

Article

Modified Approach to Reduce GCM Bias in Downscaled Precipitation: A Study in Ganga River Basin

Chetan Sharma ¹, Chandra Shekhar Prasad Ojha ¹, Anoop Kumar Shukla ¹, Quoc Bao Pham ², Nguyen Thi Thuy Linh ^{2,3}, Chow Ming Fai ⁴, Ho Huu Loc ^{5,6,*} and Tran Duc Dung ⁷

¹ Civil Engineering Department, Indian Institute of Technology, Roorkee 247667, India;

chetan.cvl@gmail.com (C.S.); cspojha1962@gmail.com (C.S.P.O.); anoopgeomatics@gmail.com (A.K.S.)

² Department of Hydraulic and Ocean Engineering, National Cheng-Kung University, Tainan 701, Taiwan; pbquoc92@gmail.com (Q.B.P.); linhntt@tlu.edu.vn (N.T.T.L.)

³ Faculty of Water Resource Engineering, Thuyloi University, Hanoi 100000, Vietnam

⁴ Civil Engineering Department, college of Engineering, Universiti Tenaga Nasional (UNITEN), Kajang 43000, Malaysia; chowmf@uniten.edu.my

⁵ Faculty of Environmental and Food Engineering, Nguyen Tat Thanh University, Ho Chi Minh City 700000, Vietnam

⁶ National Institute of Education, 1, Nanyang Walk, Singapore 637616, Singapore

⁷ Center of Water Management and Climate Change, Vietnam National University Ho Chi Minh City, Ho Chi Minh City 700000, Vietnam; dungtranducvn@yahoo.com

* Correspondence: huuloc20686@gmail.com

Received: 14 July 2019; Accepted: 15 August 2019; Published: 8 October 2019



Abstract: Reanalysis data is widely used to develop predictor-predictand models, which are further used to downscale coarse gridded general circulation models (GCM) data at a local scale. However, large variability in the downscaled product using different GCMs is still a big challenge. The first objective of this study was to assess the performance of reanalysis data to downscale precipitation using different GCMs. High bias in downscaled precipitation was observed using different GCMs, so a different downscaling approach is proposed in which historical data of GCM was used to develop a predictor-predictand model. The earlier approach is termed “*Re-Obs*” and the proposed approach as “*GCM-Obs*”. Both models were assessed using mathematical derivation and generated synthetic series. The intermodal bias in different GCMs downscaled precipitation using *Re-Obs* and *GCM-Obs* model was also checked. Coupled Model Inter-comparison Project-5 (CMIP5) data of ten different GCMs was used to downscale precipitation in different urbanized, rural, and forest regions in the Ganga river basin. Different measures were used to represent the relative performances of one downscaling approach over other approach in terms of closeness of downscaled precipitation with observed precipitation and reduction of bias using different GCMs. The effect of GCM spatial resolution in downscaling was also checked. The model performance, convergence, and skill score were computed to assess the ability of *GCM-Obs* and *Re-Obs* models. The proposed *GCM-Obs* model was found better than *Re-Obs* model to statistically downscale GCM. It was observed that *GCM-Obs* model was able to reduce *GCM-Observed* and *GCM-GCM* bias in the downscaled precipitation in the Ganga river basin.

Keywords: Ganga river basin; precipitation; general circulation models (GCM); downscaling; GCM bias; model performance

1. Introduction

General circulation models (GCM) cannot be directly used for climate studies at regional scale due to coarse spatial resolution. Wood et al. [1] suggested that products of GCMs should not be directly

used to assess local or regional scale patterns, as these are less promising, specifically for precipitation. So, downscaling procedures are adopted to bridge the gap of scale mismatch between GCM and local variable. Downscaling of coarse scale GCM to a regional scale is usually done in climate change studies with different downscaling methods. Generally, two downscaling techniques are used, i.e., (i) Statistical downscaling, i.e., based on the statistical relationship between large scale atmospheric variable and local scale observed data, and (ii) Dynamical downscaling, i.e., based on the output of regional climate models (RCM) [2–4]. Both downscaling approaches have been thoroughly discussed and documented in different literatures [4,5]. Statistical downscaling is generally used as it requires less computational resources and time than dynamical downscaling. A downscaling approach based on simple manipulation of observed data, i.e., change factor methodology, is also applied in some studies [6] but statistical downscaling methods are most widely used [7]. Pielke et al. [5] discussed different types of statistical and dynamical downscaling based on their purpose and inputs. This study may be classified into type 4 statistical downscaling in which a transfer function between present climate and large-scale atmospheric variables is used to represent the future climate with the help of earth system models representing future climate conditions.

Typically, a statistical relationship between coarse scale atmospheric variables (predictors) and finer scale local observation (predictand) is developed in statistical downscaling techniques. A developed relationship is used to downscale GCM at a local scale. Different methods have been adopted previously to develop the predictor-predictand model based on simple linear regression [8], partial least square regression [9], artificial neural networks [10], relevance vector machines [11], etc. Generally, observations of all atmospheric variables, which are required to develop a predictor-predictand relationship, are not available. So, it is a standard practice to use reanalysis data as predictors in the absence of observed atmospheric variables. Different agencies provide wide ranges of reanalysis data at a global or continental scale [12,13]. The National Center for Environment Prediction (NCEP)/National Center for Atmospheric Research (NCAR) reanalysis data [12] is mostly used in statistical downscaling. Reanalysis data have been widely used to downscale precipitation [8,10,14,15], temperature [16,17], streamflow [18], and other variables. Detailed description of previous studies related to statistical downscaling is covered in different review articles [19,20].

The variability in the values, statistical properties etc., in downscaled variable with different GCMs, is still a big issue for climate researchers. The latest assessment report of the Intergovernmental Panel on Climate Change (IPCC) emphasized the issue of uncertainty in the projection of different GCMs [21]. Uncertainty is also inherited with different GCMs due to differences in parametrization, boundary conditions, and the structure and physics used in driving GCM [19]. Sunyer et al. [22] used statistical downscaling techniques to downscale extreme rainfall events with four GCM driven RCMs and found a significant GCM-GCM bias (difference in downscaled precipitation using different GCMs) in downscaled rainfall. Hughes et al. [23] also showed high variability in downscaled rainfall using nine different GCMs. Ahmed et al. [24] reported the differences in downscaled precipitation and temperature using different GCMs. Climate change detection studies also show variability in the detection of climate change using different GCMs [8,25,26]. The bias in downscaled variables using different GCMs is reported in the latest review studies [27,28]. Jiang et al. [29] addressed the issue of high variability in downscaled products using different GCMs. Pan et al. [30] ran two regional climate models (RCMs) forced by different boundary conditions, i.e., reanalysis, GCM, and future scenarios of GCM. Large differences in downscaled precipitation were observed using different GCMs.

Bias correction methods are generally adopted to reduce this uncertainty. There have been a wide range of bias correction methods proposed by different researchers. A simple quantile-based CDF matching method has been widely used [1,31], which is based on the assumption that climate does not change over time. To overcome this assumption, Li et al. [32] proposed an equidistant CDF matching method which maps the quantile of GCM variable on corresponding value of the model and the difference between observed and model values is added to get the bias corrected value. Some researchers have also a proposed method to correct low frequency bias in a downscaled variable using

nesting techniques [33,34]. However, these techniques are unable to correct bias for defined time scales. Nguyen et al. [35] proposed frequency based bias correction method which is time independent and was found to be better than nesting techniques.

The bias correction methods are found to remove some bias and decrease uncertainty in downscaled data. In this study, a different method of downscaling GCM data is proposed which produces less GCM-GCM bias than using reanalysis data. As a case study, CMIP5 historical data of ten different GCMs was used to downscale precipitation in different regions in the Ganga river basin in India. Previous studies in the Ganga river basin show a temperature rise and increase in frequency of extreme events due to climate change [36–40].

Giorgi et al. [41] used the terms “model performance” and “model convergence” for evaluation of the model. Model performance describes the closeness of simulated data to target data and convergence shows uncertainty in different sets of simulated data. The reliability ensemble averaging (REA) technique was used to evaluate the model performance and convergence. REA considers performance in terms of bias from observed data and convergence as a distance from weighted the average of ensemble members to calculate the reliability factor of a model. The reliability of a model was defined as product of its performance and convergence. Dessai et al. [42] used a performance skill score and convergence skill score to evaluate the models. The inverse of the normalized root mean square deviation between simulated and observed data gives a performance skill score. The convergence skill score is also calculated as the inverse of root mean square deviation but between simulated data by a GCM and ensembles average data. The combined skill score was also calculated, which is a weighted product of the performance and convergence skill score, to define the reliability of a model. Most of the previous studies used deviations or root mean square deviation to evaluate the models. In this study different measures, i.e., skill score, correlation, normalized root mean square deviation (NRMSD), etc., were used to judge the reliability of both models.

This study aimed to find the answers for two questions:

- (i) Should reanalysis data be always used to develop a predictor-predictand relationship and to use the same relationship to downscale GCM? and
- (ii) Can a simple change in the downscaling approach be helpful to reduce GCM-GCM bias?

The driving ideologies of this study were that (i) there is always dissimilarity in the predictor sets of reanalysis and GCM data due to difference in their development, so predictor-predictand model based on predictors of reanalysis data might not be good enough to downscale GCMs, and (ii) a better predictor-predictand relationship of reanalysis-observed data does not assure good downscaling will occur with different GCMs. This may be better understood as follows

Say,

$$\text{Observed} = f(\text{Obs}) \quad (1)$$

$$\text{Reanalysis} = f(\text{Obs1}) \quad (2)$$

$$\text{GCM1} = f(G) \quad (3)$$

$$\text{GCM2} = f(G) \quad (4)$$

Here, ‘Obs’, ‘Obs1’ is the characteristics of observed and reanalysis data time series. As, the reanalysis data is developed using observations, Obs1 is similar to Obs. ‘G’ includes the characteristics of GCM1 and GCM2. GCM1 and GCM2 are GCM products of the same GCM model but with different times.

As a standard statistical approach, Observed, Reanalysis and GCM1 data are available for the same time period, while the GCM2 product have to downscale using the predictor-predictand relationship.

If one wishes to find the unknown value of GCM2 in accordance with the data observed, there may be two approaches:

- (i) 'Re-Obs'—Develop a relationship between Reanalysis and Observed than transfer the relationship to GCM2.
- (ii) 'GCM-Obs'—Develop a relationship between GCM1 and Observed and transfer it to GCM2.

So, it can be observed that *GCM-Obs* is a better choice as it is using the same group of variables to develop the predictor-predictand relationship which needs to be predicted.

Each GCM simulation is based on assumptions of climate systems, initial conditions, parameterizations, and numerical methods used to solve the non-linear differential equations of the fluid motion of the atmosphere and the ocean [43,44]. The reanalysis data is based on the observation data, so these are closer to the observed data than GCMs. It is obvious that the relationship between reanalysis and observed data may be better than relationship between GCM and observed data. However, this doesn't guarantee better GCM downscaling. GCM downscaling using the *Re-Obs* model would be similar to predicting a time series using the relationship of two similar time series which are different from the predicted time series. Racherla et al. [7] also concluded that a better model does not necessarily translate to better climate projections. GCM downscaling using the *GCM-Obs* model may be the better choice, as the predictor-predictand relationship between GCM and observed data already considers the GCM uncertainty, which may result in better GCM downscaled products that may be close to the observations.

Change Factor Methodologies (CFM) [6] also use a similar approach to downscale GCM. CFM only use GCM data to downscale future GCM by using additive multiplicative measures with observations. CFM is widely used across the world in climate change impact assessment studies and programs [45,46].

An effort had been made in this study to reduce the GCM bias in downscaled products using the predictor-predictand relationship and using historical GCM data itself as the predictor. This technique is referred to as *GCM-Obs*. Using historical GCM data as the predictor might provide better future projections of respective GCM, as it has already considered the inherent uncertainty of GCM in model development. Change factor methodology also has a similar approach of downscaling which uses the *GCM-Obs* relationship to downscale future projections of climate.

After the brief introduction and background of the presented study in Section 1, *GCM-Obs* logic is described using mathematical expressions and synthetic series in Section 2. The *Re-Obs* and *GCM-Obs* models are compared using a case study in the Ganga river basin. The study area and data used are described in Section 3, the methodology adopted is given in Section 4, while Section 5 describes the outcome of the study. Discussions are provided in Section 6 followed by limitations in Section 7. The study is concluded in Section 7. A methodology to develop predictor-predictand relationship is described in Appendix A.

Definition 1. *Definition of Bias: The term bias used in this study represents the differences between downscaled precipitation using different GCMs and differences between downscaled and observed precipitation.*

2. The *GCM-Obs* Logic

2.1. Mathematical Explanation

The performance of prediction model can be assessed by checking the deviations of predicted values from observed data. For example, a variable '*O*' having mean μ_O and standard deviation σ_O depend on the variable '*R*' (μ_R and σ_R). It also depends on the variable '*G*' (μ_G and σ_G). The correlation between '*O*' and '*R*' is ρ_{OR} and between '*O*' and '*G*' is ρ_{OG} . Some part of data is considered for training '*tr*' and remaining is used for testing '*ts*'. The simplest relationship between two variables can be defined by a linear regression model.

So,

$$O = f(R) \text{ and } O = f(G)$$

$$(O_{tr})_i = \alpha_{R_{tr}} + \beta_{R_{tr}}(R_{tr})_i \quad (5)$$

$$\hat{\beta}_{R_{tr}} = \rho_{OR} \frac{\sigma_R}{\sigma_O} \quad (6)$$

$$\hat{\alpha}_{R_{tr}} = \mu_O - \hat{\beta}_{R_{tr}} \mu_R \quad (7)$$

$$(O_{tr})_i = \alpha_{G_{tr}} + \beta_{G_{tr}}(G_{tr})_i \quad (8)$$

$$\hat{\beta}_{G_{tr}} = \rho_{OG} \frac{\sigma_G}{\sigma_O} \quad (9)$$

$$\hat{\alpha}_{G_{tr}} = \mu_O - \hat{\beta}_{G_{tr}} \mu_G \quad (10)$$

Here α and β are intercept and slope. $\hat{\alpha}$ and $\hat{\beta}$ are best fit values defining the relationship between independent and dependent variables. This relationship can be used to predict response of testing data 'ts'

So,

$$O_{(pR_{ts})_j} = \hat{\alpha}_{R_{tr}} + \hat{\beta}_{R_{tr}}(G_{ts})_j \quad (11)$$

$$O_{(pG_{ts})_j} = \hat{\alpha}_{G_{tr}} + \hat{\beta}_{G_{tr}}(G_{ts})_j \quad (12)$$

Here, O_{pR} and O_{pG} are prediction of G_{ts} data for O-R and O-G models using Equations (11) and (12) respectively.

