

# Precipitation of Mainland India: Copula-based bias-corrected daily CORDEX climate data for both mean and extreme values

Mayank Suman<sup>1</sup>  | Rajib Maity<sup>2</sup>  | Harald Kunstmann<sup>3,4</sup> 

<sup>1</sup>School of Water Resources, Indian Institute of Technology Kharagpur, Kharagpur, India

<sup>2</sup>Department of Civil Engineering, Indian Institute of Technology Kharagpur, Kharagpur, India

<sup>3</sup>Karlsruhe Institute of Technology (KIT), Institute of Meteorology and Climate Research (IMK-IFU), Campus Alpin – IMK-IFU, Garmisch-Partenkirchen, Germany

<sup>4</sup>Institute of Geography, University of Augsburg, Augsburg, Germany

## Correspondence

Rajib Maity, Department of Civil Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721302, India.  
Email: [rajib@civil.iitkgp.ac.in](mailto:rajib@civil.iitkgp.ac.in)

## Funding information

Department of Science and Technology, Ministry of Science and Technology, Grant/Award Number: DST/CCP/CoE/79/2017(G) and Helmholtz-Association of German Research Centres (HGF).

## Abstract

Changes in mean and extreme precipitation characteristics with changing climate may lead to an increase in frequency of hydrological extremes. For studying the impacts of the changing climate on hydrological systems, General Circulation Model (GCM)/Regional Climate Model (RCM) simulated precipitation are used. However, these products should be bias-corrected before used in hydrological simulations to predict hydrological extremes. Most of the existing bias-correction techniques suffer from either of two limitations – (a) they only reduce bias in selected precipitation quantile (either mean or extreme values), and/or (b) they exclude zero values from the analysis, even though their presence is significant in daily precipitation. In this study, a stochastic copula-based bias-correction method (Maity et al., *J. Hydrometeorol.*, 20, 2019, 595), henceforth RMPH method, is used that corrects the bias in any quantile (mean and/or extreme values) of daily precipitation including zero values. The RMPH method is applied across Indian mainland to correct bias in simulated precipitation from the Coordinated Regional Climate Downscaling Experiment (CORDEX). Due to diverse climatic conditions across India, the quality of bias-corrected precipitation is studied separately for different meteorologically homogenous regions of the country. Despite non-uniform distribution of raingauge stations for observed precipitation, the superiority of the bias-corrected precipitation (from RMPH method) in correcting

## Dataset details

The dataset with spatial resolution of 0.50° (latitude) × 0.50° (longitude) for scenario RCP2.6, RCP4.5 and RCP8.5 (till 2100) is available at <http://doi.org/10.6084/m9.figshare.9686648.v2> (Suman et al., 2019) as a collection of NetCDF files for bias-corrected mean daily precipitation; bias-corrected by using both RMPH and quantile mapping. Additionally, estimates of daily extreme precipitation (95th percentile) obtained from RMPH are also provided. The time period is 1961–2005 for historical CORDEX simulations and 2006–2100 for different RCPs. The spatial extent of these data set is the entire Indian landmass. The notes on usage of the dataset is provided in ‘Potential data set use and usage notes’ section.

Identifier: <http://doi.org/10.6084/m9.figshare.9686648.v2>

Creator: Suman, M., Maity, R. and Kunstmann, H.

Dataset correspondence: [rajib@civil.iitkgp.ac.in](mailto:rajib@civil.iitkgp.ac.in)

Title: Bias-corrected daily most expected and extreme CORDEX precipitation data set for India

Publisher: Figshare

Publication year: 2019

Conflicts of interest: The authors declare no competing financial interests.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Geoscience Data Journal* published by Royal Meteorological Society and John Wiley & Sons Ltd.

bias and retaining the seasonal variation across the country is evident when compared with tradition bias-correction approach like quantile mapping. The new bias-corrected precipitation dataset developed is particularly suited for hydrological simulations, formulating extreme event mitigation strategies and climate change adaptation strategies.

#### KEYWORDS

copula based bias-correction, CORDEX, mean and extreme precipitation

## 1 | INTRODUCTION

The characteristics of precipitation in South Asia, including India, are expected to change in a changing climate (Ghosh et al., 2012; Kitoh et al., 2013; Krishnan et al., 2016; Mohan & Rajeevan, 2017; Qiu, 2008). The region is considered a climate ‘hot spot’ due to the combination of expected strong impacts of climate change and socio-economic factors such as high population density, developing economy and low per capita income (De Souza et al., 2015). Given such challenges in the region, the change in the characteristics of precipitation (whether it is mean, or any other quantile including extreme) is expected to have extensive implications on the availability of water resources and its management. This, in turn, is expected to affect the energy and food security for the large population living in the region which is about one-sixth of the world population. As the precipitation is mostly concentrated in monsoon months, about one-eighth of the Indian mainland is prone to flash flood during this season (NDMA, 2008). Most of north Bihar, Assam, regions of Western Ghats, Gujarat regularly suffer from flash flood during the monsoon months. Many parts of coastal south India have experienced major floods in the recent times, for example Coromandel Coast flood in 2015, Kerala flood in 2018, 2019 and 2020, East and West Godavari district flood in 2019. Moreover, the flash flood events are increasing throughout the country in the recent past (Cho et al., 2016; Houze Jr et al., 2017; Prasad & Singh, 2005; Seenirajan et al., 2017; Thayyen et al., 2013; Vishnu et al., 2019). Lower than normal rainfall results in meteorological drought, which may further develop into agricultural or hydrological drought (Maity et al., 2016). About one-third of the Indian mainland is prone to droughts (Subramanya, 2013), and both the intensity and areal extent of droughts in India are increasing (Kumar et al., 2013; Mallya et al., 2016; Sharma & Mujumdar, 2017).

