Statistical downscaling of monthly precipitation over India using long short-term memory networks

Abstract

Due to the limitation of General Circulation models in capturing fine resolution variables like precipitation, a weather typing based statistical downscaling method using state of the art neural network model Long Short Term Memory (LSTM) is used for precipitation forecasting at 0.5° spatial resolution. The projections are made for the next century for 32 Indian Meteorological Subdivisions (MSD's) using 5 GCM's GFDL-ESM2M, MRI-CGCM3, IPSL-CM5A-MR, CanESM2, CNRM-CM5 for both scenarios rcp4.5 and rcp8.5. K-means clustering in combination with LSTM is used for weather classification which is identified as the most crucial step in improving model performance. Model so developed shows good results for most of the areas and is able to capture inter-site cross-correlation which is an important attribute in multisite downscaling. Our study is the first employing LSTM's for statistical downscaling to this large scale.

Keywords: GCM, Statistical downscaling, Precipitation, India. Deep learning, LSTM

Highlights:

- Regions showing diverse climatology require advance tools to capture spatial correlation.
- Deep learning is efficient in recognizing random patterns, hence giving better future simulations.
- LSTM is better than other models in capturing temporal behavior at longer leads.
- Weather states as an input help in learning extremities in data more accurately.
- Principal Component Analysis is important prior to training process to use only important features as input.

1 Introduction

2019 Indian floods and many past extreme events demand scientists to develop more robust models for flood risk mapping and weather predictions, thereby knowing the impact of climate change on hydrological processes so that right mitigation measures can be taken much before the actual occurrence. Accurate prediction of monsoons will help agriculturists plan their crops and hence boosting national economic strata.
General Circulation Models(GCMs) incorporate physical aspects of all 3 components of biosphere ,i.e, lithosphere, atmosphere and hydrosphere as inputs to mathematical models, hence simulating important large scale climatic variable patterns, such as the monsoons, seasonal shifts of temperatures, El Niño–Southern Oscillation(ENSO), Pacific Decadal Oscillation(PDO), and northern and southern annular modes (Solomon et al. 2007). But due to the limitation of having coarser temporal (seasonal to monthly) and spatial resolution (100's of km) (P. P. Mujumdar and D. Nagesh Kumar 2012), it is not possible to capture non–smooth fields especially precipitation which is a key input to models like flood mapping, watershed management etc.

This limitation of GCMs led to the development of models capable of converting low-resolution input into finer output, and the technique is known as downscaling which is divided into 2 categories dynamic and statistical downscaling.

Dynamic downscaling makes use of coarse grid GCM in providing boundary conditions to high-resolution Regional Climate Model(RCM) to get simulated outputs at a finer resolution. But due to their high computational cost and additional error, that contributed by underlying GCM makes it a second choice in comparison to statistical downscaling which works on the principle of establishing empirical relationship between predictors and predictands, comparably having less computational cost and the accuracy of results depends on choice of predictors, length of data time series and capability of model to capture uncertainties in the data, therefore often in the studies involving these, its assumed that predictor data from GCM is well simulated and relevant for study in terms of predicting predictands well, and the relationships so developed are dynamic enough to be employed into conditions of climate change (Wilby and Wigley 2000). Statistical downscaling is further subdivided into 3 types: (i) Weather classification and typing schemes (ii) Stochastic weather generator modeling (iii) Regression modeling. While Weather Typing establishes relationship with the predictand by converting predictors into weather types and Weather Generators use synthetic time series of longer lengths overcoming problems of missing data, Regression-based approach, which aims to capture linear and non–linear relationship between predictors and predictands, being simple and flexible in terms of model being used, its hyperparameters and set of predictors chosen, has been widely used for the purpose of downscaling.

Neural networks that have been developed to emulate brain functioning, are considered as universal approximators and thereby are fit for hydrological modeling where relationships can go up to different degrees of complexity. Recently neural networks have shown state of art results in sequence-related problems like: (i) Speech recognition(Hinton et al. 2012; Sainath et al. 2015) (ii) Language modeling(Mikolov et al. 2010) (iii) Machine translation(Cho et al. 2014; Luong et al. 2014) (iv) Weather forecasting(Zaytar and El Amrani 2016) etc. But primitive neural networks like Vanilla RNN's often face problems like vanishing and exploding gradients during backpropagation in solving long sequence-related problems (Pascanu et al. 2013). Long Short Term Memory neural networks(LSTM) being capable of preserving information over longer distances(Hochreiter 1997) have been exploited recently for extreme event studies. (Akbari Asanjan et al. 2018) in their study on Short Term Precipitation Forecast, used LSTM to predict Cloud Top Brightness Temperature(CTBT) which was used as one of the predictors and results show better performance over RNN, persistency and Farneback methods. (Huang et al. 2019) found that using central Pacific sea surface temperatures(SST) on daily basis at longer leads as predictors, LSTM performed better than Linear Regression Models(LR) showing that LSTM is able to capture non-linearities in daily SST better.

Since statistical downscaling is a major research area, many scientists have come up with different techniques to get better results which mainly vary in the type of model and clustering technique used. (Tripathi et al. 2006) used Support Vector Machine(SVM) showing SVM's are promising alternatives to conventional Artificial Neural Networks(ANN). (Raju and Kumar 2014) used computationally simple linear regression model in combination with fuzzy clustering to classify predictors reduced in the principal directions into weather types achieving high R^2 value. (Salvi et al. 2013) used classification and regression tree(CART) and K-means clustering to predict daily precipitation states and Kernel regression to downscale precipitation making model to be able to capture the orographic effect on rainfall in mountainous areas of the Western Ghats and northeast India. (Goyal and Ojha 2012) used ANN for predicting mean monthly minimum and maximum temperatures on lake-basin scale showing ANN's are better predictors than Linear Multiple Regression models. Other methods like clustering-based approach which uses clustering in association with K–Nearest Neighbor(KNN) (Gutiérrez et al. 2004) and sequence-to-sequence method in which present rainfall is used to predict future rainfall (Tran Anh et al. 2019) have been used for SD.

The purpose of our research is to explore state of the art LSTM model for statistical downscaling of monthly precipitation over 32 Indian Meteorological Subdivisions (MSD's) at a resolution of 0.5° for the next century and checking the robustness of the model using various statistical measures. The motivation behind the same is to explore the potential of LSTM neural network, being used for the first time on such a large scale.

The current paper is organized as follows. Data acquisition is discussed in chapter 2. The methodology and models used for the project are discussed in chapter 3. Discussions and conclusions based on the output are discussed in chapter 4.

2 Study Area and Data

32 Indian Meteorological Subdivisions (MSD) (Guhathakurta and Rajeevan 2008) (out of 36 excluding Andaman and Nicobar Islands, N.M.M.T, Jammu and Kashmir and Lakshadweep) are included in the study area which accounts for a total of 980 grid points at 0.5° spatial resolution. Because of the diverse climatology of the nation and missing/incorrect data in most regions like Jammu and Kashmir, smaller divisions have been chosen to establish better empirical relationships and thereby getting better results.

