



Selection of suitable predictors and predictor domain for statistical downscaling over the Western Himalayan region of India

Jitendra Kumar Meher¹ · Lalu Das¹

Received: 18 June 2018 / Accepted: 19 August 2019 / Published online: 30 August 2019
© Springer-Verlag GmbH Austria, part of Springer Nature 2019

Abstract

Selection of suitable predictor(s) from the NCEP/NCAR reanalysis datasets for downscaling annual and seasonal rainfall over the Western Himalayas has been carried out in the present study. Size of the domain on downscaling was also judged by considering three different sizes of domains, namely Western Himalayan region (WHR), India and South Asia. Statistical measures like spatial correlation maps, product-moment correlations, and adjusted R^2 of regression analysis were used to evaluate the skills of the predictors. Results showed predictors were sensitive to the method of analysis, choice of season, and size of the domain. A majority of the predictors exhibited stronger spatial correlations (\pm) in annual and monsoon season compared to the winter. It was found that the first principal components (PCs) of most of the predictors were consistently well correlated (R_E) with the annual and monsoon rainfall in all domains, whereas, in the winter season, none of the PCs showed such consistent results. During the monsoon season, the predictors had higher R_E values than the winter and annual time scale. Geopotential height at 850 hPa, relative humidity at 500 and 1000 hPa, and precipitation rate emerged as good predictors for downscaling precipitation over different predictor domains. On the other hand, the geopotential height at 500 and 850 hPa, v at 500 hPa, specific humidity at 500 hPa, and divergence at 850 hPa resulted as least affected predictors based on analysis of ranks of the predictors. Finally, WHR was considered as a suitable predictor domain for downscaling monsoon rainfall for the Western Himalayan region compared to other domains as ranks obtained for different predictors in this domain are not very sensitive to statistical measures used to evaluate the skills of predictors.

1 Introduction

General circulation models (GCMs) are widely used sophisticated tools to study climate as well as the large-scale upper-air features of our Earth; however, they do not give reliable information at the local scale (Hanssen-Bauer et al. 2003; Eden and Widmann 2014; Das et al. 2016; Gaur and Simonovic 2017). Direct output from the GCMs has limitations at subregional or local scale due to their scale mismatch. To overcome this scale differences, several downscaling methods have emerged to bridge the gap between the large-scale coarser resolution of

GCMs simulations and the local-scale higher-resolution information required for climate impact studies (Wilby and Wigley 1997; Huth 1999; Eden and Widmann 2014; Meher 2019). The empirical-statistical downscaling (ESD) method is one of them. The statistical downscaling involves developing empirical relationships between large-scale atmospheric predictors (for example, mean sea level pressure, geopotential height, humidity, or the wind) and local-scale surface predictands (for example temperature or rainfall at a weather station) (Hanssen-Bauer et al. 2005; Dabanlı and Şen 2017). There are three categories of statistical downscaling techniques developed so far, namely, (i) weather classification or weather typing, (ii) regression/transfer function, and (iii) weather generators (Wilby et al. 2004; Anandhi et al. 2009; Hofer 2010; Blazak 2012; Kannan and Ghosh 2013). The first two categories of approach may involve a perfect prognosis approach (Kannan and Ghosh 2013) which use empirical models that relate observation-based predictand and large-scale predictor during a common time period and then applied to simulated predictors (for example GCM scenario runs) for the future (Kannan and Ghosh 2013; Eden and Widmann

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00704-019-02980-z>) contains supplementary material, which is available to authorized users.

✉ Lalu Das
daslalu@yahoo.co.in

¹ Department of Agricultural Meteorology and Physics, Bidhan Chandra Krishi Viswavidyalaya, Mohanpur, Nadia, West Bengal 741252, India

2014). Perfect prognosis approach is based on the assumption that the relationship between simulated predictors and the predictands will remain consistent in the future (Blazak 2012).

There is no general consensus regarding appropriate selection of suitable predictor variables (Hu et al. 2013; Eden and Widmann 2014) for developing downscaling models. Selection of appropriate predictor is sensitive to the domain under consideration, predictand to be downscaled, attributes of the prevailing large-scale circulation, seasonality, and the topographic context, etc. (Anandhi et al. 2008; Anandhi et al. 2009; Forland et al. 2011). Earlier studies on predictor selection (Wilby et al. 1999; Wilby et al. 2004; Ghosh and Mjujundar 2006; Anandhi et al. 2009; Shashikanth and Ghosh 2013; Salvi and Ghosh 2013) reported that the suitable predictors must have the following characteristics: Firstly, statistical features of the predictors need to be well reproduced by GCMs and reanalysis data products. Secondly, they ought to be strongly correlated with the considered predictand. Thirdly, they should be physically and/or conceptually sensible. It is often advised to experiment with different geographical domain (preferably larger than the targeted predictand domain) while selecting a suitable predictor for a target region (Wilby and Wigley 2000; Sauter and Venema 2011) because a smaller domain over the target region may fail to capture the strongest correlation between predictand and the predictor (Wilby and Wigley 2000). In the case of precipitation downscaling for a particular location, the optimal predictor domain should be selected in such a way that all domains should capture the mechanism that leads to the formation of precipitation over that location. For example, rainfall downscaling study by Anandhi et al. (2008) and Hu et al. (2013) screened various predictors from the NCEP/NCAR (National Centers for Environmental Prediction/National Center for Atmospheric Research) reanalysis datasets on the basis of predictor's role in generating monsoon rainfall over the "Malaprabha river basin of India" and "Yellow River source region of China" respectively.

The major portion of annual rainfall (70–80%) over the Indian subcontinent occurs in the south-west monsoon season from June to September due to the large-scale monsoonal wind circulation. Similarly, the winter precipitation over the Western Himalayan region (WHR) of India occurs in the cooler winter season during December to February due to another wind flow originated from the extratropical region which is commonly known as western disturbances (Dimri et al. 2015; Dimri et al. 2016; Das and Meher 2019). Lower tropospheric planetary waves over mid-latitudes play a significant role in generating monsoon rainfall over India (Bawiskar 2005) and its neighboring regions like Western Himalayas (Priya et al. 2016; Meher 2019) and the Indus basin (Saeed et al. 2013). Studies by some scholars showed that meridional (or *v* wind) velocities (Bawiskar et al. 2005; Parthasarathy et al. 1991) and mean sea level pressure

(Douville 2006) are the major parameters which play a significant role in the occurrence of all-India monsoon rainfall. The mean sea level pressure (mslp) can be directly linked to the south-west monsoon rainfall over India (Douville 2006) through a pressure gradient developed between the Thar Desert (low pressure) and the Bay of Bengal (high pressure) during the active phase of the south-west monsoon season. The monsoonal circulation intensifies over the Indian region with the increase of pressure gradient, causing the increased moisture advection. Saeed et al. (2013) reported that geopotential height over central Asia could be used as a potential predictor to serve as a precursor for the rainfall in the upper Indus basin region. Pervez and Henebry (2014) reported that precipitation over two major river basins of South Asia namely the Ganges and the Brahmaputra were significantly influenced by various predictors like geopotential height, *u* wind (wind flow from east to west or across lines of latitude), *v* wind (wind traveling from south to north or across lines of longitude) (at 850 and 1000 hPa pressure level), and specific humidity (at 500 and 1000 hPa pressure level) while the influence of air temperature was found to be poor. Dimri et al. (2016) reported that the mid-tropospheric circulation due to zonal wind (*u*-wind) and geopotential at 500 and 850 hPa play a crucial role in convergence for triggering the winter precipitation over the Western Himalayas. Statistical metrics for evaluating independent predictors against a predictand is not firmly established in the literature. We have reviewed the works of several scholars (see Table 1) leading to the selection of suitable predictors for different statistical downscaling studies and found that scatter plots, partial correlation, and stepwise regression are some of the commonly used tools to select a suitable subset of predictors from the reanalysis datasets. For more reviews on predictor used in different downscaling studies, readers are advised to follow the work by Anandhi et al. (2008). In the present study, authors have reviewed numerous literatures (mentioned in the above paragraphs and Table 1) which have pointed out different upper-air large-scale predictors and surface variables that can be possibly taken as suitable predictors to downscale precipitation over different regions of South Asian domain and Indian region (excluding the north-western part) in specific. The Western Himalayan region of India is one of the less monitored regions of the globe in terms of downscaling studies. In the present paper, the authors have taken the initiative to select some of the suitable predictors that will help for the future downscaling studies over this region. The purpose of the present paper is to choose suitable predictors for reliably predicting station level precipitation over the WHR using some conventional statistical techniques such as correlation maps, EOF-based variance analysis, and calculating correlation coefficient. All the predictors were exhaustively evaluated through several statistical measures over three different

Table 1 A brief literature review on the selection of suitable predictors using the NCEP/NCAR datasets in different domains over India and South Asia. Bold references show that the work has been carried out nearby the Western Himalayas. See [Appendix](#) for the long name of the variables