Prediction and hydrological simulation of these extreme events may help in formulating an effective mitigation strategy. Currently, future precipitation estimates for different climate change scenarios, designated by different Representative Concentration Pathways (RCPs), are provided by several Regional Climate Models (RCMs) under the Coordinated Regional Climate Downscaling Experiment (CORDEX) project. However, the RCMs are known to produce systematic

under- or over-estimation of precipitation, limiting their applicability in hydrological studies. Moreover, many RCMs tend to show drizzle effect (generate too many wet days with light rainfall), underestimate heavy rainfall values and produce an incorrect seasonal variation (Christensen et al., 2008; de Elía et al., 2017; Fowler et al., 2007; Maraun et al., 2017; Schmidli et al., 2006; Teutschbein & Seibert, 2010). Hence, RCM-simulated precipitation requires bias-correction before being utilized in hydrological models. A review of different bias-correction methodologies can be found in literature (Lafon et al., 2013; Pierce et al., 2015; Teutschbein & Seibert, 2010). However, most of the bias-correction methods suffer from one of following two limitations: (a) they reduce bias in a selected precipitation quantile (e.g., either mean or extreme values) and (b) they exclude zero values from the analysis, even though their presence is significant in daily precipitation. Copula-based bias-correction scheme (RMPH model) was developed by Maity et al. (2019) that reduces the bias in the entire range of the precipitation (including mean and extreme) and corrects the seasonality of RCM simulations. In this study, a bias-corrected precipitation dataset is generated for the entire Indian mainland using the aforementioned RMPH model. The quality of bias-corrected precipitation is investigated over different hydro-meteorological regions across the study area.

While comparing with Quantile Mapping (QM) method, the bias-corrected product from the RMPH model is better for both the mean and extreme precipitation values across most parts of the Indian mainland. In case of extreme precipitation, QM is found to over-estimate the extreme precipitation during dry period and under-estimate the extreme precipitation during monsoon months. Hence, the bias-corrected dataset from RMPH model is expected to be more useful for many hydrological modelling and climate change-related studies.

## 2 | METHODS

The RMPH model uses the entire range of daily precipitation values including zeros to correct bias in any quantile of precipitation with proper consideration to local climatic factors. The study area (Indian mainland) is climatologically diverse; hence, the quality of the bias-corrected data products

is investigated separately for different meteorologically homogeneous regions of study area.

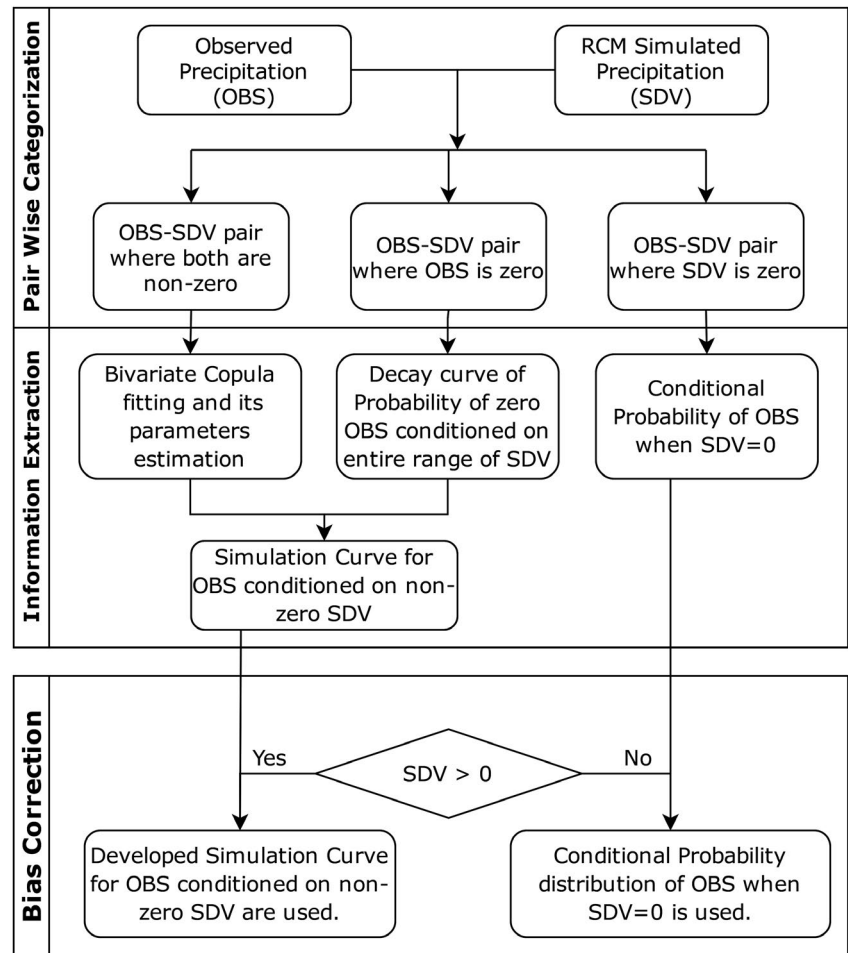
An overview of RMPH model is shown in Figure 1. The model utilizes bivariate copulas to model the association between RCM-simulated downscaled values (SDV) of precipitation from CORDEX and observed precipitation (OBS). The conditional distribution from the copula-based joint distribution is modified to consider the presence of zero precipitation. This is carried out in the following steps: first, the SDV-OBS pair in the dataset are divided into three categories – (a) Pairs in which both SDV and OBS are non-zero, (b) Pairs in which OBS is zero and (c) Pairs in which SDV is zero. From these categories, three sets of information are obtained during calibration of the RMPH model.

