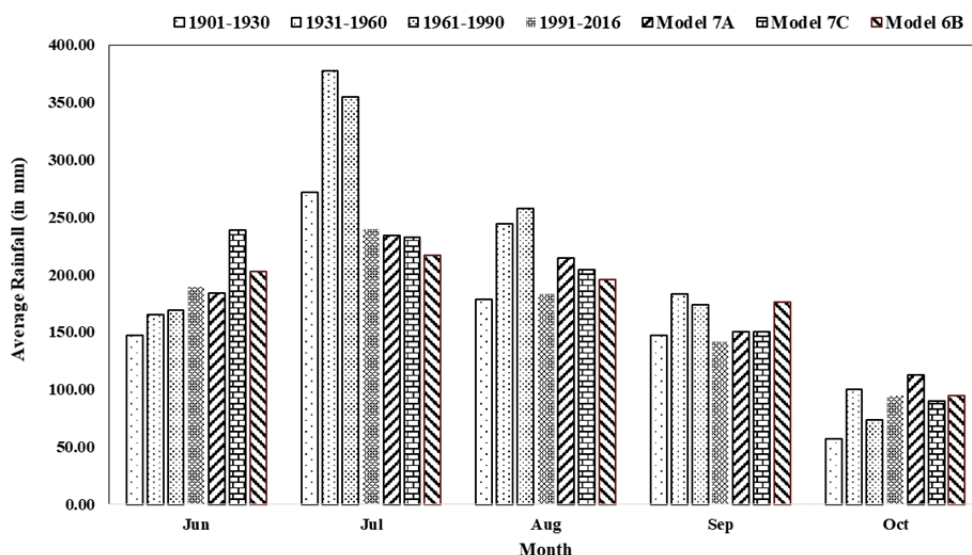
Figure 4. Changes in future average rainfall for monsoon months.

Table 7. Percentage increase in future rainfall (2021-2050) (%).

Month	Model 7A	Model 7C	Model 6B
Jan	325.26	565.97	341.06
Feb	226.05	231.97	296.94
Mar	259.53	256.86	257.38
Apr	160.05	118.03	42.44
May	172.18	396.57	69.95
Jun	9.93	42.62	21.07
Jul	-24.70	-25.18	-30.17
Aug	-0.71	-5.50	-9.39
Sep	-7.19	-7.15	8.98
Oct	38.27	10.29	15.93
Nov	54.58	63.29	11.43
Dec	156.38	205.11	64.11

7. CONCLUSION

India's major source of income is generated through its agriculture. A majority of the Indian population relies solely on their agricultural produce to meet their daily needs. In such a case the unexpected turn of rainfall patterns causes havoc on the livelihood of many families. Apart from this, sudden changes in the rainfall trends lead to extreme drought or flooding conditions as well. Thus, an accurate estimate of the rainfall that would occur for the coming years is necessary to take the appropriate actions required. The current study helps in studying the rainfall pattern for the near future (2021-2050) using 5 different GCMs.

The authors' have deployed Artificial Neural networks as their primary downscaling technique along with Distribution Based Scaling (DBS) as suggested by Rana, et al. [9] for comparison purposes. The models formulated for the ANNs were cause effect models in which the inputs were varied as per their Average Mutual Information (AMI) values with the observed rainfall. Various models were formulated for each of the 5 GCMs. All of these models pointed towards ANN having a superior performance as a downscaling tool as compared to DBS. Although a three layered network was unable to give the desired results, adding 1 additional hidden layer improved the results by a large margin. Also, the best models for each GCM were the ones having all the causative parameters (maximum and minimum temperature, humidity, pressure at sea level and wind speed) along with the precipitation as input. For each of the parameters, the 4 grid points surrounding Pune were given as input.

A detailed RMSE analysis of the results obtained proved that the models trained and tested using ANNs were not oscillating and that the 3 of the 5 GCMs were more suitable for the current study area. These GCMs were

IPSL_CMA5, CNRM_CM5 and NorESMI_1. The best models from these GCMs (Model 7A, 7C and 6B) were then used to forecast the rainfall for the future years (2021-2050).

The future rainfall trends show that the monsoon season that Pune has currently, might not hold true in the future. The average rainfall shows a 200-500% increase for the month of January, 200-300% increase for February and March and 100-150% increase for April and December. Whereas, for the months of June-September the rainfall seems to be decreasing by approximately 20-30%. Thus, pointing towards rainfall events occurring throughout the year instead of just from June-September.

A gist of the contributions from the current work are thus enumerated below:

1. The core innovation of the current study lies in its methodology of using the physical parameters obtained by GCMs at 4 surrounding grid points as input (for training the ANNs) to downscale the parameter at the desired location, thus eliminating 2 major steps of re-gridding the data and performing standardization over the raw GCM data. The output in this case would be the observed values at the desired location.