The quality of predicted response can be checked in terms of deviation from actual value as below

$$O_{(pR_{ts})_j} - (O_{ts})_j = \hat{\alpha}_{R_{tr}} + \hat{\beta}_{R_{tr}}(G_{ts})_j - \alpha_{R_{ts}} - \beta_{R_{ts}}(R_{ts})_j \quad (13)$$

or

$$O_{(pR_{ts})_j} - (O_{ts})_j = (\hat{\alpha}_{R_{tr}} - \alpha_{R_{ts}}) + (\hat{\beta}_{R_{tr}}(G_{ts})_j - \beta_{R_{ts}}(R_{ts})_j) \quad (14)$$

or

$$O_{(pR_{ts})_j} - (O_{ts})_j = (\hat{\alpha}_{R_{tr}} - \alpha_{R_{ts}}) + (G_{ts})_j \left(\hat{\beta}_{R_{tr}} - \beta_{R_{ts}} \left(\frac{R_{ts}}{G_{ts}} \right)_j \right) \quad (15)$$

and

$$O_{(pG_{ts})_j} - (O_{ts})_j = \hat{\alpha}_{G_{tr}} + \hat{\beta}_{G_{tr}}(G_{ts})_j - \alpha_{G_{ts}} - \beta_{G_{ts}}(G_{ts})_j \quad (16)$$

or

$$O_{(pG_{ts})_j} - (O_{ts})_j = (\hat{\alpha}_{G_{tr}} - \alpha_{G_{ts}}) + (\hat{\beta}_{G_{tr}}(G_{ts})_j - \beta_{G_{ts}}(G_{ts})_j) \quad (17)$$

or

$$O_{(pG_{ts})_j} - (O_{ts})_j = (\hat{\alpha}_{G_{tr}} - \alpha_{G_{ts}}) + (G_{ts})_j (\hat{\beta}_{G_{tr}} - \beta_{G_{ts}}) \quad (18)$$

Comparing the Equations (11) and (14), assuming deviation in intercept terms for O-R and O-G models almost equal, the deviation in prediction from actual value is

$$O_{(pR_{ts})_j} - (O_{ts})_j \propto \left(\hat{\beta}_{R_{tr}} - \beta_{R_{ts}} \left(\frac{R_{ts}}{G_{ts}} \right)_j \right) \quad (19)$$

and

$$O_{(pG_{ts})_j} - (O_{ts})_j \propto (\hat{\beta}_{G_{tr}} - \beta_{G_{ts}}) \quad (20)$$

It can be noted that difference between best fit value of slope $\hat{\beta}$ and the normal slope β for the same variable will certainly be lower than the value in Equation (15). To better understand this, we assume a specific case when $\hat{\beta}_R = \beta_R$ and $\hat{\beta}_G = \beta_G$, i.e., there is already best fit relationship between dependent and independent variables. So, from Equations (15) and (16)

$$O_{(pR_{ts})_j} - (O_{ts})_j \propto \hat{\beta}_{R_{tr}} \left(1 - \left(\frac{R_{ts}}{G_{ts}} \right)_j \right) \tag{21}$$

$$O_{(pG_{ts})_j} - (O_{ts})_j \propto 0 \tag{22}$$

So, the O-G model is expected to perform better than the R-G model to predict the response of unknown values of variable G.

In support of the above statement, an example of statistical downscaling using *Re-Obs* and *GCM-Obs* models is also shown in Supplementary Information. As simple statistical downscaling using multiple-linear-regression (MLR) is applied to downscale precipitation at grid point—28.25° latitude and 73.25° longitude. Monthly Global Precipitation Climatology Center (GPCC) data was used at observed precipitation (details of GPCC data is provided in Section 3). Only few predictor variables are selected to keep the example simple. A summary of the data and method is presented in Table 1.

Table 1. Summary of Data/Methods.

Property	Observed (Predictand)	Reanalysis (NCEP/NCAR) (Predictor)			GCM (CMCC-CMS) (Predictor)		
Variables	GPCC Observed	Near Surface Air Temperature (Temp)	Geopotential Height (Gph)	Specific Humidity (SHum)	Near Surface Air Temperature (Temp)	Geopotential Height (Gph)	Specific Humidity (SHum)
Predictor-Predictand Model		Multiple Linear Regression					
Time		January-1948 to December-2005					
Training data		January-1948 to December-1994					
Testing data		January-1995 to December-2005					

The statistical downscaling using both *Re-Obs* and *GCM-Obs* models is compared in Table 2.

Table 2. Comparison of downscaled product with Observed using *Re-Obs* and *GCM-Obs* models.

Performance Measure	<i>Re-Obs</i>	<i>GCM-Obs</i>
Correlation Coefficient	0.53	0.67
Root Mean Square Error	27.57	24.42

It may be observed from the results that the *GCM-Obs* model performed better than the *Re-Obs* model to downscale GCM data. It can also be observed that the *Re-Obs* model showed a better predictor-predictand relationship than *GCM-Obs* because reanalysis data is closer to observations but the *GCM-Obs* model performed better to downscale GCM.

2.2. Explanation Using Synthetic Series

The logic of using historical GCM data itself instead of reanalysis data is also explained by considering two different sets of synthetic time series: (i) Set1—time series having no seasonal component i.e., annual or seasonal values and (ii) Set2—a time series which has some cyclic effects due to inter-annual patterns or natural atmospheric oscillations, i.e., ENSO etc.

Each set of series consists of observed, reanalysis, and historical GCM data for the same length. It is to be noted that both reanalysis and GCM data are correlated to some degree (even very small) with observed data. So,

$$Obs = f_1(\rho_1, \rho_2 \dots, Re1, Re2 \dots) \tag{23}$$

$$Obs = f_2(\rho_1, \rho_2 \dots, GCM1, GCM2 \dots) \tag{24}$$

Here, *Obs*, *Re1*, *Re2*, *GCM1* and *GCM2* are time series of observed, reanalysis data for first variable, reanalysis data for second variable, GCM data of first variable and GCM data of second variable respectively. ρ_1, ρ_2 are cross correlations of reanalysis and GCM variables with observed data series.

The values of variables at time *i* for Set1 can be expressed as

$$Obs_i = \mu_{Obs} + \sigma_{Obs}N(0, 1^2)_j \tag{25}$$

$$Re1_i = \mu_{Re1} + \sigma_{Re1} \left(N(0, 1^2)_j \rho_{Obs-Re1} + N(0, 1^2)_{j+1} \sqrt{(1 - \rho_{Obs-Re1}^2)} \right) \tag{26}$$

$$Re2_i = \mu_{Re2} + \sigma_{Re2} \left(N(0, 1^2)_j \rho_{Obs-Re2} + N(0, 1^2)_{j+2} \sqrt{(1 - \rho_{Obs-Re2}^2)} \right) \tag{27}$$

$$GCM1_i = \mu_{GCM1} + \sigma_{GCM1} \left(N(0, 1^2)_j \rho_{Obs-GCM1} + N(0, 1^2)_{j+3} \sqrt{(1 - \rho_{Obs-GCM1}^2)} \right) \tag{28}$$

$$GCM2_i = \mu_{GCM2} + \sigma_{GCM2} \left(N(0, 1^2)_j \rho_{Obs-GCM2} + N(0, 1^2)_{j+4} \sqrt{(1 - \rho_{Obs-GCM2}^2)} \right) \tag{29}$$

Here, μ , σ and ρ are mean, standard deviation and cross correlation of reanalysis/GCM with observed data. $N(0, 1^2)_j$ is the *j*th standard normal random number.

Set2 is generated considering the seasonality. Here we have considered seasonality composed of three sinusoidal wave forms and the data series are expressed as

$$Obs_i = \sin(2\pi n_1 i) + \sin(2\pi n_2 i) + \sin(2\pi n_3 i) + \mu_{Obs} + \sigma_{Obs}N(0, 1^2)_j \tag{30}$$

$$Re1_i = A_1 \sin(2\pi n_4 i) + A_2 \sin(2\pi n_5 i) + A_3 \sin(2\pi n_6 i) + \mu_{Re1} + \sigma_{Re1} \left(N(0, 1^2)_j \rho_{Obs-Re1} + N(0, 1^2)_{j+1} \sqrt{(1 - \rho_{Obs-Re1}^2)} \right) \tag{31}$$

$$Re2_i = A_4 \sin(2\pi n_7 i) + A_5 \sin(2\pi n_8 i) + A_6 \sin(2\pi n_9 i) + \mu_{Re2} + \sigma_{Re2} \left(N(0, 1^2)_j \rho_{Obs-Re2} + N(0, 1^2)_{j+2} \sqrt{(1 - \rho_{Obs-Re2}^2)} \right) \tag{32}$$

$$GCM1_i = A_6 \sin(2\pi n_{10} i) + A_7 \sin(2\pi n_{11} i) + A_8 \sin(2\pi n_{12} i) + \mu_{GCM1} + \sigma_{GCM1} \left(N(0, 1^2)_j \rho_{Obs-GCM1} + N(0, 1^2)_{j+3} \sqrt{(1 - \rho_{Obs-GCM1}^2)} \right) \tag{33}$$

$$GCM2_i = A_9 \sin(2\pi n_{13} i) + A_{10} \sin(2\pi n_{14} i) + A_{11} \sin(2\pi n_{15} i) + \mu_{GCM2} + \sigma_{GCM2} \left(N(0, 1^2)_j \rho_{Obs-GCM2} + N(0, 1^2)_{j+4} \sqrt{(1 - \rho_{Obs-GCM2}^2)} \right) \tag{34}$$

Here, *A* is the amplitude and *n* is frequency of the wave form.

Typical pattern of synthetic time series for Set1 and Set2 are shown in Figure 1a,b respectively. Different values of parameters considered for Set1 and Set2 are given in Tables 3 and 4 respectively.

Table 3. Parameters of series for Set1.

Data	Mean (μ)	Standard Deviation (σ)	Cross Correlation with Observed (ρ)
<i>Obs</i>	1.0	0.5	1
<i>Re1</i>	5	0.3	0.8
<i>Re2</i>	2	0.8	0.8
<i>GCM1</i>	7	0.7	0.6, 0.8
<i>GCM2</i>	3	0.2	0.6, 0.8

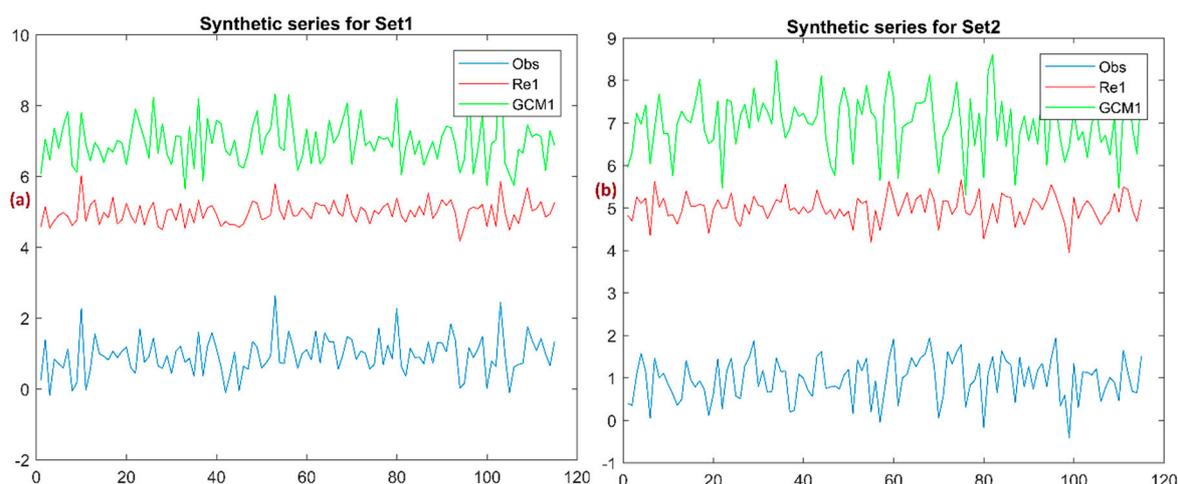


Figure 1. Typical patterns of Synthetic series for (a) Set1 and (b) Set2. Re2 and GCM2 are not shown for better clarity.

Table 4. Parameters of series for Set2.

Data	Mean	Standard Deviation	Seasonal Component						Correlation with Observed
			Wave 1		Wave 2		Wave 3		
			Amplitude	Frequency	Amplitude	Frequency	Amplitude	Frequency	
Obs	1	0.5	1	10	1	25	1	40	1
Re1	5	0.3	1.1	11	1.1	23	0.9	45	0.8
Re2	2	0.8	1.2	11	1.2	23	0.8	45	0.8
GCM1	7	0.7	1.5	15	2.1	20	1.9	60	0.6, 0.8
GCM2	3	0.2	1.6	15	2.5	20	2.1	60	0.6, 0.8

The parameters of synthetic series of reanalysis data are considered to be closer to the observed as it is assumed that reanalysis data shows better similarity than corresponding GCM with observed data because the former is based on observations while the latter is generated while assuming an ocean-atmospheric relationship.

1000 series of sample size 115 for each variable for Set1 and Set2 are generated and tested with statistical downscaling with *GCM-Obs* and a formal approach using reanalysis data (*Re-Obs*). To reduce complexity, a simple linear regression model is used to develop the relationship between predictors, i.e., *GCM/reanalysis* and predictand Observed data. Around 60% of the data is used for training the model and the rest is used for testing. The predictor-predictand relationship is used to downscale/predict the testing data. The predicted response of both models is tested with observed data in terms of mean, standard deviation, skewness, lag1 autocorrelation, correlation coefficient, and root-mean-square-error (RMSE).

The predicted response of test data with both models for Set1 is compared with observed data and shown in Figure 2. All the measures to test the prediction with observed data are calculated for testing period. It is found that there is a huge bias in the mean and standard deviation of the predicted response with the *Re-Obs* model, while the *GCM-Obs* model shows a high degree of similarity. Root mean square error shows a true deviation of predicted values with the observed data. The *GCM-Obs* model shows a RMSE close to 1 while the *Re-Obs* model shows a very high RMSE of the order of 25–30. Little to no effect was found on skewness and lag1 autocorrelation of prediction with selection of the prediction model. Both of them are found to predict similar skewness and lag1 autocorrelation with the observed data.

Figure 3 shows performance of the *GCM-Obs* and *Re-Obs* models for Set2. Here also, the *GCM-Obs* models shows good agreement in terms of mean and standard deviation. RMSE of around 30 is found for the *Re-Obs* model while for the *GCM-Obs* model, it is found to be below 1. Both models show same

degree of similarity in skewness and lag1 autocorrelation with observed data. It can be seen that higher correlation is found in predicted response with observed data for both models for a higher correlation of GCM data in generated synthetic series.