The first set of data are converted to reduced variate (non-exceeding probability) using a best-fit marginal probability distribution function. The best-fit distribution is selected from a pool of parametric probability distribution functions on the basis of two criteria – (a) the fitted distribution should pass the chi-square ( $\chi^2$ ) test at 5% level of significance and (b) it should have the lowest Bayesian Information Criteria (Schwarz, 1978; Wit et al., 2012). Nine parametric probability functions, namely, Exponential, Normal, Log Normal, Inverse Gaussian, Gamma, Beta, Generalized Pareto,

Logistic and Log logistic, are used for selecting the best-fit distribution. If no probability distribution satisfies both of the aforementioned criteria simultaneously, then a nonparametric Gaussian kernel-based estimate of probability distribution (hereinafter, nonparametric distribution) is used as best-fit probability distribution function. The relationship between the reduced variate of OBS and SDV is modelled using the most suitable copula function out of four bivariate copulas, namely, Clayton, Frank, Gumbel and Gaussian. These copulas are selected as they have different tail dependence characteristics (Maity et al., 2019; Mao et al., 2015). The best-fit copula function is selected based on the smallest Cramér-von Mises statistics (Genest et al., 2009). The conditional distribution of observed precipitation given non-zero SDV is obtained from the fitted copula function:

$$F_{X_1/X_2}(x_1|x_2) = \frac{\partial C(F_{X_1}(x_1), F_{X_2}(x_2))}{\partial F_{X_2}(x_2)} \quad (1)$$

where  $C(F_{X_1}(x_1), F_{X_2}(x_2))$  represents the selected copula function of the reduced variate of OBS ( $F_{X_1}(x_1)$ ) and SDV ( $F_{X_2}(x_2)$ ), respectively, obtained from the first set of data.  $F_{X_1/X_2}(x_1|x_2)$  is the conditional distribution of OBS ( $x_1$ ) conditioned on SDV ( $x_2$ ), which can be used to predict observed



**FIGURE 1** Methodological overview: copula-based bias-correction method or RMPH model developed by Maity et al. (2019). Model calibration is carried out through ‘Pair-Wise Categorization’ and ‘Information Extraction’ modules, which is used for the generation of bias-corrected products through ‘Bias Correction’ module

precipitation given non-zero SDV. However, RCM simulations may generate high simulated precipitation for low observed precipitation, and it may also produce incorrect seasonality (de Elia et al., 2017; Maraun et al., 2017). Further, many RCM outputs show drizzle effect, *i.e.*, many days with simulated low precipitation when observed precipitation is zero. To correct this type of bias, the obtained conditional distribution is required to be updated to include the probability of no/less precipitation. This information (probability of no observed precipitation, *i.e.*, zero OBS, given SDV) is obtained from the second set of data (*i.e.*, pairs where OBS is zero). The probability of getting zero OBS is expected to decrease with increasing SDV. An exponentially decaying function is used to model this probability in the following form:

$$Y = ae^{bX} \quad (2)$$

where  $Y$  is the probability of zero OBS and  $X$  is the value of SDV. The parameters  $a$  and  $b$  are estimated during calibration. For a decaying curve,  $b$  should be negative. The probability of zero precipitation obtained for different SDV (say  $p$ , the value of  $Y$ ) for a specific value of  $X$ , *i.e.*, SDV, is obtained in Equation (2). It is then used to update the conditional distribution of OBS given positive SDV as follows:

$$\tilde{F}_{X_1/X_2}(x_1|x_2) = \begin{cases} p & \text{for } x_1 = 0, x_2 > 0 \\ p + (1-p)F_{X_1/X_2}(x_1|x_2) & \text{for } x_1 > 0, x_2 > 0 \end{cases} \quad (3)$$