2.1 Data Extraction

2.1.1 Predictor Selection

Predictor selection is an important step to ensure that our model performs well both on present and future scenarios. The choice of predictors varies from region to region depending upon atmospheric circulation patterns and predictand to be downscaled. Based on source of predictors, Statistical Downscaling can be further divided into 3 types: (i) Model Output Statistics (from GCM outputs) (ii) Perfect Prognosis (from reanalysis datasets) (iii) Surface Variable Based (from large scale surface observations) (Tatli et al. 2005). For our study, Reanalysis data has been used to train the model and GCM data to make future projections. Training period from 1951 – 2005 and the testing period from 2006 – 2100 have been chosen for the study.

[Table 1]

So based on the literature review in Table 1, predictors chosen are temperature, specific humidity, u wind and mean sea level pressure all at 1000 hpa and geopotential height at 500 hpa and are same for all MSD's. These predictors are available for entire study duration, are simulated well by all the GCM's and are well correlated with the predictand. All the datasets are at monthly scale.

2.1.2 Datasets

National Center for Environmental Prediction (NCEP) Reanalysis-I pressure level data for all the above-mentioned variables at a resolution of 2.5° delimited by latitudes -5° - $42.5^\circ N$ and longitudes 60° - $120^\circ E$ covering entire India and 0.5° resolution precipitation data (1240 grid points) has been extracted from Indian Meteorological Department (IMD). Both the datasets have been extracted for the training period. Historical data for the training period and future data for 2 scenarios $rcp4.5$ and $rcp8.5$ for the testing period considering ensemble $r1i1p1$ have been extracted for the GCM's for the variables denoted by symbols tas (temperature), uas (u wind), $huss$ (specific humidity), psl (mean sea level pressure), zg (geopotential height). Five GCM models chosen for the study are mentioned in Table 2.

[Table 2]

All the five models have been previously been used in some of the notable Indian studies. The first three models are a part of eleven GCMs used by (Raju and Kumar 2014) in their study to rank GCMs through a multicriterion decision-making outranking method and GFDL-ESM2M is one of the six selected to be most suitable for India. CanESM2 and CNRM-CM5 have been used in (Shashikanth et al. 2017) and (Chaudhuri and Srivastava 2017) respectively. Some of the important properties of models are discussed as follows. GFDL-ESM2M model is based on combination of atmospheric and oceanic circulation models, and incorporate interactive biogeochemistry including the carbon cycle (Dunne et al. 2012, 2013). IPSL-CM5A-MR includes representation of tropospheric and stratospheric chemistry including aerosols (Dufresne et al. 2013) and MRI-CGCM3 is composed of ocean-ice, aerosol and atmosphere-land models (Yukimoto et al. 2012). CanESM2 consists of atmosphere-ocean model, ocean carbon and terrestrial carbon model (Christian et al. 2010). CNRM-CM5 consists of atmosphere, ocean, land and sea-ice models (Voldoire et al. 2013).

3 Methodology

Before feeding the data to the model it is important to choose the right preprocessing steps since they greatly affect the model learning process. Various preprocessing steps like interpolating GCM dataset to NCEP grid points using Inverse Distance Weighting Interpolation (IDW) to bring the predictor set to a common reference, bias correcting using quantile mapping considering NCEP data as reference to remove bias due to the assumptions made in the simulation of GCM data and further removing bias in mean and variance of both NCEP and GCM data using standardization, selecting the features having only a certain amount of spearman correlation with the precipitation data, and finally using dimensionality reduction to squeeze the data with little loss in information. Next using the weather typing approach, rainfall states for the future period are predicted based on the classification of present rainfall using K-means clustering. Finally, the reduced variables and weather states together as predictors are used to downscale precipitation.

[Figure 1]

3.1 Preprocessing

3.1.1 Interpolation

NCEP reanalysis data being the output of the high-resolution climate model can be considered as an output from ideal GCM (P. P. Mujumdar and D. Nagesh Kumar 2012) and hence is considered as a reference for our study. Since GCM's that we have taken have different spatial resolutions as compared to NCEP data, therefore it is important to interpolate them to the same scale in order to ensure spatial consistency while using the data as an input to the model. Inverse distance weighting method (IDW) which assumes that closer values are more related than farther values with its function, has been used and the number of neighbors ($T_{s,i}$) influencing a given point (T_t) have been kept constant equal to 8. The method can be expressed as:

$$T_t = \frac{\sum_{i=1}^{8} T_{s,i} w_i}{\sum_{i=1}^{8} w_i}$$

where $w_i = 1/{d_i}^2$ and d_i is the distance of neighbors on the source grid to the point on the reference grid. Interpolated data can now be bias-corrected.

3.1.2 Bias Correction

Due to incomplete knowledge of geophysical processes, different sets of realisations (initial states), initialisations (parameterisations) and physics (underlying empirical formulas) are assumed before running the GCM simulation resulting in differences between observed and modeled time series, hence this bias in data needs to be removed before using the data for future hydrologic projections. Quantile mapping aims to match cumulative distribution function (CDF) (which gives the probability that a random variable will take value less than or equal to desired value) of reference and modeled time series, and then inversing back the transformation, i.e, replacing the GCM values with the NCEP values having equal CDF to get the corrected modeled values.

After CDF's of the data has been matched as shown in figures 2 and 3, GCM and NCEP predictors are further corrected for the bias in the mean and variance by subtracting mean and dividing standard deviation of the data from the period 1961 – 1990 (World Meteorological Organization baseline period) from the respective datasets.

[Figures 2 and 3]

The duration so chosen is of sufficient duration to establish a reliable climatology, and not too long, nor too contemporary to include a strong global change signal (Wilby et al. 2004). Standardized value for k^{th} predictor variable at time t can be expressed as:

$$v_{stan,t}(k) = \frac{v_t(k) - \mu_{v,1961-1990}(k)}{\sigma_{v,1961-1990}(k)}$$

Where $v_t(k)$ is the original value of the k^{th} predictor at time t, $\mu_{v,1961-1990}(k)$ and $\sigma_{v,1961-1990}(k)$ are the mean and standard deviation of the k^{th} predictor for the baseline period.

3.2 Predictor Selection

After data has been bias-corrected, a 6 * 6 grid surrounding a particular MSD evenly from all 4 directions is chosen (figure 4) and only 180 (6 * 6 * 5) predictors are chosen for that MSD. Out of these, only the predictors having sufficient absolute spearman correlation (greater than or equal to 0.4) with average rainfall for all the IMD grid points in that region are screened for further analysis, for example for East Madhya Pradesh 125 predictors out of the total are found to have sufficient table 4. This is done to identify the most correlated features in order to remove redundancy in the dataset.