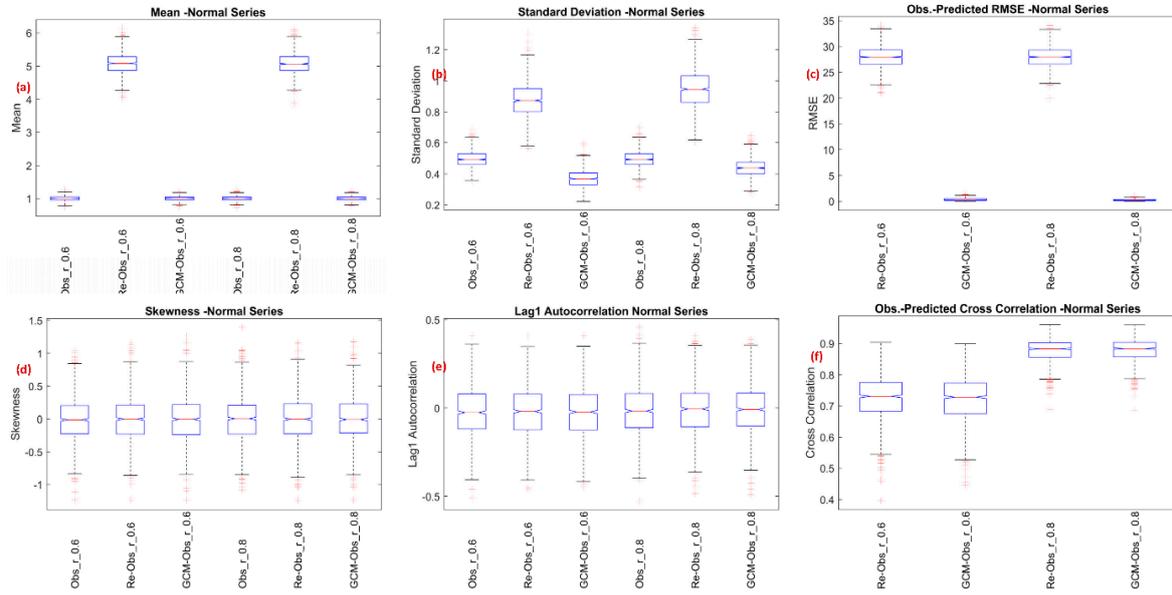


Figure 2. Comparison of predicted data with observed values using the *Re-Obs* and *GCM-Obs* models for Set1. Similarity of prediction with observed is shown by (a) Mean (b) Standard Deviation (c) RMSE (d) Skewness (e) Lag1 autocorrelation and (f) Cross correlation. Suffix 0.6 and 0.8 on X-axes labels show the correlation between observed and GCM data considered for synthetic series generation.

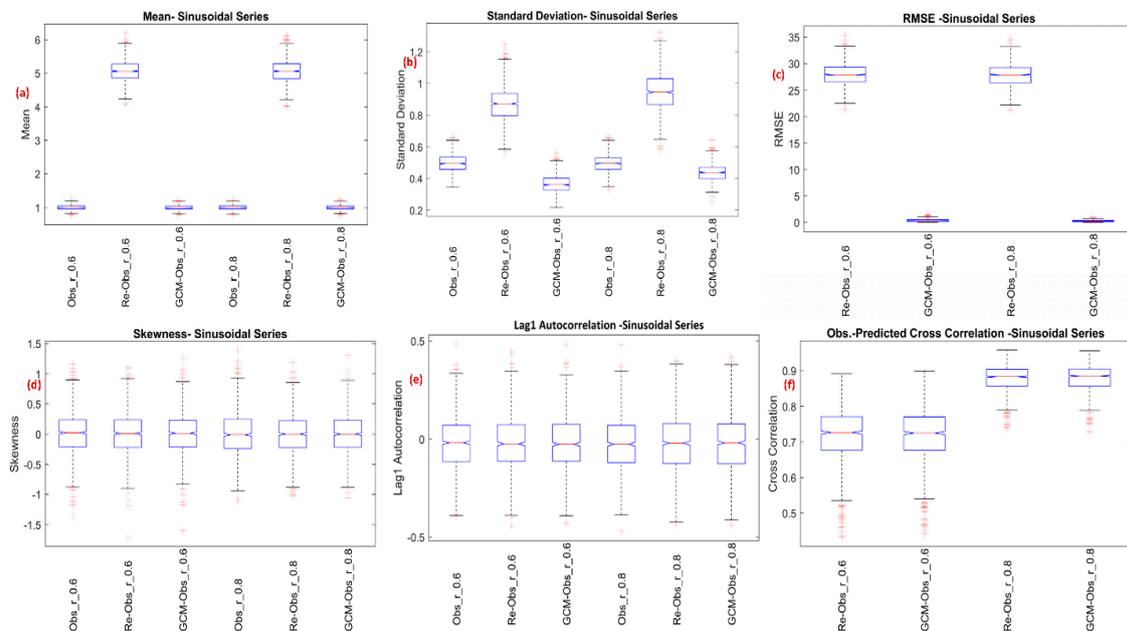


Figure 3. Comparison of predicted data with observed values using the *Re-Obs* and *GCM-Obs* models for Set2. Similarity of prediction with observed is shown by (a) Mean (b) Standard Deviation (c) RMSE (d) Skewness (e) Lag1 autocorrelation and (f) Cross correlation. Suffix 0.6 and 0.8 on x-axes labels show correlation between observed and GCM data considered for synthetic series generation.

This shows that the *GCM-Obs* model should perform better than the *Re-Obs* model to reduce bias and produce better predictions.

3. The *GCM-Obs* model: Case Study

The *GCM-Obs* model was found to be better than the *Re-Obs* model for downscaling GCM data while considering synthetic series in the previous section. Here, the *GCM-Obs* model was used to downscale precipitation at different locations in the Ganga river basin and was compared with the *Re-Obs* model.

3.1. Study Area and Data

The confluence of the rivers Alaknanda and Bhagirathi at Devprayag in the Uttarakhand district is the beginning of the river Ganga. Gangotri glacier is the primary source of the Ganga river and is also the originating place of River Bhagirathi. The terminus of Gangotri glacier is at Gaumukh in Uttarakhand. Gaumukh is considered as the true source of the Ganga river. The Ganga river travels around 2525 km from its origin at Gaumukh to terminus at the Bay of Bengal [47]. The catchment area of Ganga river basin is around 1,086,000 km² which lies between latitudes 22°30' N to 31°30' N and longitudes 73°30' to 89° E to and falls in four countries: India, Nepal, Bangladesh, and Tibet (China) [48].

The Ganga river basin map is shown in Figure 4. The elevation in the basin ranges from near to mean sea level (MSL) to about 8000 m MSL. Five regions/zones of the Ganga river basin were considered representing different climate types defined by an updated Koppen-Geiger global climate classification [49]. These are:

- i. Zone 1: Northern snow dominant subtropical highlands—29.5°–30° N and 79.5°–80° E at grid interval 0.5° × 0.5°
- ii. Zone 2: Western hot semi-arid zone—25°–25.5° N and 74°–74.5° E at grid interval 0.5° × 0.5°
- iii. Zone 3: Central subtropical humid zone—23.5°–24° N and 81°–81.5° E at grid interval 0.5° × 0.5°
- iv. Zone 4: Eastern tropical monsoon zone—22.5°–23° N and 87.5°–88° E at grid interval 0.5° × 0.5°
- v. Zone 5: Eastern subtropical highlands—27.5°–28° N and 87.5°–88° E at grid interval 0.5° × 0.5°



Figure 4. Study area in the Ganga river basin. Zones are shown by solid yellow rectangles and location of districts are shown by dashed yellow rectangles.

The Global precipitation climatology center (GPCC) is a leading agency providing global monthly precipitation high resolution gridded observed data at a grid interval of $0.5^\circ \times 0.5^\circ$ [50]. Other gridded data set, i.e., from the Climate Research Unit (CRU) [51,52] have also been used which agrees with the GPCC data set. GPCC precipitation data for the years 1948 to 2005 were used as gridded observed data for zones 1 to 5 in this study.

Observed precipitation data provided by India Water Portal (available at indiawaterportal.org) was also used to check the effect of different observed data providing agency on downscaled data. Monthly precipitation data for two districts in The Ganga river basin, i.e., Uttarkashi district (lies in latitude $30^\circ 30' N$ to $31^\circ N$ and longitude $78^\circ 25' E$ to $78^\circ 75' E$) in state Uttarakhand, in which the Ganga river originates and Darjeeling district (lies in latitude $26^\circ 30' N$ to $27^\circ N$ and longitude $88^\circ 25' E$ to $88^\circ 75' E$) in west Bengal state, which receives considerably higher rainfall, is obtained from India Water Portal (IWP) for the years 1948 to 2002. The location of the study area in the Ganga river basin is also shown in Figure 4.

Mean monthly, global NCEP/NCAR reanalysis data [12] of six climatological variables, i.e., temperature (ta), geopotential height (zg), specific humidity (hus), zonal and meridional wind components (Ua and Va respectively) and mean sea level pressure (psl), at grid interval of $2.5^\circ \times 2.5^\circ$ for the years 1948 to 2015 were used to develop a predictor-predictand model.

Coupled Model Inter-comparison Project-5 (CMIP5) historical mean monthly data of six climatological variables (as considered in reanalysis data) available at different grid intervals depending on specific GCM for year 1948 to 2010 were obtained from ESGF website (available at <https://pcmdi9.llnl.gov/projects/cmip5>). Brief description of GCMs used in the study is given in Table 5.

Table 5. GCMs used in this study and descriptions.

S.N.	GCM	Modeling Center	Horizontal Atmospheric Resolution (Lon × Lat)
1	CMCC-CESM	Euro-Mediterranean Center on Climate Change, Italy	3.75×3.75
2	CMCC-CM		0.75×0.75
3	CMCC-CMS		1.88×1.88
4	FGOALS-g2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences and CESS, Tsinghua University	2.81×2.81
5	FGOALS-s2		2.81×1.66
6	GFDL-CM3	US Dept. of Commerce/NOAA/Geophysical Fluid Dynamics Laboratory	2.0×2.5
7	HadGEM-AO	National Institute of Meteorological Research, Seoul, South Korea	1.3×1.9
8	INMCM4	Institute for Numerical Mathematics, Moscow, Russia	1.5×2.0
9	MIROC-ESM	Japan Agency for Marine-Earth Science and Technology, Kanagawa, Japan), AORI (Atmosphere and Ocean Research Institute, The University of Tokyo, Chiba, Japan), and NIES (National Institute for Environmental Studies, Ibaraki, Japan)	2.8×2.8
10	MIROC-ESM-CHEM		

Detailed descriptions for CMIP5 products of GCMs, their modeling centers, and resolution can be found in previous studies [53,54].

3.2. Methodology

The first objective of the study was to check whether reanalysis data was best suited to downscale GCM data. A predictor-predictand model was developed using reanalysis data as the predictor and observed data as the predictand in the *Re-Obs* model. The *GCM-Obs* model was developed using historical data of GCM itself as the predictor and observed data as the predictand. Both reanalysis and GCM variable data were re-gridded to the grid interval of $0.5^\circ \times 0.5^\circ$ to match observed data grids. Predictor variables considered at different pressure levels in milibar (mbar) in the development of predictor-predictand models are tabulated in Table 6.

Principal component analysis (PCA) of predictor data is done to reduce the dimension and to remove inter-collinearity [55]. Multiple linear regression (MLR) and Artificial Neural Network (NN) methods were used to develop predictor-predictor relationships. The Artificial Neural Network method was found to be better than MLR, so the NN method is used for downscaling GCM data.

Detailed description of predictor-predictand model development is given in Appendix A. Downscaling of historical GCM data is done with developed *Re-Obs* and *GCM-Obs* predictor-predictand models. The number of training, validation/testing, and downscaling years considered in model development and downscaling are given in Table 7.

Table 6. Pressure levels (in mbar) for development of predictor-predictand relationship.

Variable	Zone-1	Zone-2	Zone-3	Zone-4	Zone-5	Uttarkashi	Darjeeling
ta	600	500	600	600	600	600	600
zg	500	400	700	925	500	500	500
hus	600	500	600	600	600	600	600
ua	600	500	600	600	600	500	600
va	600	500	600	600	600	500	600
psl	-	-	-	-	-	-	-

Table 7. Division of data for development of predictor-predictand relationship and downscaling of GCM.

S.N.	Observed Data Source	Total Years	Training Years	Validation/Testing Years	Downscaling Years
1	GPCC	58	37	11	10
2	IWP	55	37	10	8

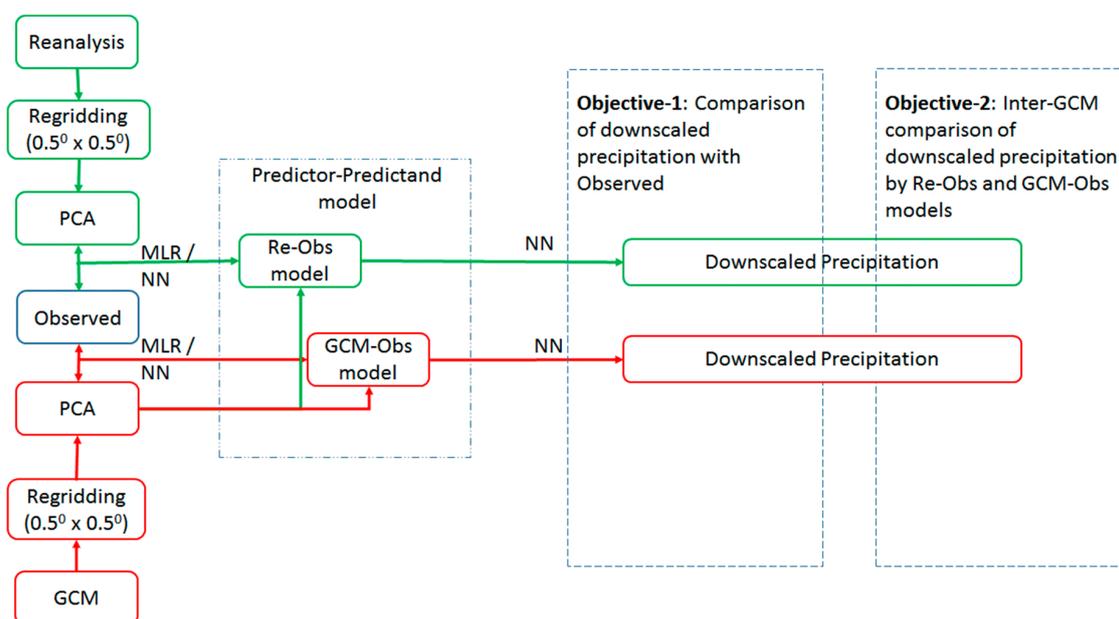


Figure 5. Flow chart of methodology used in this study. PCA indicates principal component analysis. MLR/NN represents development of predictor-predictand model using multiple linear regression and artificial neural network methods, respectively.

Evaluation of both models was done in terms of performance and convergence capabilities. Performance of the model was checked by assessing the similarity of downscaled precipitation using different GCMs with observed precipitation in terms of normalized standard deviation, correlation, skill score, and normalized root mean square deviation (NRMSD). Convergence skill of model is inversely proportional to GCM uncertainty; that is, a model with lower GCM uncertainty shows better convergence and vice-versa. Measures like normalized root mean square deviation and correlation

between downscaled precipitations using different GCMs were used to analyze the similarity or uncertainty of both models. Measures to define performance and convergence capabilities of both models are discussed in detail in the results and discussions section. Step-by-step procedures followed in this study are shown in Figure 5.