Reference	Study region	Predictor variables used from NCEP/NCAR reanalysis data	Method of selection	Downscaled climate variable
Anandhi et al. (2008)	Malaprabha river basin, India	t0925, t0700, t0500, t0200, z0925, z0500, z0200, z0925, z0850, u0925, u0200, prw and sp.	Monsoon rainfall influence on probable predictors	Precipitation
Goyal and Ojha (2010)	Pichola lake catchment, Rajasthan, India	t0200, t0500, t0925, u0200, u0925, v0200, v0925, z0200 and z0925	Partial correlation and scatter plot between predictor and predictand	Precipitation
Ojha et al. (2010)	Pichola lake region in Rajasthan state in India	t0925, t0500, t0200, z0500, z0200, u0925, u0200, v0925, v0200	Partial correlation and scatter plot between predictor and predictand	Precipitation
Huang et al. (2011)	Yangtze river basin, China	mslp, sf, su, sv, sz, u0500, v0500, u0850, v0850, z0500, z0850, r0500, r0850, r, s, t2 m	Correlation and partial correlation analysis among the predictand and the individual predictors	Precipitation
Goyal and Ojha (2012)	Pichola lake catchment, Rajasthan, India	t0200, t0500, t0925, u0200, u0925, v0200 and v0925	Partial correlation and scatter plot between predictor and predictand	Temperature
Mahmood and Babel (2013)	Jhelum basin, Pakistan, and India	mslp, r, s, r500, and t2m	Correlation and partial correlation analysis among the predictand and the individual predictors, <i>P</i> -values	Temperature and Precipitation
Hu et al. (2013)	Yellow river basin, China	Air temp, s, mslp, u and v at various pressure levels (Exact levels were not mentioned in the article)	Correlation analysis between predictor and predictand	Precipitation
Devak and Dhanya (2014)	Mahanadi river basin, India	t0200, t0500, t0700, t0925, z0200, z0500, z0925, u0200, u0925, v0200 and v0925	On the basis of monsoon rainfall influence on probable predictors	Precipitation
Pervez and Henebr-y (2014)	The Ganges and the Brahmaputra river basins, South Asia	z0500, z0850, v0850, v0500, v1000, s0500, s0850, s1000, u500, u0850, u1000, mslp, t2 m, precipitation	Correlation coefficients and literature review regarding influence of snow and monsoon precipitation on predictors	Precipitation
Akhter et al. (2019)	Seven homogeneous zones of India	z500, z850, z1000, u500, u850, u1000, v500, v850, v1000, sh500, s850, s1000, r500, r850, r1000, ta500, t850, t1000, mslp, prw	Variance explained by different principal components; multiple linear regressions of combined EOFs.	Precipitation

domains to ensure reliable choice of suitable predictors for estimating annual and seasonal rainfall over WHR through statistical downscaling techniques. With this background information, the present study was carried out with the following objectives:

1. To show how different techniques and domain size are sensitive towards selecting appropriate predictors over WHR.
2. To select domain wise potential predictor on the basis of ranks obtained from the predictor-predictand relationship.

The rest of the description on the present investigation has been divided into three major sections. Section 2 of the paper gives a short description of the study region, the data used, and the detailed methodology used in the present work. Section 3 of the article provides a detailed account of the results and discussion. The key messages or the conclusions from the present work are inscribed in Sect. 4.

2 Data used and methodology of predictor selection

The study area is the Western Himalayan region of India (WHR), extending between 28°42' to 33°12'N and 75°34' to 81°05'E and comprises the two northern states of India namely Himachal Pradesh and Uttarakhand. Monthly gridded rainfall data (Pai et al. 2014) from the India Meteorological Department (IMD) was used in the present study. The area-averaged rainfall over the two states mentioned above was treated as the reference predictand. Besides the observational data, 24 numbers of large-scale atmospheric variables extracted from the NCEP/NCAR reanalysis dataset (Kalnay et al. 1996) on a 2.5° × 2.5° grid over the same time period as the observation data (1951–2005) were engaged for the present study. These variables include geopotential height, zonal and meridional wind speeds, specific humidity, relative humidity, divergence at various pressure levels, vorticity, wind speed, sea level pressure, precipitable water content, precipitation rate, and air temperature (see Table 2 for details).

2.1 Spatial region and correlation mapping

As downscaling results are sensitive to the size of the predictor domain, the developing downscaling model considering the different size of domains may provide more reliable information for policymaking. Forland et al. (2011) showed that smaller predictor domain is more reliable than the larger ones; however, the GCMs have a minimum skillful scale, and the local state is expected to depend on ambient large-scale conditions. To justify which size of domain will provide reliable downscaling results, three different sizes of domains namely (i) South Asia (10°S–40°N, 20°–120°E), (ii) India (8–38°N, 68–98°E) and (iii) the Western Himalayan region (27–38°N, 72–82°E) have been considered in the present work. For each domain, we separately tested how well different predictor variables can reproduce the observed feature of rainfall over the WHR. First, the following statistical analyses were performed over the bigger domain of South Asia for selecting the suitable predictor(s) and thereby, the same procedure was continued for another two domains as stated above.

Spatial correlation or pattern correlation coefficient has been a commonly used metric for quantifying the similarities between predictands and spatial patterns of the predictors (Srinivasan et al. 1995; Parding et al. 2019). It is quantified through the calculation of correlation coefficient between predictor's data at each grid point

and the predictand for a common time period. Spatial correlation maps were used to visualize those regions on the map where the correlation coefficients were higher than the other regions.

2.2 Multiple regression and temporal correlation

Predictor selection often requires a transformation of the raw predictors into a useful form because the information in the nearby grid boxes in the predictor data is not independent of each other (Maraun 2010). Empirical orthogonal function (EOF) analysis, or more generally principal component analysis, is a prominent technique for reducing higher dimensional fields (for example raw predictors) into a set of orthogonal basis vectors that are linearly independent (uncorrelated) to each other (Lorenz 1963; Hannachi et al. 2007). One merit of EOF analysis is that the orthogonal basis vectors reduce the problem of co-variability in subsequent regression analysis, and a small set of components capture most of the variability (often > 90% in its first seven vectors) through a lower dimensional representation of the original data (Huth 1999; Maraun 2010).

First of all, we subtracted the long-term (1951–2005) mean of the observed rainfall from the raw rainfall data (or predictand). In case of predictor variables, the long-term mean was subtracted

Table 2 Twenty-four NCEP/NCAR large-scale variables taken for the present study is given in the first column. For simplicity, the variables are represented by a variable I.D. in the second column

NCEP variables	Variable I.D.
Monthly mean geopotential height 500	z0500
Monthly mean geopotential height 850	z0850
Monthly mean geopotential height 1000	z1000
Monthly mean u wind 500	u0500
Monthly mean u wind 850	u0850
Monthly mean u wind 1000	u1000
Monthly mean v wind 500	v0500
Monthly mean v wind 850	v0850
Monthly mean v wind 1000	v1000
Monthly mean of specific humidity 500	s0500
Monthly mean of specific humidity 850	s0850
Monthly mean of specific humidity 1000	s1000
Monthly mean of relative humidity 500	r0500
Monthly mean of relative humidity 850	r0850
Monthly mean of relative humidity 1000	r1000
Monthly mean of divergence 500	∇500
Monthly mean of divergence 850	∇0850
Monthly mean of vorticity 1000	1000
Monthly mean of vorticity 0.21	
Monthly mean wind Speed at sigma level 0.995 from daily wind speed	w
Sea level pressure 0.21	mslp
Monthly mean air temperature at sigma level 0.995	`t
Monthly mean of precipitable water content	prw
Monthly mean of precipitation rate	pr

from each grid point. In the present study, we retained the principal components (PCs) of the seven leading EOFs of each of the predictor variables to estimate the total variance explained by each of the EOFs. The percentage of explained variance of the N^{th} EOF can be defined as the ratio between the eigenvalue of the N^{th} EOF to the sum of all eigenvalues of all the EOFs taken together (Wilks 2011; Lorenzo-Seva 2013). The higher-order EOFs (beyond those explaining 90%) associated with negligible variance represent noise and are not expected to add any value to the regression used in the downscaling.

Backward elimination is a special case of stepwise regression. In this study, the whole process of backward elimination was carried out using the R-statistical package (R Core team 2002). For each predictor variables, backward elimination begins with seven leading EOFs in the model, and at each step, different EOFs were eliminated from the model one at a time. The final model or the best-fit model includes only those EOFs which produce a minimum AIC (Akaike information criterion) value and eliminating any one of these EOFs that did not result in a lower AIC (Ripley 2002). We have identified the better-performing predictors by observing the adjusted R^2 values from the best-fit model; the larger the value of the adjusted R^2 , the better the ability of the variable to act as a suitable predictor (Hofer et al. 2010).

In another exercise, a stepwise regression was carried out between the area-averaged data of observation (predictand) and the area-averaged data of large-scale predictors in all three predictor domains. Before carrying out the multiple regressions, all the area-averaged datasets were standardized using the linearly detrended method to avoid spurious results associated with accidental trends. The best-fit model with the highest value of adjusted R^2 has been taken to select the suitable combination of predictors over different predictor domain. The fitted values of the models were compared with the observational data using different agreement indices (d-index and Pearson correlation coefficient) and error indices, i.e., normalized root mean squared error (NRMSE). Details of these indices were mentioned in Meher et al. (2017) and Meher and Das (2019).

The correlation analysis between the predictors and the predictand is carried out in two different ways as follows.

1. The linear relation between different predictor and predictand was analyzed using the Pearson's product-moment correlation between the area-averaged predictor (X) and the area-averaged predictand (Y) for different seasons (T) (annual, monsoon, and winter). For simplicity, we have termed this correlation as the R_A in the whole document.
2. The Pearson's product-moment correlation was also used to quantify the linear relation between the predictand and each of the leading EOFs of the predictor variables. For simplicity, we have termed this correlation as R_E in the whole document.

Each of the method mentioned above was repeated for four different time periods of varying temporal resolution (25 years [1981–2005], 35 years [1971–2005], 45 years [1961–2005], and 55 years [1951–2005]) to put more confidence in the selection of suitable predictors.

2.3 Ranking of predictor variables and sensitivity analysis

The overall aim of the ranking approach is to scrutinize the top/bottom-ranked (1/24) predictors in all domains irrespective of season. Ranking of predictors was carried out for the three selected domains using the results obtained from the three methods (i.e., R^2 , R_A , and R_E) as discussed in Sect. 2.2. The ranking of predictors for a single domain (say India) and for a particular season (say monsoon) was carried out as follows.

2.3.1 Ranking for R^2 and R_A values

1. Firstly, we have taken four time periods of different temporal resolution as mentioned in Sect. 2.2 and calculated the values of R^2 and R_A for the predictor variables.
2. Secondly, ranks (1–24) were given to the predictor variables based on their absolute values. Therefore, the highest correlation value of a predictor implies a top-ranked (1) predictor, whereas the lowest correlation value of a predictor implies a bottom-ranked predictor (24).
3. Final rank (lie between 1 and 24) of a predictor is calculated by taking the arithmetic mean of its ranks obtained in all the four-time period.