[Figure 4]

[Table 4]

3.3 Principal Component Analysis (PCA)

The predictor dataset being large and highly correlated faces from problems like multidimensionality and multicollinearity. Multidimensionality increases computation time and requires more memory resources. Having more dimensions also increases the sparseness of data and results in addition of noise (features not related to the context of the study) to the model resulting from overfitting during training and due to data being multicollinear, small changes in one variable can result in erratic changes in overall data. PCA being a vector space transform helps in reducing the data by finding important features amongst the predictors with little tradeoff in overall variance by mapping n-dimensional predictor space to m dimensions (m <= n) which are the first m eigenvectors arranged in decreasing order of eigenvalues of the correlation matrix considering data to be of the form number of timesteps * number of predictors. First m dimensions accounting for 98 % variance are only chosen for the study. For East Madhya Pradesh out of 125 predictors, the first 8 dimensions have variance equal to the desired value.

3.4 Rainfall State Estimation

Achieving accuracy in multisite downscaling demands to capture spatial variability along with temporal variability which is challenging for heterogenous variables like precipitation which is a result of complex interaction majorly between 2 biospheric components land and sea. To overcome this we use the concept of rainfall states as introduced by (Kannan and Ghosh 2013). For a particular month, every region is assigned a single representative state based on the rainfall magnitudes of all the points falling in that region and the state so assigned is relative to other months falling in the time series, thereby dividing entire time series into clusters (groups) where every point (time step) belonging to the same cluster resembles more amongst themselves than the members of other clusters. This is achieved via K–means clustering (MacQueen, 1967) which helps in identifying natural groups in data by classifying a set of n (timesteps) d dimensional observations (precipitation at all the points) (t_1, t_2, ..., t_n) into K classes{S_1, S_2, ..., S_K} such that $S_i{}^{th}$ cluster has T_i number of vectors and each t_i belongs to exactly one cluster ($\sum_{i=1}^{K} T_i = n$). The objective of clustering is to minimize the sum of distance of all the observations from their respective cluster centroids (μ_i), i.e, finding:

$$\min_S \sum_{i=1}^{k} \sum_{t \, \epsilon \, S_i} \left\| t - \mu_i \right\|^2$$

where $\mu_i{}^{(k)} = \left(\frac{1}{T_k}\right) \sum_{i=1}^{T_k} t_i{}^{(k)}$ for k = 1…..K

The optimal number of clusters is found using the Silhouette index which is a measure of how close each point in one cluster is to points in neighboring clusters. A higher value indicates that the point is well classified. The value of K is varied from 3 to 9 considering a minimum of 3 states (dry, semi-wet and wet). Future rainfall states are calculated using NCEP predictors and present rainfall states as input to LSTM for model training and then classifying future timesteps based on GCM predictors as input.

[Figure 5]

3.5 Long Short Term Memory Neural Networks (LSTM)

3.5.1 Introduction

LSTM's have the ability to store information outside the normal flow of the recurrent network in a gated cell. Each cell has 3 gates that are responsible for managing the flow of information through the network. First information from the previous cell is passed through forget-gate (f_t) which helps in forgetting the information not required in further time steps by applying matrix operations like element-wise multiplication with weights and addition of bias and then passing it through logistic sigmoid function (σ) which converts the value of resulting vector in range of (0, 1) depending on the degree to which information is remembered by the gate. The weights are adjusted via gradient descent during backpropagation. This helps in optimizing the learning process by forgetting less useful information. The same mechanism is followed by other gates using different sets of weights and bias. Forget gate output can be expressed as:

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right)$$

where x_t is the input at the present time step, W_f, U_f are adjustable weights for forget gate, b_f is the bias vector for forget gate, h_{t-1} is the previous cell output (hidden state).

Next, input-gate(i_t) ensures that only important information is added to the cell state. It is a two-step process and is done in combination with another vector $\hat{c}$ and forget-gate output f_t. First, the input-gate output can be expressed as:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$

where x_t is the input at the present time step, W_i, U_i are adjustable weights for input-gate, b_i is the bias vector for input-gate, h_{t-1} is the hidden state. Next $\hat{c}$ is calculated using the same methodology as above but has an output in range (-1, 1) making use of hyperbolic tangent (tanh) as activation function:

$$\hat{c} = tanh(W_{\hat{c}} x_t + U_{\hat{c}} h_{t-1} + b_{\hat{c}})$$

where x_t is the input at the present time step. $W_{\hat{c}}$, $U_{\hat{c}}$ are adjustable weights for $\hat{c}$, $b_{\hat{c}}$ is the bias vector for $\hat{c}$, h_{t-1} is the previous hidden state. Both f_t and i_t being in range (0, 1) decide the degree to which previous cell information (cell state) and present information are to be incorporated in the present cell state which can be expressed as:

$$c_t = f_t \times c_{t-1} + i_t \times \hat{c}$$

The cell state is transferred within the cells and is responsible for remembering long term dependencies. The process of reading, storing and writing only relevant information via linear mathematical operations helps in avoiding the gradients to be too low or too high making LSTM more superior over traditional RNN's.

Finally, the task of identifying relevant present cell state information and returning it as output is done by output gate which can be expressed as:

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$$

where x_t is the input at the present time step, W_o, U_o are adjustable weights for output gate b_o is the bias vector for output-gate, h_{t-1} is the previous hidden state. The output of the gate is in range (0, 1) and therefore decides the amount of cell state to be returned as output as follows:

$$h_t = \tanh(c_t) \times o_t$$

3.5.2 Model development and downscaling process

Both rainfall state estimation and future precipitation projection have been done using a 3 layer LSTM model with 0.5 dropout having 30 neurons in every layer followed by a dense layer. The model has been implemented using Keras with Tensorflow backend. Learning rate equal to 0.001, batch-size equal to 5, epochs equal to 120 have been found as best hyper-parameters for both the models. The activation function used for the LSTM layer is the default 'tanh' function and for dense layer, the classifier has 'softmax' activation and regressor has 'linear' activation. Adaptive Moment Estimation (ADAM) optimizer (Kingma and Ba 2015) is used for the learning process. For classifier, 'categorical-cross-entropy' and for regressor 'mean absolute error' have been chosen as loss functions during the training process and are validated for 'accuracy' and 'mean squared error' respectively. The complete information for all the hyper-parameters can be found in Keras online documentation. With the above-mentioned parameters, the model attains a constant validation loss after a few epochs.

The training period (1951-2005) has been split into 3 parts training, validation and testing accounting for 70%, 15%, 15% of the total NCEP data. Models are tested for their performance on GCM's historical dataset with respect to IMD observations based on the goodness of fit parameters R^2 where:

$$R = \frac{n\left(\sum P_m P_o\right) - \sum P_m \sum P_o}{\sqrt{\left(n \sum {P_m}^2 - \left(\sum P_m\right)^2\right)\left(n \sum {P_o}^2 - \left(\sum P_o\right)^2\right)}}$$

and NSE (Nash Sutcliffe Model Efficiency):

$$NSE = 1 - \frac{\sum (P_m - P_o)^2}{\sum (P_0 - \overline{P_o})^2}$$

for the regression model where P_m is modeled precipitation, P_o is observed precipitation, $\overline{P_o}$ is the mean of observed precipitation and n is the length of time series. The summation is taken over the entire time series. Accuracy, i.e, percentage of modeled results matching with observed results is considered for the classification model. The results for 5 regions spanning all 5 parts of India are given in Table 4 and 5.