4. Results

Re-Obs and *GCM-Obs* models were used to downscale historical GCM data at each grid point of zones 1 to 5, in the Uttarakhand and Darjeeling districts. Downscaled precipitation using both models were compared with observed precipitation for the model performance assessment. Zonal averaged values of observed and downscaled precipitation were considered for comparison in each zone. Time series of observed and downscaled precipitation by the *Re-Obs* model with CMCC-CMS GCM and the *GCM-Obs* model with GFDL-CM3 GCM for zone 1 and the Darjeeling district is shown in Figures 6 and 7, respectively. A considerable difference in downscaled precipitation when using *Re-Obs* and models can be observed. The coefficient of determination (R^2) between observed and downscaled precipitation by *Re-Obs* and *GCM-Obs* models varies from 0.5 to 0.69 and 0.8 to 0.81, respectively. Higher variability between observed and downscaled precipitation can be seen by with the *Re-Obs* model than with the *GCM-Obs* model for Darjeeling. Similar results were also found for other regions.

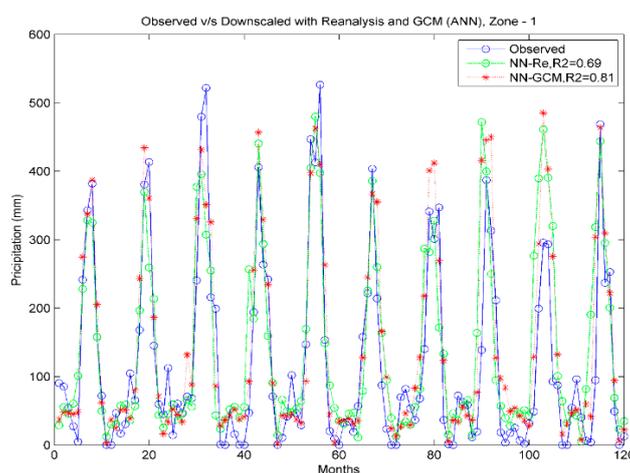


Figure 6. Simulated v/s observed zonal averaged precipitation with CMCC-CMS GCM by *Re-Obs* and the *GCM-Obs* models for zone-1. Notations—NN-Re and NN-GCM represents precipitation downscaled using ANN method with *Re-Obs* and the *GCM-Obs* models, respectively. R^2 is the coefficient of determination between observed and downscaled precipitation.

Along with a comparison of downscaling performance by both models for zone-1 to 5, Uttarkashi and Darjeeling districts are also represented by Taylor diagrams [56] in Figures 8–13, respectively. A Taylor diagram is a useful plot to concisely show the degree of similarity between observed and modeled data. In this study, a Taylor diagram is used to show the relative performances of both methods to downscale precipitation with different GCMs. Radial lines from origin show the correlation between observed and downscaled precipitation. X and Y axis indicate the normalized standard deviation, which is computed by the following formula:

$$\text{Normalized standard deviation (NSD)} = \frac{\text{Standard deviation of downscaled precipitation}}{\text{Standard deviation of observed precipitation}} \quad (35)$$

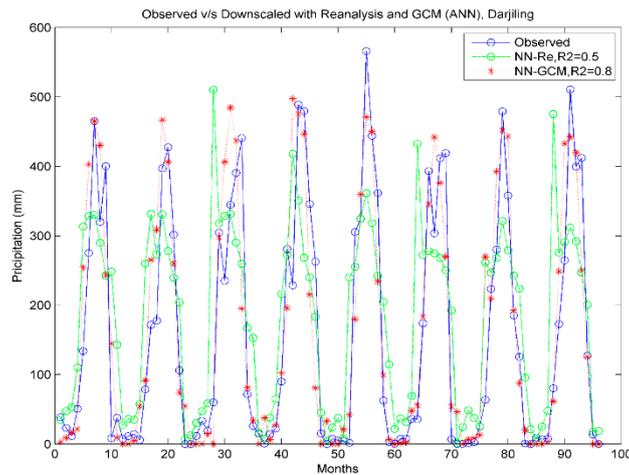


Figure 7. Modeled v/s observed zonal averaged precipitation with *Re-Obs* and the *GCM-Obs* models for Darjeeling. GFDL-CM3 GCM is used for the *GCM-Obs* model. Notations are the same as shown in Figure 6.

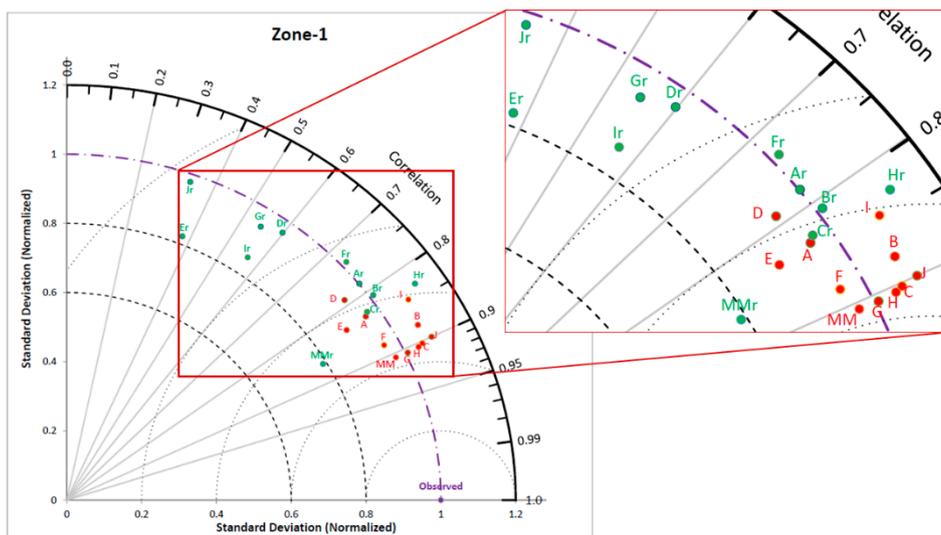


Figure 8. Taylor diagram for comparison of downscaled performances by *Re-Obs* and *GCM-Obs* methods for zone 1. Notations A to J and MM are used for different GCMs and Multi-model ensemble average (A—CMCC-CESM, B—CMCC-CM, C—CMCC-CMS, D—FGOALS-g2, E—FGOALS-s2, F—GFDL-CM3, G—HadGEM-AO, H—INMCM4, I—MIROC-ESM, J—MIROC-ESM-CHEM, MM—Multi-model ensemble average). Notations with subscript ‘r’ and that is green coloured are used for defining downscaled with the *Re-Obs* model. Downscaled with the *GCM-Obs* model is shown in red and with no subscript in notations.

NSDs are represented by two sets of concentric arches of circles having centers at origin and observed data point. NSD and correlation coefficient for observed data is always unity and marked at unit correlation and unit NSD in Taylor diagrams. Points in green and red colour represent downscaled precipitation with *Re-Obs* and *GCM-Obs* models, respectively. GCM Multi-model ensemble averaged

precipitation is also calculated by adding more weight to the highly correlated GCM downscaled precipitation with observed precipitation using the following formula.

$$MM_t = \left(\frac{\sum_{i=1}^n r_i DP_i}{\sum_{i=1}^n r_i} \right)_t \tag{36}$$

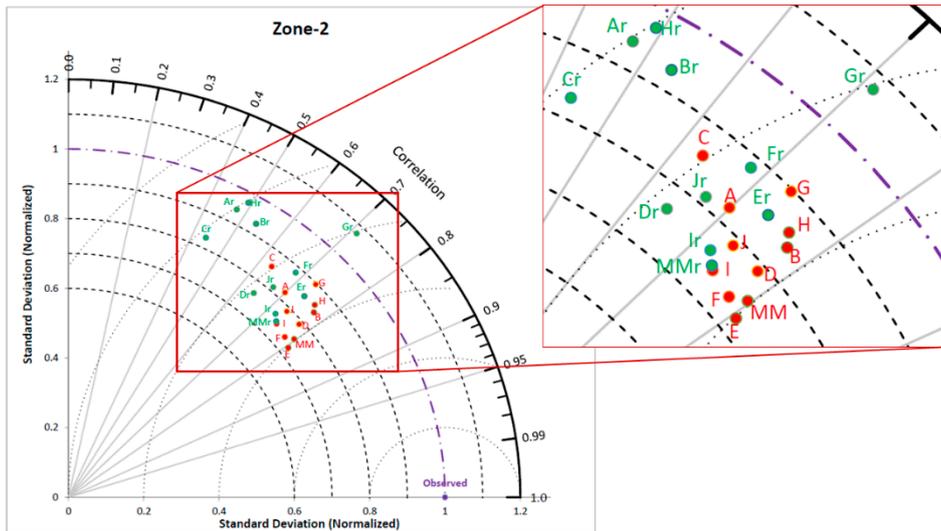


Figure 9. Taylor diagram for downscaling with *Re-Obs* and *GCM-Obs* models for zone 2. Notations are same as those used in Figure 8.

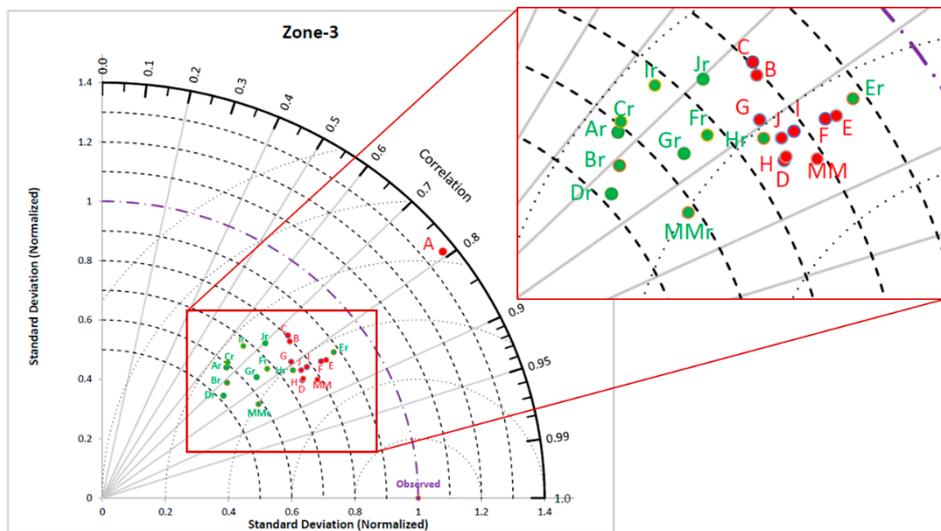


Figure 10. Taylor diagram for downscaling with *Re-Obs* and *GCM-Obs* models for zone 3. Notations are same as the ones used in Figure 8.

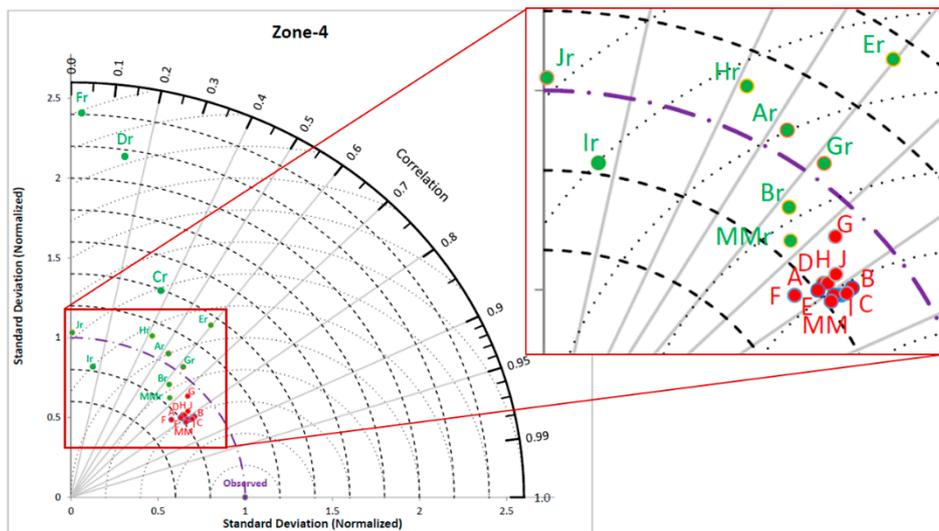


Figure 11. Taylor diagram for downscaling with *Re-Obs* and *GCM-Obs* models for zone 4. Notations are same as the ones used in Figure 8.

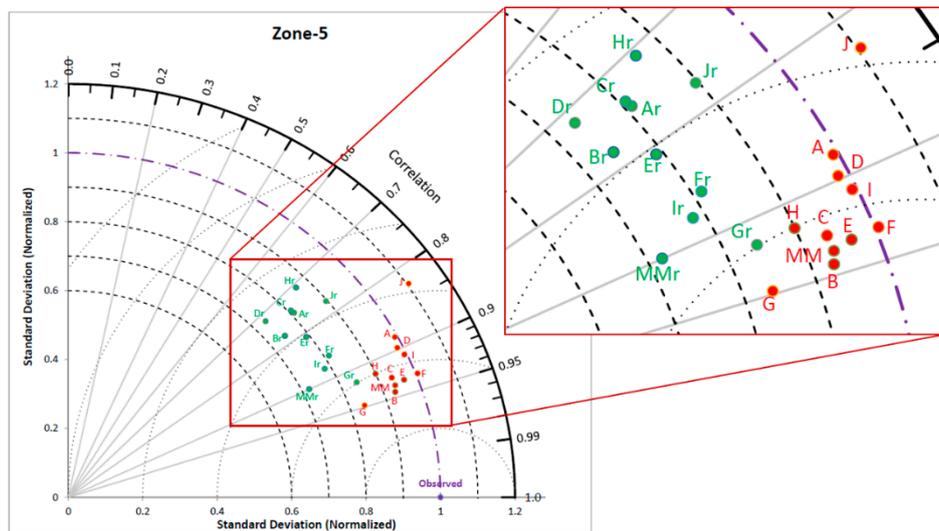


Figure 12. Taylor diagram for downscaling with *Re-Obs* and *GCM-Obs* models for zone 5. Notations are same as used in Figure 8.

Here, MM_t is the GCM Multi-model ensemble averaged precipitation at time t , r_i is correlation between downscale precipitation (DP) and observed precipitation for particular GCM, $i = 1, 2, \dots, n$ number of GCMs (here 'n' is 10).

Notations used in this study to represent precipitation downscaled by *Re-Obs* and *GCM-Obs* models with different GCMs are given in Table 8.

Figure 8 shows the Taylor diagram for zone 1. Downscaled precipitation determined by the *Re-Obs* model with MIROC-ESM-CHEM and FGOALS-s2 GCMs is found to be the least correlated with observed precipitation, as it has a correlation coefficient (r) of 0.36 and 0.38 respectively. The *GCM-Obs* model is showing a significant improvement in similarity of downscaled precipitation with observed precipitation compared to the *Re-Obs* model. The coefficient of correlation improved to 0.9 and 0.83 using the *GCM-Obs* model with MIROC-ESM-CHEM and FGOALS-s2 GCMs, respectively. NSD values by the *GCM-Obs* model are considerably better than the *Re-Obs* model, which shows less bias in observed and GCM downscaled precipitation. Similarly, Taylor diagrams for zone 2 to 5 also show improved similarity in downscaled precipitation by the *GCM-Obs* model than the *Re-Obs* model.

