Theor Appl Climatol
https://doi.org/10.1007/s00704-017-2307-2

ORIGINAL PAPER

Statistical downscaling of precipitation using long short-term memory recurrent neural networks

Saptarshi Misra[1] · **Sudeshna Sarkar**[1] · **Pabitra Mitra**[1]

Received: 13 December 2016 / Accepted: 20 October 2017

Abstract Hydrological impacts of global climate change on regional scale are generally assessed by downscaling large-scale climatic variables, simulated by General Circulation Models (GCMs), to regional, small-scale hydrometeorological variables like precipitation, temperature, etc. In this study, we propose a new statistical downscaling model based on Recurrent Neural Network with Long Short-Term Memory which captures the spatio-temporal dependencies in local rainfall. The previous studies have used several other methods such as linear regression, quantile regression, kernel regression, beta regression, and artificial neural networks. Deep neural networks and recurrent neural networks have been shown to be highly promising in modeling complex and highly non-linear relationships between input and output variables in different domains and hence we investigated their performance in the task of statistical downscaling. We have tested this model on two datasets—one on precipitation in Mahanadi basin in India and the second on precipitation in Campbell River basin in Canada. Our autoencoder coupled long short-term memory recurrent neural network model performs the best compared to other existing methods on both the datasets with respect to temporal cross-correlation, mean squared error, and capturing the extremes.

This work was supported by MHRD, Govt. of India and Indian Institute of Technology, Kharagpur.

✉ Saptarshi Misra
saptarshimisra2011@gmail.com

Sudeshna Sarkar
sudeshna@cse.iitkgp.ernet.in

Pabitra Mitra
pabitra@cse.iitkgp.ernet.in

[1] Department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur, West Bengal, India

1 Introduction

Neural networks, especially deep neural networks, are a powerful class of machine learning models. The tremendous impact of global warming and climate change are already observed throughout the world in all spheres of ecosystems, be it terrestrial or aquatic. Water is the lifeline of our society and hence assessing the effects of climate change on hydro-meteorology is of utmost importance. General Circulation Models (GCMs) provide the most reliable simulations of the global climate systems, and they provide present and future time series of climate variables for the entire globe (Prudhomme et al. 2002; Intergovernmental Panel on Climate Change - Task Group on Scenarios for Climate Impact Assessment 1999). Though they are capable of capturing large-scale circulations and smoothly varying fields such as mean sea level pressure well enough, they often fail to capture the non-smooth fields such as precipitation, temperature, soil moisture, etc. on a regional scale (Hughes and Guttorp 1994). GCMs have a large spatial resolution (say, $3.75^\circ \times 3.75^\circ$ in case of coupled global circulation model (CGCM2)). Downscaling is therefore necessary to model these regional-scale hydrometeorological variables.

The downscaling methods can be primarily classified into dynamic and statistical downscaling. While dynamic downscaling involves deriving a high-resolution RCM (Regional Climate Model) from the coarse-grid GCM, statistical downscaling deals with developing a linear or non-linear relationship between the large scale atmospheric variables (predictor) to the small-scale surface variable of

Published online: 28 November 2017

interest (precipitation in our case). The statistical downscaling models are based on the major assumptions that regional climate is largely affected by the global-scale circulation patterns (von Storch 1995, 1999) and the relationship between the predictor and the predictand variables is invariant under future climate scenarios. A detailed review of the downscaling approaches can be found in the works of von Storch (1995), Wilby and Wigley (1997), Murphy (1999), Haylock et al. (2006), Hanssen-Bauer et al. (2005), Christensen and et al (2007), and Bi et al. (2015).

Neural networks, especially deep neural networks, have drawn a lot of interest due to their success in solving some of the most computationally difficult problems with highly non-linear relationships between input and output variables. The success of neural networks is mainly attributed to their ability to learn hierarchical representations, unlike traditional machine learning models that build up on hand-engineered features. Long short-term memory, an improved version of recurrent neural networks, performs exceptionally well in modeling temporal dependencies, in tasks with inherent temporal dependencies such as image captioning, video captioning, sequence learning, translation in natural language processing, etc.

The recent advances of deep learning has helped in solving complex computational problems in all fields (computer vision, image processing, and related fields). A few such applications in climate data mining using deep learning techniques (Liu et al. 2014; Gope et al. 2016) has motivated us to use deep neural networks and Long Short-Term Memory Recurrent Neural Networks (RNN-LSTMs) as models for statistical downscaling of precipitation. As Kannan and Ghosh (2013) and Mandal et al. (2016) have used the lag-1 precipitation states as conditioning input for their model, we have used the precipitation state as one of the inputs in our neural network models. The details of the method are given in the next sections. For model performance evaluation, the results obtained by our proposed method are compared with that from the existing non-parametric kernel regression and beta regression methods and our model (RNN-LSTM) is found to perform considerably better compared to these methods. Details regarding the case study areas and the datasets used are given in the Section 2, followed by methodology and the results obtained.

2 Related work

The statistical downscaling methodologies developed so far can principally be grouped into three categories, namely, weather classification or weather typing (e.g., Hay et al. 1991; Bardossy and Plate 1992; Corte-Real et al. 1995; Conway and Jones 1998), regression or transfer function (Murphy 1999; von Storch et al. 1993; Bardossy et al. 1995), and weather generators (Hughes et al. 1993; Hughes and Guttorp 1994; Wilks 1999). The weather typing and transfer function based approaches, generally known as perfect-prognosis downscaling, establish a relationship between large-scale climate variables and regional-scale predictand variable (Wilks 2006). The selection of large-scale predictors (Wilby and Wigley 1997; Wilby et al. 1998) and developing a statistical model linking the large-scale with the small-scale variables are the main constituents of this class of downscaling techniques. Linear regression (Karl et al. 1990), generalized linear model (GLM) (Dobson 2001), the generalized additive model (Hastie and Tibshirani 1990), and vector GLM (Yee and Wild 1996) are some models belonging to this category. The weather generators, on the other hand, are statistical models generating random sequences of climate variables, which preserves statistical properties of observed weather (Richardson 1981; Richardson and Wright 1984; Wilks 1998). Some other noteworthy statistical downscaling models are CRF model (Raje and Mujumdar 2009), non-homogeneous hidden Markov models (NHMMs) (Hughes and Guttorp 1994; Hughes et al. 1999).

Despite reasonable advancements in the development of statistical downscaling techniques, especially for simulation of precipitation, challenges still exist in simulating precipitation series with realistic spatial and temporal dependencies and capturing extremities in precipitation. Modeling spatial dependence in a nonstationary climate is one of the most significant challenges in downscaling (Yang et al. 2005). This problem was addressed well by the recently developed methods of nonparametric kernel regression (Kannan and Ghosh 2013) and beta regression conditioned on rainfall states (Mandal et al. 2016).

3 Case study area considered

Two case study areas are considered to test the model. Details of the case study areas are mentioned below.

Case study area 1: The Campbell river is a significant river located on the Vancouver island in the British Columbia province of Canada. It is both snow and rain-fed. Details regarding the river can be found in the work of Mandal et al. (2016). The location of the Campbell River basin in Canada and the exact downscaling locations are depicted in Fig. 1 and Table 1 respectively.

Case study area 2: The Mahanadi is a major peninsular river in East Central India, flowing through the states of Orissa and Chattisgarh and acting as a lifeline for millions of people residing there. The details regarding the Mahanadi river is discussed in the works of Kannan and Ghosh (2011, 2013). The location of the Mahanadi basin

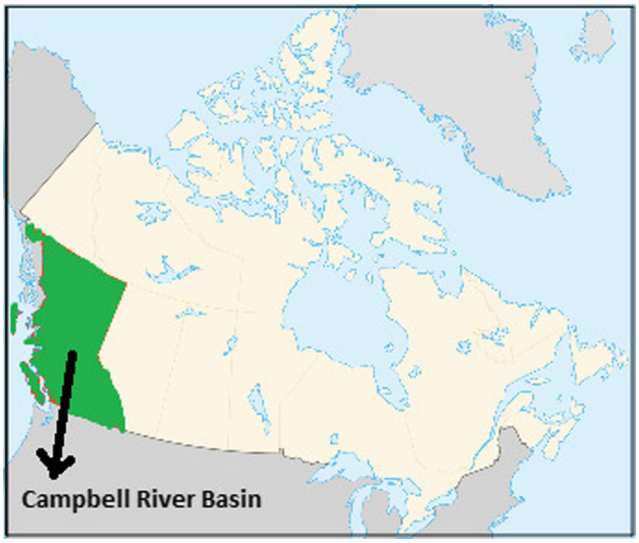


Fig. 1 Location of Campbell river basin in Canada

in India is pointed out in Fig. 2. The exact downscaling locations used for our study in sync with Kannan and Ghosh (2013)'s work are depicted in Table 2.

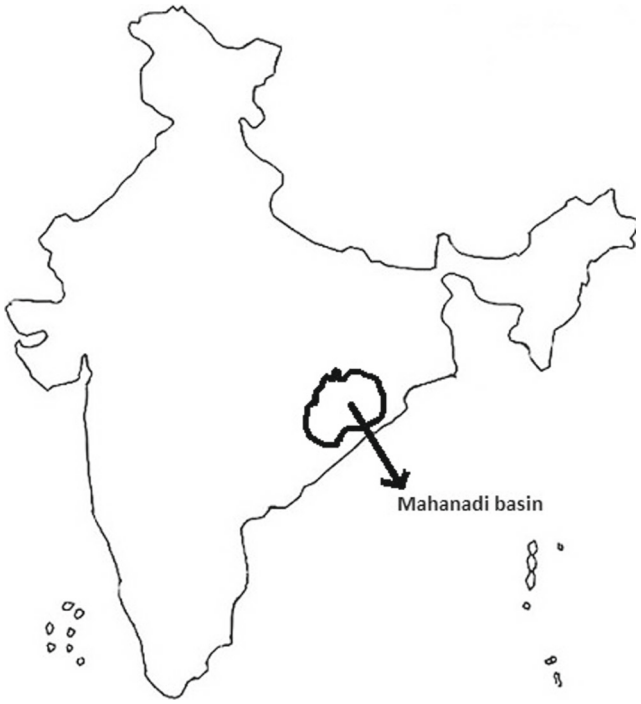


Fig. 2 Location of Mahanadi river basin in India

4 Climatic variables used

The climatic variables of the two datasets used for testing our proposed statistical downscaling model are discussed subsequently.