Similar steps were followed for the other two domains in the winter season and annual time scale.

2.4 Ranking of R_E values

1. Firstly, we calculated the R_E values of all the predictors for the first seven leading EOFs in the four different time period as mentioned above.
2. Repeated step 2 as mentioned in Sect. 2.3.1.
3. An aggregate value of the ranks was calculated for each predictor using the arithmetic sum of their ranks obtained in all the four-time period for all the seven numbers of EOFs.
4. For a particular predictor, the values obtained in step 3 were summed together to arrive at a final rank.

Similar steps were followed for the other two domains in the winter season and annual time scale.

The overall rank of a predictor variable is calculated irrespective of the ranks obtained in all the season (for example, see Table 3). The sensitivity of the ranks was tested to ensure different methods used in the study are meaningful and to

check the consistency of the predictors. The sensitivity of the ranks was analyzed through two different methods.

1. Comparison of overall rank obtained by each predictor in different domains. Box plots were used to visualize the sensitivity of the ranks.
2. Comparison of overall ranks with rank calculated excluding only R^2 value, ranks calculated excluding only R_E values, and ranks calculated excluding only R_A values over South Asia, India, and the WHR. The ranks are calculated considering all the season and annual scale.

3 Results and discussion

3.1 Analysis of spatial correlation maps

Figure 1 shows correlation maps of two randomly selected predictors ($z0500$ and $r1000$) over the South Asia domain during annual, monsoon, and winter season. Correlation maps

of all the other predictors are shown in Fig. S1–S3 of the supplementary material. In all the three figures, predictors like u , v , and ∇ at all pressure levels and exhibited scattered patches of positive and negative correlations over and around the Indian subcontinent. Some predictors like z at all pressure level and $mslp$ had negative correlations (< -0.35) over the South Asian region, while $v1000$, $s0500$, and $s0850$, r at all pressure level, prw , and pr gave mostly positive correlations (> 0.40). Over the target-predictand domain (i.e., the WHR), predictors like $u0850$ and t showed positive correlations in all the season and annual scale, but they failed to reproduce the same in other regions. The predictors had stronger positive or negative correlation in annual and monsoon season than in the winter season. It was observed that both u and v were well correlated with the predictand. In this regard, Satyanarayana and Srinivas (2008) reported that u responds to heating in the monsoon trough in North India, while v has more local effects. Hence, together, u and v are responsible for the convergence of moisture and therefore related to precipitation over India. On the basis of correlation maps, Sinha et al. (2013) found that s , u , and v (at different levels) over different domains around India are potential predictors (from NCEP/NCAR) to predict

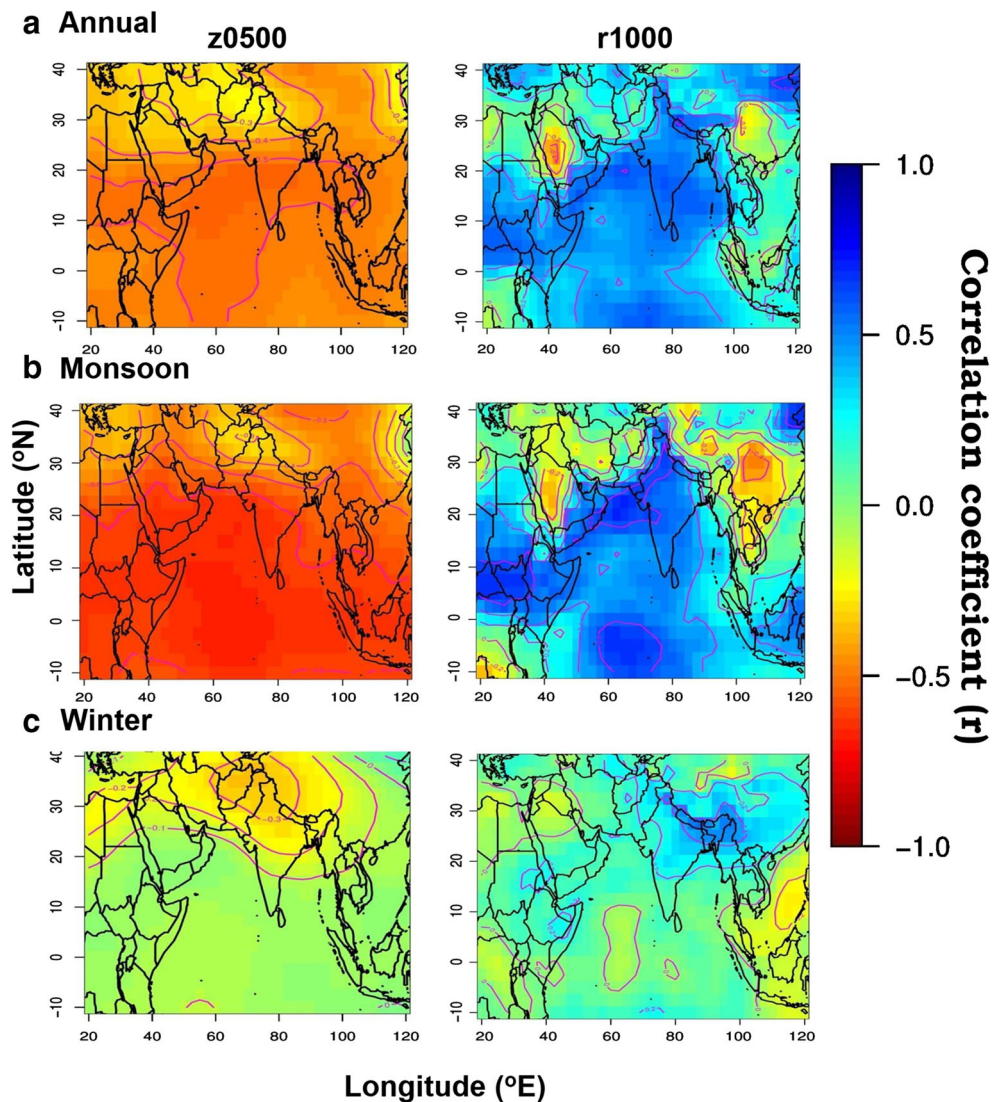
Table 3 Suitable combination of predictors selected using multiple regressions of all the 24 predictors in different predictor domain and time scale. Agreement (d-index and R^2) and error indices (normalized root mean squared error, NRMSE) between observation and modeled

data were also calculated at each of the predictor domain in two different seasons (monsoon and winter) as well as in the annual time scale. Cells in italics represent predictors that are common in all domains in different seasons and annual scale

	South Asia	India	WHR
Variables selected			
Annual	$z0850, v0850, \nabla 0850, prw$	$z0500, z0850, u850, u1000, v0500, v0850, s0850, s1000, r0850, r1000, prw$	$u0850, u1000, s0850, s1000, r1000, , mslp, t, prw$
Monsoon	$z0850, u0500, v0850, v1000, s1000, r0500, r0850, \nabla 0850, 1000, w, mslp, prw$	$z0500, z0850, z1000, u0850, v0850, v1000, 1000, , w, mslp$	$z0500, z1000, v0500, 1000, mslp, prw$
Winter	$z1000, u0850, u1000, v0850, v1000, s1000, r0850, r1000, \nabla 0850, 1000, mslp$	$u1000, v1000, s0500, r0500, r0850, r1000$	$z0500, s0500, s1000, r0500, r0850, r1000, \nabla 0850, , mslp, t, prw, pr$
Adjusted R^2 values of multiple regression			
Annual	0.18	0.30	0.50*
Monsoon	0.52*	0.61*	0.58*
Winter	0.30	0.40*	0.61*
Index of agreement (d-Index)			
Annual	0.60*	0.78*	0.86*
Monsoon	0.88*	0.90*	0.87*
Winter	0.30	0.30	0.24
Correlation Coefficient (r)			
Annual	0.50*	0.66*	0.76*
Monsoon	0.79*	0.82*	0.79*
Winter	-0.05	-0.12	-0.20
NRMSE			
Annual	172.3	111.5	82.6
Monsoon	76.2	68.3	76.7
Winter	183.6	184.9	168.6

*Significant at 5%

Fig. 1 Correlation maps of two different predictors in annual, monsoon, and winter time scale over the South Asian domain. Here, the correlations were calculated between the total annual/monsoon/winter rainfall of the observational data and the aggregated annual/monsoon/winter values of the predictors at each grid point location. For all other predictors, refer to Figs. S1 to S3 of the supplementary article