The expression can be used to generate a family of conditional distribution curves ( $\% \tilde{F}_{X_1/X_2}(x_1|x_2)$ ). These conditional distribution/simulation curves estimated during calibration can be used for simulating OBS given any non-zero SDV. Additionally, for the case when SDV is zero, the mixed distribution of observed values is estimated from the last set of data pairs as following.

$$F_{X_1/X_2=0}(x) = \begin{cases} M & \text{for } x_1 = 0, x_2 = 0 \\ M + (1-M)G_X(x) & \text{for } x_1 > 0, x_2 = 0 \end{cases} \quad (4)$$

where  $M$  and  $G_X(x)$  are the probability mass for OBS being zero and the cumulative probability distribution of non-zero OBS given zero SDV, respectively. Hence, from Equations (1), (3) and (4), a set of simulation curves for OBS given any value of SDV is obtained. Next, the most optimal quantile is ascertained to correct the bias in the best possible way by minimizing the mean absolute error between the same month-wise statistics (mean or extreme) of OBS and the bias-corrected SDV (henceforth, bias-corrected values; BCV). The value of the most optimal quantile varies spatio-temporally due to the varying nature of the bias for different regions (stemming from climatic/geographical differences) and different RCMs (due to varying

drizzling effect and overestimated extreme precipitation across RCMs). Hence, the estimated most optimal quantile differs spatially, and it is considered a model parameter. The most optimal quantile is estimated during the calibration period, and it is used for simulating BCV during the validation/future period. Furthermore, the model is run seasonally as characteristics of bias may change seasonally (given some RCMs are poor in reproducing seasonality).

Next, the QM method is also used, and the quality of bias-corrected products from the RMPH and QM models are compared. The QM reduces the bias in different quantiles on the assumption that the probability distribution of bias-corrected precipitation does not change when compared to the probability distribution of the observed precipitation (or the probability distribution remains stationary with time). For applying the QM, the best-fit probability distribution function is fitted for OBS and SDV during the calibration period. The best-fit probability distributions for OBS and SDV are selected in the same way as they were selected in the case of RMPH model. As stated earlier also, the bias-corrected SDV precipitation by QM method (hereafter QMC) is also provided in the dataset for comparison.

To explore the quality of the bias-corrected precipitation, the correspondence between SDV and OBS is compared with that between BCV and OBS (or QMC and OBS). Four statistical measures, namely, coefficient of determination ( $R^2$ ), refined index of agreement (Dr), unbiased root-mean-square error (uRMSE) and mean absolute error/distance (MAE) are utilized to quantify the correspondence between SDV, BCV and QMC with the OBS dataset (Maity et al., 2016). Additionally, Taylor diagrams (Taylor, 2001) are analysed for the correspondence between the variables. The coefficient of determination is a measure of fraction of variability of OBS explained by other series, and it ranges from 0 to 1 (for best possible association). The refined index of agreement is a measure of association conceptualized as the mean absolute distance between two series scaled by the mean absolute deviation of OBS from its mean. It varies from  $-1$  to  $1$ . uRMSE is the root-mean-square error (RMSE) between the 'deviation from the mean' series obtained from two data series. Hence, the uRMSE is a measure of the bias in the variability of two data series. The mean absolute error/distance (MAE) quantifies the total bias (arising from both – bias in mean and bias in variability) between two series. Lower values of uRMSE and MAE indicate better correspondence.