[Table 4 and 5]

4 Observations and Conclusions

Accuracy of the model depends on how well the predictor variables have been simulated by the GCM's. Taking a weighted mean over different GCM's can help in getting satisfactory results. For our study we have taken similar weight for all the GCM's. Results have been observed both temporally(using boxplots) and spatially(using spatial plots).

A boxplot displays minimum, median(middle line in the box), and maximum of the dataset along with the outliers showing how the data is distributed. Figure 6 compares the distribution of average rainfall of all the GCM's with respect to observed value for 32 chosen MSD's for both rcp4.5 and rcp8.5 scenarios.

[Figure 6]

The plots indicate a maximum increase in south-western(Coastal Karnataka, Konkan and Goa) with lesser change in central, western and north-eastern India whereas other regions either show slight decrease or non-monotonous variation in rainfall.

Figure 7 depicts spatial variation in rainfall. Further considering observed value as reference, error with respect to GCM simulation for historical period is shown in figure 8. With more than 90% of points receiving less than 250 mm/month rainfall, maximum change can be observed in south western, central and north eastern regions amongst present and future pattern. Overall error is bounded between -100 to 25 mm/month for more than 90% of points with maximum error in regions of heavy rainfall like eastern, south western and some parts of northern and central India.

[Figure 7 and 8]

The model developed lacks in capturing extreme events which is a limitation of extreme event classification tasks. In order to get better results for all the regions, it is important to identify potential predictors specific to that region since predictors relative to the given predictand vary across time and space. Also by increasing the number of clusters during the training process, the model showed better results but due to lack of subjectivity, we have limited the choice for number of clusters based on the validation index. Other techniques like passing input for multiple time-steps (projecting present-day precipitation using predictors from present and past time scales), making projections separately for dry and wet seasons, taking more correlated predictors can be potential improvements to be considered for future work.

In short there are always uncertainties associated with future climate variables and the inheritance of present relationship to future is also not a valid assumption, still downscaling remains the most trusted tool to study impact of climate change on various hydrological processes.

5 References

Akbari Asanjan A, Yang T, Hsu K, et al (2018) Short-Term Precipitation Forecast Based on the PERSIANN System and LSTM Recurrent Neural Networks. J Geophys Res Atmos 123:12,543-12,563. https://doi.org/10.1029/2018JD028375

Chaudhuri C, Srivastava R (2017) A novel approach for statistical downscaling of future precipitation over the Indo-Gangetic Basin. J Hydrol 547:21–38. https://doi.org/10.1016/j.jhydrol.2017.01.024

Cho K, Van Merriënboer B, Gulcehre C, et al (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv Prepr arXiv14061078

Christian JR, Arora VK, Boer GJ, et al (2010) The global carbon cycle in the Canadian Earth system model (CanESM1): Preindustrial control simulation. J Geophys Res Biogeosciences 115:1–20. https://doi.org/10.1029/2008JG000920

Dufresne JL, Foujols MA, Denvil S, et al (2013) Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5

Dunne JP, John JG, Adcroft AJ, et al (2012) GFDL's ESM2 global coupled climate-carbon earth system models. Part I: Physical formulation and baseline simulation characteristics. J Clim 25:6646–6665. https://doi.org/10.1175/JCLI-D-11-00560.1

Dunne JP, John JG, Shevliakova S, et al (2013) GFDL's ESM2 global coupled climate-carbon earth system models. Part II: Carbon system formulation and baseline simulation characteristics. J Clim 26:2247–2267. https://doi.org/10.1175/JCLI-D-12-00150.1

Goyal MK, Ojha CSP (2012) Downscaling of surface temperature for lake catchment in an arid region in India using linear multiple regression and neural networks. Int J Climatol 32:552–566. https://doi.org/10.1002/joc.2286

Guhathakurta P, Rajeevan M (2008) Trends in the rainfall pattern over India. Int J Climatol 28:1453–1469. https://doi.org/10.1002/joc.1640

Gutiérrez JM, Cofino AS, Cano R, Rodríguez MA (2004) Clustering methods for statistical downscaling in short-range weather forecasts. Mon Weather Rev 132:2169–2183. https://doi.org/10.1175/1520-0493(2004)132<2169:CMFSDI>2.0.CO;2

Hinton G, Deng L, Yu D, et al (2012) Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Process Mag 29:

Hochreiter S (1997) Long Short-Term Memory. 1780:1735–1780

Huang A, Vega-Westhoff B, Sriver RL (2019) Analyzing El Niño–Southern Oscillation Predictability Using Long-Short-Term-Memory Models. Earth Sp Sci 6:212–221. https://doi.org/10.1029/2018EA000423

Kannan S, Ghosh S (2013) A nonparametric kernel regression model for downscaling multisite daily precipitation in the Mahanadi basin. Water Resour Res 49:1360–1385. https://doi.org/10.1002/wrcr.20118

Kingma DP, Ba JL (2015) Adam: A method for stochastic optimization. 3rd Int Conf Learn Represent ICLR 2015 - Conf Track Proc 1–15

Luong M-T, Sutskever I, Le Q V, et al (2014) Addressing the rare word problem in neural machine translation. arXiv Prepr arXiv14108206

MacQueen J, others (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. pp 281–297

Mikolov T, Karafiát M, Burget L, et al (2010) Recurrent neural network based language model. In: Eleventh annual conference of the international speech communication association

P. P. Mujumdar and D. Nagesh Kumar (2012) More Information - Www.Cambridge.Org/9781107018761

Pascanu R, Mikolov T, Bengio Y (2013) On the difficulty of training recurrent neural networks. In: International conference on machine learning. pp 1310–1318

Raju KS, Kumar DN (2014) Ranking of global climate models for India using multicriterion analysis. Clim Res 60:103–117. https://doi.org/10.3354/cr01222

Sainath TN, Kingsbury B, Saon G, et al (2015) Deep convolutional neural networks for large-scale speech tasks. Neural Networks 64:39–48

Salvi K, Kannan S, Ghosh S (2013) High-resolution multisite daily rainfall projections in India with statistical downscaling for climate change impacts assessment. 118:3557–3578. https://doi.org/10.1002/jgrd.50280

Shashikanth K, Sukumar P, Professor A (2017) Indian Monsoon Rainfall Projections for Future Using GCM Model Outputs Under Climate Change. 10:1501–1516

Solomon S, Qin D, Manning M, et al (2007) Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC. Cambridge university press

Tatli H, Dalfes HN, Menteş ŞS (2005) Surface air temperature variability over Turkey and its connection to large-scale upper air circulation via multivariate techniques. Int J Climatol 25:331–350. https://doi.org/10.1002/joc.1133

Tran Anh D, Van SP, Dang TD, Hoang LP (2019) Downscaling rainfall using deep learning long short-term memory and feedforward neural network. Int J Climatol. https://doi.org/10.1002/joc.6066

Tripathi S, Srinivas V V., Nanjundiah RS (2006) Downscaling of precipitation for climate change scenarios: A support vector machine approach. J Hydrol 330:621–640. https://doi.org/10.1016/j.jhydrol.2006.04.030