4.1 Dataset 1 [Campbell basin]

The daily precipitation data (0.1° latitude × 0.1° longitude) for 40 years from 1961 to 2013 at ten locations covering the entire Campbell river basin have been obtained from the ANUSPLIN dataset, Environment Canada (Hutchinson and Xu 2013). The ANUSPLIN data is developed using "thin plate smoothing splines" algorithm. This method interpolates climate variables as a function of latitude, longitude, and elevation.

Following the methodology of Wilby et al. (1999) as mentioned in the previous section, daily maximum and minimum temperature ($Tmax$ and $Tmin$), mean sea level pressure ($mslp$), specific humidity at 500 hpa (hus), zonal wind velocity ($Uwind$), and meridional wind velocity ($Vwind$) are used as predictors for the downscaling task. Due to unavailability of sufficient data for the predictor variables, predictor data is extracted from the NCEP/NCAR (National Centers for Environmental Prediction/National Center for Atmospheric Research) reanalysis dataset (Kalnay et al. 1996) for 53 years from 1961 to 2013. NCEP/NCAR dataset

Table 1 Precipitation stations in Campbell river basin, Canada

Station	Latitude (°N)	Longitude (°W)	Station abbreviation
Elk R ab Campbell Lk	49.85	125.8	ELK
Eric creek	49.6	125.3	ERC
Gold R below Ucona R	49.7	126.1	GLD
Heber river near gold river	49.82	125.98	HEB
John hart substation	50.05	125.31	JHT
Quinsam R at agronaut Br	49.93	125.51	QIN
Quinsam R nr Campbell R	50.03	125.3	QSM
Salmon R ab Campbell div	50.09	125.67	SAM
Strathcona dam	49.98	125.58	SCA
Wolf river upper	49.68	125.74	WOL

Table 2 Precipitation stations in Mahanadi river basin, India

Station ID	Latitude (°N)	Longitude (°E)
1	20.5	82.5
2	20.5	83.5
3	20.5	84.5
4	21.5	81.5
5	21.5	82.5
6	21.5	83.5
7	22.5	82.5
8	22.5	83.5

is a combination of actual and model-forecasted gridded data at 2.5° × 2.5° resolution.

So, the predictor variables are $Tmax$, $Tmin$, $mslp$, hus, $uwind$, $vwind$ and the predictand variable is the Campbell basin region precipitation data extracted from the gridded ANUSPLIN precipitation data for Canada (0.1° latitude × 0.1° longitude). Data from 1961 to 1990 is used for training and data from 1991 to 2013 is used for testing the model.

4.2 Dataset 2 [Mahanadi basin]

The gridded daily rainfall data (1° latitude × 1° longitude) developed by Rajeevan et al. (2006) for the whole of Indian subcontinent has been obtained from Indian Meteorological Department (IMD). This data is based on actual recorded rainfall data at 1803 gauging stations all over India, which had a minimum of 90% data availability for the period of 1951–2003. Unevenly spaced daily rainfall data was interpolated by Rajeevan et al. (2006) to form a regular, n-dimensional array using a numerical interpolation method proposed by Shephard (1968). This method uses an ensemble of simple, local interpolation functions that match suitably at their boundaries. More details of this interpolation scheme can be found in the works of Shephard (1968) and Rajeevan et al. (2006).

The selection of appropriate predictor variables is one of the most challenging tasks, and there are contradicting opinions for this. However, in general, as reported in the literature (Wilby et al. 1999; Wetterhall et al. 2005), the predictors should be reliably simulated by GCMs, easily available from the archives of GCM outputs, and highly correlated with the local surface variable of interest (rainfall in this case). There are several methods like partial mutual information(PMI) criteria (Sharma 2000), Variable Convergence Score (VCS) (Johnson and Sharma 2009). But, as these methods may exclude predictors which may be important for future climate scenarios, following Kannan and Ghosh (2011), we adopt a conventional method suggested by Wilby et al. (1999). As historical data for climate variables are not available uniformly, we use the NCEP/NCAR Reanalysis data. Following Kannan's work, we have selected surface air temperature(AIR), mean sea level pressure (MSLP), horizontal component of wind velocity (UWind), vertical component of wind velocity (VWind), and surface specific humidity (SHUM) as predictor variables.

To account for the physical processes, such as low pressure area over northern and central India and the movement of air current from the Indian Ocean through Bay of Bengal toward the low-pressure area during the summer monsoon, following Kannan and Ghosh (2011), we have extracted surface air temperature, mean sea level pressure, surface specific humidity, horizontal and vertical components of wind velocity for the region spanning the latitudes 7.5° N–35° N and longitudes 70° E–97.5° N covering 144 grid points (on 2.5° × 2.5° grid) for the monsoon months of June, July, August and September from 1951 to 2000.

So, the predictor variables are AIR, MSLP, SHUM, UWind, VWind, and the predictand variable is the Mahanadi basin region rainfall data extracted from the gridded all India IMD rainfall data (1° latitude × 1° longitude). This data has been truncated into two parts—data from 1951 to 1980 is used for training our model and that from 1981 to 2000 is used for validating the model.

5 Methodology for statistical downscaling using LSTM

The task is to perform statistical downscaling of precipitation for two regions—Mahanadi basin and Campbell basin using a LSTM based neural network model and compare the results of this model with those of a suite of existing approaches like non-parametric kernel regression (Kannan and Ghosh) and beta regression (Mandal, Srivastav and Simonovic).

On Mahanadi basin dataset, the existing state of the art statistical downscaling model, to the best of our knowledge, is the non-parametric kernel regression based model developed by Kannan and Ghosh (2013). On Campbell basin dataset, the existing state of the art model, to the best of our knowledge, is the beta regression based model developed by Mandal, Srivastav and Simonovic (2016). An overview of these models is presented in Fig. 3.

The basic steps involved in the non-parametric kernel regression based statistical downscaling model developed by Kannan and Ghosh (2013) are as follows:

- Standardization of NCEP/NCAR variables is performed.
- Principal component analysis is used to reduce the size of this predictor data.
- kmeans clustering is used to classify days based on precipitation of all stations in a region. The classified days are grouped into three categories—heavy, medium, and low. These three categories are named as precipitation states. This helps to capture the spatial dependency among the stations by considering a common precipitation state for an entire region.
- A decision tree based CART model (Breiman et al. 1984) is developed using the principal components of NCEP/NCAR data and lag-1 NCEP/NCAR data, and the precipitation state of the river basin as input and actual precipitation value of each station as output. The following functional form is assumed for training: $s_t = \mathrm{f}(p_t, p_{t-1}, s_{t-1})$ where s_t is the precipitation state and p_t is the set of predictor variables on tth day and s_{t-1} and p_{t-1} are the precipitation state and the set of predictor variables on $(t-1)$th day.

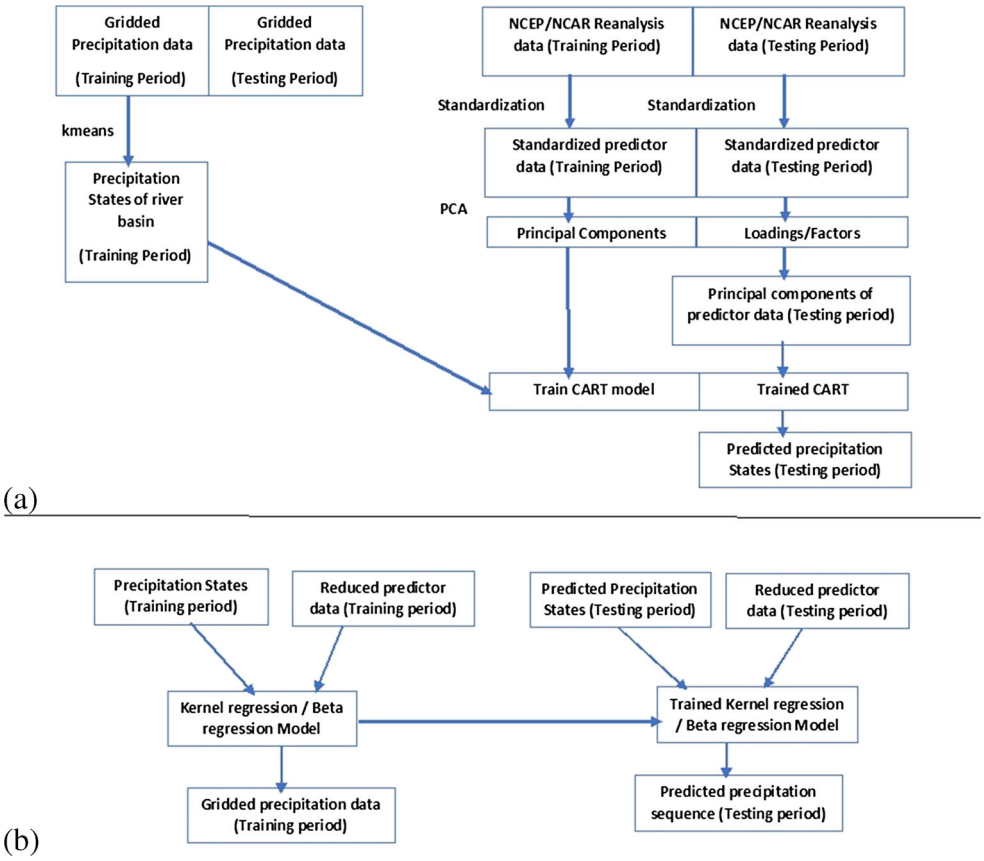


Fig. 3 Overview of the downscaling algorithm proposed by Kannan and Ghosh (2013) and Mandal et al. (2016)