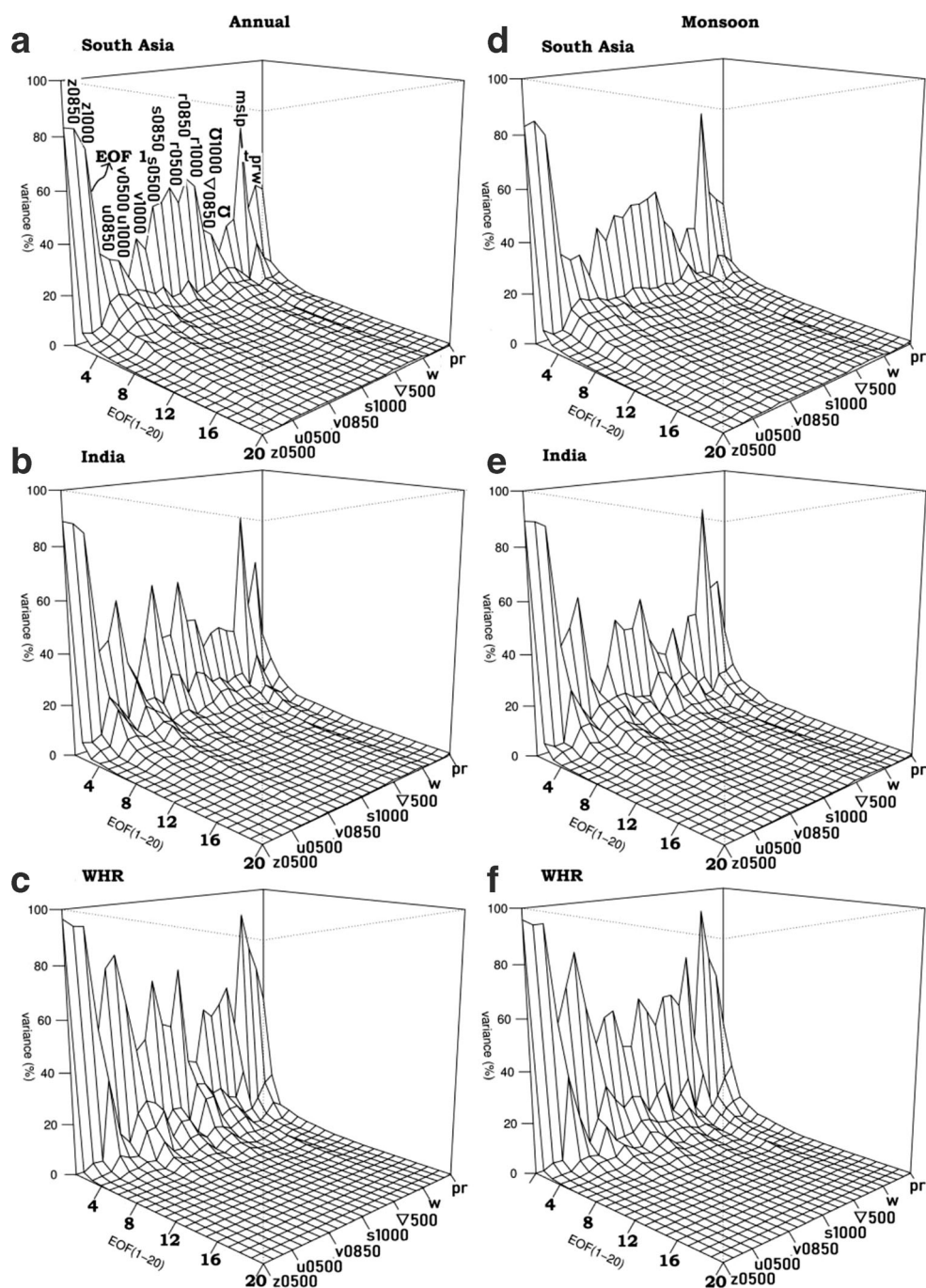
Indian monsoon rainfall. The results of the present work are almost similar to the finding of Sinha et al. (2013).

3.2 Analysis of variance explained by EOFs and multiple regression

Figure 2 shows the variance explained by first 20 leading EOFs for all the predictors over the three study domains during annual time scale and monsoon season. The variance shown for the EOFs are the average value of the variance shown by respective predictors in four different time periods of varying temporal resolution as given in Sect. 2.2. It can be seen that the variance explained by higher-order EOFs (e.g., 8–20) were negligible as compared to the first seven leading EOFs. Similar results were found for the winter season (not shown). In most of the cases, the first seven leading EOFs together explained > 90% variance of the raw data. First, few EOFs are expected

to explain a major portion of the variance compared to the variance explained by the rest of the higher-order EOFs. It is also found that the explained variance of the first EOF of most of the predictors was higher in WHR than in the Indian and South Asian region (not shown here). A similar type of result was reported by Akhter et al. (2019) where they found the explained variance of the downscaling model in the monsoon season was higher in the Western Himalayan region (Also known as the North mountainous India in their paper) compared to four other homogenous rainfall zones of India. In the present study, we have taken first seven leading EOFs in the multiple regression processes so that most of the regional and subregional variance can be incorporated in the selection of suitable predictors. Figure 3 shows the R^2 values obtained for each predictor over different domains during annual, monsoon, and winter time frame. Over the South Asian region, the mean of the R^2 values obtained for all the predictors was

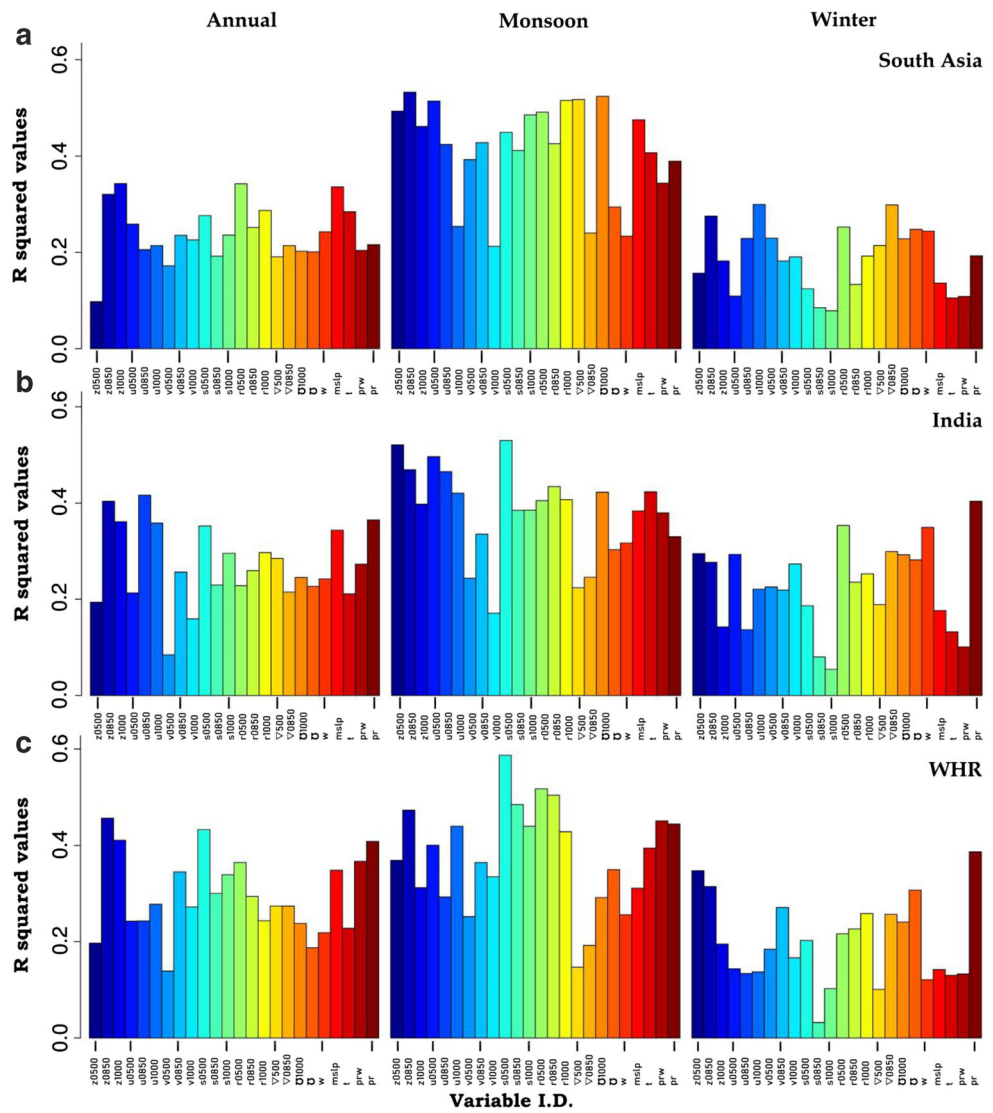
Fig. 2 Variance explained by first 20 leading EOFs is shown for all the predictors over the selected study domains during annual time scale and monsoon season. The variances shown for the EOFs are the average value of the variances shown by respective predictors in four different time periods of varying temporal resolution (25 years [1981–2005], 35 years [1971–2005], 45 years [1961–2005], and 55 years [1951–2005]). It can be seen that the variance explained by higher-order EOFs (e.g., 8–20) were negligible (~ 0 –5%) as compared to the first seven leading EOFs. Similar results were found for the winter season (not shown here)



higher during monsoon (0.41) season than annual (0.24) and winter (0.18) seasons. Similar results were also found in the other two regions namely whole India and WHR. In all the seasons, the average R^2 of all the predictors over the WHR was less than the Indian and South Asian region whereas over the Indian and South Asian regions, they were close to each other. Some predictors such as z0850 and z1000, s0500 and r0500, mslp, and pr gave higher R^2 values for most of the cases (season and domain). The predictors having the highest value of R^2 on the annual

time scale were z1000 (0.34), u0850 (0.42), and z0850 (0.46) over South Asia, India, and WHR respectively. In the monsoon season, the predictors having the highest value of R^2 were the z0850 (0.53) over South Asia and s0500 over both India (0.53) and WHR (0.59). Similarly, in the winter season, the predictors having the highest value of R^2 were u1000 (0.30) over South Asia and pr over both India (0.41) and WHR (0.39). The results obtained in the monsoon season are consistent with the findings of Akhter et al. (2019) in their predictor selection study over the seven

Fig. 3 R^2 from the regression between the predictand and the predictors over **a** South Asia (upper), **b** India (middle), and **c** Western Himalayan region (lower) for the annual total, monsoon total, and winter total precipitation calculated using backward elimination method. Seven leading empirical orthogonal functions (EOFs) were used against the observation in the backward elimination method and the R^2 values were calculated from the best model fit



homogenous regions of India that reported that downscaling model with s500, s850, s1000, and prw was able to explain more than 70% of the observed rainfall variance over the Western Himalayan region whereas predictors like ta500 and u1000 have explained little about observed variance. ∇ and \square parameters at different pressure level have shown poor skills in all the selected predictor domains. An EOF-based downscaling study by Nicholas and Battisti (2012) found that the most skilful predictors from the NCEP/NCAR data were all combinations of low-level specific humidity and one or more other fields at the same level over China, which was consistent with our finding over South Asia and the WHR. Pervez and Henebry (2014) reported that the predictors like z, s, u0500, u0850, and u1000, mslp, and w0500 gave higher explained variance of the observation in the Ganges–Brahmaputra basin of the South Asian region, which supports the findings of the present study.