2. Earlier works, as described earlier, have used linear measures such as coefficient of correlation to choose the causative parameters. This work suggests the use of Average Mutual Information (AMI), which is much better suited for non-linear parameters such as rainfall. Using the AMI value between rainfall and causative parameters, various models were formulated by adding causative parameters as input one by one. This methodology again is an innovative approach to check the effect of missing parameters while training the neural networks.

3. While modelling of the neural networks, 9 different training algorithms were tried and tested out of which LM was finalized. Also, it was seen that the ANNs having a single hidden layer were unable to give the desired results and hence 2 hidden layers (4 layered network) were used for the present study. Additionally, the normalization range, epochs, number of hidden neurons in each layer, all these parameters were finalized after rigorous trial and error. Such detailing of ANN models is seldom seen in the past works.

4. To make sure that the trained models were not giving oscillating results, a detailed decade-wise RMSE analysis was carried out for the past 116 years and plotted, hence keeping a check on the accuracy of results at all times.

5. The current work also shows that ANNs are able to model precipitation with sufficient accuracy even when some of the causative parameters are not given as input while training. Hence, showing efficiency of the neural networks, even in cases where there is missing data.

6. Numerous works have been conducted over India, but precipitation being a location specific parameter, needs to be modelled accordingly. No such work has been carried out for the city of Pune in Maharashtra. Furthermore, none of the earlier works give us a detailed forecast of the future monthly precipitation trends. The current work throws light on how the so-called monsoon season might not hold true for the upcoming years.

7. A shift in the monsoon season brings about the scope for changes in the policies for various agricultural and water management activities, hence making the current work practically applicable.

REFERENCES

[1]   R. Kripalani, A. Kulkarni, S. Sabade, and M. Khandekar, "Indian monsoon variability in a global warming scenario," *Natural Hazards*, vol. 29, pp. 189-206, 2003.

[2]   H. Ueda, A. Iwai, K. Kuwako, and M. E. Hori, "Impact of anthropogenic forcing on the Asian summer monsoon as simulated by eight GCMs," *Geophysical Research Letters*, vol. 33, p. L06703, 2006.Available at: https://doi.org/10.1029/2005gl025336.

[3]   M. Stowasser, H. Annamalai, and J. Hafner, "Response of the South Asian summer monsoon to global warming: Mean and synoptic systems," *Journal of Climate*, vol. 22, pp. 1014-1036, 2009.Available at: https://doi.org/10.1175/2008jcli2218.1.

[4]   S. Sabade, A. Kulkarni, and R. Kripalani, "Projected changes in South Asian summer monsoon by multi-model global warming experiments," *Theoretical and Applied Climatology*, vol. 103, pp. 543-565, 2011.Available at: https://doi.org/10.1007/s00704-010-0296-5.

[5]   R. Krishnan, T. Sabin, D. Ayantika, A. Kitoh, M. Sugi, H. Murakami, A. Turner, J. Slingo, and K. Rajendran, "Will the South Asian monsoon overturning circulation stabilize any further?," *Climate Dynamics*, vol. 40, pp. 187-211, 2013.Available at: https://doi.org/10.1007/s00382-012-1317-0.

[6]   M. Lal, G. A. Meehl, and J. M. Arblaster, "Simulation of Indian summer monsoon rainfall and its intraseasonal variability in the NCAR climate system model," *Regional Environmental Change*, vol. 1, pp. 163-179, 2000.Available at: https://doi.org/10.1007/s101130000017.

[7]   G. A. Meehl and J. M. Arblaster, "Mechanisms for projected future changes in South Asian monsoon precipitation," *Climate Dynamics*, vol. 21, pp. 659-675, 2003.Available at: https://doi.org/10.1007/s00382-003-0343-3.

[8]   K. Rupakumar, A. K. Sahai, K. K. Kumar, S. K. Patwardhan, P. K. Mishra, J. V. Revadekar, K. Kamala, and G. B. Pant, "High- resolution climate change scenarios for India for the 21st century," *Current Science*, vol. 90, pp. 334–345, 2006.

[9]   A. Rana, K. Foster, T. Bosshard, J. Olsson, and L. Bengtsson, "Impact of climate change on rainfall over Mumbai using distribution-based scaling of global climate model projections," *Journal of Hydrology: Regional Studies*, vol. 1, pp. 107-128, 2014.Available at: https://doi.org/10.1016/j.ejrh.2014.06.005.

[10]   S. N. Londhe and S. B. Charhate, "Towards modelling of streamflow using soft tools," *IAHS-AISH Publication*, vol. 331, pp. 245-253, 2009.