### 3 | DESCRIPTION OF DATA USED

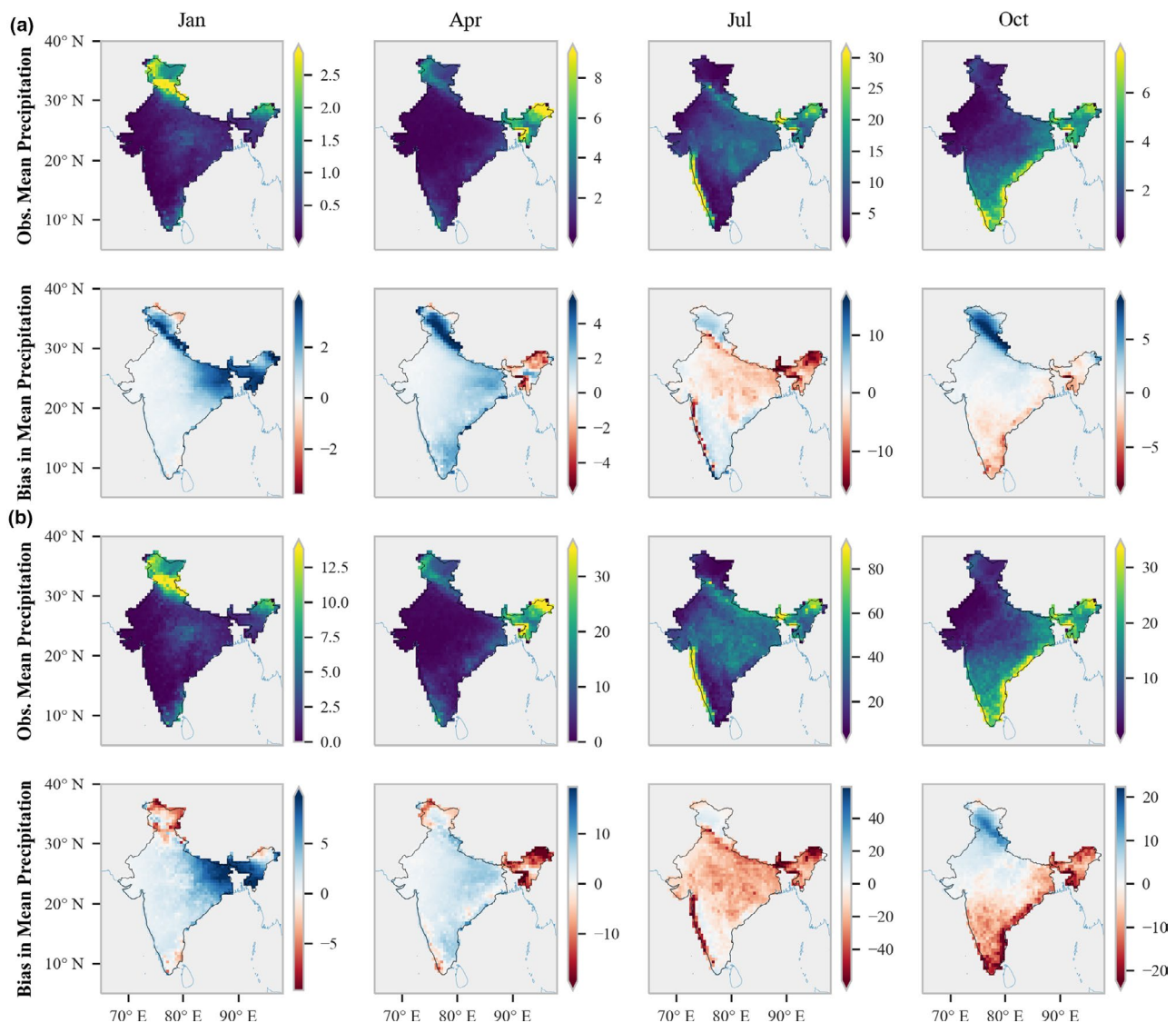
The quality of the bias-corrected products may vary due to diverse climatic conditions throughout the Indian mainland. Hence, to explore the spatial variation, four locations from each of the seven meteorologically homogenous divisions in

India (Kothawale & Rajeevan, 2017) are selected randomly. It should be noted that due to inadequate data reliability, hilly regions were not considered by Kothawale & Rajeevan, 2017, and five meteorologically homogenous divisions regions were demarcated. However, due to data availability for hill regions in this study, these regions are considered for analysis in this study (Hilly regions in north and north-east India are termed Hilly region 1 and 2, respectively). The study region is shown in Figure S2, and the details of these selected points are presented in Table S1 of the Supporting Information.