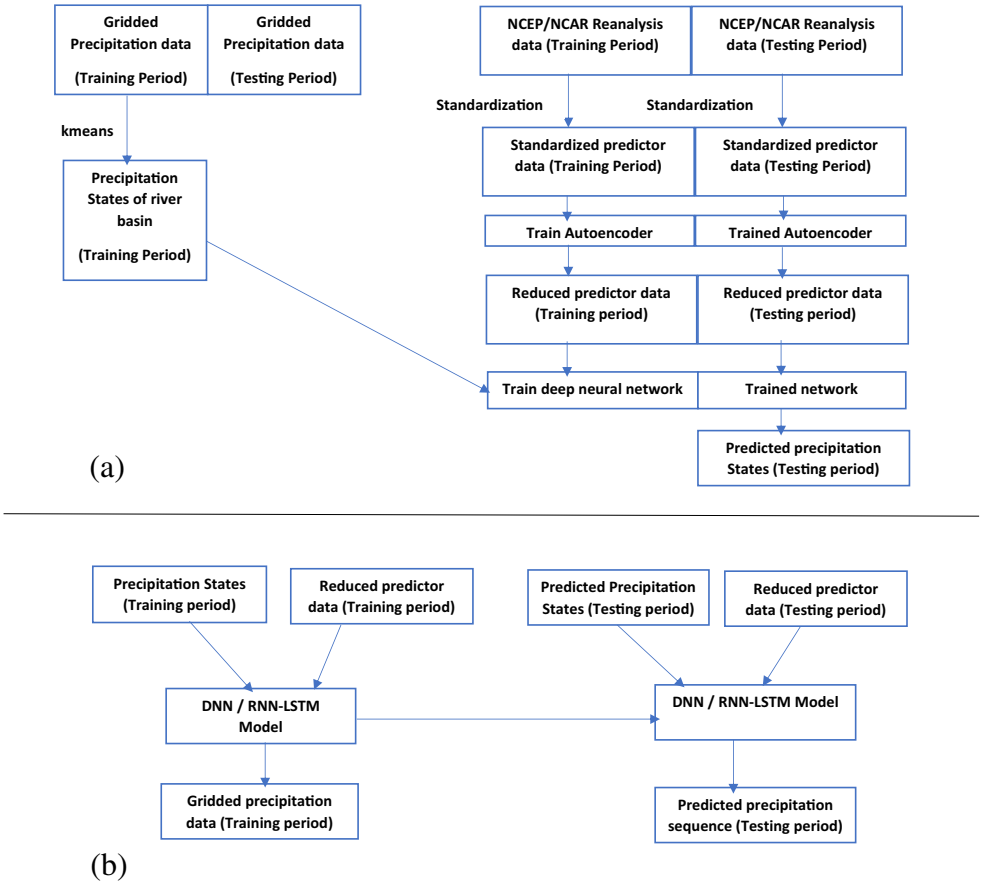
- Next, a non-parametric kernel regression model is used to predict the multisite precipitation. The predictors used to develop this model are current day principal components of NCEP predictor data and current day precipitation state of the river basin and the predictand is the present day precipitation at each station.

In the beta regression-based statistical downscaling methodology developed by Mandal et al. (2016), all the steps discussed above are exactly same except the use of beta regression model instead of kernel regression in the last step.

In the next two paragraphs, we give a brief overview of the LSTM based model developed by us.

The first part [part 1] (Fig. 4a) of developing the model deals with generation of a common precipitation state for capturing the spatial dependencies in the precipitation pattern of a particular region, following the work of Kannan and Ghosh (2013). But, we have incorporated few changes in order to provide better predictability to the model. Unlike PCA which is a linear mapping method for feature reduction, autoencoders (Hinton et al. 2006) capture non-linearity in the data much better. Unlike PCA which is just a linear transformation of the feature space, autoencoders are supervised machine learning models, which can produce meaningful features by capturing the information content in the predictor data properly. Hence, we have used autoencoder instead of PCA to reduce the standardized predictors. Deep neural networks are much more powerful machine learning models than the decision tree based CART model. So, instead of the CART model used previously, we have used a deep neural network with 2 hidden layers. Architectural details of the deep neural network are presented in Section 5.2.1.

In the second part [part 2] (Fig. 4b), we have replaced the previously used regression models like kernel regression and beta regression with neural network based models. The overall basic steps in developing the model are shown in Fig. 3b. As neural networks have been very successful in capturing complex input-output relationships, we investigated their performance in statistical downscaling. We have used recurrent neural networks as they have been proved to be successful in modeling temporal dependencies in various tasks with high accuracy and our hypothesis is that climatic patterns have signatures in the past. The whole methodology has two parts—the first part (Fig. 4a) is concerned with the generation of daily precipitation states (heavy, medium, and low) using a deep neural network (DNN) and the second part (Fig. 4b) deals with the generation of the daily precipitation sequences for a specific location using deep neural network (DNN)/ long short term memory recurrent neural



network (RNN-LSTM) model. The following Sections 5.1 and 5.2 describe the detailed steps for developing the model:

5.1 Generation of daily precipitation states

This is the first part in the development of our model, which partially follows Kannan and Ghosh (2013) and Mandal et al. (2016). It is a weather typing approach. The daily precipitation state, as referenced in literature, represents the approximate range of precipitation in a particular region on a given day. Instead of CART model as used in the previous works, we have used a deep neural network with two layers to build a relationship between the continuous-valued independent variables extracted from NCEP/NCAR and the categorical dependent variable (precipitation in our case). This neural net acts as a classifier and predicts the precipitation state. Unsupervised kmeans algorithm is used along with this deep neural network classifier model in this step.

The overview of this procedure is briefly stated as follows:

i) The NCEP/NCAR predictor variables in the training period are standardized by subtracting mean of the data and dividing by its standard deviation in order to reduce the systematic bias among the variable means and standard deviations.

ii) The NCEP/NCAR variables in the testing period are also standardized.

iii) The k-means clustering method (McQueen 1967) is used to classify the days in the training period into precipitation states (representing the precipitation pattern of the entire region). Number of clusters used is three following Kannan and Ghosh (2013) and Mandal et al. (2016).

iv) A single hidden layer autoencoder is trained on the standardized predictor data in the training period to reduce its dimension and remove multicollinearity from the data.

v) The deep neural network classifier model built using the principal components of NCEP/NCAR data and the precipitation states in the river basin is used to predict the next day rainfall state from the previous day rainfall state using the following function:

$s_t = f(p_t, p_{t-1}, s_{t-1})$ where s_t is the precipitation state and p_t is the set of predictor variables on tth day and s_{t-1} and p_{t-1} are the precipitation state and the set of predictor variables on $(t-1)$th day.

The precipitation states are one hot encoded. The inputs to the neural net are the one hot encoded lag-1 precipitation states and the reduced predictor variables for the same day and previous day and the output is the one hot encoded precipitation state.

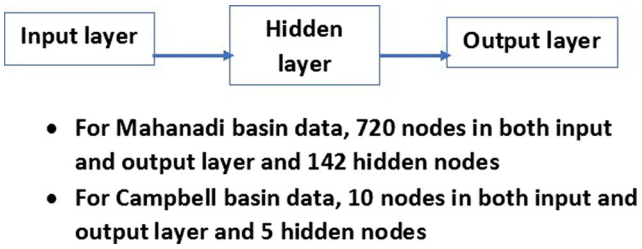


Fig. 5 Basic architecture of autoencoder used in our study

vi) The developed neural classifier is applied to the reduced NCEP/NCAR predictor data in the testing period to predict rainfall states in the river basin in the testing period.

The autoencoder architecture is shown in Fig. 5. For Mahanadi basin data, it has 142 hidden nodes and for Campbell data, it has 5 hidden nodes. Out of 720 dimensions for each input instance, 142 input dimensions of each input sample are extracted using the autoencoder for the Mahanadi region and out of 10 variables for 10 locations for each input instance, 5 input dimensions are extracted using the autoencoder for the Campbell River basin region of Canada.

The neural net classifier's architecture is depicted in Fig. 6. For Mahanadi basin data, it has 142 (reduced same day predictor data) + 142 (reduced lag-1 predictor data) + 3 (one-hot encoded rainfall state) = 287 input nodes, 3 (current day one-hot encoded predicted precipitation state) as output and 150, 70 hidden nodes on hidden layers 1 and 2. Similarly, for Campbell data, $5 + 5 + 3 = 13$ input nodes and 10, 5 output nodes are there in the neural network.

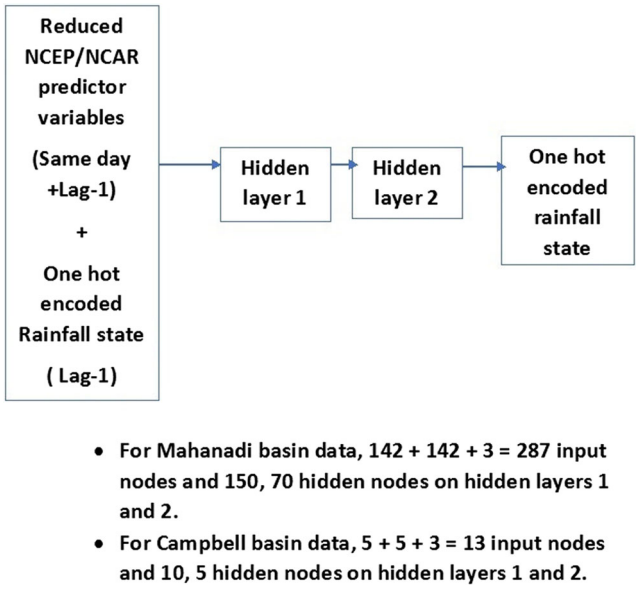


Fig. 6 Basic architecture of deep neural network classifier used in our study

Instead of directly performing regression, the use of the precipitation states as an additional information helps to preserve the spatial correlation in the predictand (here, precipitation). It also helps to give better results in finding the actual precipitation series as it adds an extra information to the input.

5.2 Multisite precipitation generation

This is part 2 of our downscaling model. First, we will discuss the intuition behind the usage of LSTM in this part of our statistical downscaling model. Recurrent Neural Networks are neural networks with loops in them so that data from hidden layer at previous timestep is fed as input to the next step. In theory, RNNs can preserve long-term dependencies at an arbitrarily large historical timeframe. But Hochreiter and Schmidhuber (1997) and Bengio et al. (2015) showed that RNNs suffer from several problems like vanishing gradient and exploding gradient while preserving long-term dependencies. Hence, training RNNs become difficult. The problem arises when there are long-term dependencies with some input features at a considerably earlier timeframe—then the training time and hardware requirements to get the desired efficacy from the system increases exponentially. In order to make the system feasible to train, we need to decide how much earlier data we need to keep and how much we need to dispose. This problem was solved by a modified version of RNNs- Long Short-Term Memory (LSTM) networks. LSTMs consist of a memory cell that can preserve long-term information and a set of gates - input gate and forget gate which is used to control how much information to store and how much to discard. A basic LSTM cell is depicted in Fig. 7.