Figures 13 and 14 represent the Taylor diagram for district Uttarakhand and Darjeeling, respectively, for which IWP observed data was used. Figures 13 and 14 also show more significant improvements in downscaled precipitation with the *GCM-Obs* model than with the *Re-Obs* model. Taylor diagrams for most of the study area show that downscaled precipitation in the *GCM-Obs* model with a different GCM is in the form of cluster.

Skill score [56] is a popular method used to check the skill of model to simulate the results close to target data. Skill score increases with increase in correlation between simulated and observed data. It also increases when variance of modeled data approaches near to observed data. Skill score is defined as follows:

$$S_{score} (\%) = \frac{4(1 + R)^K}{(\sigma + \frac{1}{\sigma})^2(1 + R_0)^K} \times 100 \tag{37}$$

Table 8. Notations used in study for downscaling with different GCMs by *Re-Obs* and *GCM-Obs* models.

S.N.	GCM/Observed	Notation for the <i>Re-Obs</i> Model	Notation for the <i>GCM-Obs</i> Model
1	CMCC-CESM	Ar	A
2	CMCC-CM	Br	B
3	CMCC-CMS	Cr	C
4	FGOALS-g2	Dr	D
5	FGOALS-s2	Er	E
6	GFDL-CM3	Fr	F
7	HadGEM-AO	Gr	G
8	INMCM4	Hr	H
9	MIROC-ESM	Ir	I
10	MIROC-ESM-CHEM	Jr	J
11	Multi-model ensemble average	MMr	MM
12	Observed	Obs	Obs

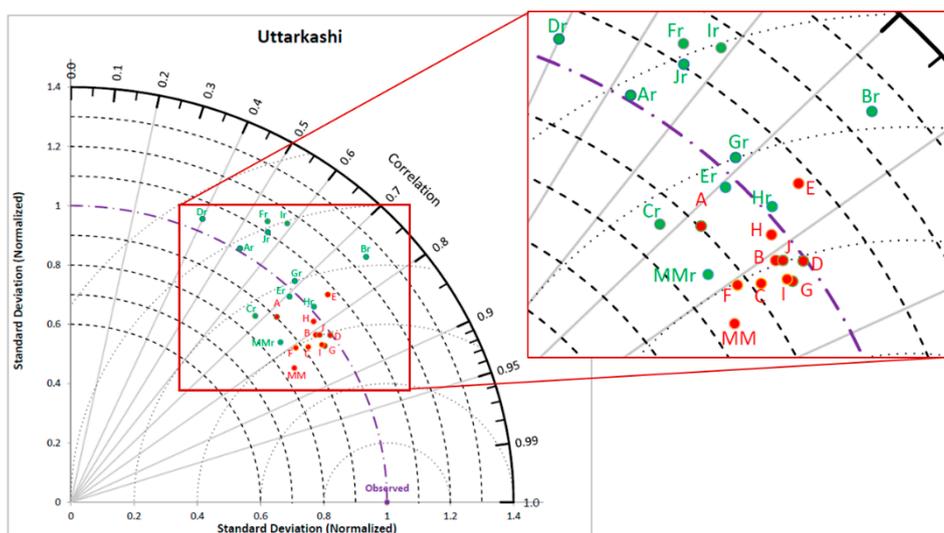


Figure 13. Taylor diagram for downscaling with *Re-Obs* and *GCM-Obs* models for Uttarkashi. Notations are same as the ones used in Figure 8.

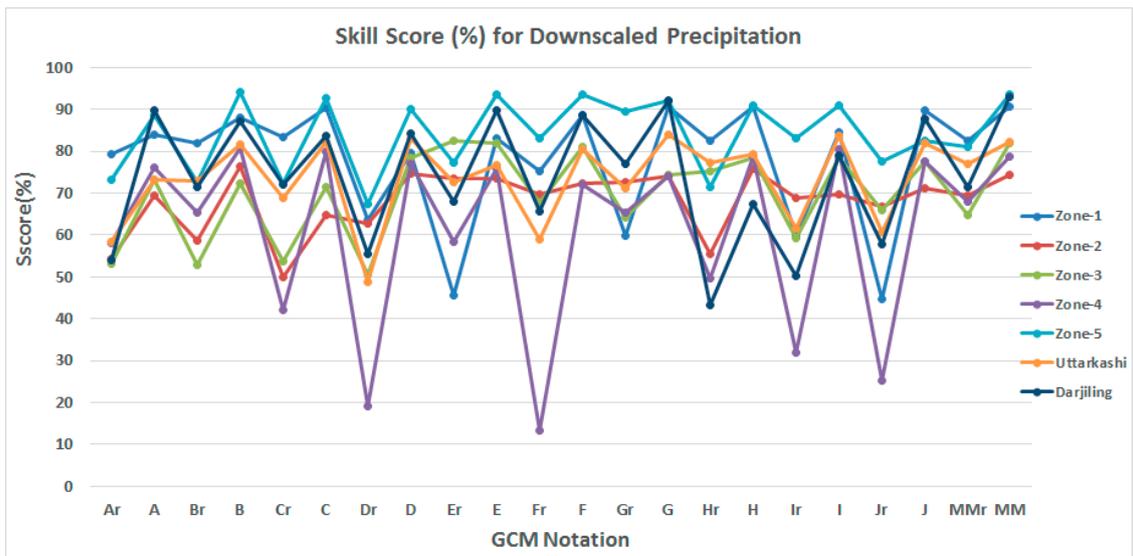


Figure 15. Skill Scores (%) of *Re-Obs* and *GCM-Obs* models for downscaling precipitation with different GCMs. GCM notations are same as the ones used in Figure 8.

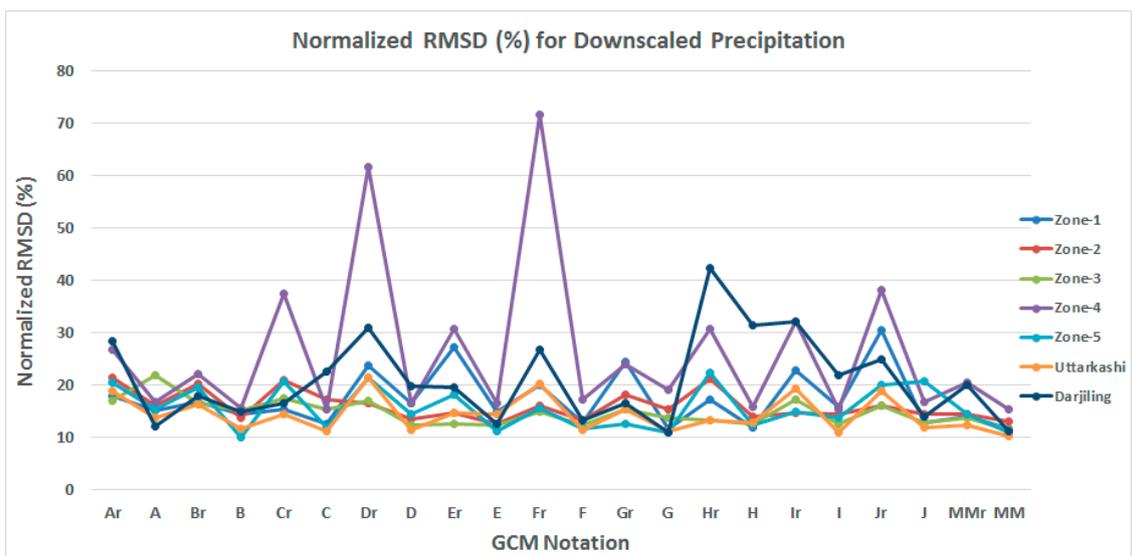


Figure 16. Normalized Root Mean Square Difference (%) of GCMs for downscaling precipitation with *Re-Obs* and *GCM-Obs* models. GCM notations are same as used in Figure 8.

Correlation coefficient (r) is a good measure to check linear similarity between two datasets. Higher value of r shows higher similarity. The Pearson correlation coefficient is adopted in this study, which is computed as follows:

$$r_{X,Y} = \frac{\sum_{t=1}^N (x_t - \bar{x}) \sum_{t=1}^N (y_t - \bar{y})}{\sqrt{\sum_{t=1}^N (x_t - \bar{x})^2} \sqrt{\sum_{t=1}^N (y_t - \bar{y})^2}} \tag{39}$$

Here, $r_{X,Y}$ is Pearson correlation coefficient between datasets X and Y , \bar{x} and \bar{y} are mean of datasets X and Y respectively, x_t and y_t are values of datasets at time $t = 1, 2, \dots, N$.

Correlation coefficient between downscaled precipitation with different GCMs by the *Re-Obs* and *GCM-Obs* models are presented in the form of a correlation matrix. Correlation between downscaled and observed precipitation is also shown. The diagonal of the correlation matrix shows a histogram of respective data series. The correlation matrix of downscaled precipitation for zone 4 by the *Re-Obs* and *GCM-Obs* models is shown in Figures 17 and 18 respectively. Similar matrix is also shown for Uttarkashi district in Figures 19 and 20.

A correlation matrix with the *Re-Obs* model for zone 4 and Uttarkashi shows a significant difference in downscaled precipitation. Four GCMs, i.e., FGOALS-g2, GFDL-CM3, MIROC-ESM and MIROC-ESM-CHEM, show almost no correlation with other GCMs, the Multi-model ensemble average, and observed data. A correlation matrix involving the *GCM-Obs* model shows a significant improvement in the correlation between GCMs. GCMs shows high correlation among themselves with a correlation coefficient of 0.88 to 0.97 with the *GCM-Obs* model instead of almost 0 to 0.73 with the *Re-Obs* model. This shows a significant reduction in GCM-GCM bias. Similar results were also found for other regions.

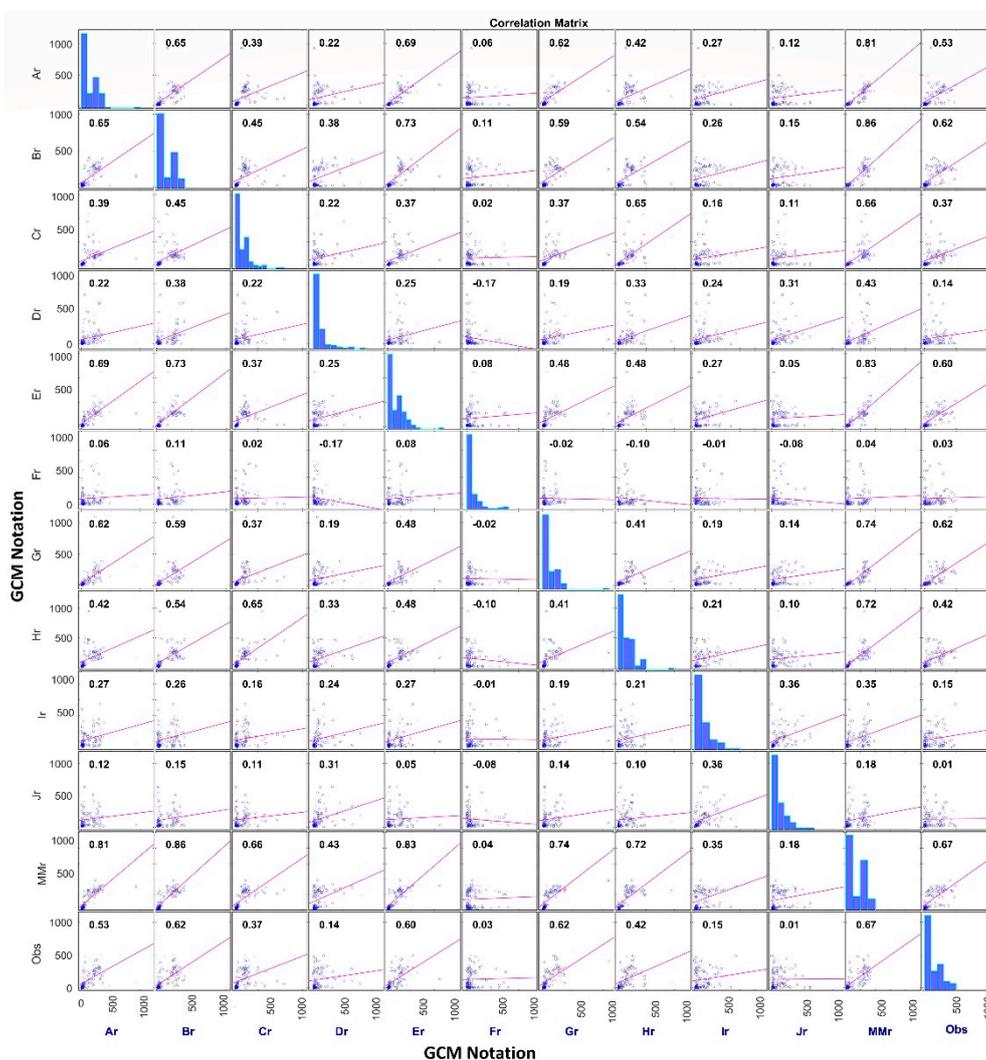


Figure 17. Correlation matrix of downscaled precipitation for different GCMs with the *Re-Obs* model for Zone 4. Notations are given in Table 5. Numbers in scatter plot show the correlation coefficient between two GCMs. Diagonal of matrix represents histogram of observed precipitation and downscaled precipitation for a particular GCM.

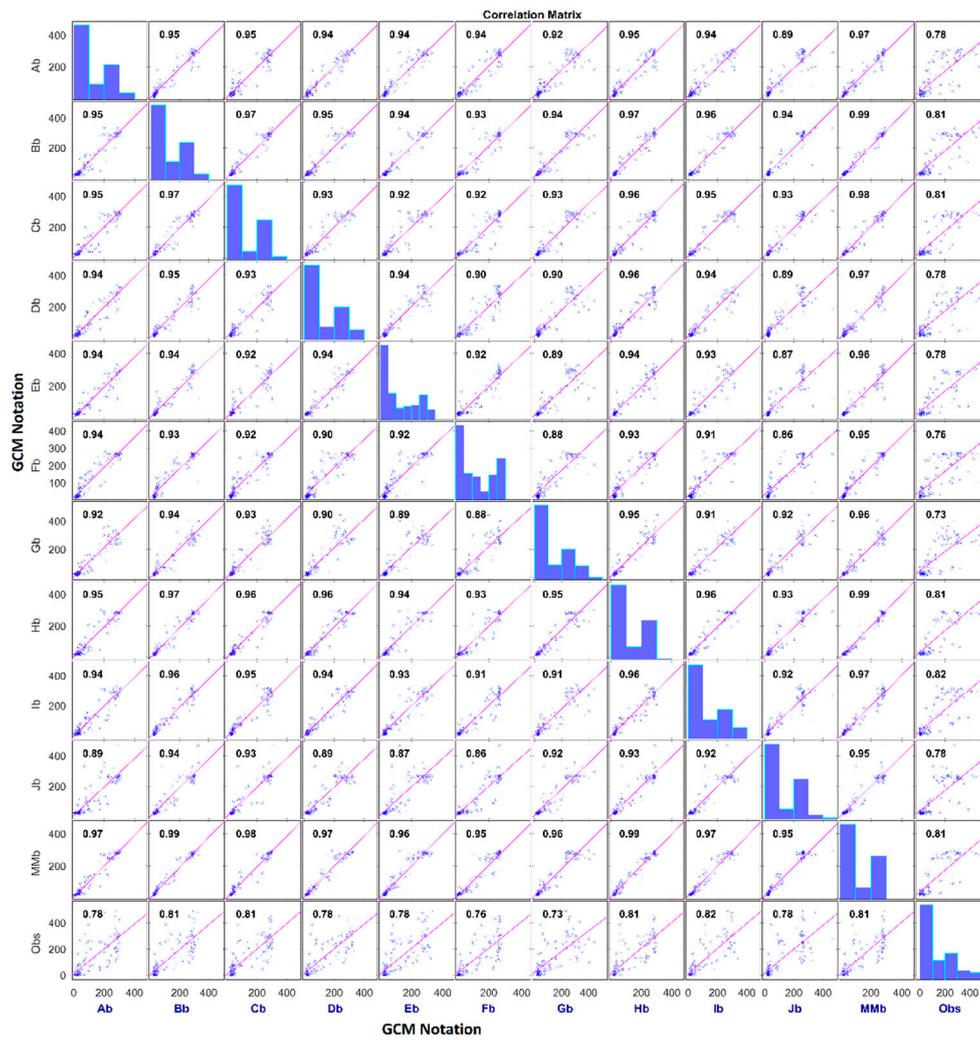


Figure 18. Correlation matrix of downscaled precipitation for different GCMs with the *GCM-Obs* model for Zone-4. Markings are the same ones as given in Figure 17.