Voldoire A, Sanchez-Gomez E, Salas y Mélia D, et al (2013) The CNRM-CM5.1 global climate model: Description and basic evaluation. Clim Dyn 40:2091–2121. https://doi.org/10.1007/s00382-011-1259-y

Wilby RL, Charles SP, Zorita E, et al (2004) Guidelines for use of climate scenarios developed from statistical downscaling methods. Support Mater Intergov Panel Clim Chang available from DDC IPCC TGCIA 27:

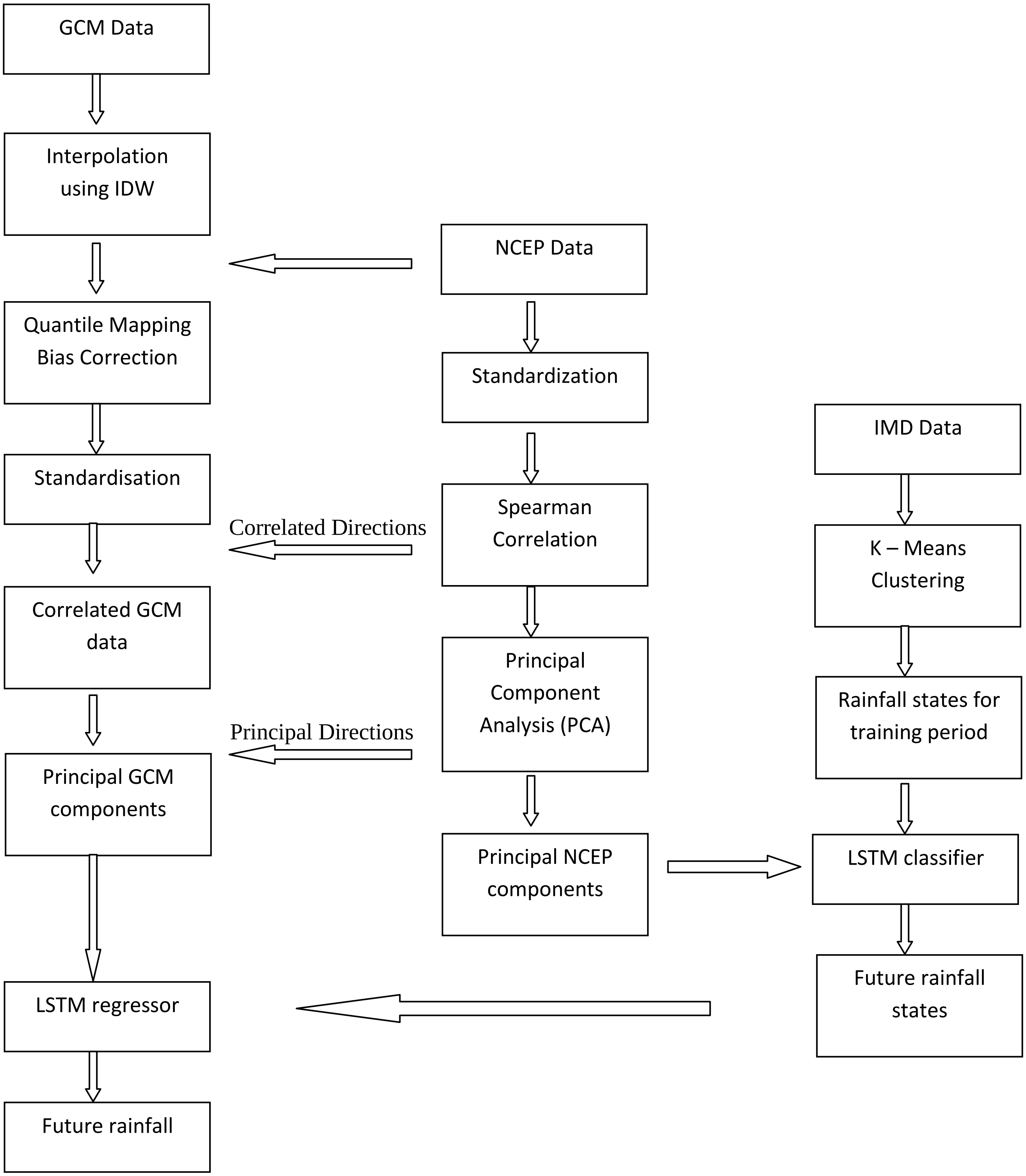
Wilby RL, Wigley TML (2000) Precipitation predictors for downscaling: Observed and general circulation model relationships. Int J Climatol 20:641–661. https://doi.org/10.1002/(SICI)1097-0088(200005)20:6<641::AID-JOC501>3.0.CO;2-1

Yukimoto S, Adachi Y, Hosaka M, et al (2012) A new global climate model of the Meteorological Research Institute: MRI-CGCM3: -Model description and basic performance-. J Meteorol Soc Japan 90:23–64. https://doi.org/10.2151/jmsj.2012-A02

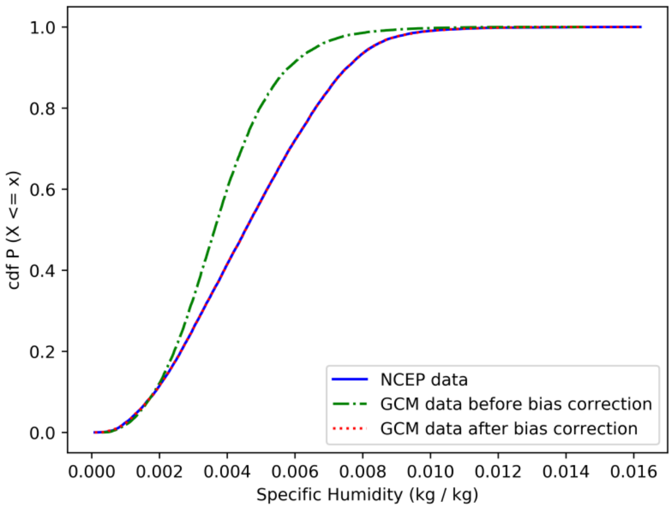
Zaytar MA, El Amrani C (2016) Sequence to sequence weather forecasting with long short-term memory recurrent neural networks. Int J Comput Appl 143:7–11

LIST OF FIGURES

Figure 1 Flowchart for multisite downscaling showing different models, datasets and operations involved.



Figures 2 and 3. Comparison of CDF's and true values for NCEP and historical CanESM2 data for a grid point before and after bias correction.