[11]   S. Londhe, P. Dixit, S. Shah, and S. Narkhede, "Infilling of missing daily rainfall records using artificial neural network," *ISH Journal of Hydraulic Engineering*, vol. 21, pp. 255-264, 2015.Available at: https://doi.org/10.1080/09715010.2015.1016126.

[12]   S. Londhe and S. Shah, "Evaluation of pan evaporation model developed using ANN. In Development of Water Resources in India," ed Cham: Springer, 2017, pp. 221-231.

[13]   S. N. Londhe and S. Shah, "A novel approach for knowledge extraction from artificial neural networks," *ISH Journal of Hydraulic Engineering*, vol. 25, pp. 269-281, 2019.Available at: https://doi.org/10.1080/09715010.2017.1409667.

[14]   O. Fistikoglu and U. Okkan, "Statistical downscaling of monthly precipitation using NCEP/NCAR reanalysis data for Tahtali River Basin in Turkey," *Journal of Hydrologic Engineering*, vol. 16, pp. 157-164, 2011.Available at: https://doi.org/10.1061/(asce)he.1943-5584.0000300.

[15]   N. Chithra, S. G. Thampi, S. Surapaneni, R. Nannapaneni, A. Reddy, and J. D. Kumar, "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," *Theoretical and Applied Climatology*, vol. 121, pp. 581-590, 2015.Available at: https://doi.org/10.1007/s00704-014-1257-1.

[16]   E. D. Swain, J. Gomez-Fragoso, and S. Torres-Gonzalez, "Projecting impacts of climate change on water availability using artificial neural network techniques," *Journal of Water Resources Planning and Management*, vol. 143, p. 04017068, 2017.Available at: https://doi.org/10.1061/(asce)wr.1943-5452.0000844.

[17]   P. Sarzaeim, O. Bozorg-Haddad, A. Bozorgi, and H. A. Loáiciga, "Runoff projection under climate change conditions with data-mining methods," *Journal of Irrigation and Drainage Engineering*, vol. 143, p. 04017026, 2017.Available at: https://doi.org/10.1061/(asce)ir.1943-4774.0001205.

[18] J. Abdullahi and G. Elkiran, "Prediction of the future impact of climate change on reference evapotranspiration in Cyprus using artificial neural network," *Procedia Computer Science*, vol. 120, pp. 276-283, 2017.Available at: https://doi.org/10.1016/j.procs.2017.11.239.

[19] T. M. F. Rabezanahary, M. Rahaman, and J. Zhai, "Assessment of the future impact of climate change on the hydrology of the Mangoky River, Madagascar using ANN and SWAT," *Water*, vol. 13, p. 1239, 2021.Available at: https://doi.org/10.3390/w13091239.

[20] V. Gowariker, V. Thapliyal, R. Sarkar, G. Mandal, and D. Sikka, "Parameteric and power regression models: New approach to long range forecasting of monsoon rainfall in India," *Mausam*, vol. 40, pp. 115-122, 1989.Available at: https://doi.org/10.54302/mausam.v40i2.2033.

[21] N. K. Bose and P. Liang, *Neural network fundamentals with graphs, algorithms, and applications: Electrical and Computer Engineering Series*: Tata McGraw-Hill, 1998.

[22] P. D. Wasserman, *Advanced methods in neural computing*. New York, 264: Van Nostrand Reinhold, 1993.

[23] The ASCE Task Committee, "Artificial neural networks in hydrology, II: Hydrologic applications," *Journal of Hydrologic Engineering*, vol. 5, pp. 124–137, 2000.

[24] H. R. Maier and G. C. Dandy, "Application of artificial neural networks to forecasting of surface water quality variables: Issues, applications and challenges. In Artificial neural networks in hydrology," ed Dordrecht: Springer, 2000, pp. 287-309.

[25] C. Dawson and R. Wilby, "Hydrological modelling using artificial neural networks," *Progress in Physical Geography*, vol. 25, pp. 80-108, 2001.Available at: https://doi.org/10.1191/030913301674775671.

[26] P. Jain and M. Deo, "Neural networks in Ocean engineering," *Ships and Offshore Structures*, vol. 1, pp. 25-35, 2006.

[27] S. N. Londhe and V. Panchang, "ANN techniques: A survey of coastal applications," in *Advances In Coastal Hydraulics*, ed, 2018, pp. 199-234.

[28] C. W. Dawson, R. J. Abrahart, and L. M. See, "HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts," *Environmental Modelling & Software*, vol. 22, pp. 1034-1052, 2007.Available at: https://doi.org/10.1016/j.envsoft.2006.06.008.