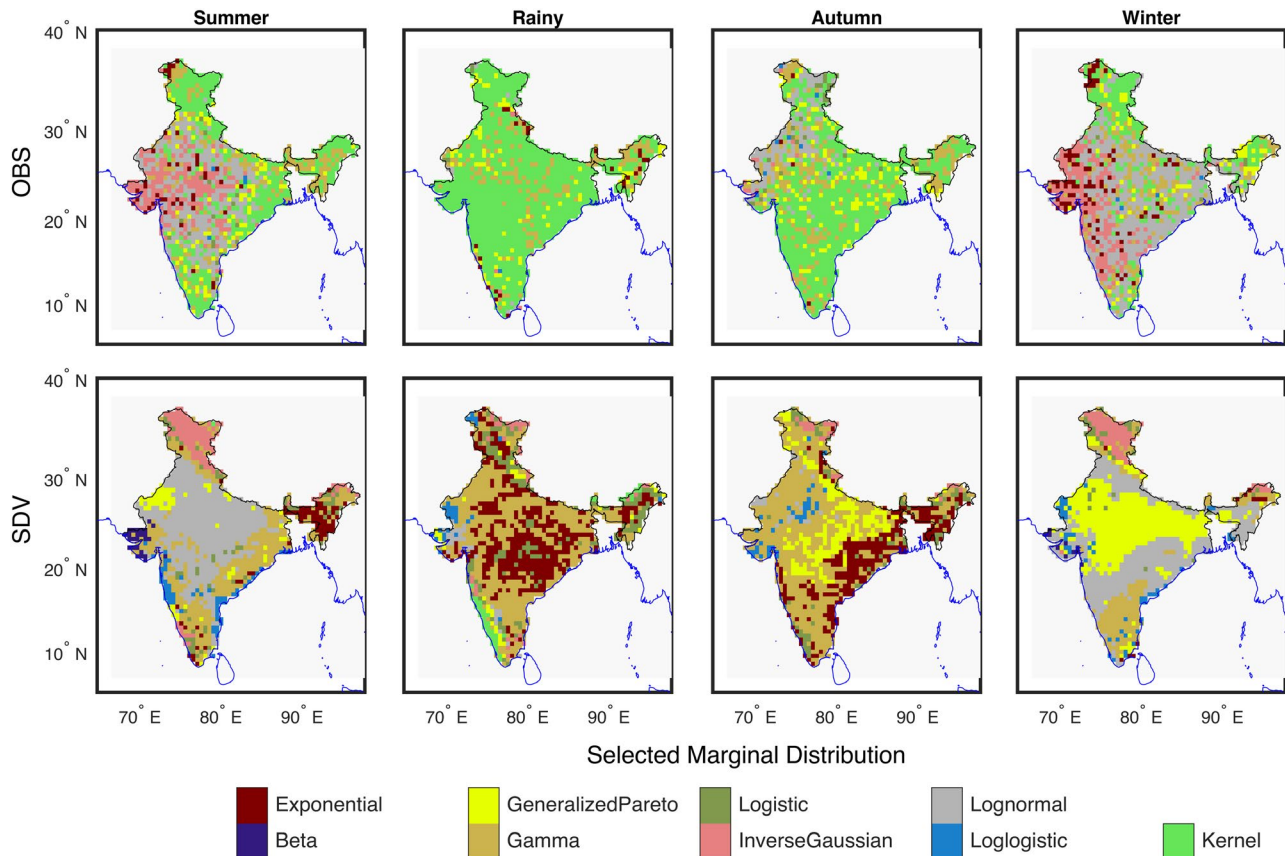
The observed daily precipitation for 1901–2017 at a spatial resolution of  $0.25^\circ \times 0.25^\circ$  is obtained from India Meteorological Department (IMD; Pai et al., 2014). The CORDEX (Girogi & Gutowski Jr, 2015) precipitation outputs from seven models (Table S2 of the Supporting Information) for the study region are obtained from Earth

System Grid Federation (ESGF; Cinquini et al., 2014). It may be noted that the primary aim of this study is to produce a bias-corrected future daily precipitation dataset for hydrological simulations. Hence, the dataset should have finer spatial resolution. As GCMs have a coarse resolution, a downscaling technique is required for getting finer scaled details. We can apply statistical downscaling, but as CORDEX is a coordinated project for downscaling, we used the CORDEX products. Additionally, though it is not always guaranteed, some RCMs do add value to GCM simulations (Singh et al., 2017). However, in general, outputs from RCMs are used to achieve a fine-resolution bias-corrected products that will be useful for further hydrological studies.

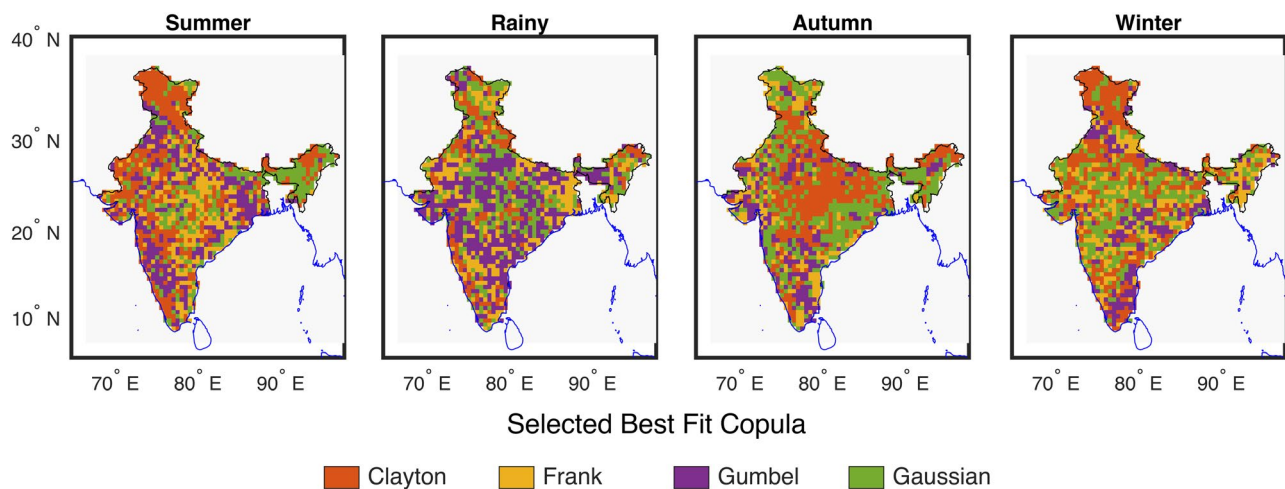
The spatial resolution of CORDEX data varies; however, most of them have a resolution close to  $50 \text{ km} \times 50 \text{ km}$ , and hence, they are regridded to a common resolution of  $0.50^\circ \times$