3.3 Analysis of product-moment correlation coefficient

Figure 4 shows the correlations between area-averaged predictand data and area-averaged predictor datasets for all domains during annual, monsoon, and winter timescale. Predictors like u0850, v0850, u1000, v1000, s, r at all pressure levels, prw, and pr had high positive correlation coefficients ($R_A > 0.4$) in all domains during annual and monsoon time frame, whereas in winter, the R_A values were less (< 0.3). These predictors gave higher R_A values over the Indian domain (average $R_A = 0.54$) compared to the other two domains (average $R_A = 0.46$) during annual and monsoon time scale, whereas during winter season, the R_A values over WHR (average $R_A = 0.28$) was higher than over South Asia (average $R_A = 0.06$) and India (average $R_A = 0.18$) domain. The predictors having a strong positive correlation with the predictand were s0500 (average $R_A = 0.66$), r0500 (average $R_A = 0.60$),

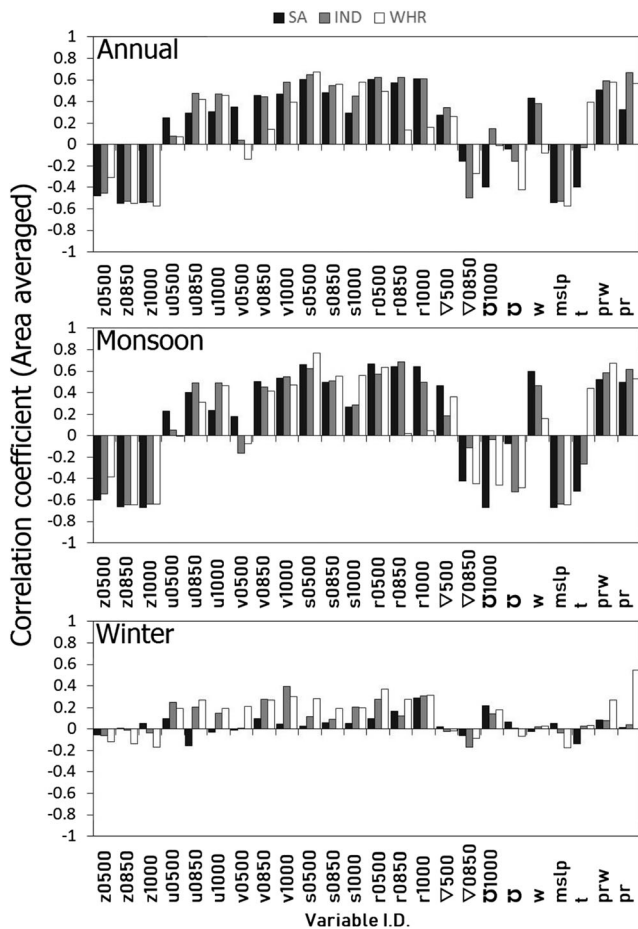


Fig. 4 Comparison of the correlation coefficient between the observed rainfall and the area-averaged predictor variables over the South Asia (Black bars), India (gray bars), and Western Himalayan region (white bars) during annual, monsoon, and winter seasons. The figure shows the correlation coefficients were higher in monsoon season than the annual and winter time scale

and prw (average $R_A = 0.57$) in all the three domains during annual and monsoon time frame. In the winter season, the predictor having high positive correlation was r1000 ($R_A = 0.29$), v1000 ($R_A = 0.40$), and pr ($R_A = 0.55$) over South Asia, India, and WHR, respectively. Similarly, some predictors like z at all pressure levels and mslp gave strong negative correlation (average $R_A = -0.55$) with the predictand over all the domains in all the season and annual time scale. These predictors exhibited higher R_A values over the South Asia domain (average $R_A = -0.59$) than the other two domains (average $R_A = -0.55$). Similar results were also found for the winter season.

Figure 5 shows the correlation coefficients (R_E) between area-averaged predictand and seven leading EOFs of each predictor field over the three selected domain. The first EOF (and sometimes the second EOF) of most of the predictors were consistently well correlated with the predictand in annual and monsoon timescale over all domains, whereas, in the winter season, none of the EOFs have shown such consistent

results. In all the seasons and annual time scales, the first EOF of u0850, u1000, v1000, s at all pressure level, r0500, prw, and pr had strong positive correlation with the predictand over all domains. It was also found that these predictors had nominally higher correlation over the WHR domain (Average R_E is 0.54 in annual and 0.60 in monsoon) than over the South Asian (average R_A is 0.53 in annual and 0.58 in monsoon) and Indian domain (average R_E is 0.53 in annual and 0.49 in monsoon) during the same time frame. In the winter season, the correlations shown by the first EOF of these predictors were though positive but their values were low, i.e., < 0.20 . The leading EOF of z at all pressure levels, ∇0850, , and mslp had strong negative correlation with the predictand in all the domains during annual (average $R_E = -0.51$), monsoon

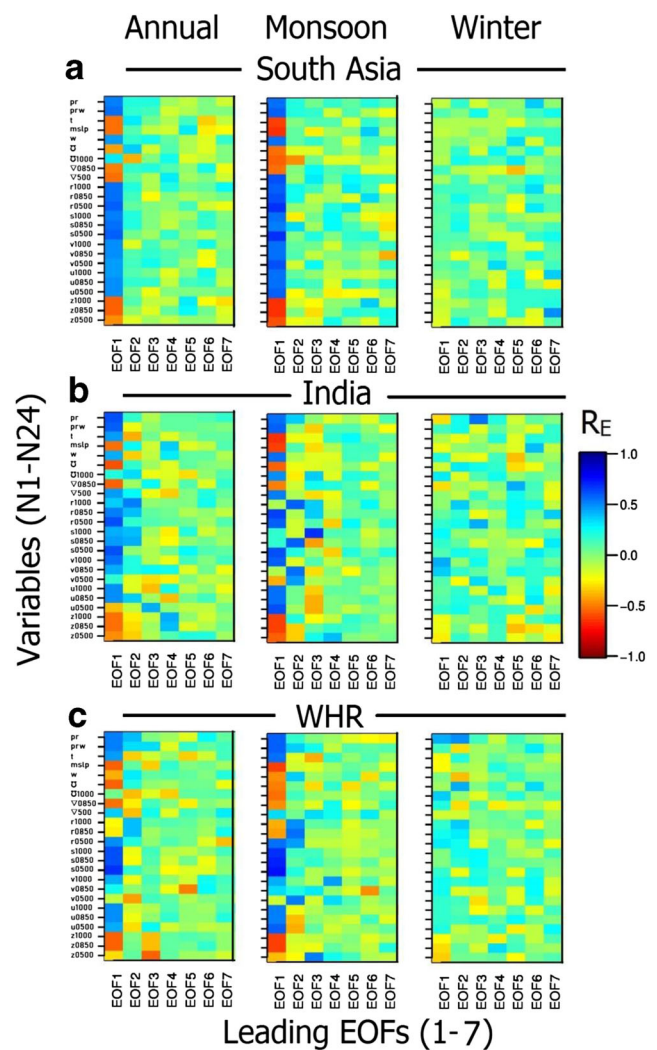


Fig. 5 Comparison of the correlation coefficient (R_E) between the observed rainfall and the seven leading EOFs of different predictor variables over **a** South Asia, **b** India, and **c** Western Himalayan region during annual, monsoon, and winter season. Figures show the correlation coefficients are either highly positive or negative for the first or second EOFs. There are several cases in which higher-order EOFs also showed good correlation with the observation

(average $R_E = -0.58$), and winter (average $R_E = -0.17$) time frame. In general, it was found that during the monsoon season, the predictors were having higher R_E values than the winter and annual time scale. Besides the first EOF, there were also other EOFs which gave a good correlation (both positive and negative) with the predictand in different domains and a different season, but a general statement cannot be written for these correlations. Hence, we have included the seven leading EOFs while ranking different predictors in the subsequent sections.

3.4 Ranking of predictors and sensitivity analysis

Figure 6, 7, and 8 show the ranking of all the predictors over South Asia, India, and WHR domains respectively. The three statistical metrics R_A , R_E , and R^2 were used to evaluate the final rank (1 to 24) of each predictor over all domains. In all the study domains, rank 1 of a predictor denotes the best predictor whereas rank 24 denotes a poor predictor. In the case of R_A and R_E , we have used their absolute values in evaluating the ranks. The final rank of a predictor was calculated by ranking the total sum of all the ranks obtained in all the seasons.

Over the South Asia region, the top-ranked predictors were r1000 (rank 1), r0500 (rank 2), and z0850 (rank 3) whereas the bottom-ranked predictors were v1000 (rank 23 each) and s0850 (rank 22). Over the India domain, the top-ranked predictors were z0850 (rank 1), r1000 (rank 2), and r0850 (rank 3), whereas the bottom-ranked predictors were v1000 (rank 24), ∇ 0500 (rank 23), and v0500 (rank 22). Similarly, over the WHR domain, the top-ranked predictors were pr (rank 1), r0500 (rank 2), and s0500 (rank 3), whereas the bottom-ranked predictors were v0500 (rank 24), ∇ 0500 (rank 23), and w (rank 22). In a separate study, it was reported that downscaling models with precipitable water (prw) and specific humidity predictors have shown good validation results compared to other predictors over the WHR of India. In general, the predictors which have shown very poor performance were the v0500, v1000, and ∇ 0500, whereas the well-performed predictors were z0850, r500, r1000, and pr. These are the predictors which acquired either top ranks (1–4) or bottom rank (2–24) at least in two out of three domains.