NRMSD was again used to judge the convergence skill of both models in downscaled precipitation with different GCMs. To show the closeness of downscaled precipitation with different GCMs under the *Re-Obs* and *GCM-Obs* models, a *NRMSD* matrix is prepared as follows:

$$(NRMSD)_{X,Y}(\%) = \frac{1}{(\theta_{obs_max} - \theta_{obs_min})} \sqrt{\frac{\sum_{t=1}^N (\theta_{X,t} - \theta_{Y,t})^2}{N}} \times 100 \tag{40}$$

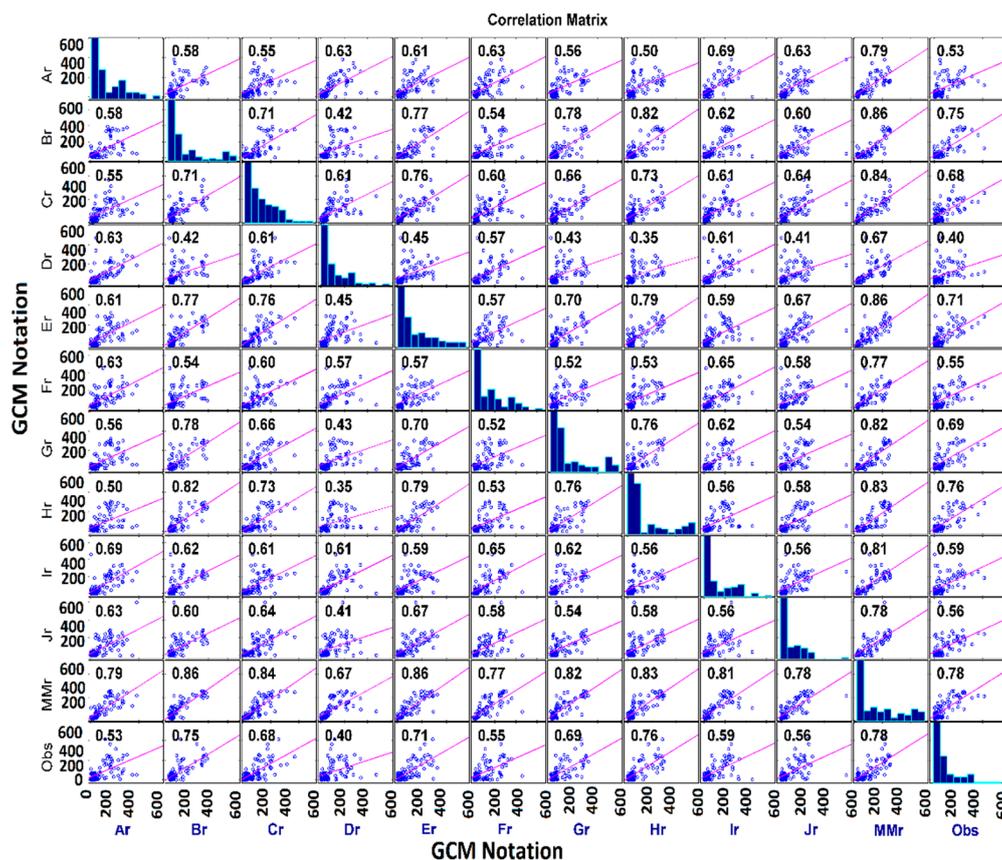


Figure 19. Correlation matrix of downscaled precipitation for different GCMs with the *Re-Obs* model for Uttarkashi. Markings are the same as the ones given in Figure 17.

Here, $(NRMSD)_{X,Y}$ is normalized root mean square deviation between downscaled precipitation for GCM_X and GCM_Y , $X, Y = 1, 2, \dots, 10$, $\theta_{X,t}$ and $\theta_{Y,t}$ are downscaled precipitation of GCM_X and GCM_Y at time t , $t = 1, 2, \dots, N$ number of months, θ_{obs_max} and θ_{obs_min} is maximum and minimum observed precipitation respectively.

The pattern of the *NRMSD* matrix of downscaled precipitation by both models for Zone 4 and Uttarkashi are shown in Figures 21 and 22, respectively. Two GCMs, i.e., FGOALS-g2 and GFDL-CM3, downscaled by the *Re-Obs* model show the highest variability with other GCMs and observed data for Zone 4 and Uttarkashi. The *NRMSD* value of the order of 70%–90% with the *Re-Obs* model is significantly reduced to 5%–15% by the *GCM-Obs* model for zone 4. Similarly, for Uttarkashi, the *NRMSD* of 15%–30% using *Re-Obs* model is reduced to 5–15% using *GCM-Obs* model. A significant reduction in the *NRMSD* value can also be seen with observed data using the *GCM-Obs* model. It can also be seen that the spatial resolution of GCM is not an influencing parameter in downscaling. A mixed pattern of downscaling performance is achieved on downscaling using coarser gridded CMCC-CESM to obtain relatively finer gridded CMCC-CM GCMs. Similar results were also found for other regions.

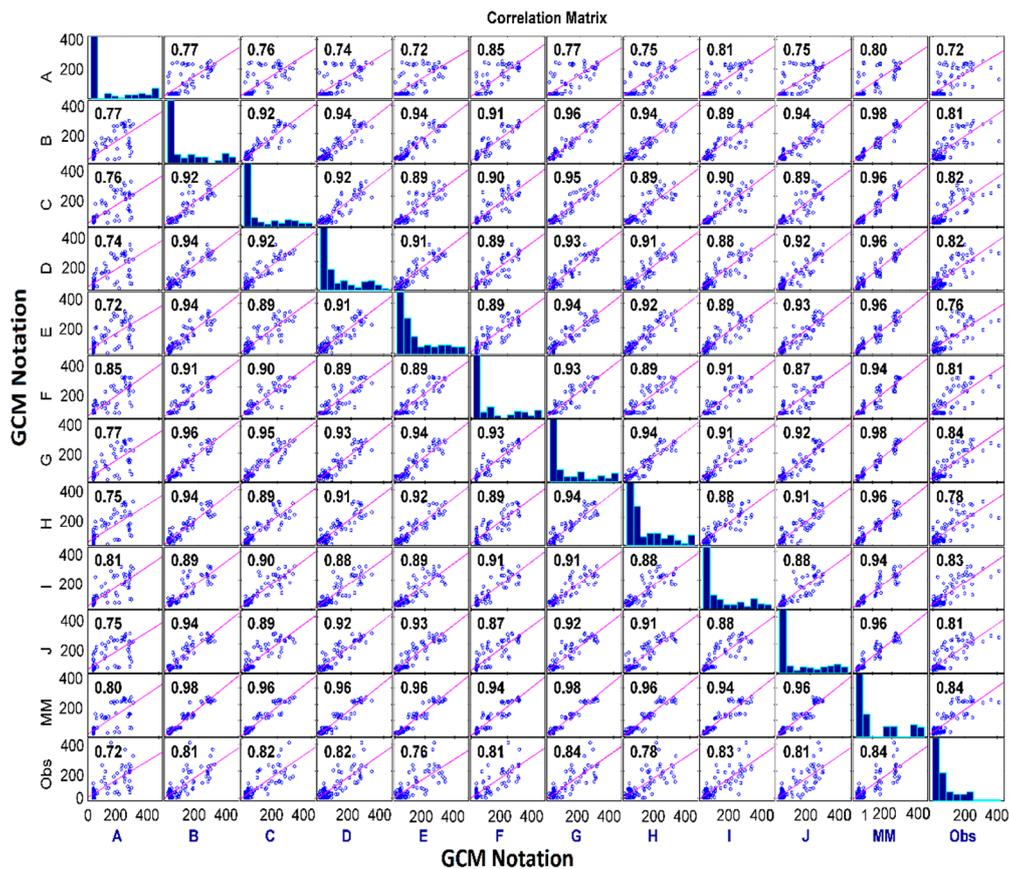


Figure 20. Correlation matrix of downscaled precipitation with different GCMs by the GCM-Obs model for Uttarkashi. Markings are the same as the ones given in Figure 17.

It can be noted from the results above that the GCM-Obs model performs better than the Re-Obs model in statistical downscaling. However, bias correction methods are generally adopted with the Re-Obs model to reduce GCM bias. Lie et al. [32] proposed Equidistant CDF matching (EDCDFm) bias correction method which is widely used to correct bias in monthly precipitation and temperature. The same method is used to correct the bias in downscaled precipitation using the Re-Obs and GCM-Obs models. The performance of both models considering EDCDFm bias correction method for zone-4 is presented by NRMSD matrix and correlation plots.

The correlation matrix for the bias corrected Re-Obs and GCM-Obs are shown in Figures 23 and 24, respectively. It can be observed from the Figures 17 and 23 that the bias correction method improved the downscaled precipitation. However, bias correction also improved the downscaled precipitation using the GCM-Obs model. So, the bias correction method improved the inter GCM and GCM-observed correlation and the GCM-Obs model performed better than the Re-Obs model.

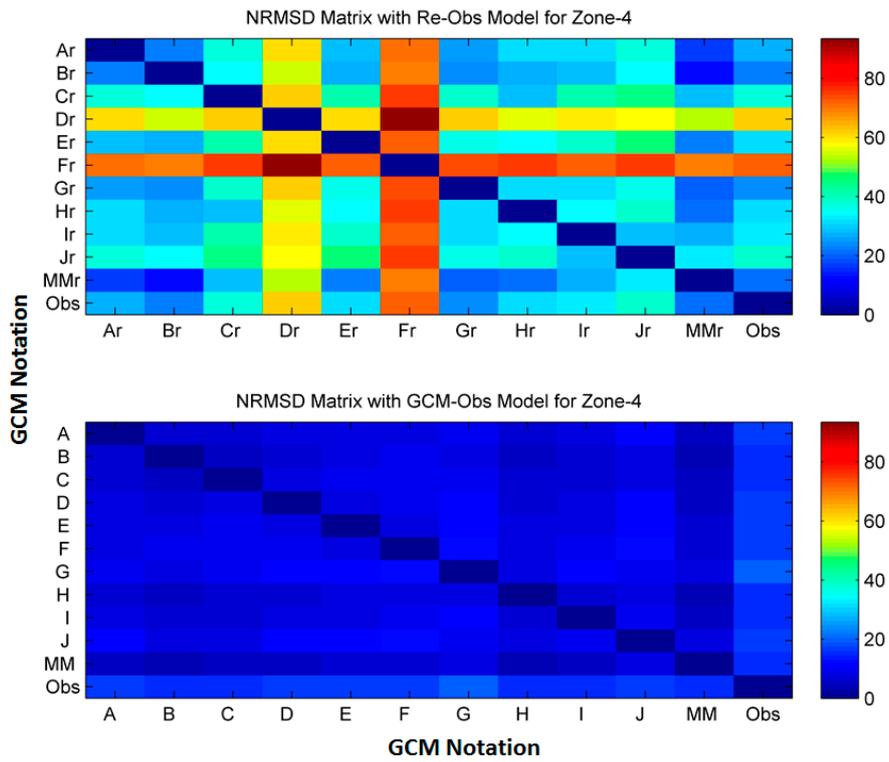


Figure 21. NRMSD matrix for downscaled precipitation for different GCMs with *Re-Obs* and the *GCM-Obs* model for zone 4. The coloured bar represents the NRMSD percentage. GCM notations are the same as the ones marked in Figure 17.

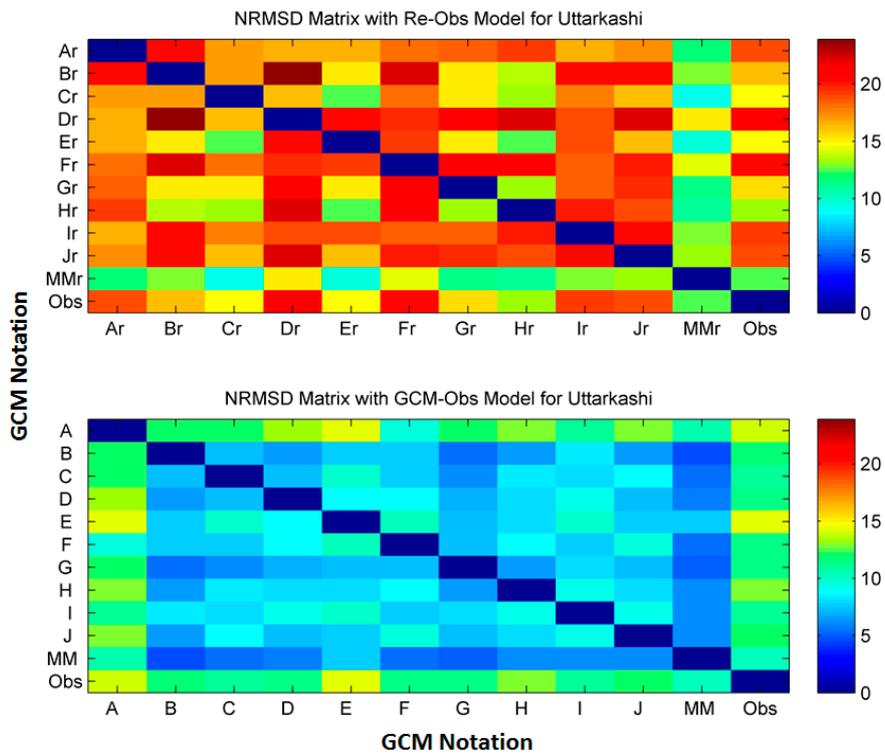


Figure 22. NRMSD matrix for downscaled precipitation for different GCMs with *Re-Obs* and the *GCM-Obs* model for Uttarkashi. The coloured bar represents the NRMSD percentage. GCM notations are the same as the ones marked in Figure 17.