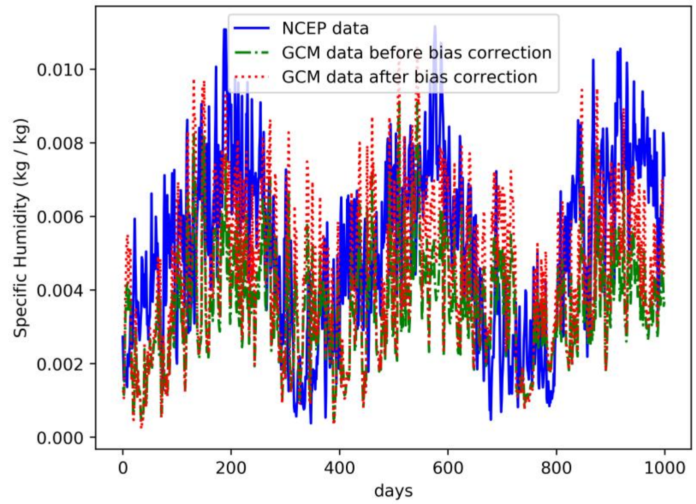


Figure 4. Thirty-Six NCEP grid points considered for choosing predictor variables having sufficient correlation with precipitation for East Madhya Pradesh subdivision of India.

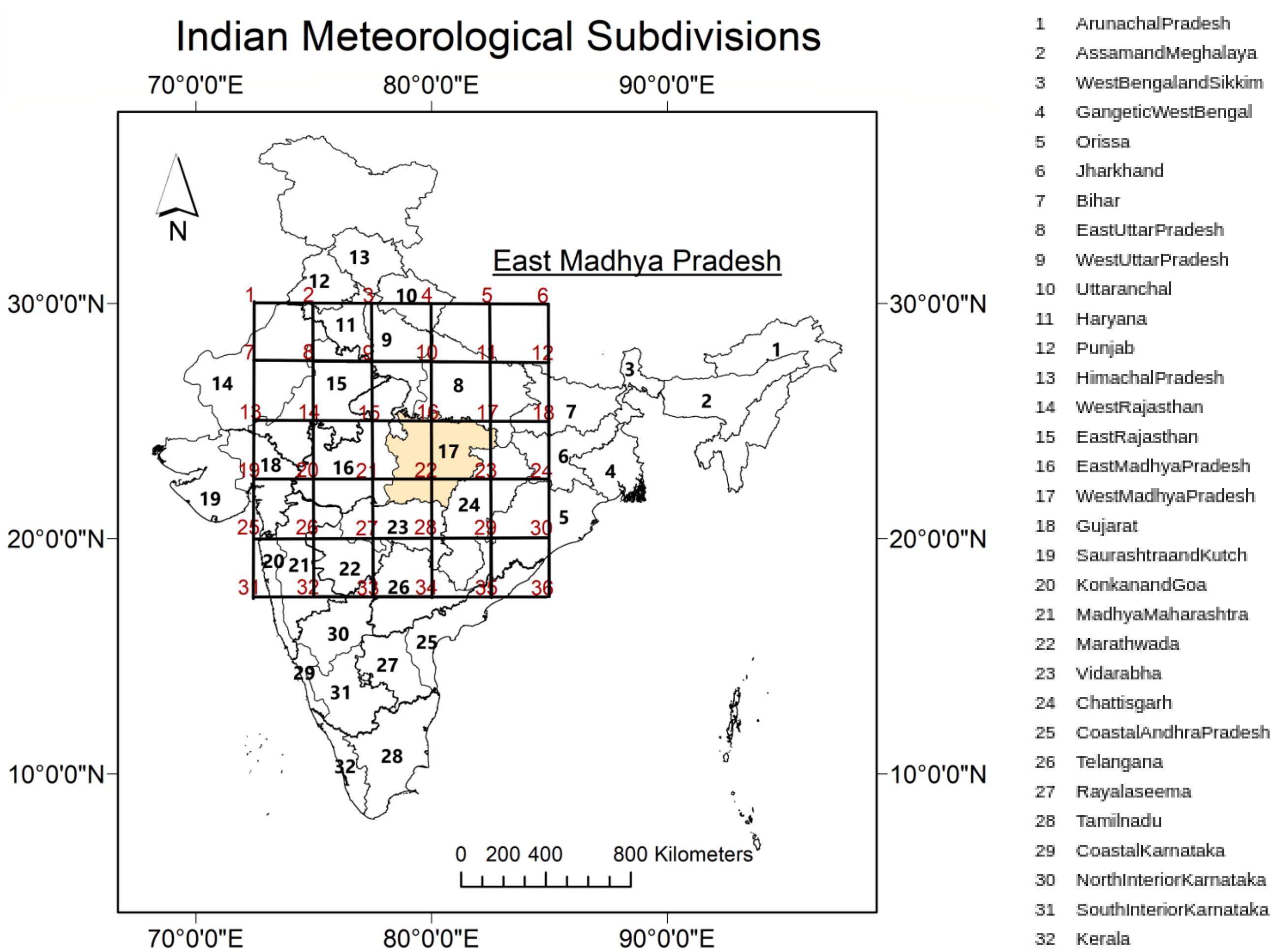


Figure 5. Silhouette Score for different number of clusters of weather-types for East Madhya Pradesh Subdivision of India

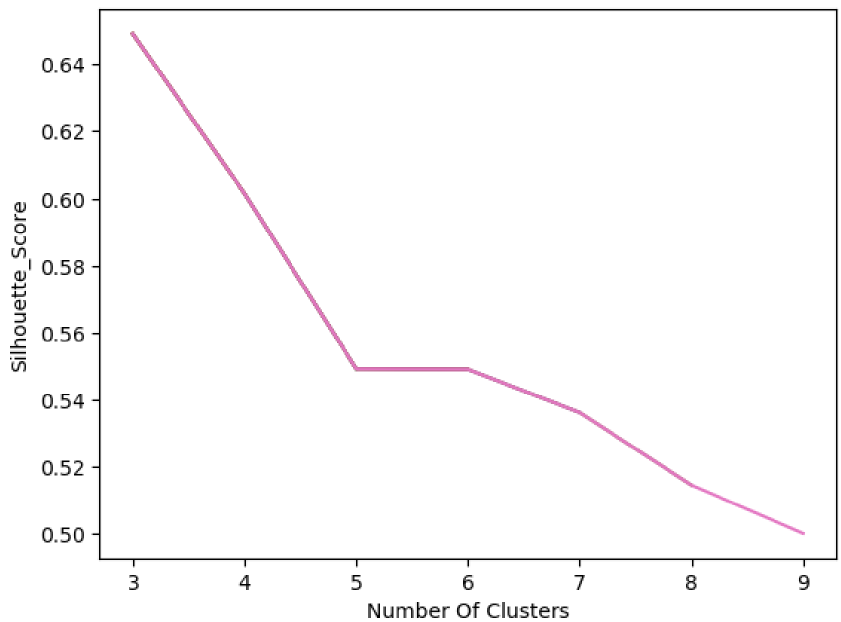


Figure 6. Boxplots for (a) Spatial average observed rainfall for training period (1951-2005), (b) Spatial average simulated rainfall for testing period (2006-2060), (c) Spatial average simulated rainfall for testing period (2061-2100) for rcp4.5 and rcp8.5 scenarios respectively.