**FIGURE 2** Bias in mean and extreme precipitation (in mm/day) throughout India for selected months. (a) Monthly mean observed precipitation (first row) and corresponding bias in ensemble mean CORDEX precipitation (second row) for a selected month during calibration period, (b) Similar to a, but for monthly extreme precipitation (monthly 95<sup>th</sup> percentile)



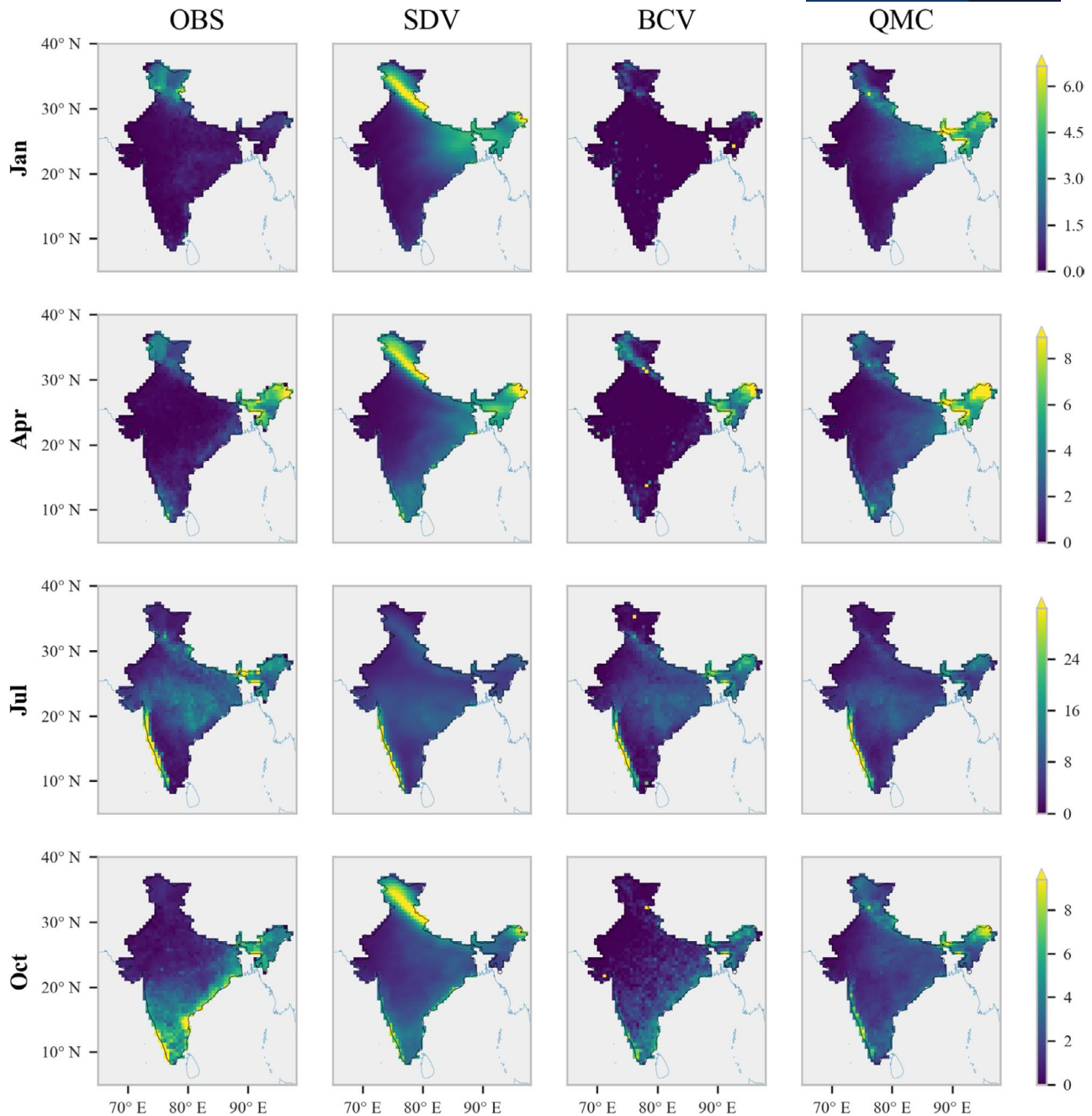
**FIGURE 3** Spatio-temporal distribution of selected best-fit marginal distribution for OBS and SDV during model calibration for REMO2009 RCM forced by MPI-ESM-LR GCM



**FIGURE 4** Spatio-temporal distribution of selected copula functions during model calibration for REMO2009 RCM forced by MPI-ESM-LR GCM. The best-fit copula is selected on the basis of the smallest Cramér-von Mises statistics

0.50° using the inverse distance-weighting method. To match the spatial resolution, the observed precipitation data are up-scaled from  $0.25^\circ \times 0.25^\circ$  to  $0.50^\circ \times 0.50^\circ$ . The CORDEX data are available for 1961–2100, of which 1961–2005 is historical simulation and the future period (2006–2100) is simulated using different RCPs. The CORDEX data from the historical period (1961–2005) are divided into a calibration

period (1961–1990), and a validation period (1991–2005). Bias-corrected estimates of daily precipitation (both most expected and extreme condition) are calculated for different RCPs in the future time period (2006–2100) and are provided in the dataset. It should be noted that the 95<sup>th</sup> percentile of daily precipitation is taken as a threshold for extreme precipitation. Selection of spatially varying extreme precipitation



**FIGURE 5** Spatio-temporal distribution of monthly mean OBS, SDV, BCV and QMC during validation period for four representative months. The SDV signifies ensemble mean CORDEX precipitation

threshold is desirable in climatically diverse region like Indian mainland.