The recent successes of LSTMs has inspired us to apply this to the statistical downscaling problem which has an

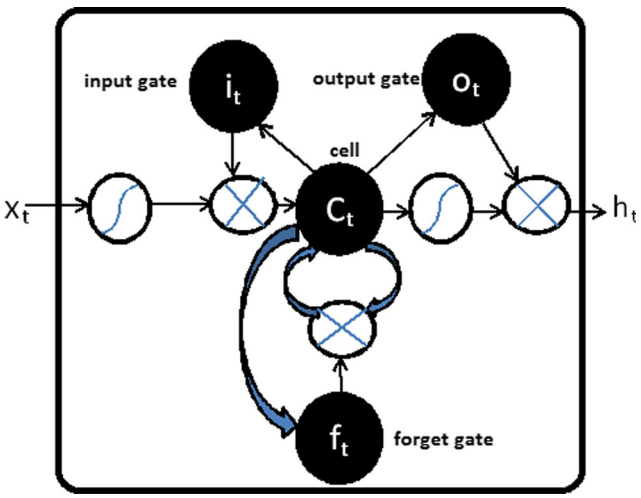


Fig. 7 Basic repeating module in a standard RNN-LSTM

inherent temporal component attached to it. LSTMs, a special variant of recurrent neural networks, are finding renewed interest in the machine learning and the entire scientific community currently as a result of its exceptionally successful applications in a wide variety of applications involving sequential data. LSTMs have achieved state of the art results in sequence related problems like speech recognition (Hinton et al. 2012; Dahl et al. 2012), language modeling (Mikolov et al. 2010), machine translation (Sutskever et al. 2014; Bahdanau et al. 2014), etc. It has also proved to be successful in some of its recent manifestations in the area of weather forecasting (Zaytar and El Amrani 2016).

Our hypothesis is based on the fact that climate data may have signatures at a considerably earlier time. This characteristic can be captured well by LSTM networks.

In this part of developing the downscaling model, we use DNN and LSTM for prediction of actual multisite precipitation series. The inputs to the models are the predicted current day one-hot encoded precipitation states and the current day reduced NCEP/NCAR variables and the output is the predicted current precipitation at each station.

5.2.1 Deep neural network

In the previous literature in statistical downscaling, single-hidden layer neural network was already used. We have used two-layer deep neural network instead, making the model more robust and achieving a better performance. The basic structure of the multilayer neural net used in our study is depicted in Fig. 8.

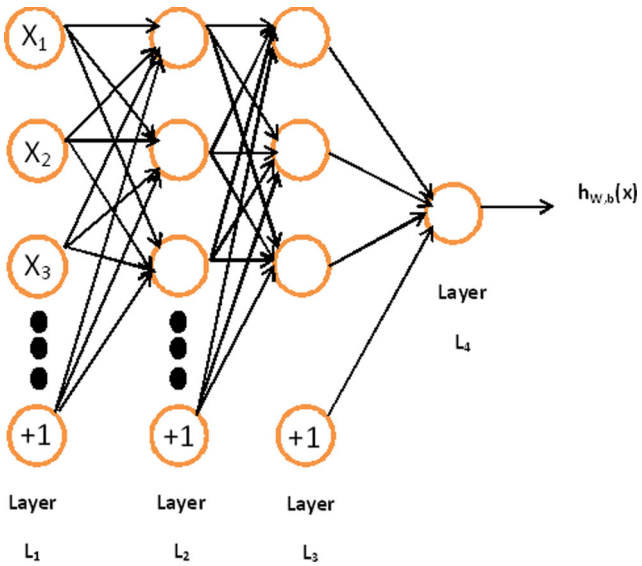


Fig. 8 Multilayer neural network architecture used in our study [L_1: 145 nodes (Mahanadi basin data), 8 nodes (Campbell basin data). L_2: 60 nodes (Mahanadi basin data), 15 nodes (Campbell basin data). L_3: 30 nodes (Mahanadi basin data), 5 nodes (Campbell basin data). L_4: 8 nodes (Mahanadi data), 10 nodes (Campbell data)]

The learning is done using Adaptive Moment Estimation algorithm (Adam), a variant of stochastic gradient descent algorithm. Details of this algorithm can be found in Kingma and Ba (2015). The authors have shown in their paper that this method performs better compared to other versions of stochastic gradient descent algorithm. A multilayer neural network can learn highly complex functions—hence we tested its performance.

The deep neural network model (henceforth referred to as DNN) is constructed using the following function: $R_t = \mathrm{f}(p_t, s_t)$ where R_t is the rainfall value, s_t is the precipitation state and p_t is the set of predictor variables on t^{th} day.

There are $142 + 3 = 145$ input nodes in the neural network for the Mahanadi region data of India and $5 + 3 = 8$ input nodes for the Campbell river basin data as shown in Fig. 4. For the Dataset 1, 15 and 5 nodes are considered in the first and second hidden layers and for the Dataset 2, 60 and 30 nodes are considered in the first and second hidden layers as these configurations give the best possible result. For both cases, number of nodes in the output layer is equal to the number of precipitation stations in the river basin - 10 for Campbell basin data and 8 for Mahanadi basin data. We build a single neural network for predicting precipitation at all locations in a single river basin. tanh activation function is used in the first two activation layers and linear activation function is used in the last activation layer for Dataset 1 and log-sigmoid activation function is used in all the activation ayers for Dataset 2. Adam training function (a variant of stochastic gradient descent algorithm) is used in both cases.

5.2.2 RNN and LSTM

The basic architecture of the LSTM model used in our study and a block diagram for a simplified version of the same are shown in Figs. 9 and 10 respectively. The LSTM model is constructed using the following function: $R_t = f(p_t, s_t, R_{t-1})$ where R_t is the rainfall value, s_t is the

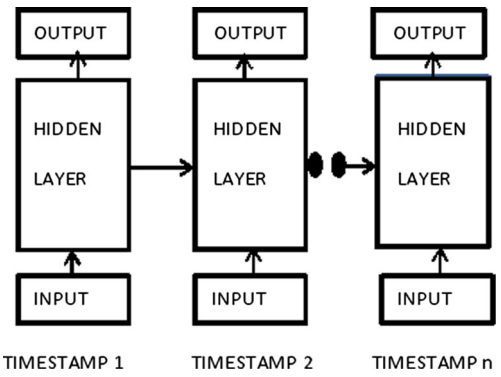


Fig. 9 Basic architecture of RNN-LSTM used in our study [At each timestep, input vector of size 8 for Campbell basin dataset and size 145 for Mahanadi basin dataset and output vectors of size 10 and 8 for Campbell and Mahanadi data are used respectively]

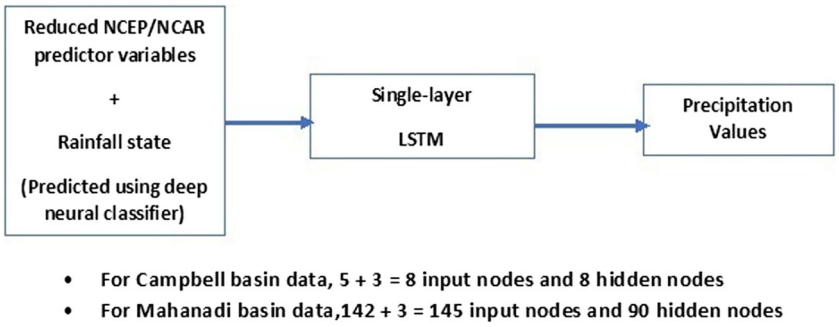


Fig. 10 Block diagram of LSTM model used in our study

precipitation state and p_t is the set of predictor variables on tth day and R_{t-1} is the rainfall value on $(t-1)$th day.

So, there are $142 + 3 = 145$ input nodes in the neural network for the Mahanadi region data of India and $5+3 = 8$ input nodes for the Campbell river basin data. We consider a LSTM with 1 hidden layer and lag-1 feedback loop. For the Dataset 1, 8 hidden nodes are considered and for dataset 2, 90 hidden nodes are considered as these configurations give the optimum result. For both the cases, number of nodes in the output layer is equal to the number of precipitation stations in the river basin - 10 for Campbell basin data and 8 for Mahanadi basin data. We build a single neural network for predicting precipitation at all locations in a particular basin. Linear activation function is used at the final layer for both the datasets. The Adam training function is used in both cases. The details of the Adam method can be found in the work of Kingma and Ba (2015).

Use of the common precipitation state for the entire river basin as an input helps to maintain the spatial correlation pattern in the multisite precipitation sequence.

5.3 Model application

In order to compare our model with their work, we have used precipitation data for the monsoon months of June, July, August, and September (JJAS) for Mahanadi basin data and for Campbell river basin data, we have used precipitation data for the whole year.

For the second part, we use DNN and RNN-LSTM models to generate the daily precipitation series using NCEP/NCAR predictor variables and predicted rainfall state (which we get as output from Part a) of the model) of the same day as inputs.

Both the models are implemented in Theano (Bergstra et al. 2010; Bastien et al. 2012) using Keras library. Each epoch takes about 15–20 s on a multicore CPU machine on an average.

The Pearson product-moment correlation coefficient and mean squared error between the predicted precipitation series and the observed precipitation series for the test period are calculated for both the datasets to test the efficiency of the model.

For the null hypothesis test used to test whether the means of the predicted and the observed precipitation series are similar and for computing the actual means and standard deviation of the predicted data, we denormalize the precipitation dataset.

Both our models give considerable improvement in terms of all metrics considered and the RNN-LSTM performs the best overall. It captures the spatio-temporal variability and extremities in the precipitation data quite well.