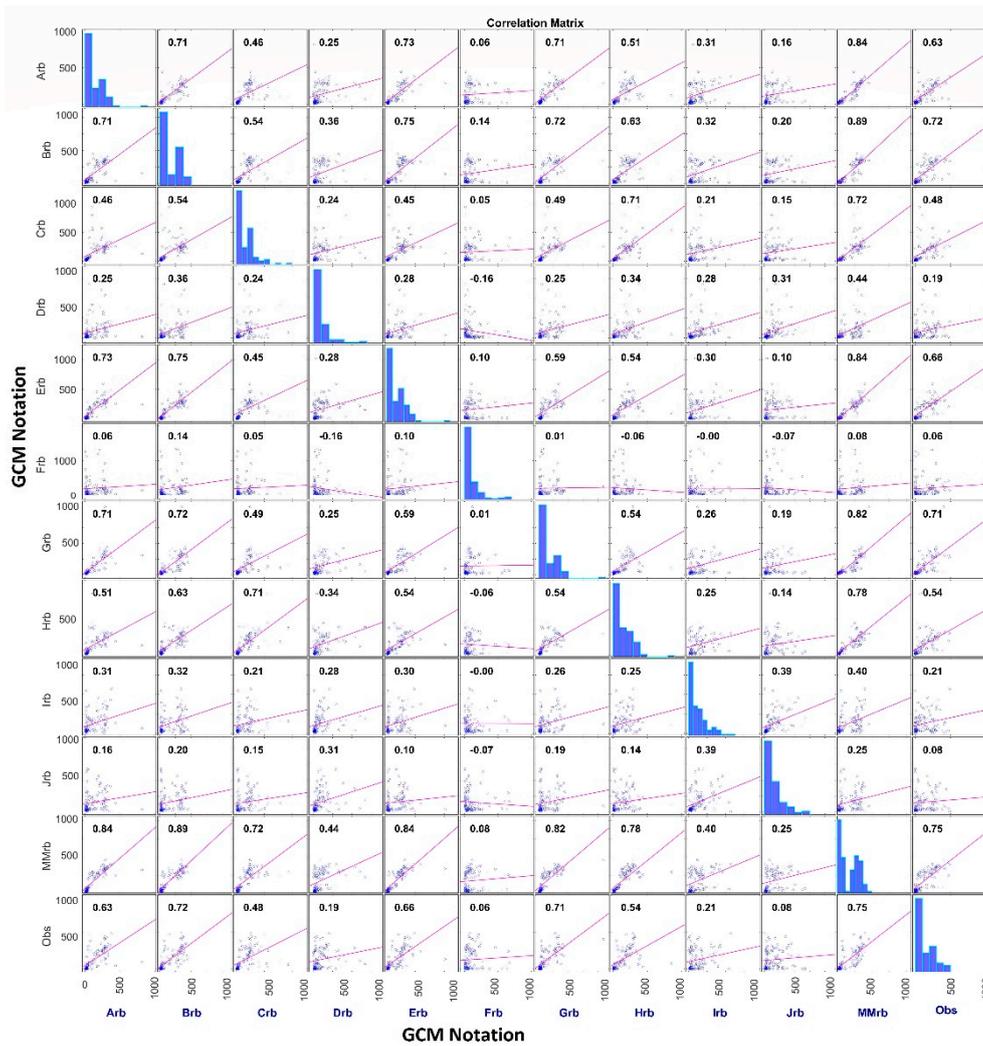


Figure 23. Correlation matrix of downscaled precipitation for different GCMs with the *Re-Obs* model for Zone 4. Notations are similar to Figure 17 except the GCM notations followed by the letter 'b' indicates corrected bias.

The *NRMSD* matrix of bias corrected downscaled precipitation using *Re-Obs* and *GCM-Obs* models is shown in Figure 25. Here also, the bias correction method improved the downscaling performance of both models, and the *GCM-Obs* model was found to be better than the *Re-Obs* model.

Measures adopted to judge the convergence skill of both models indicate the *GCM-Obs* model's capability to reduce GCM bias, and show a better convergence skill than the *Re-Obs* model with or without using bias correction methods. Overall, it can be conveyed that the *GCM-Obs* model performs better than the *Re-Obs* model.

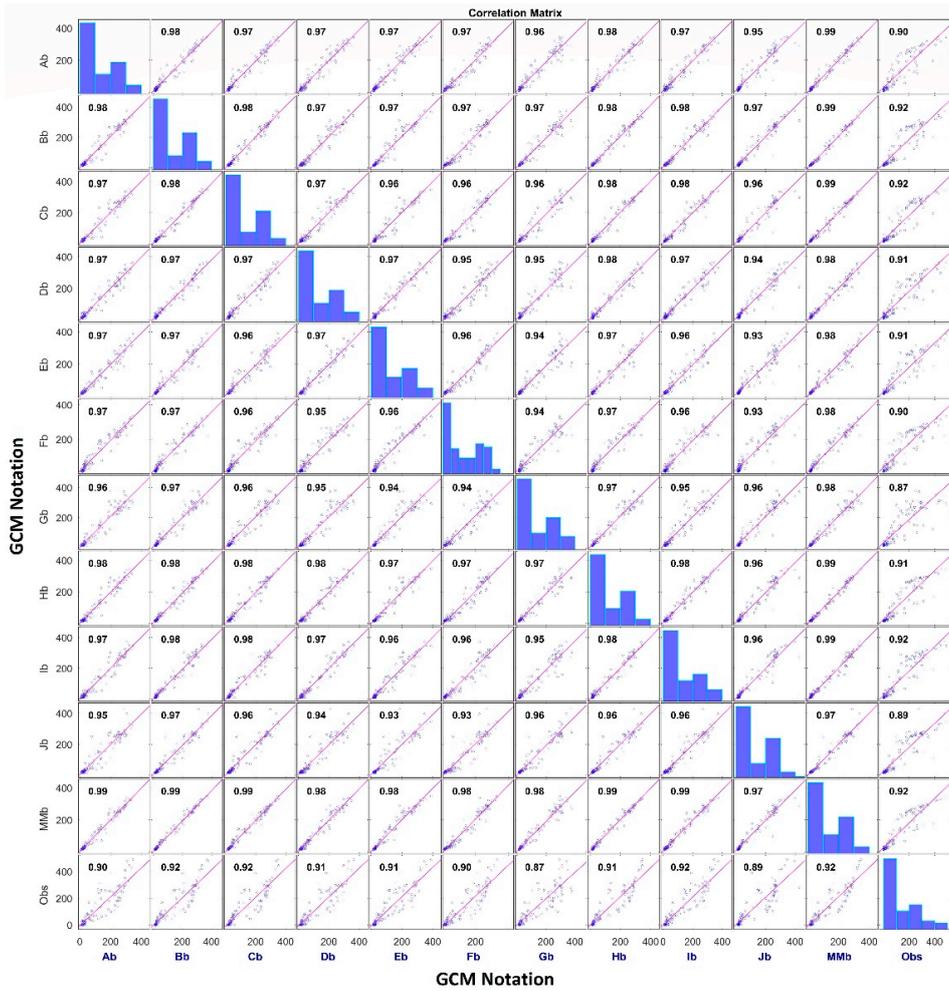


Figure 24. Correlation matrix of downscaled precipitation for different GCMs with the GCM-Obs model for Zone 4. Notations are similar to for Figure 23.

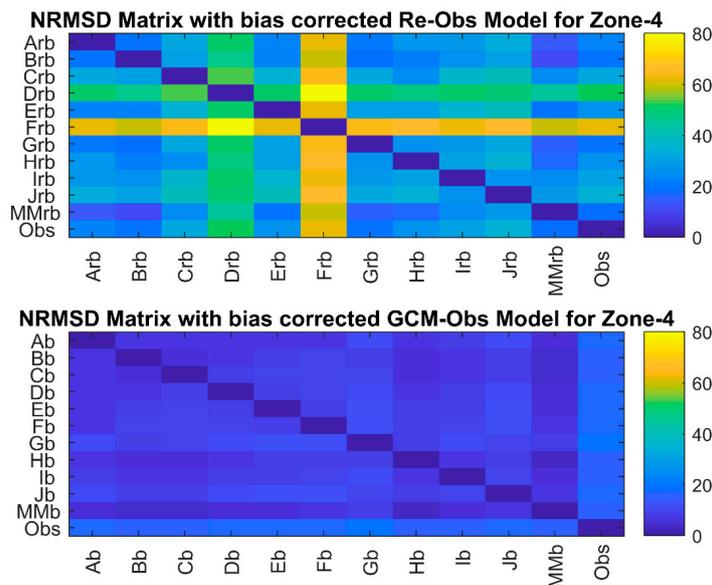


Figure 25. NRMSD matrix for downscaled precipitation using Re-Obs and GCM-Obs models post bias correction. Notations are similar to in Figure 23.

5. Discussion

This article discusses the choice of a better statistical downscaling model among *Re-Obs* and *GCM-Obs* to downscale GCM data. The former *Re-Obs* model is adopted by the majority of researchers. However, the bias in the GCM downscaled product from observations and the differences in GCM downscaled product using different GCMs is major concern of the climate researchers [8,15,21]. The *Re-Obs* and *GCM-Obs* models are compared using three methods: (i) Mathematical derivation, (ii) Synthetic Series and (iii) a case study considering real observed data. The Performance of the *GCM-Obs* model was found to be better than the *Re-Obs* model for statistically downscaling GCM data. It may be argued that if the reanalysis data is closer to the observed data, it should be able to better downscale GCM, but the *Re-Obs* model could produce better results if the reanalysis products are to be downscaled. However, if the goal is to downscale GCM, the obvious choice for the predictor-predictand model should be the *GCM-Obs* model, as it already considers the characteristics of the GCM that will be downscaled.

The performance of *Re-Obs* and the *GCM-Obs* model for statistically downscaling precipitation in different regions in the Ganga river basin was checked using different skill scores [35,41]. The case study indicates that the *GCM-Obs* model could be better choice for statistically downscaling GCM. The *GCM-Obs* model can be used to downscale different CMIP5 experiments [57], as reported in the IPCC AR5 for assessing different climate states while considering different assumptions.

6. Limitations of the Study

The study aimed to propose a different downscaling approach to improve the similarity of the downscaled variable and observed data and to reduce variability in downscaled data using different GCMs. Downscaling can be further improved by using a different set of predictor variables, downscaling methods, and other techniques/methods. In this study ten different GCMs are considered and the results can be verified with other GCMs as well. However, the results are likely to be in agreement with this study. Uncertainty is inherent with GCM due to different boundary conditions, equations, methods and other factors in the development of different GCMs. GCM uncertainty cannot be fully removed, but in this study an effort has been made to reduce the variability in downscaled variables using a different downscaling approach. Observed data is obtained from renowned agencies which take utmost care in sampling and production in datasets, but still errors in the data cannot be fully ruled out.

7. Conclusions

Statistical downscaling of precipitation in different regions in the Ganga river basin was carried out with ten different GCMs along with the *Re-Obs* and *GCM-Obs* models. The predictor-predictand relationship used to downscale GCM at a local scale was developed using reanalysis and historical GCM data as predictors in *Re-Obs* and *GCM-Obs* models, respectively. Different measures were adopted to judge the relative performances of *Re-Obs* and *GCM-Obs* models to downscale precipitation with different GCMs.

Downscaled precipitation with different GCMs by the *Re-Obs* and *GCM-Obs* models showed significant differences in each region of the study area. Although the predictor-predictand model showed good connection between reanalysis and observed data, but same model could not better simulate the downscaling of GCM. Higher variance and lesser correlation between modeled and observed precipitation was shown by the *Re-Obs* model. Development of the predictor-predictand relationship with historical data of GCM itself as the predictor and observed data as the predictand showed a high similarity and less variability in downscaled and observed precipitation. The skill score of downscaled precipitation also showed a more significant improvement with the *GCM-Obs* model than with the *Re-Obs* model. Datasets of downscaled precipitation with different GCMs using the *GCM-Obs* model fall near to each other in the form of clusters, as represented by the Taylor diagram.

Intercomparison of downscaled precipitation with different GCMs by the *Re-Obs* model also showed higher GCM-GCM bias, as indicated by the *NRMSD* correlation matrix. The *GCM-Obs* model significantly reduced GCM bias in downscaled precipitation for all regions and all GCMs. The *GCM-Obs* model showed its robustness in downscaling with different GCMs for all regions and both sets of observed data, i.e., GPCC and IWP. The *GCM-Obs* model was found to be more reliable in terms of performance and convergence skill than the *Re-Obs* model. Multi-model ensemble average precipitation data showed better resemblance with observed data than most of the individual GCM. It is also found that spatial resolution of the GCM does not have a considerable effect on performance and convergence skill for both models. The bias correction method also improves the downscaling performance of both *Re-Obs* and *GCM-Obs*.

It can be said that using historical GCM data to develop the predictor-predictand relationship is a better choice to simulate the precipitation and to reduce GCM uncertainty. The *GCM-Obs* model is robust against bias due to different data observing agencies. Predictor-predictand model development using historical data of GCM itself can be applied to downscale other atmospheric variables, i.e., temperature, humidity, evapotranspiration etc. Improvement in performance and reduction in GCM bias is also expected to downscale other variables under the *GCM-Obs* model, because precipitation is in least agreement with GCM in downscaling [58]. The bias correction measures may still be used to further improve the quality of the downscaled variable. This study will be helpful for climate change researchers to develop better downscaling models and to be more certain when they decide the ranges of downscaled atmospheric variables.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4441/11/10/2097/s1>.

Author Contributions: In this research work, data collection, their analysis and manuscript preparation were carried out by C.S. under the supervision of C.S.P.O. A.K.S., Q.B.P., N.T.T.L., C.M.F., H.H.L. and T.D.D. helped in fine tuning and improving the paper.

Funding: This research received no external funding.

Acknowledgments: The authors would like to appreciate the financial support received from Bold 2025 grant coded RJO: 10436494 by Innovation & Research Management Center (iRMC), Universiti Tenaga Nasional (UNITEN), Malaysia. In addition, the authors are thankful to Global Precipitation Climatology Center (GPCC) for providing the gridded precipitation and wet day frequency data.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A Development of Optimum Predictor-Predictand Model

Predictor-predictand relationship between coarse scale reanalysis/GCM predictors and local scale predictand was developed using multiple linear regression (MLR) and artificial neural network (NN) methods. Six predictor variables, i.e., temperature (*ta*), geopotential height (*zg*), specific humidity (*hus*), zonal and meridional wind components (*Ua* and *Va* respectively) and mean sea level pressure (*psl*) at different pressure levels (as given in Table 2) were used to develop predictor-predictand relationship. Predictor variables which have lesser correlation with observed data, i.e., correlation coefficient < 0.5, were not used in model development to eliminate least influencing datasets. Principal component analysis (PCA) was used to reduce the dimension of predictor variables and to reduce the collinearity among themselves. However, first six principal components show more than 95% variability, number of principal components to be used in predictor-predictand model is decided based on optimum model. The number of principal components which gave highest value of adjusted coefficient of determination (R^2_{adj}) for testing data sets was said to be optimum model. Instead of normal R^2 , R^2_{adj} was used because normal R^2 keeps on increasing by increase in number of independent variable but R^2_{adj} impose penalty to independent variables not having significant contribution in model development. R^2_{adj} is calculated as follows

$$R^2_{adj} = 1 - (1 - R^2) \frac{n - 1}{n - p - 1} \quad (A1)$$

Here, R^2 is coefficient of determination, n is number of samples and p is number of explanatory variables.