Figure 9 shows the sensitivity analysis of ranks obtained by each of the predictors in different domains. The final ranks of the predictor were used in this analysis as they were prepared taking all the seasons and annual fields into consideration. It can be seen that five predictors (z0500, z0850, v0500, s0500, and ∇ 0850) were independent of the method and season of predictor selection, as their ranks had lower standard deviation than the other predictors. The overall ranks obtained for z0850 were consistently good and never exceeded > 5 in any of the domains. Similarly, the ranks obtained by v0500 were consistently poor and never fell below < 21 . One predictor, i.e.,

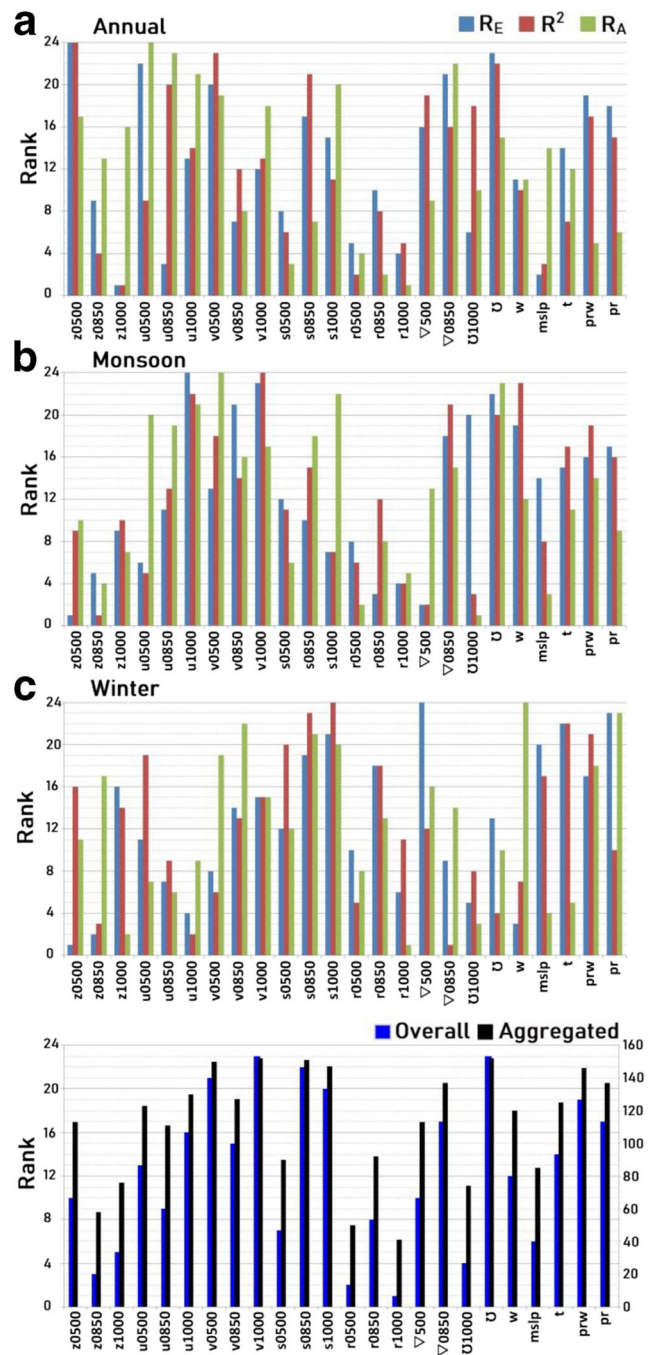


Fig. 6 Ranking of 24 predictor variables in the South Asia domain during **a** annual, **b** monsoon, and **c** winter. Here, rank 1 shows a better performing variable, whereas rank 24 denoted a poor performance of the variable. Bottom panel of the figure shows the final ranks obtained for all the predictors considering all the three time scales taken in a–c. The aggregated rank for a particular predictor was calculated by the arithmetic sum of the ranks obtained in all season and annual scale. Ranking of the aggregated ranks was represented through overall rank

z0500, consistently showed a rank which lies between 10 and 13 in all the domains. It was found that other than these predictors, all others were sensitive to the domain under consideration. For example, predictor like pr ranked 1 and 4 over the WHR and India domains, respectively, but was ranked 17

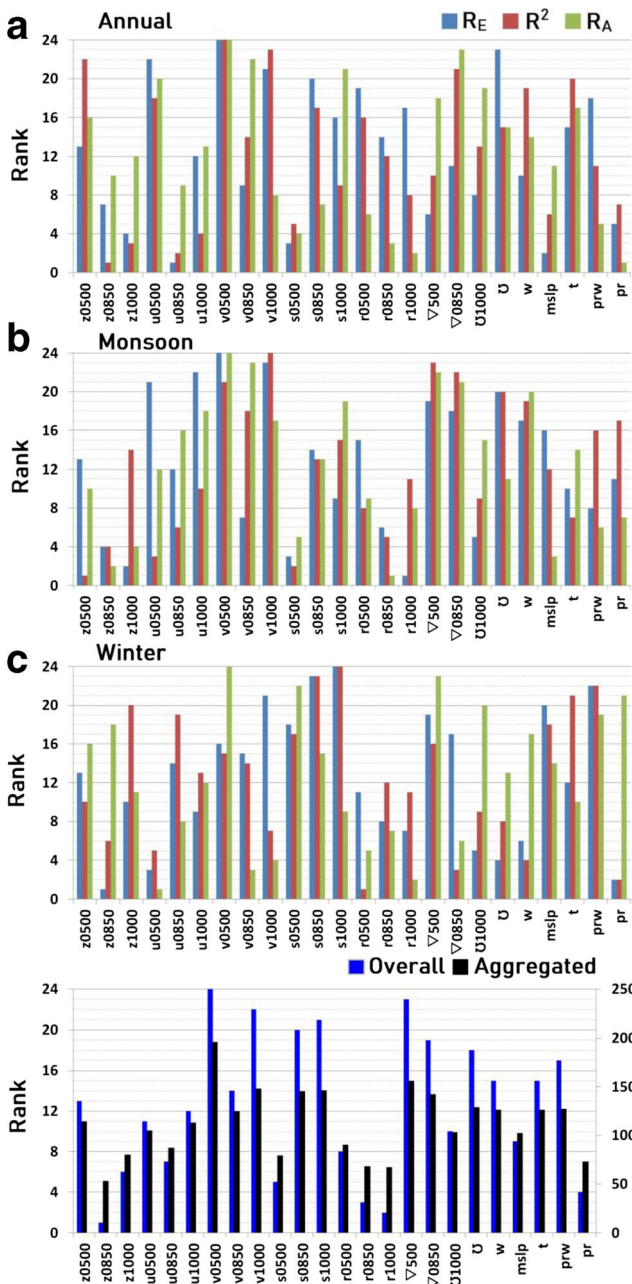


Fig. 7 Same as Fig. 6, but for India domain

over South Asia. Similarly, s1000 showed a poor rank of 20 and 21 over the South Asian and Indian regions respectively whereas it showed a better rank (7) over the WHR. Figure 10 shows the second way of analyzing the sensitivity of the ranks where we compared the ranks obtained through different methods like overall rank, rank calculated excluding only R^2 method, ranks calculated excluding only R_E method, and ranks calculated excluding only R_A method over the three selected domains. It is revealed that in each of the three domains, there is a significant correlation (between 0.90 and 0.95, at 1% level) between the overall ranks and the ranks obtained after excluding different methods; hence, the

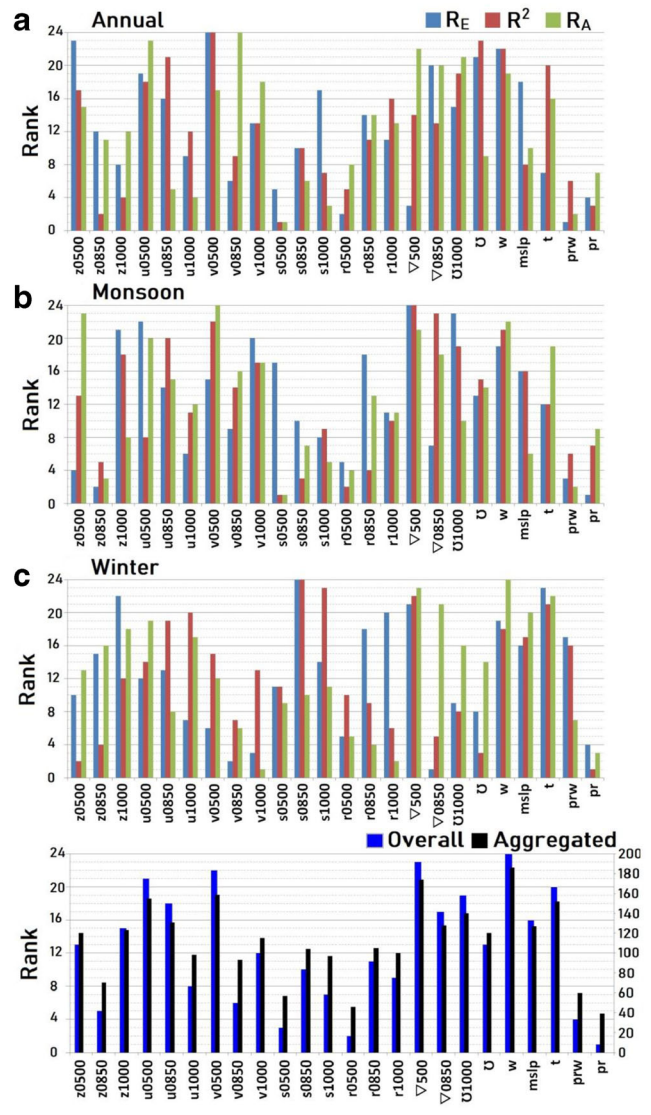


Fig. 8 Same as Fig. 6, but for WHR domain

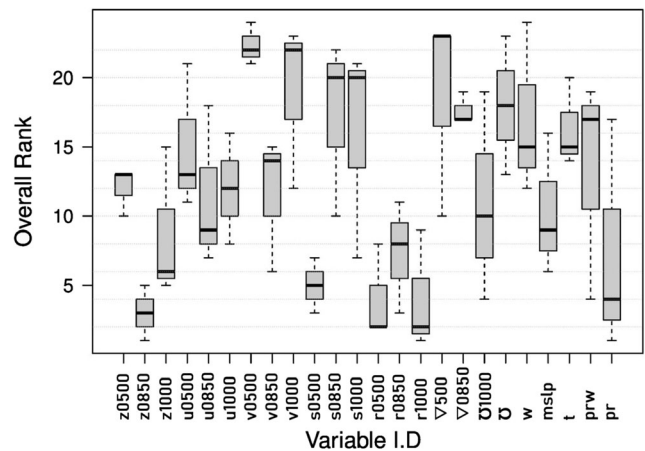


Fig. 9 Sensitivity analysis of ranks in different domains. The plots were generated using the overall rank obtained by each of the predictor variables over different domains namely South Asia, India, and Western Himalayan region