6 Results and discussions

The aim of this study is to demonstrate the efficiency of the proposed RNN-LSTM based statistical downscaling model. Using RNN-LSTM model, we compute correlation coefficient, mean squared error, etc. by varying the number of iterations for which the model is run and the best values are considered. Performance evaluation of the model is done based on comparing the following parameters with the historically observed data:

- Temporal mean and standard deviation
- Mean squared error
- Temporal and Spatial cross-correlation
- Basin averaged annual and monthly precipitation
- Skill score of the model
- Ability to downscale extreme events
- Computational scalability of the model

Table 3 Brief description of models used for comparison

Abbreviation	Description
Dataset 1	Precipitation in Campbell Basin
Dataset 2	Precipitation in Mahanadi Basin
BR	Beta regression
KR	Kernel regression
DNN	Deep neural network
RNN-LSTM	Recurrent neural network with long short-term memory
LSTMWS	LSTM without states

6.1 Comparison of statistical characteristics—mean, standard deviation and normalized mean squared error

Some abbreviations used for the ease of our comparative analysis are depicted in Table 3. Means and standard deviation of the observed precipitation data for the test period (1991–2013 for Dataset 1 and 1981–2000 for Dataset 2) are compared with the means and standard deviation of the model-simulated precipitation time series for the same time period. This is shown in Tables 4 and 5 for Dataset 1 and Dataset 2 respectively. Next, Student's *t* test is performed to check whether the means of the predicted and the observed precipitation time series are similar at some significance level (Here, 5% significance level is considered for both datasets). The results for this test are depicted in Tables 4 and 5. We can see from here that for both Dataset 1 and Dataset 2, the null hypothesis at 5% significance level is accepted for all stations for both our DNN and RNN-LSTM

Table 4 Mean and standard deviation of observed and simulated precipitation series (mm) and mean squared error and temporal correlation between them [DATA 1]

	Downscaling location										
	ELK	ERC	GLD	HEB	JHT	QIN	QSM	SAM	SCA	WOL	Average
Mean											
Observed	6.01	5.91	7.29	7	4.56	5.45	4.53	5.5	5.47	6.31	5.80
BR	5.4	7.2	7.37	6.65	3.95	3.49	3.94	4.27	4	7.16	5.34
KR	9.00	9.08	10.50	9.77	6.02	7.71	6.00	7.69	7.49	9.61	8.29
DNN	5.87	6.35	7.44	7.32	3.83	5.11	3.94	4.79	4.71	6.65	5.58
RNN-LSTM	6.22	6.18	7.22	6.59	4.65	5.39	4.62	5.39	5.30	6.65	5.82
LSTMWS	6.95	6.75	7.08	8.21	6.48	4.72	5.48	6.77	6.36	6.65	6.74
Standard deviation											
Observed	10.22	10.22	12.7	12.4	8.19	9.84	8.20	9.63	9.79	10.81	12.24
BR	10.58	11.65	10.5	12.8	7.89	9.62	7.63	10.02	8.8	9.8	9.93
KR	12.9	13.02	15.76	14.93	9.26	11.67	9.25	11.47	11.45	13.87	12.36
DNN	10.23	9.99	11.53	12.03	6.03	8.51	5.97	7.49	7.82	9.96	8.96
RNN-LSTM	9.75	10.48	13.08	12.13	7.09	9.57	9.12	9.21	9.28	10.55	10.02
LSTMWS	8.46	7.58	15.08	14.13	9.32	10.46	10.25	8.14	9.89	11.43	10.77
Mean squared error											
BR	0.0085	0.0054	0.0046	0.0038	0.0035	0.0040	0.0039	0.0036	0.0058	0.0043	0.0047
KR	0.0132	0.0102	0.0100	0.0060	0.0083	0.0056	0.0065	0.0042	0.0128	0.0107	0.0175
DNN	0.0015	0.0020	0.0014	0.0012	0.0029	0.0014	0.0029	0.0017	0.0017	0.0015	0.0018
RNN-LSTM	0.0013	0.0019	0.0014	0.0012	0.0021	0.0012	0.0020	0.0012	0.0012	0.0014	0.0015
LSTMWS	0.0043	0.0057	0.0054	0.0039	0.0065	0.0049	0.0053	0.0041	0.0032	0.0034	0.0047
Correlation coefficient between observed and simulated precipitation series[1991–2013]											
BR	0.6999	0.6804	0.7086	0.6925	0.6312	0.6842	0.6299	0.6997	0.6858	0.6906	0.6802
KR	0.5697	0.4737	0.5389	0.5674	0.5282	0.5678	0.4685	0.6075	0.4480	0.3938	0.5163
DNN	0.7153	0.7993	0.7178	0.7185	0.8631	0.7171	0.7615	0.7216	0.8158	0.8187	0.7649
RNN-LSTM	0.7624	0.8186	0.7501	0.7760	0.8744	0.7468	0.7741	0.7354	0.8222	0.8250	0.7884
LSTMWS	0.6624	0.6153	0.5145	0.6967	0.6456	0.6864	0.6421	0.6344	0.7421	0.7735	0.6613
Hypothesis test results for predicting means of predicted and observed rainfall series											
BR	–	–	X	–	–	–	–	–	–	X	
KR	X	X	X	–	–	–	–	–	–	X	
DNN	–	–	–	–	–	–	–	–	–	–	
RNN-LSTM	–	–	–	–	–	–	–	–	–	–	
LSTMWS	–	–	–	–	–	–	–	–	–	–	

[X]Rejection of null hypothesis

[–]Acceptance of null hypothesis

Table 5 Mean and standard deviation of observed and simulated precipitation series (mm) and mean squared error and temporal correlation between them [DATA 2]

	Downscaling location								
	1	2	3	4	5	6	7	8	Average
Mean									
Observed	8.75	9.76	9.46	8.07	7.20	9.30	9.72	10.77	9.14
BR	9.86	7.67	7.73	7.15	9.59	10.49	8.94	6.68	6.81
KR	9.20	9.21	8.28	8.24	9.11	10.17	11.04	9.98	9.52
DNN	8.41	8.35	8.44	7.22	5.83	9.81	10.94	9.79	9.37
RNN-LSTM	9.22	9.23	11.68	6.92	7.38	9.43	9.54	9.63	9.13
LSTMWS	7.23	6.57	6.44	6.36	4.79	5.98	6.46	7.33	6.39
Standard deviation									
Observed	15.80	19.36	20.85	15.71	12.14	15.63	18.44	15.35	16.66
BR	12.56	18.65	15.5	16.8	12.97	18.62	17.63	16.02	16.09
KR	14.66	16.16	14.18	14.40	13.85	16.60	17.88	18.26	15.75
DNN	14.68	17.15	16.44	15.32	12.83	16.21	20.54	14.79	15.99
RNN-LSTM	16.31	16.89	17.85	14.67	13.44	15.17	19.30	13.35	16.15
LSTMWS	16.31	16.89	17.85	14.67	13.44	15.17	19.30	13.35	16.15
Mean squared error									
BR	0.0080	0.0034	0.0046	0.0035	0.0032	0.0052	0.0069	0.0077	0.0047
KR	0.0109	0.0041	0.0060	0.0042	0.0045	0.0077	0.0092	0.0132	0.0059
DNN	0.0043	0.0027	0.0028	0.0026	0.0025	0.0041	0.0065	0.0068	0.0040
RNN-LSTM	0.0029	0.0025	0.0016	0.0035	0.0019	0.0022	0.0031	0.0060	0.0029
LSTMWS	0.0085	0.0045	0.0036	0.0049	0.0038	0.0042	0.0041	0.0054	0.0097
Correlation coefficient between observed and simulated precipitation series [1981–2000]									
BR	0.6989	0.6544	0.5784	0.6827	0.6312	0.6732	0.6299	0.5997	0.6435
KR	0.7087	0.6453	0.5809	0.6624	0.7266	0.7225	0.6686	0.6708	0.6732
DNN	0.7183	0.6643	0.5978	0.7085	0.7331	0.7235	0.6705	0.6786	0.6868
RNN-LSTM	0.7476	0.6720	0.6174	0.7237	0.7343	0.7265	0.6760	0.6994	0.6996
LSTMWS	0.6537	0.5432	0.5899	0.6018	0.6245	0.5463	0.4965	0.5873	0.5804
Hypothesis test results for predicting means of predicted and observed rainfall series									
BR	–	–	–	–	X	–	–	–	
KR	–	X	–	–	X	–	–	–	
DNN	–	–	–	–	–	–	–	–	
RNN-LSTM	–	–	–	–	–	–	–	–	
LSTMWS	–	–	–	–	–	–	–	–	

[X]Rejection of null hypothesis

[–]Acceptance of null hypothesis

models. This has not been achieved by any other method like KR or BR till now.

Mean squared error is a very good measure of accuracy for any regression problem. The same thing applies to our problem as basically we are trying to solve a regression problem. Normalized mean squared error is mean squared error in [0,1]. The RNN-LSTM model gives significant improvement in the normalized mean squared error compared to the BR, KR, and DNN models. These results are given in Tables 4 and 5.

6.2 Temporal and spatial correlation

The temporal correlation between the observed and predicted precipitation series for all the locations for Dataset 1 and Dataset 2 are depicted in Tables 4 and 5. Figures 11

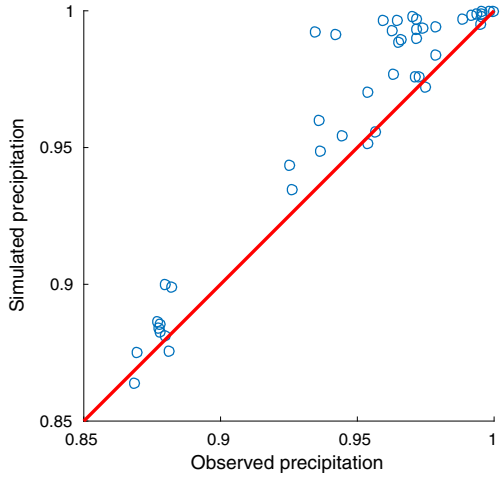


Fig. 11 Interstation correlation coefficient for RNN-LSTM model (Campbell basin data)

and 12 represents the scatter plot of interstation correlation coefficients computed from model-simulated daily precipitation series and observed precipitation series for all station pairs using the RNN-LSTM model we have used for statistical downscaling. From the results and the plot, it can be concluded that the temporal and spatial variability of precipitation is modeled well enough by our proposed RNN-LSTM model for both Dataset 1 and Dataset 2.