References

1. Wood, A.W.; Leung, L.R.; Sridhar, V.; Lettenmaier, D.P. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Clim. Chang.* **2004**, *62*, 189–216. [[CrossRef](#)]
2. Rockel, B.; Castro, C.L.; Pielke, R.A., Sr.; von Storch, H.; Leoncini, G. Dynamical downscaling: Assessment of model system dependent retained and added variability for two different regional climate models. *J. Geophys. Res. Atmos.* **2008**, *113*. [[CrossRef](#)]
3. Castro, C.L.; Pielke, R.A., Sr.; Leoncini, G. Dynamical downscaling: Assessment of value retained and added using the Regional Atmospheric Modeling System (RAMS). *J. Geophys. Res. Atmos.* **2005**, *110*. [[CrossRef](#)]
4. Lo, J.C.-F.; Yang, Z.-L.; Pielke, R.A. Assessment of three dynamical climate downscaling methods using the Weather Research and Forecasting (WRF) model. *J. Geophys. Res. Atmos.* **2008**, *113*. [[CrossRef](#)]
5. Pielke, R.A.; Wilby, R.L. Regional climate downscaling: What's the point? *Eos Trans. Am. Geophys. Union* **2012**, *93*, 52–53. [[CrossRef](#)]
6. Anandhi, A.; Frei, A.; Pierson, D.C.; Schneiderman, E.M.; Zion, M.S.; Lounsbury, D.; Matonse, A.H. Examination of change factor methodologies for climate change impact assessment. *Water Resour. Res.* **2011**, *47*, 1–10. [[CrossRef](#)]
7. Racherla, P.N.; Shindell, D.T.; Faluvegi, G.S. The added value to global model projections of climate change by dynamical downscaling: A case study over the continental U.S. using the GISS-ModelE2 and WRF models. *J. Geophys. Res. Atmos.* **2012**, *117*. [[CrossRef](#)]
8. Mondal, A.; Mujumdar, P.P. On the basin-scale detection and attribution of human-induced climate change in monsoon precipitation and streamflow. *Water Resour. Res.* **2012**, *48*, 1–18. [[CrossRef](#)]
9. Goyal, M.K.; Ojha, C.S.P. Application of PLS-Regression as Downscaling Tool for Pichola Lake Basin in India. *Int. J. Geosci.* **2010**, *1*, 51–57. [[CrossRef](#)]
10. Vu, M.T.; Aribarg, T.; Supratid, S.; Raghavan, S.V.; Liang, S. Statistical downscaling rainfall using artificial neural network: Significantly wetter Bangkok? *Theor. Appl. Climatol.* **2016**, *126*, 453–467. [[CrossRef](#)]
11. Ghosh, S.; Mujumdar, P.P. Statistical downscaling of GCM simulations to streamflow using relevance vector machine. *Adv. Water Resour.* **2008**, *31*, 132–146. [[CrossRef](#)]
12. Kalnay, E.; Kanamitsu, M.; Kistler, R.; Collins, W.; Deaven, D.; Gandin, L.; Iredell, S.; Saha, S.; White, G.; Zhu, Y.; et al. The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Am. Meteorol. Soc.* **1996**, *77*, 437–471. [[CrossRef](#)]
13. Dee, D.P.; Uppala, S.M.; Simmons, A.J.; Berrisford, P.; Poli, P.; Kobayashi, S.; Andrae, U.; Balmaseda, M.A.; Balsamo, G.; Bauer, P.; et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **2011**, *137*, 553–597. [[CrossRef](#)]
14. Charles, S.P.; Bari, M.A.; Kitsios, A.; Bates, B.C. Effect of GCM bias on downscaled precipitation and runoff projections for the Serpentine catchment, Western Australia. *Int. J. Climatol.* **2007**, *27*, 1673–1690. [[CrossRef](#)]
15. Sharma, C.; Arora, H.; Ojha, C.S.P. Assessment of the Effect of Climate Change on Historical and Future Rainfall in Uttarakhand. In Proceedings of the Hydro-2015 International Conference, Roorkee, India, 17–19 December 2015.
16. Duhan, D.; Pandey, A. Statistical downscaling of temperature using three techniques in the Tons River basin in Central India. *Theor. Appl. Climatol.* **2015**, *121*, 605–622. [[CrossRef](#)]
17. Goyal, M.K.; Ojha, C.S.P.; Burn, D.H. Nonparametric statistical downscaling of temperature, precipitation, and evaporation in a semiarid region in India. *J. Hydrol. Eng.* **2011**, *17*, 615–627. [[CrossRef](#)]
18. Okkan, U.; Inan, G. Statistical downscaling of monthly reservoir inflows for Kemer watershed in Turkey: Use of machine learning methods, multiple GCMs and emission scenarios. *Int. J. Climatol.* **2015**, *35*, 3274–3295. [[CrossRef](#)]
19. Fowler, H.J.; Blenkinsop, S.; Tebaldi, C. Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling. *Int. J. Climatol.* **2007**, *27*, 1547–1578. [[CrossRef](#)]
20. Wilby, R.L.; Charles, S.P.; Zorita, E.; Timbal, B.; Whetton, P.; Mearns, L.O. Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods. *Zenodo* **2004**. [[CrossRef](#)]

21. Stocker, T.F.; Qin, D.; Plattner, G.K.; Tignor, M.; Allen, S.K.; Boschung, J.; Nauels, A.; Xia, Y.; And, V.B.; Midgley, P.M. *IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2013.
22. Sunyer, M.A.; Madsen, H.; Ang, P.H. A comparison of different regional climate models and statistical downscaling methods for extreme rainfall estimation under climate change. *Atmos. Res.* **2012**, *103*, 119–128. [[CrossRef](#)]
23. Hughes, D.A.; Mantel, S.; Mohobane, T. An assessment of the skill of downscaled GCM outputs in simulating historical patterns of rainfall variability in South Africa. *Hydrol. Res.* **2014**, *45*, 134–137. [[CrossRef](#)]
24. Ahmed, K.F.; Wang, G.; Silander, J.; Wilson, A.M.; Allen, J.M.; Horton, R.; Anyah, R. Statistical downscaling and bias correction of climate model outputs for climate change impact assessment in the U.S. northeast. *Glob. Planet. Chang.* **2013**, *100*, 320–332. [[CrossRef](#)]
25. Bonfils, C.; Santer, B.D.; Pierce, D.W.; Hidalgo, H.G.; Bala, G.; Das, T.; Barnett, T.P.; Cayan, D.R.; Doutriaux, C.; Wood, A.W.; et al. Detection and Attribution of Temperature Changes in the Mountainous Western United States. *J. Clim.* **2008**, *21*, 6404–6424. [[CrossRef](#)]
26. Reeves, J.; Chen, J.; Wang, X.L.; Lund, R.; Lu, Q.Q. A Review and Comparison of Change point Detection Techniques for Climate Data. *J. Appl. Meteorol. Climatol.* **2007**, *46*, 900–915. [[CrossRef](#)]
27. Agarwal, A.; Babel, M.S.; Maskey, S.; Shrestha, S.; Kawasaki, A.; Tripathi, N.K. Analysis of temperature projections in the Koshi River Basin, Nepal. *Int. J. Climatol.* **2015**, *36*, 266–279. [[CrossRef](#)]
28. Barria, P.; Walsh, K.J.E.; Peel, M.C.; Karoly, D. Uncertainties in runoff projections in southwestern Australian catchments using a global climate model with perturbed physics. *J. Hydrol.* **2015**, *529*, 184–199. [[CrossRef](#)]
29. Jiang, Y.; Kim, J.B.; Still, C.J.; Kerns, B.K.; Kline, J.D. Erratum: Inter-comparison of multiple statistically downscaled climate datasets for the Pacific Northwest, USA. *Sci. Data* **2018**, *5*, 180050. [[CrossRef](#)] [[PubMed](#)]
30. Pan, Z.; Christensen, J.H.; Arritt, R.W.; Gutowski, W.J.; Takle, E.S.; Otieno, F. Evaluation of uncertainties in regional climate change simulations. *J. Geophys. Res. Atmos.* **2001**, *106*, 17735–17751. [[CrossRef](#)]
31. Cayan, D.R.; Maurer, E.P.; Dettinger, M.D.; Tyree, M.; Hayhoe, K. Climate change scenarios for the California region. *Clim. Chang.* **2007**, *87*, 21–42. [[CrossRef](#)]
32. Li, H.; Sheffield, J.; Wood, E.F. Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *J. Geophys. Res. Atmos.* **2010**, *115*, D10101. [[CrossRef](#)]
33. Johnson, F.; Sharma, A. A nesting model for bias correction of variability at multiple time scales in general circulation model precipitation simulations. *Water Resour. Res.* **2012**, *48*. [[CrossRef](#)]
34. Mehrotra, R.; Sharma, A. An improved standardization procedure to remove systematic low frequency variability biases in GCM simulations. *Water Resour. Res.* **2012**, *48*. [[CrossRef](#)]
35. Nguyen, H.; Mehrotra, R.; Sharma, A. Correcting for systematic biases in GCM simulations in the frequency domain. *J. Hydrol.* **2016**, *538*, 117–126. [[CrossRef](#)]
36. Immerzeel, W. Historical trends and future predictions of climate variability in the Brahmaputra basin. *Int. J. Climatol.* **2008**, *28*, 243–254. [[CrossRef](#)]
37. Dharmaveer, S.; Jain, S.K.; Dev, G.R. Trend in Observed and Projected Maximum and Minimum Temperature over N-W Himalayan Basin. *J. Mt. Sci.* **2015**, *12*, 417–433.
38. Dhar, O.N.; Nandargi, S. Flood study of the Himalayan tributaries of the Ganga river. *Meteorol. Appl.* **2002**, *9*, 63–68. [[CrossRef](#)]
39. Bharati, L.; Lacombe, G.; Gurung, P.; Jayakody, P. *The Impacts of Water Infrastructure and Climate Change on the Hydrology of the Upper Ganges River Basin*; International Water Management Institute: Colombo, Sri Lanka, 2011; ISBN 9789290907442.
40. Sharma, C.; Ojha, C.S.P. Climate Change Detection in Upper Ganga River Basin. In Proceedings of the International Conference on Sustainable Technologies for Intelligent Water Management, Roorkee, India, 16–19 February 2018.
41. Giorgi, F.; Mearns, L.O. Calculation of Average, Uncertainty Range, and Reliability of Regional Climate Changes from AOGCM Simulations via the “Reliability Ensemble Averaging” (REA) Method. *J. Clim.* **2002**, *15*, 1141–1158. [[CrossRef](#)]
42. Dessai, S.; Lu, X.; Hulme, M. Limited sensitivity analysis of regional climate change probabilities for the 21st century. *J. Geophys. Res.* **2005**, *110*, D19108. [[CrossRef](#)]

43. Evans, J.; McGregor, J.; McGuffie, K. Chapter 9-Future Regional Climates. In *The Future of the World's Climate*, 2nd ed.; Henderson-Sellers, A., McGuffie, K., Eds.; Elsevier: Boston, MA, USA, 2012; pp. 223–250. ISBN 978-0-12-386917-3.
44. Sen Gupta, A.; McNeil, B. Chapter 6-Variability and Change in the Ocean. In *The Future of the World's Climate*, 2nd ed.; Henderson-Sellers, A., McGuffie, K., Eds.; Elsevier: Boston, MA, USA, 2012; pp. 141–165. ISBN 978-0-12-386917-3.
45. George, G. *The Impact of Climate Change on European Lakes | Glen George*; Springer: Dordrecht, The Netherlands. Available online: <https://www.springer.com/gp/book/9789048129447> (accessed on 4 August 2019).
46. The Impacts of Climate Change and Urbanisation on Drainage in Helsingborg, Sweden: Combined sewer system-ScienceDirect. Available online: <https://www.sciencedirect.com/science/article/pii/S0022169407002910> (accessed on 4 August 2019).
47. India-Wris Water Resources Information System of India Website. Available online: <http://india-wris.nrsc.gov.in/wrpinfo/index.php?title=Ganga> (accessed on 9 June 2015).
48. NIH National Institute of Hydrology Website. Available online: http://www.nih.ernet.in/rbis/basinmaps/ganga_about.htm (accessed on 12 May 2015).
49. Peel, M.C.; Finlayson, B.L.; McMahon, T.A. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 1633–1644. [[CrossRef](#)]
50. Udo, M.S.; Andreas, B.; Peter, F.; Meyer-Christoffer, A.; Bruno, R.; Ziese, M. *GPCC Full Data Reanalysis Version 6.0 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-Based and Historic Data*; Global Precipitation Climatology Centre: Offenbach, Germany, 2011. [[CrossRef](#)]
51. Harris, I.; Jones, P.D.; Osborn, T.J.; Lister, D.H. Updated high-resolution grids of monthly climatic observations-the CRU TS3.10 Dataset. *Int. J. Climatol.* **2014**, *34*, 623–642. [[CrossRef](#)]
52. Sharma, C.; Ojha, C.S.P. Statistical Parameters of Hydrometeorological Variables: Standard Deviation, SNR, Skewness and Kurtosis. In *Advances in Water Resources Engineering and Management*; Select Proceedings of TRACE 2018; Springer: Singapore, 2019; Volume I, p. 257. ISBN 9811381801.
53. Koutroulis, A.G.; Grillakis, M.G.; Tsanis, I.K.; Papadimitriou, L. Evaluation of precipitation and temperature simulation performance of the CMIP3 and CMIP5 historical experiments. *Clim. Dyn.* **2016**, *47*, 1881–1898. [[CrossRef](#)]
54. Ahmadalipour, A.; Rana, A.; Moradkhani, H.; Sharma, A. Multi-criteria evaluation of CMIP5 GCMs for climate change impact analysis. *Theor. Appl. Climatol.* **2017**, *128*, 71–87. [[CrossRef](#)]
55. Mondal, A.; Mujumdar, P.P. On the detection of human influence in extreme precipitation over India. *J. Hydrol.* **2015**, *529*, 1161–1172. [[CrossRef](#)]
56. Taylor, K.E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* **2001**, *106*, 7183–7192. [[CrossRef](#)]
57. Taylor, K.E.; Stouffer, R.J.; Meehl, G.A. An Overview of CMIP5 and the Experiment Design. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 485–498. [[CrossRef](#)]
58. Johnson, F.; Sharma, A. Measurement of GCM skill in predicting variables relevant for hydroclimatological assessments. *J. Clim.* **2009**, *22*, 4373–4382. [[CrossRef](#)]