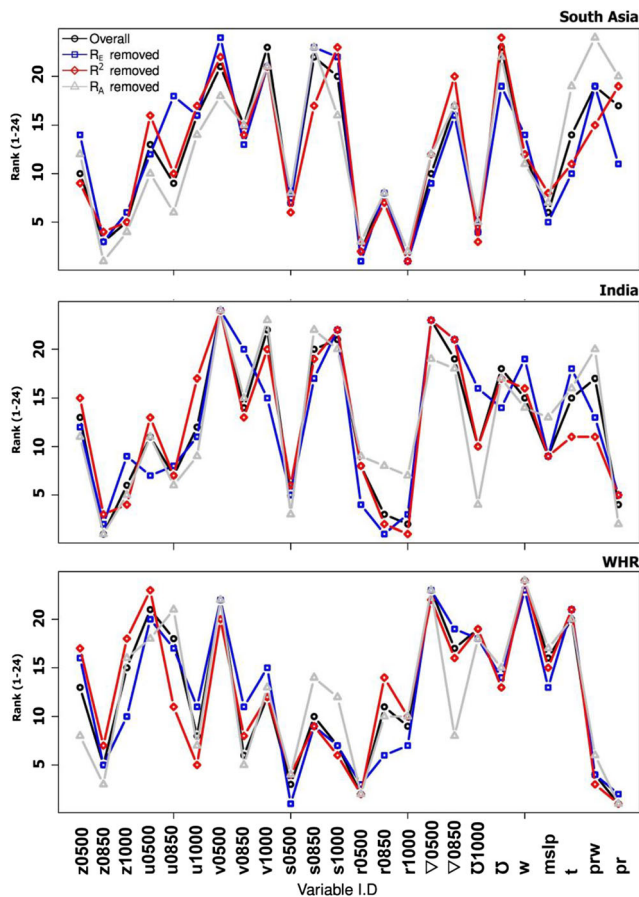


Fig. 10 Comparison of ranks obtained through different methods like overall rank, rank calculated excluding only R² values, ranks calculated excluding only R_E values, and ranks calculated excluding only R_A values over South Asia, India, and WHR. The ranks are calculated considering all the seasons and annual scales

methods used in this study are very effective and meaningful in selecting suitable predictors over the South Asia domain. Over the South Asia domain, eight predictors (z1000, v0850, v1000, s0500, r at all pressure levels, and 1000) have shown their overall ranks were not varied more than ± 3 after exclusion of any of the methods taken in this study. Similar results were also obtained for five predictors (z0850, u0850, v0500, s1000, and pr) over the India domain and eight predictors (v0500, v1000, r0500, and ∇0500, 1000, w, t, pr) over the WHR domain.

3.5 Analysis of multiple regressions of the raw data

Table 3 shows the selected combination of variables obtained from the backward multiple regression between all the 24 variables (i.e., predictors) taken in this study and the observational data (i.e., predictand). Table 3 also shows calculated values of various statistical metrics between the observational data and the best-fit model data. Variables which have maximum occurrence in different best-fit models were v0850, s1000, r0850, r1000, 1000,

mslp, and prw. The adjusted R² values over the smaller predictor domain of WHR were significant (at 5%) and higher than other regions, while monsoon was the season in which the R² values were significantly higher than annual and winter time scale. Calculated d-index (and correlation coefficient) values were 0.60 < d-index < 0.90 (and 0.50 < r < 0.82) for the annual and monsoon time scale and < 0.31 (and very low negative values) in the winter, which showed that model-fitted values using the selected combination of variables were close and in good agreement with the observation during the annual and monsoon seasons whereas poor in winter season over all the selected predictor domain. The calculated NRMSE values showed that the model-fitted data were characterized by low normalized error with observation in annual and monsoon seasons whereas high error in the winter season.

4 Conclusions

The major conclusions from the present study were outlined as follows:

- The predictors examined exhibited stronger positive or negative spatial correlation with the observed regionally averaged rainfall (the reference predictand) in annual and monsoon season than in the winter season. In all the selected domains, the mean of the regression coefficient values obtained for all the predictors was higher during monsoon than annual and winter seasons. In all the season, the average R² values of all the predictors over the WHR were less than the Indian and South Asian region whereas over the Indian and South Asian regions, they were close to each other.
- Predictors like u0850, u1000, v0850, v1000, s and r at all pressure levels, prw, and pr indicated high positive correlation coefficients (calculated through areal average method, R_A) in all the selected domains during annual and monsoon time frame whereas in winter, the R_A values were less. These predictors also had higher R_A values over the Indian domain as compared to the other two domains during annual and monsoon time scale, whereas during the winter season, the R_A values over WHR were higher than South Asia and India domain.
- First EOFs of most of the predictors were consistently well correlated (R_E) with the predictand in annual and monsoon timescale over all the selected domains, whereas, in the winter season, none of the EOFs have shown such consistent results. In general, it was found that during the monsoon season, the predictors were having higher R_E values than the winter and annual time scale.

- WHR predictor domain as mentioned in this study can be taken as a potential predictor domain for downscaling monsoon rainfall for the Western Himalayan region. Whereas, the statistical analysis of predictor selection for winter season rainfall over the Western Himalayan region was associated with poor findings (low agreement with observation). Hence, extreme care must be taken while downscaling winter rainfall over the Western Himalayan region.
- Predictors like z0500, z0850, v0500, s0500, and ∇ 0850 were independent of the method, season, or size of the domain.

Acknowledgments The authors would like to thank India Meteorological Department, and NCEP/NCAR for providing the required data for the present study. JKM would like to thank all the members of climate simulation lab, Bidhan Chandra Krishi Viswavidyalaya, West Bengal, for their constant support in preparing the work. The authors would like to thank the anonymous reviewers for their critical comments on the present work and improvements suggested by them.

Appendix

NB numbers associated with each variable shows the pressure level at that hPa

z = geopotential height	t2 m = air temperature at 2 m
mslp = mean sea level pressure	sf = surface airflow strength
t = air temperature	su = surface zonal wind velocity
u = zonal wind velocity	sv = surface meridional wind velocity
v = meridional velocity	s = surface vorticity
prw = precipitable water	sw = surface wind direction
sp = surface pressure	s ∇ = surface divergence
s = specific humidity	sr = surface relative humidity
r = relative humidity	ss = surface specific humidity

References

- Akhter J, Das L, Meher JK, Deb A (2019) Evaluation of different large-scale predictor-based statistical downscaling models in simulating zone-wise monsoon precipitation over India. *Int J Climatol* 39(1): 465–482
- Anandhi A, Srinivas VV, Nagesh Kumar D, Nanjundiah RS (2009) Role of predictors in downscaling surface temperature to river basin in India for IPCC SRES scenarios using support vector machine. *Int J Climatol* 29:583–603. <https://doi.org/10.1002/joc.1719>
- Anandhi A, Srinivas VV, Nanjundiah RS, Kumar DN (2008) Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine. *Int J Climatol* 28:401–420. <https://doi.org/10.1002/joc.1529>
- Bawiskar SM, Chipade MD, Puranik PV, Bhide UV (2005) Energetics of lower tropospheric planetary waves over mid latitudes: precursor for Indian summer monsoon. *J Earth Syst Sci* 114:557–564. <https://doi.org/10.1007/bf02702031>
- Blazak A (2012) Statistical downscaling of precipitation projections in Southeast Queensland catchments. PhD diss., University of Southern Queensland. Available at https://eprints.usq.edu.au/23571/1/Blazak_2012_whole.pdf (Assessed on 29 May 2017)
- Das L, Meher JK, Dutta M (2016) Construction of rainfall change scenarios over the Chilka lagoon in India. *Atmos Res* 182:36–45. <https://doi.org/10.1016/j.atmosres.2016.07.013>
- Das L, Meher JK (2019) Drivers of climate over the Western Himalayan region of India: A review. *Earth Sci Rev*. <https://doi.org/10.1016/j.earscirev.2019.102935>
- Dabanlı İ, Şen Z (2017) Precipitation projections under GCMs perspective and Turkish Water Foundation (TWF) statistical downscaling model procedures. *Theor Appl Climatol* 132(1-2):153–166. <https://doi.org/10.1007/s00704-017-2070-4>
- Devak M, Dhanya CT (2014) Downscaling of precipitation in Mahanadi basin, India. *Int J Civil Eng Res* 5:111–120
- Dimri AP, Yasunari T, Kotlia BS, Mohanty UC, Sikka DR (2016) Indian winter monsoon: present and past. *Earth-Sci Rev* 163:297–322. <https://doi.org/10.1016/j.earscirev.2016.10.008>
- Dimri AP, Niyogi D, Barros AP, Ridley J, Mohanty UC, Yasunari T, Sikka DR (2015) Western disturbances: a review. *Rev Geophys* 53:225–246. <https://doi.org/10.1002/2014RG000460>
- Douville H (2006) Impact of regional SST anomalies on the Indian monsoon response to global warming in the CNRM climate model. *J Clim* 19(10):2008–2024. <https://doi.org/10.1175/JCLI3727.1>