6.3 Basin averaged annual and monthly precipitation

The basin averaged annual and monthly precipitation shows better correlation with the observed precipitation in the testing period compared to the existing methods, which is depicted in Figs. 13 and 14. For Campbell basin dataset,

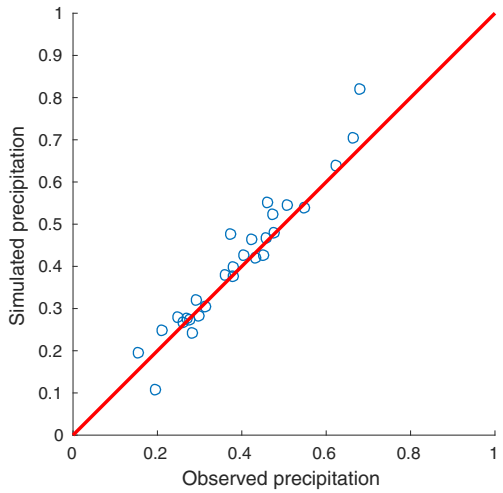
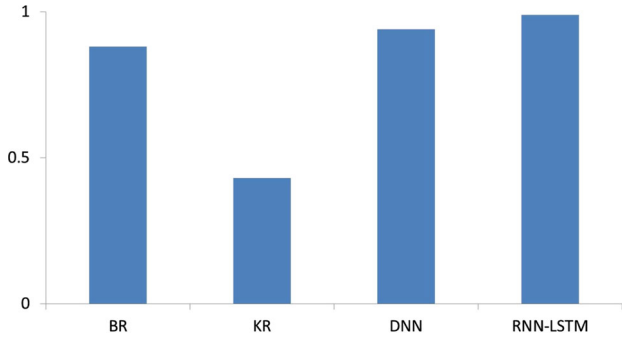
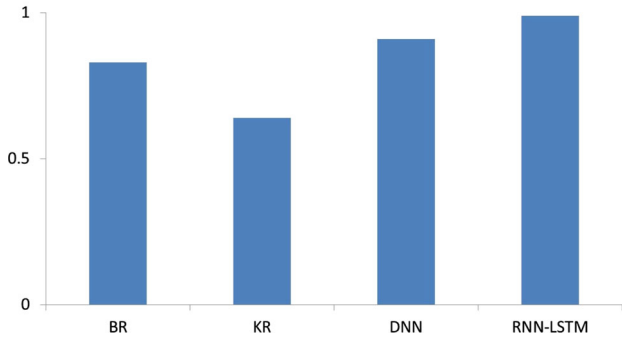


Fig. 12 Interstation correlation coefficient for RNN-LSTM model (Mahanadi basin data)



(a) Correlation between annual predicted and observed precipitation



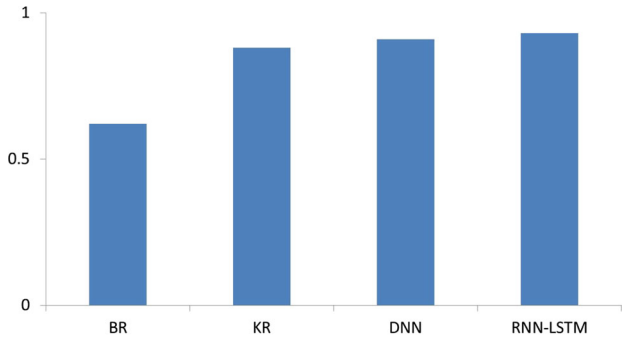
(b) Correlation between monthly predicted and observed precipitation

Fig. 13 Temporal correlation for spatially averaged rainfall over Campbell river basin (1991–2013)

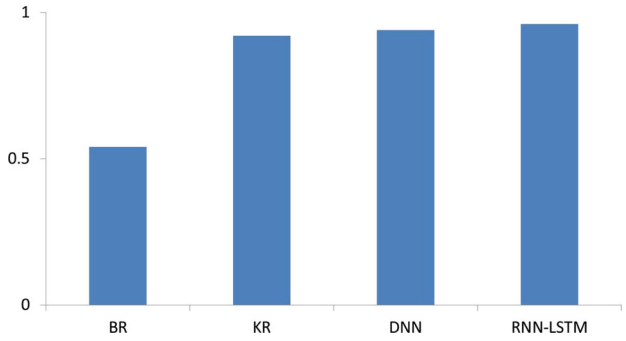
the DNN and RNN-LSTM models shows correlations of 0.94 and 0.99 for annual precipitation and 0.91 and 0.93 for monthly precipitation whereas BR and KR models shows correlations of 0.88 and 0.43 for annual precipitation and 0.83 and 0.64 for monthly precipitation. For Mahanadi basin dataset, the DNN and RNN-LSTM models shows correlations of 0.90 and 0.93 for annual precipitation and 0.94 and 0.96 for monthly precipitation whereas BR and KR models shows correlations of 0.62 and 0.88 for annual precipitation and 0.54 and 0.92 for monthly precipitation.

6.4 Skill score of the model

We use the skill score presented by Perkins et al. (2007) that measures how similar two probability density functions are from a range of 0 to 1 where 1 corresponds to identical distributions. It is one of the effective measures to capture the ability of the model to properly simulate the output data distribution. The skill scores for predictions of different models are compared with the LSTM model for each station in Dataset 1 and Dataset 2 in Tables 6 and 7 respectively. LSTM performs better than most of the models on most stations, as seen from the results depicted in these tables.



(a) Correlation between annual predicted and observed precipitation



(b) Correlation between monthly predicted and observed precipitation

Fig. 14 Temporal correlation for spatially averaged rainfall over Mahanadi river basin (1981–2000)

6.5 Ability to downscale extreme events

The ability of a statistical downscaling model to estimate extreme events is assessed using some climate indices. Four such climate indices from ClimDEX (http://clim-dex.org) has been chosen following Burger et al. (2011) and Vandal et al. (2017) to encompass a range of extremes. These are:

- CWD—Maximum length of wet spell, i.e., maximum no. of consecutive wet days with precipitation $\geq$ 1mm.
- R20—Maximum no. of consecutive days with precipitation $\geq$ 20mm.
- RX5day—Monthly maximum consecutive 5 day precipitation.
- SDII—Daily intensity index = Annual total / no. of days with precipitation $\geq$ 1m.

These metrics are computed on both observed and downscaled precipitation for both datasets. Then we find a correlation between these and also use a skill score to ascertain the capability of the model in reproducing the statistical distributions properly. These results are depicted for Dataset 1 and Dataset 2 in Tables 8 and 9 respectively. LSTM emerges as a clear winner here too.

Table 7 Skill score [Dataset 2]

	Downscaling location								
	1	2	3	4	5	6	7	8	Average
BR	0.84	0.72	0.59	0.61	0.67	0.75	0.78	0.79	0.75
KR	0.75	0.77	0.71	0.74	0.82	0.69	0.75	0.83	0.79
DNN	0.87	0.76	0.74	0.62	0.85	0.77	0.77	0.82	0.81
RNN-LSTM	0.97	0.79	0.96	0.80	0.81	0.80	0.70	0.84	0.85

6.6 Computational scalability of the model

The average time taken to train the LSTM model is 8 s per epoch for Dataset 1 and 15 s per epoch for Dataset 2 on a 12-core Intel Xeon CPU. The model converges within 150 epochs. So, the average total training time taken is 20 min for Dataset 1 and 37.5 min for Dataset 2. It will speed up significantly if a GPU is used. In comparison, the BR, KR AND DNN methods, on an average, take 30, 45, and 15 min for Dataset 1 and 36, 20, and 27 min for Dataset 2 for 150 epochs for all stations. So, the model scales quite well. For large datasets also, the time taken will be negligible if only a single GPU is used.

6.7 Uncertainty analysis of the model

It should be ascertained that the downscaled outputs can represent the current precipitation distribution reasonably well. Only then, we can have full confidence and reliability on the climate scenarios downscaled from the GCM outputs. In other words, the ability of the outputs predicted from

Table 6 Skill score [Dataset 1]

	Downscaling location										
	ELK	ERC	GLD	HEB	JHT	QIN	QSM	SAM	SCA	WOL	Average
Mean											
BR	0.64	0.61	0.67	0.59	0.72	0.65	0.72	0.66	0.57	0.82	0.67
KR	0.67	0.60	0.62	0.62	0.55	0.46	0.52	0.62	0.49	0.73	0.59
DNN	0.77	0.71	0.77	0.79	0.76	0.69	0.77	0.72	0.71	0.92	0.73
RNN-LSTM	0.91	0.91	0.89	0.90	0.94	0.93	0.94	0.93	0.93	0.91	0.92

Table 8 Correlation between climate index scores of observed and predicted precipitation on a daily scale [Dataset 1]

Climate indices				
	CWD	R20	RX5day	SDII
BR	0.84	0.76	0.82	0.71
KR	0.78	0.85	0.75	0.66
DNN	0.93	0.93	0.89	0.84
RNN-LSTM	0.94	0.93	0.97	0.92

the downscaling model to properly represent the baseline climate is a necessary condition to have good amount of confidence on the reliability of the climate change anomalies computed from the GCM scenario runs. Hence, the primary objective of the uncertainty analysis is to measure the performance of the downscaling method in reproducing the mean value and standard deviation of observed meteorological variables properly when provided with climate predictors for the baseline period.

Following Dibike et al. (2007), two complementary methods have been used in our study to analyze the uncertainty of the statistical downscaling model's output-hypothesis testing and confidence intervals.

The hypothesis testing method used in this study is the Wilcoxon Signed Rank test (Wilcoxon 1945), is a non-parametric method used to test the null hypothesis of no median difference in a pair of samples. It involves calculating the test statistic and p value for the null hypothesis, followed by either accepting or rejecting the hypothesis at a given significance levels based on the p value. The p value is the probability that the null hypothesis is wrongly rejected when it is actually true. (type 1 error). A s value of 0.05 which corresponds to 5% significance level is used in this study. Small p values suggest that the null hypothesis is unlikely to be true and the null hypothesis is rejected when $p < 0.05$. In this study, the analysis was performed by comparing observed precipitation with the corresponding downscaled precipitation. The hypothesis tests are repeated 100 times for monthly precipitation of each station from which we calculate the rejection percentage of the simulated precipitation values.