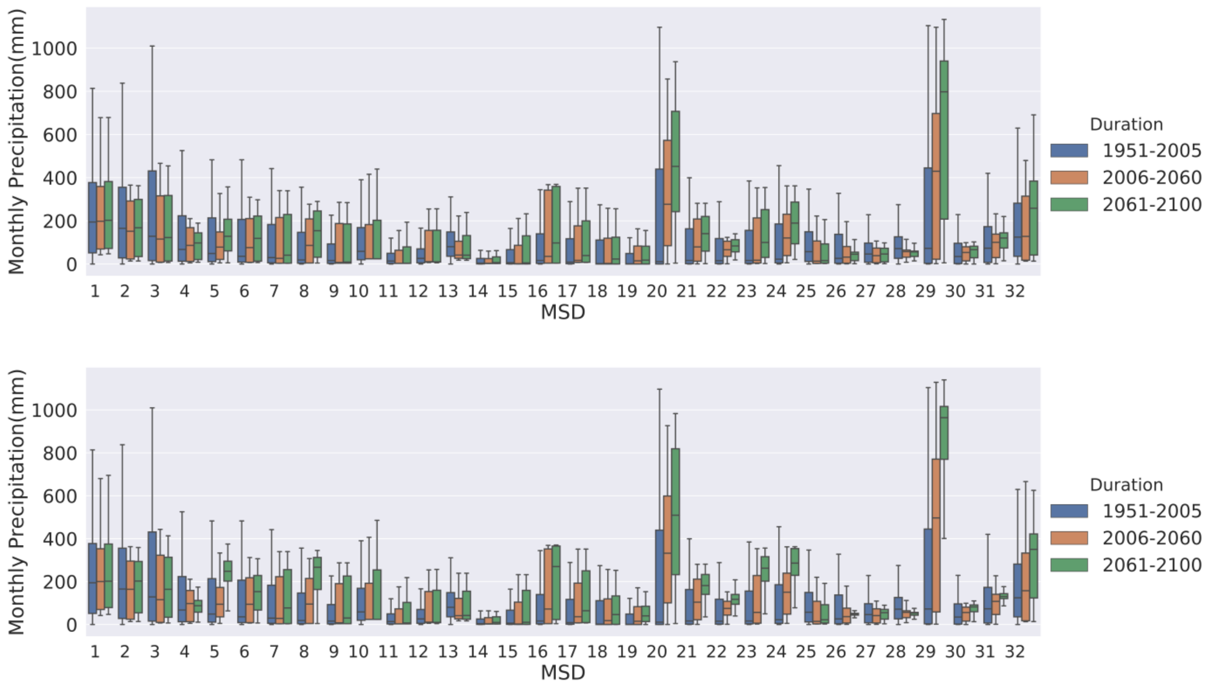


Figure 7. Spatial Plots For Average Monthly Rainfall for period 1951-2005 for plots (a) and (b) and for period 2006-2100 for plots (c) and (d).

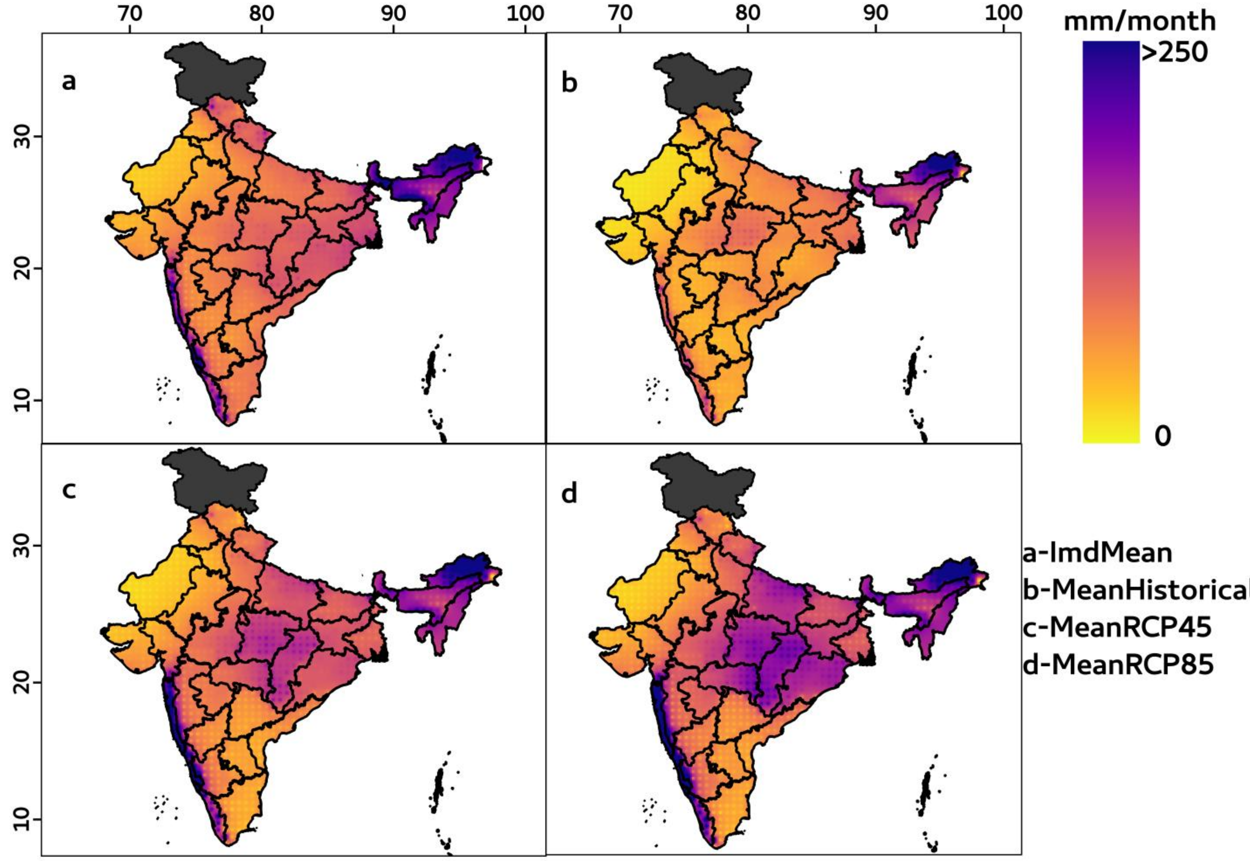
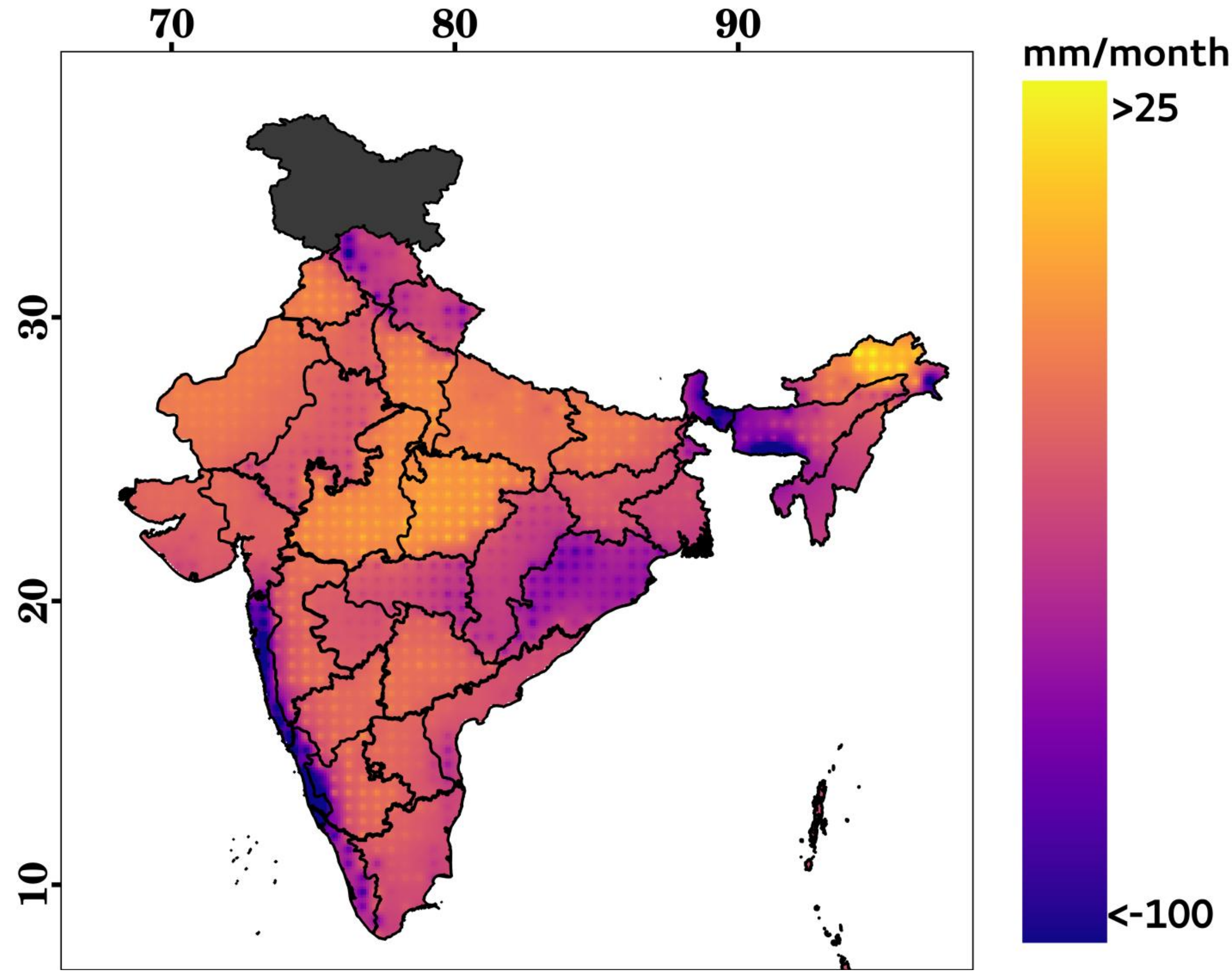