- Eden JM, Widmann M (2014) Downscaling of GCM-simulated precipitation using model output statistics. *J Clim* 27(1):312–324. <https://doi.org/10.1175/JCLI-D-13-00063.1>
- Førland EJ, Benestad R, Hanssen-Bauer I, Haugen JE, Skaugen TE (2011) Temperature and precipitation development at Svalbard 1900–2100. *Adv Meteorol*. <https://doi.org/10.1155/2011/893790>
- Gaur A, Simonovic SP (2017) Application of physical scaling towards downscaling climate model precipitation data. *Theor Appl Climatol* 132(1-2):287–300. <https://doi.org/10.1007/s00704-017-2088-7>
- Ghosh S, Mujumdar PP (2006) Future rainfall scenario over Orissa with GCM projections by statistical downscaling. *Curr Sci* 90(3):396–404
- Goyal MK, Ojha CSP (2012) Downscaling of surface temperature for lake catchment in an arid region in India using linear multiple regression and neural networks. *Int J Climatol* 32(4):552–566. <https://doi.org/10.1002/joc.2286>
- Goyal MK, Ojha CSP (2010) Evaluation of various linear regression methods for downscaling of mean monthly precipitation in arid Pichola watershed. *Nat Res Forum* 1(01):11–18. <https://doi.org/10.4236/nr.2010.11002>
- Hannachi A, Jolliffe IT, Stephenson DB (2007) Empirical orthogonal functions and related techniques in atmospheric science: a review. *Int J Climatol* 27(9):1119–1152. <https://doi.org/10.1002/joc.1499>
- Hanssen-Bauer I, Førland EJ, Haugen JE, Tveito OE (2003) Temperature and precipitation scenarios for Norway: comparison of results from dynamical and empirical downscaling. *Clim Res* 25(1):15–27
- Hanssen-Bauer I, Achberger C, Benestad RE, Chen D, Førland EJ (2005) Statistical downscaling of climate scenarios over Scandinavia. *Clim Res* 29(3):255–268
- Hofer M, Mölg T, Marzeion B, Kaser G (2010) Empirical-statistical downscaling of reanalysis data to high-resolution air temperature and specific humidity above a glacier surface (Cordillera Blanca, Peru). *J Geophys Res Atmos* 115(D12). <https://doi.org/10.1029/2009JD012556>
- Hu Y, Maskey S, Uhlenbrook S (2013) Downscaling daily precipitation over the Yellow River source region in China: a comparison of three statistical downscaling methods. *Theor Appl Climatol* 112(3-4):447–460. <https://doi.org/10.1007/s00704-012-0745-4>
- Huang J, Zhang J, Zhang Z, Xu C, Wang B, Yao J (2011) Estimation of future precipitation change in the Yangtze River basin by using statistical downscaling method. *Stoch Env Res Risk A* 25(6):781–792
- Huth R (1999) Statistical downscaling in Central Europe: evaluation of methods and potential predictors. *Clim Res* 13(2):91–101
- Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y (1996) The NCEP-NCAR 40-year reanalysis project. *Bull Am Meteorol Soc* 77:437–471. [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2)
- Kannan S, Ghosh S (2013) A nonparametric kernel regression model for downscaling multisite daily precipitation in the Mahanadi basin. *Water Resour Res* 49(3):1360–1385. <https://doi.org/10.1002/wrcr.20118>
- Lorenz E (1956) Empirical orthogonal functions and statistical weather prediction. Scientific Report No. 1, Statistical Forecasting Project, Massachusetts Institute of Technology, Department of Meteorology, Cambridge, Mass., 49 pp
- Lorenzo-Seva U (2013) How to report the percentage of explained common variance in exploratory factor analysis Available at http://psico.fcep.urv.es/utilitats/factor/documentacion/Percentage_of_explained_common_variance.pdf (Assessed on 29 Sep 2017)
- Mahmood R, Babel MS (2013) Evaluation of SDSM developed by annual and monthly sub-models for downscaling temperature and precipitation in the Jhelum basin, Pakistan and India. *Theor Appl Climatol* 113(1-2):27–44
- Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Sauter T, Themeßl M, Venema VK (2010) Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. *Rev Geophys* 48(3). <https://doi.org/10.1029/2009RG000314>
- Meher JK (2019) Estimation of rainfall statistics over the Western Himalaya region through empirical-statistical downscaling. Doctoral dissertation. Department of Agricultural Meteorology and Physics, Bidhan Chandra Krishi Viswavidyalaya
- Meher JK, Das L (2019) Gridded data as a source of missing data replacement in station records. *J Earth Syst Sci* 128(3). <https://doi.org/10.1007/s12040-019-1079-8>
- Meher JK, Das L, Akhter J, Benestad RE, Mezghani A (2017) Performance of CMIP3 and CMIP5 GCMs to simulate observed rainfall characteristics over the Western Himalayan region. *J Clim* 30:7777–7799. <https://doi.org/10.1175/JCLI-D-16-0774.1>
- Nicholas RE, Battisti DS (2012) Empirical downscaling of high-resolution regional precipitation from large-scale reanalysis fields. *J Appl Meteorol Climatol* 51(1):100–114. <https://doi.org/10.1175/JAMC-D-11-04.1>
- Ojha CS, Goyal MK, Adeloje AJ (2010) Downscaling of precipitation for lake catchment in arid region in India using linear multiple regression and neural networks. *Int J Climatol* 4(1):122–136. <https://doi.org/10.1002/joc.2286>
- Pai DS, Sridhar L, Rajeevan M, Sreejith OP, Satbhai NS, Mukhopadhyay B (2014) Development of a new high spatial resolution (0.25° × 0.25°) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *Mausam* 65(1):1–18
- Parding KM, Benestad R, Mezghani A, Erlandsen HB (2019) Statistical projection of the North Atlantic storm tracks. *J Appl Meteorol Climatol* 58(7):1509–1522
- Pervez MS, Henebry GM (2014) Projections of the Ganges–Brahmaputra precipitation—downscaled from GCM predictors. *J Hydrol* 517:120–134. <https://doi.org/10.1016/j.jhydrol.2014.05.016>
- Parthasarathy B, Kumar KR, Deshpande VR (1991) Indian summer monsoon rainfall and 200-mbar meridional wind index: Application for long-range prediction. *Int J Climatol* 11(2):165–176
- Priya P, Krishnan R, Mujumdar M, Houze RA (2016) Changing monsoon and midlatitude circulation interactions over the Western Himalayas and possible links to occurrences of extreme precipitation. *Clim Dyn* 49:2351–2364. <https://doi.org/10.1007/s00382-016-3458-z>
- R Core Team (2002) R: A Language and Environment for Statistical Computing. R Core Team R Foundation for Statistical Computing, Vienna, Austria
- Ripley BD (2002) Modern applied statistics with S. Springer-Verlag, New York. <https://doi.org/10.1007/978-0-387-21706-2>
- Saeed F, Hagemann S, Saeed S, Jacob D (2013) Influence of mid-latitude circulation on upper Indus basin precipitation: the explicit role of irrigation. *Clim Dyn* 40(1-2):21–38. <https://doi.org/10.1007/s00382-012-1480-3>
- Salvi K, Ghosh S (2013) High-resolution multisite daily rainfall projections in India with statistical downscaling for climate change impacts assessment. *J Geophys Res-Atmos* 118(9):3557–3578. <https://doi.org/10.1002/jgrd.50280>
- Satyanarayana P, Srinivas VV (2008) Regional frequency analysis of precipitation using large-scale atmospheric variables. *J Geophys Res-Atmos* 113(D24). <https://doi.org/10.1029/2008JD010412>
- Sauter T, Venema V (2011) Natural three-dimensional predictor domains for statistical precipitation downscaling. *J Clim* 24(23):6132–6145. <https://doi.org/10.1175/2011JCLI4155.1>
- Shashikanth K, Ghosh S (2013) Fine Resolution Indian Summer Monsoon Rainfall Projection with statistical Downscaling. *Int J Chem Environ Biol Sci* 1(4):615–618
- Sinha P, Mohanty UC, Kar SC, Dash SK, Robertson AW, Tippett MK (2013) Seasonal prediction of the Indian summer monsoon rainfall

- using canonical correlation analysis of the NCMRWF global model products. *Int J Climatol* 33(7):1601–1614. <https://doi.org/10.1002/joc.3536>
- Srinivasan G, Hulme M, Jones CG (1995) An evaluation of the spatial and interannual variability of tropical precipitation as simulated by GCMs. *Geophys Res Lett* 22(16):2139–2142
- Wilby R, Wigley T (2000) Precipitation predictors for downscaling: observed and general circulation model relationships. *Int J Climatol* 20:641–661
- Wilby RL, Charles SP, Zorita E, Timbal B, Whetton P, Mearns LO (2004) Guidelines for use of climate scenarios developed from statistical downscaling methods. Supporting material of the Intergovernmental Panel on Climate Change. Available http://www.ipcc-data.org/guidelines/dgm_no2_v1_09_2004.pdf (Assessed on 29 May 2017)
- Wilby RL, Wigley TML (1997) Downscaling general circulation model output: a review of methods and limitations. *Progress in physical geography* 21(4): 530–548. <https://doi.org/10.1177/030913339702100403>
- Wilby RL, Hay LE, Leavesley 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(1):67–91. [https://doi.org/10.1016/S0022-1694\(99\)00136-5](https://doi.org/10.1016/S0022-1694(99)00136-5)
- Wilks DS (2011) *Statistical methods in the atmospheric sciences*. 100. Academic press
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.