Table 9 Correlation between climate index scores of observed and predicted precipitation on a daily scale [Dataset 2]

Climate Indices				
	CWD	R20	RX5day	SDII
BR	0.69	0.65	0.73	0.57
KR	0.70	0.66	0.71	0.59
DNN	0.81	0.82	0.86	0.74
RNN-LSTM	0.86	0.81	0.89	0.79

A confidence interval is an estimate of uncertainty regarding the true value of a statistic. Sampling with replacement is used to estimate confidence intervals for the statistics of a distribution. In this analysis, Bootstrap simulation (Efron and Tibshirani 1993), a non-parametric technique is used to estimate confidence intervals. Bootstrapping is a method used to generate a pseudo population of a test statistic by re-sampling from the original data set. The bootstrap method takes random samples, known as pseudosamples, with replacement from the original one repeatedly. The statistics in question (mean values and standard deviations in our case) is then evaluated for each pseudo-sample. Then, each time series of observed precipitation is paired with each of the corresponding downscaled precipitation taken from the 100 simulations. 1000 boot strap samples are generated from each such pairs to calculate two sets of statistic for each of the bootstrap sample pairs—the differences between the mean and the differences between the standard deviation values of the observed and simulated variables. Then the bootstrap percentile method is used to calculate the confidence interval. It involves ranking the one thousand statistics calculated from the bootstrap samples followed by selecting the statistic corresponding to the appropriate percentile (5th percentile for the lower confidence limit and 95th percentile for the upper confidence limit) for the confidence interval required (90% in our case). This procedure is repeated for each of the one hundred simulation outputs from our LSTM model. The overall upper bound and lower bound of the statistic's 90% confidence interval is then calculated by averaging the upper bound and lower bounds of the hundred simulations, respectively. All the above steps are repeated for both the datasets. The ideal simulation result is identified as the one with the smallest confidence interval and which includes zero between its upper and lower confidence limits.

Our LSTM model performs reasonably well on both of these measures and hence can be relied upon to a large extent for future climate scenario projections.

The Wilcoxon signed rank hypothesis test results for testing difference in means of both datasets on a daily scale are presented in Tables 10 and 11.

Table 10 Error due to exclusion of each predictor variable [Dataset 1]

Variable excluded	Average mean squared error	Average temporal correlation
None	0.0015	0.7884
Tmax	0.0018	0.6874
Tmin	0.0012	0.6375
mslp	0.0017	0.6997
Uwind	0.0010	0.5975
Vwind	0.0025	0.5597

Table 11 Error due to exclusion of each predictor variable [Dataset 2]

Variable excluded	Average mean squared error	Average temporal correlation
None	0.0029	0.6996
Air temperature	0.0035	0.6547
Mean sea level pressure	0.0037	0.6386
Specific humidity	0.0045	0.5795
Horizontal wind velocity	0.0046	0.5866
Vertical wind velocity	0.0037	0.6674

The confidence interval test results for difference between means and standard deviation of predicted and observed precipitation for station WOL in Dataset 1 and station id 1 in Dataset 2 on a monthly scale are presented in Figs. 15, 16, 17, and 18 respectively.

In order to understand how much significance each predictor variable individually holds in precipitation prediction, we select each predictor variable, individually remove them from the model input one at a time and run the models. The corresponding mean squared error and correlation between predicted and observed precipitation are compared with the original model's mean squared error and correlation. This analysis is preformed for both Dataset 1 and Dataset 2 and are demonstrated in Tables 10, 11, 12, and 13 respectively.

7 Key contributions

We have used an autoencoder coupled deep neural network model for predicting the precipitation state, the weather

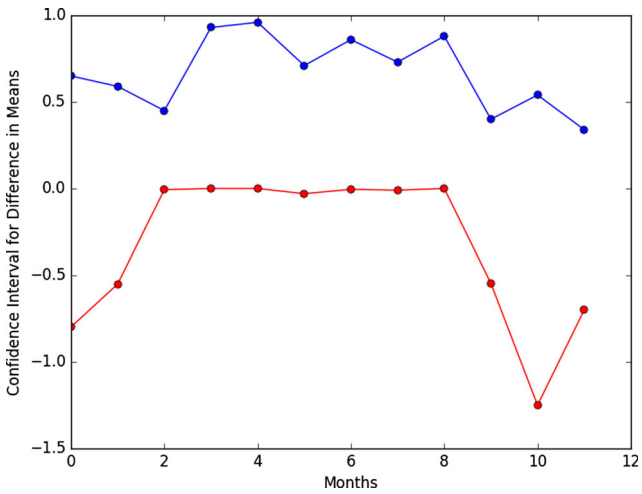


Fig. 15 90% confidence interval for rejected percentage in Wilcoxon signed test that the difference in means between observed and downscaled precipitation is zero for station WOL in Campbell dataset

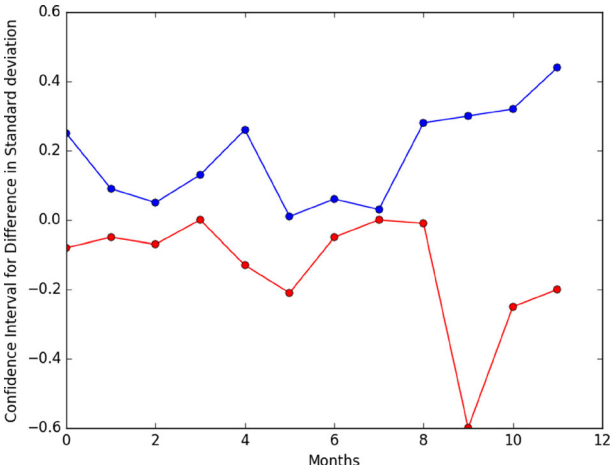


Fig. 16 90% confidence interval for rejected percentage in Wilcoxon signed test that the difference in standard deviation between observed and downscaled precipitation is zero for station WOL in Campbell dataset

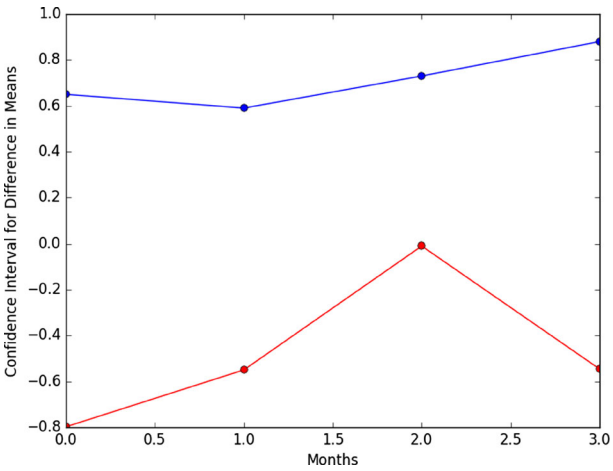


Fig. 17 90% confidence interval for rejected percentage in Wilcoxon signed test that the difference in means between observed and downscaled precipitation is zero for station id 1 in Mahanadi dataset

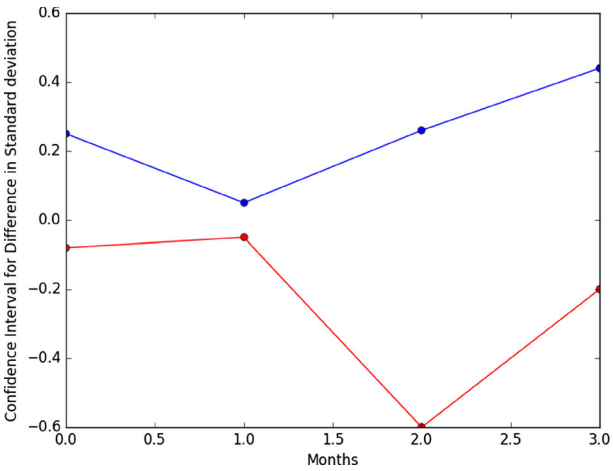


Fig. 18 90% confidence interval for rejected percentage in Wilcoxon signed test that the difference in standard deviation between observed and downscaled precipitation is zero for station id 1 in Mahanadi dataset

Table 12 Rejected percentage of null hypothesis (median difference in observed and downscaled precipitation is zero) in Wilcoxon signed rank test at 5% significance level [Dataset 1]

	Downscaling location									
	ELK	ERC	GLD	HEB	JHT	QIN	QSM	SAM	SCA	WOL
BR	5	0.12	1.7	0.02	0	2.1	0	0.23	0.19	0
KR	1.04	2.15	0.24	1.03	0	0	2.09	0.18	0.12	0
DNN	0	0.07	0.09	0	0	0	0	0.05	0.07	0.01
RNN-LSTM	0	0.05	0.04	0	0	0	0	0.03	0.07	0

typing part of the entire algorithm. This shows superior performance over the existing PCA + CART approach of Kannan and Ghosh. While the PCA + CART approach gave 63.81 accuracy in prediction of precipitation states for Mahanadi basin, our neural network model gives 75.2% accuracy. For the Campbell basin data, the CART model gave 73.5% accuracy while the proposed neural network model gives 82.4% accuracy. Autoencoders have been used for the first time in this task, to the best of our knowledge.

The LSTM based model used for predicting the actual precipitation series is also the first of its kind, to be used in a statistical downscaling task, as per our knowledge. It captures the temporal aspect of the statistical downscaling problem well enough causing a considerable improvement in the performance, measured in terms of metrics such as mean squared error and cross-correlation.

Another important contribution of this work is the usage of a single neural network model for predicting multisite precipitation. All the previous works like Kannan and Ghosh's non-parametric kernel regression based model, Mandal, Srivastav and Simonovic's beta regression based model used separately trained models for each station while predicting multisite precipitation. This is the first ever work, to the best of our knowledge, where a single model predicts multisite precipitation, thereby speeding up the entire training process and that too, with an improved level of accuracy.

Table 13 Rejected percentage of null hypothesis (median difference in observed and downscaled precipitation is zero) in Wilcoxon signed rank test at 5% significance level [Dataset 2]

	Downscaling location							
	1	2	3	4	5	6	7	8
BR	2.18	1.27	1.36	0.20	1.17	0.15	0.09	0.07
KR	2.23	0.35	2.27	0.16	0.24	1.19	0.15	0.09
DNN	0.12	0.23	0.25	0.15	0.04	0.07	0.16	0.08
RNN-LSTM	0.11	0.21	0.15	0.05	0.09	0.06	0.13	0.07

8 Conclusion and future scope of work

We see that the RNN-LSTM model shows satisfactory improvements over the existing perfect prognosis statistical downscaling techniques in capturing the temporal correlation conforming with our hypothesis. On Campbell basin dataset, it provides significant improvements of around 16% in correlation and 68% in mean squared error over the existing state of the art beta regression model.

The spatial correlation between the stations is captured satisfactorily by our model

On the Campbell river basin dataset, the model gives better results and shows more improvement than the Mahanadi basin dataset, most probably due to the use of actual station precipitation data which was available for that region.