Figure 8. Spatial Plot for error in mean rainfall for simulated and IMD observed rainfall for period 1951-2005



LIST OF TABLES

Table 1. Different predictors chosen for precipitation downscaling over Indian landmass in other studies

Author	Study Area	Predictors Chosen	Predictand
(Tripathi, Srinivas, & Nanjundiah, 2006)	29 Meteorological Indian Subdivisions (MSD)	Air temperature, relative humidity, specific humidity, geo-potential height, zonal, vertical and meridional wind velocities at 1000 mb, 850 mb, 500 mb and 200 mb pressure levels	Monthly precipitation
(Ghosh & Mujumdar, 2006)	Orissa	Mean sea level pressure, 500 mb geopotential height	Monthly precipitation
(Salvi, Kannan, & Ghosh, 2013)	7 Meteorological homogeneous zones	Temperature, pressure, specific humidity, u wind and v wind all at surface level	Daily precipitation

Table 2. GCM models chosen for the study

Serial Number	Model Name	Spatial Resolution	Modelling Centre
1	GFDL-ESM2M	2.0225 * 2.5	NOAA Geophysical Fluid Dynamics Laboratory
2	IPSL-CM5A-MR	1.268 * 2.5	Institut Pierre-Simon Laplace
3	MRI-CGCM3	1.113 * 1.125	Meteorological Research Institute
4	CanESM2	2.8 * 2.8	Canadian Centre for Climate Modelling and Analysis
5	CNRM-CM5	1.4 * 1.4	Centre National de RecherchesMéteéorologiques/Centre Européen de Recherche et Formation Avanceées en Calcul Scientifique

Table 3. List of identified predictors and their domain for East Madhya Pradesh sub-division of India

Variable	Grid Points
Specific Humidity at 1000 hpa	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36
U-Wind Speed at 1000 hpa	3, 9, 10, 18, 20, 21, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36
Mean Sea Level Pressure at 1000 hpa	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36
Temperature at 1000 hpa	1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 31, 32, 36
Geopotential Height at 500 hpa	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16, 17, 18, 31, 32, 33, 34, 35, 36

Table 4. Different skill scores of the model evaluated on NCEP dataset for testing portion of the training period (last 15% of 1951-2005)

Region	Total IMD grid points	Total predictors(NCEP principal components + weather types)	Accuracy (%)	R^2	NSE
East Madhya Pradesh	49	8 + 3	85.86	0.784	0.76
Uttaranchal	21	9 + 3	84.85	0.73	0.635
South Interior Karnataka	30	11 + 3	86.87	0.487	0.432
Assam and Meghalaya	37	9 + 3	75.76	0.677	0.623
Konkan and Goa	14	10 + 3	86.87	0.862	0.844

Table 5. Different skill scores of the model evaluated on GCM dataset for the training period (1951-2005)

Region	Evaluation Metrics	Model				
		CanESM2	CNRM-CM5	GFDL-ESM2M	IPSL-CM5A-MR	MRI-CGCM3
East Madhya Pradesh	Accuracy	0.807	0.83	0.81	0.82	0.82
	R^2	0.56	0.634	0.57	0.611	0.61
	NSE	0.47	0.579	0.523	0.57	0.6
Uttaranchal	Accuracy	0.841	0.886	0.84	0.84	0.81
	R^2	0.523	0.56	0.48	0.5	0.54
	NSE	0.5	0.53	0.45	0.47	0.477
South Interior Karnataka	Accuracy	0.73	0.725	0.727	0.73	0.71
	R^2	0.33	0.386	0.38	0.335	0.374
	NSE	0.286	0.325	0.34	0.28	0.311
Assam and Meghalaya	Accuracy	0.67	0.74	0.737	0.747	0.74
	R^2	0.55	0.578	0.569	0.58	0.59
	NSE	0.46	0.495	0.495	0.496	0.5
Konkan and Goa	Accuracy	0.77	0.81	0.76	0.757	0.78
	R^2	0.48	0.64	0.487	0.57	0.63
	NSE	0.366	0.55	0.4	0.4	0.49