On the Mahanadi river basin dataset also, the model performs well enough but not as good as the Campbell river basin dataset. The most probable reason is the lack of actual station precipitation data and presence of highly interpolated data.

There is a lot of scope to extend our present work. It is possible to extend the process to GCMs and see how well the model works on them and what predictions it gives and whether they are at par with the predictions given by the other models used on this data previously. We can extract the principal components of the data by other methods. We can also use some other model or an ensemble of different models and find out if there is an improvement in prediction accuracy, especially in the Mahanadi region.

Our study shows that the rainfall sequence has long term dependencies with long-term predictor variables. That is probably why our model performs better than the other methods, which we have compared with where the authors had used short-term dependencies on predictors, mainly lag-1.

Acknowledgements We would like to thank Subimal Ghosh of IIT Bombay and Kannan Shanmugham of IBM India Pvt. Ltd., Bangalore, India for sharing the codes and data for their work on kernel regression based statistical downscaling and Slobodan Simonovic and Sohom Mandal of the University of Western Ontario for sharing the codes and

data for their work on beta regression based statistical downscaling. We would also like to thank Ministry of Human Resource Development, India and Indian Institute of Technology, Kharagpur for funding our work as a part of the project "Artificial Intelligence for Societal needs" and the Indian Meteorological Department for providing us the Indian rainfall data.

References

Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. ICLR 2015

Bardossy A, Plate EJ (1992) Space-time model for daily rainfall using atmospheric circulation patterns. Water Resour Res 28(5):1247–1259

Bardossy A, Duckstein L, Bogardi I (1995) Fuzzy rule-based classification of atmospheric circulation patterns. Int J Climatol 15:1087–1097

Bastien F, Lamblin P, Pascanu R, Bergstra J, Goodfellow I, Bergeron A, Bouchard N, Warde-Farley D, Bengio Y (2012) Theano: new features and speed improvements. Deep Learning and Unsupervised Feature Learning, NIPS 2012 Workshop

Bengio Y, Goodfellow I, Courville A (2015) Deep learning. Book in Preparation for MIT Press

Bergstra J, Breuleux O, Bastien F, Lamblin P, Pascanu R, Desjardins G, Turian J, Warde-Farley D, Bengio Y (2010) Theano: a CPU and GPU math expression compiler. Scipy, vol. 4, p. 3. Austin

Bi EG, Gachon P, Vrac M, Monette F (2015) Which downscaled rainfall data for climate change impact studies in urban areas? Review of current approaches and trends. Theoretical and Applied Climatol. Springer

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Brooks/Cole Publishing, Monterey

Burger G, Murdock TQ, Werner AT, Sobie SR, Cannon AJ (2011) Downscaling extremes - an intercomparison of multiple statistical methods for present climate. J Clim 25:4366–4388

Christensen JH et al (2007) Climate change 2007: the physical science basis-contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge Univ. Press, Cambridge

Conway D, Jones PD (1998) The use of weather types and air flow indices for GCM downscaling. J Hydrol 212/213:348–361

Corte-Real J, Zhang X, Wang X (1995) Downscaling GCM information to regional scales: A non-parametric multivariate regression approach. Clim Dyn 11:413–424

Dahl G, Yu D, Deng L, Acero A (2012) Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. IEEE Trans Audio Speech Lang Process 20(1):30–42

Dibike YB, Gacon P, Hilaire AS, Ouarda T, Nguyen V (2007) Uncertainty analysis of statistically downscaled temperature and precipitation regimes in Northern Canada. Theoretical and Applied Climatology. Springer

Dobson AJ (2001) An introduction to generalized linear models, 2nd edn. Chapman and Hall, London

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman & Hall/CRC

Gope S, Sarkar S, Mitra P, Ghosh S (2016) Early prediction of extreme rainfall events: a deep learning approach. ICDM 2016:154–167

Hanssen-Bauer I, Achberger C, Benestad R, Chen D, Forland E (2005) Statistical downscaling of climate scenarios over Scandinavia. Clim Res 29(3):255–268

Hastie TJ, Tibshirani RJ (1990) Generalized additive models. Chapman and Hall, London

Hay LE, McCabe GJ, Wolock DM, Ayers MA (1991) Simulation of precipitation by weather type analysis. Water Resour Res 27(4):493–501

Haylock MR, Cawley GC, Harpham C, Wilby RL, Goodess CM (2006) Downscaling heavy precipitation over the UK: a comparison of dynamical and statistical methods and their future scenarios. Int J Climatol 26:1397–1415

Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. SCIENCE, vol. 13, 2006

Hinton G, Deng L, Yu D, Dahl GE, Mohamed A, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN et al (2012) Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. Signal Process Mag IEEE 29(6):82–97

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comp 9(8):1735–1780

Hughes JP, Guttorp P (1994) A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. Water Resour Res 30(5):1535–1546

Hughes JP, Lettenmair DP, Guttorp P (1993) A stochastic approach for assessing the effect of changes in synoptic circulation patterns on gauge precipitation. Water Resour Res 29(10):3303–3315

Hughes JP, Guttorp P, Charles SP (1999) A non-homogeneous hidden Markov Model for precipitation occurence. J R Stat Soc Ser C Appl Stat 48(1):15–30

Hutchinson MF, Xu T (2013) Anusplin Version 4. 4 User Guide, 52p

Intergovernmental Panel on Climate Change - Task Group on Scenarios for Climate Impact Assessment (1999) Guidelines on the use of scenario data for climate impact and adaptation assessment. Version 1, 69 pp

Johnson F, Sharma A (2009) Measurements of GCM skill in predicting variables relevant for hydroclimatological variables. J Clim 22:4373–4382. https://doi.org/10.1175/2009JCLI2681.1.

Kalnay E et al. (1996) The NCEP/NCAR 40-years reanalysis project. Bull Am Meteorol Soc 77(3):437–471

Kannan S, Ghosh S (2011) Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output. Stochastic Environ Res Risk Assess 25:457–474. https://doi.org/10.1007/s00477-010-0415-y

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

Karl TR, Wang WC, Schlesinger ME, Knight RW, Portman D (1990) A method of relating general circulation model simulated climate to observed local climate. Part I: seasonal statistics. J Clim 3:1053–1079

Kingma D, Ba J (2015) Adam: a method for stochastic optimization. ICLR

Liu JN, Hu Y, You JJ, Chan PW (2014) Deep neural network based feature representation for weather forecasting. In: Proceedings on the international conference on artificial intelligence (ICAI). The steering committee of the world congress in computer science, computer engineering and applied computing (WorldComp), 2014, p 1

Mandal S, Srivastav RK, Simonovic SP (2016) Use of beta regression for statistical downscaling of precipitation in the Campbell River basin, British Columbia. J Hydrol 538:49–62

McQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley symposium on mathematical statistics and probability, vol 1. Univ. of Calif. Press, Berkeley, pp 282–297

Mikolov T, Karafiat M, Burget L, Khudanpur S (2010) Recurrent neural network based language model. INTERSPEECH 2010

Murphy JM (1999) An evaluation of statistical and dynamical techniques for downscaling local climate. J Clim 12(8):2256–2284

Perkins SE, Pitman AJ, Holbrook NJ, McAneney J (2007) Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. J Clim 20(17):4356–4376

Prudhomme C, Reynard N, Crooks S (2002) Downscaling of global climate models for flood frequency analysis: where are we now? Hydrol Processes 16:1137–1150

Raje D, Mujumdar PP (2009) A conditional random field-based downscaling method for assessment of climate change impact on multisite daily precipitation in the Mahanadi basin. Water Resour Res 45:W10404. https://doi.org/10.1029/2008WR007487

Rajeevan M, Bhate J, Kale JD, Lal B (2006) High resolution daily gridded rainfall data for the Indian region: analysis of break and active monsoon spells. Curr Sci 91(3):296–306

Richardson CW (1981) Stochastic precipitation of daily precipitation, temperature, and solar radiation. Water Resour Res 17(1):182–190

Richardson CW, Wright DA (1984) A model for generating daily weather variables, USDA Agric. Res. Serv. Rep. ARS-8, US Dep. Agric. Res. Serv. Temple, Tex

Sharma A (2000) Seasonal to interannual rainfall probabilistic forecasts for improved water supply management. Part 1: a strategy for system predictor identification. J Hydrol 239:249–258

Shephard D (1968) A two-dimensional interpolation function for irregularly-spaced data. In: Proceedings of the 1968 23rd ACM national conference, (ACM'68), vol 517. ACM, New York, p 524, https://doi.org/10.1145/800186.810616

Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. Adv Neural Inf Process Syst 3104–3112

Vandal T, Kodra E, Ganguly S, Michaelis A, Nemani R, Ganguly AR (2017) DeepSD: generating high resolution climate change projecions through single image super-resolution. KDD

von Storch H (1995) Inconsistencies at the interface of climate impact studies and global climate research. Meteorol Z 4 NF:72–80

von Storch H (1999) The global and regional climate system. In: von Storch H, Floscr G (eds) Anthropogenic climate change. Springer, Berlin, pp 3–36

von Storch H, Zorita E, Cubash U (1993) Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. J Clim 6:1161–1171

Wetterhall F, Halldin S, Xu C (2005) Statistical precipitation downscaling in central Sweden with the analogue method. J Hydrol 306:174–190

Wilcoxon F (1945) Individual comparisons by ranking methods. Biometrics 1:80–83

Wilby R, Hassan H, Hanaki K (1998) Statistical downscaling of hydrometeorological variables using general circulation model output. J Hydrol 205(1–2):1–19

Wilby RL, Wigley TML (1997) Downscaling general circulation model output: a review of methods and limitations. Prog Phys Geogr 21:530–548

Wilby RL, Hey LE, Leavesly GH (1999) A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River Basin, Colorado. J Hydrol 225:67–91

Wilks DS (1998) Multisite generalization of a daily stochastic precipitation generation model. J Hydrol 210:178–191

Wilks DS (1999) Multisite downscaling of daily precipitation with a stochastic weather generator. Clim Res 11:125–136

Wilks DS (2006) Statistical methods in the atmospheric sciences, 2nd edn. Academic, Amsterdam, p 627

Yee TW, Wild CJ (1996) Vector generalized additive models. J R Stat Soc Ser B 58:481–493

Yang C, Chandler RE, Isham VS, Wheater HS (2005) Spatial-temporal rainfall simulation using generalized linear models. Water Resor Res 41:W11415. https://doi.org/10.1029/2004WR003739

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