## 4 | TECHNICAL VALIDATION

### 4.1 | Bias reduction in CORDEX using the RMPH model

The RMPH model is applied seasonally for four selected seasons: Summer (March–May), Rainy (June–August), Autumn

(September–November) and Winter (December–February). Bias correction is carried out separately for different CORDEX models. Hence, underlying details of the RMPH model can only be studied with respect to the individual CORDEX model. However, after bias correction, the ensemble mean of bias-corrected CORDEX precipitation (ensemble mean BCV) is calculated and used for assessing the quality of bias correction in terms of both mean and extreme values.

The bias in monthly mean precipitation for every CORDEX simulation is calculated throughout the region for analysis of spatio-temporal characteristics of the bias.

**TABLE 1** Correspondence of monthly mean most expected BCV, QMC and ensemble mean CORDEX (SDV) with respect to monthly mean OBS precipitation during the (a) calibration period, and (b) validation period.

Location code	SDV				BCV				QMC			
	$R^2$	Dr	MAE	uRMSE	$R^2$	Dr	MAE	uRMSE	$R^2$	Dr	MAE	uRMSE
(a)												
CN1	0.55	0.67	1.84	2.63	0.47	0.72	1.55	2.73	0.55	0.69	1.74	2.75
CN2	0.44	0.62	2.51	3.48	0.43	0.74	1.74	3.33	0.40	0.64	2.41	3.62
CN3	0.32	0.55	2.99	3.54	0.48	0.72	1.87	3.05	0.34	0.60	2.69	3.63
CN4	0.62	0.70	2.10	2.74	0.58	0.74	1.87	2.82	0.60	0.70	2.11	2.92
H11	0.00	-0.32	5.20	3.00	0.05	0.53	1.66	2.37	0.00	0.48	1.84	2.54
H12	0.17	0.50	3.89	5.07	0.51	0.65	2.71	3.92	0.15	0.55	3.49	5.16
H13	0.00	-0.37	3.87	2.23	0.00	0.57	1.05	1.76	0.01	0.48	1.29	1.92
H14	0.18	0.18	4.69	3.58	0.39	0.64	2.02	3.05	0.17	0.56	2.52	3.60
H21	0.27	0.61	5.79	8.39	0.44	0.66	4.99	7.14	0.27	0.60	5.83	8.24
H22	0.49	0.63	4.53	6.24	0.45	0.67	3.99	5.63	0.50	0.66	4.14	5.73
H23	0.56	0.67	2.98	4.00	0.55	0.71	2.61	3.67	0.57	0.67	3.01	3.94
H24	0.28	0.61	5.98	9.03	0.43	0.65	5.38	7.77	0.32	0.63	5.79	8.61
NE1	0.33	0.61	3.29	4.18	0.58	0.73	2.31	3.31	0.42	0.63	3.14	4.03
NE2	0.17	0.56	3.18	4.32	0.48	0.68	2.26	3.41	0.24	0.59	2.97	4.13
NE3	0.44	0.64	3.11	4.31	0.53	0.72	2.43	3.89	0.43	0.65	3.03	4.44
NE4	0.60	0.66	4.54	5.93	0.63	0.75	3.28	4.79	0.61	0.65	4.59	5.91
NW1	0.39	0.62	1.04	1.51	0.38	0.67	0.90	1.52	0.38	0.64	0.99	1.55
NW2	0.39	0.60	1.03	1.51	0.36	0.68	0.83	1.57	0.39	0.65	0.91	1.54
NW3	0.37	0.54	0.74	1.12	0.33	0.69	0.49	1.14	0.33	0.66	0.55	1.04
NW4	0.44	0.67	1.22	2.15	0.44	0.73	1.02	2.16	0.42	0.73	1.01	2.05
PE1	0.59	0.72	5.97	10.93	0.67	0.78	4.56	8.35	0.58	0.73	5.69	9.59
PE2	0.21	0.56	1.75	2.36	0.41	0.69	1.23	1.98	0.22	0.62	1.50	2.29
PE3	0.01	0.49	2.68	3.79	0.24	0.64	1.90	3.17	0.02	0.51	2.56	3.74
PE4	0.48	0.68	3.96	6.09	0.55	0.71	3.50	5.48	0.50	0.69	3.80	5.81
WC1	0.72	0.73	2.18	3.19	0.64	0.77	1.91	3.07	0.71	0.73	2.18	3.11
WC2	0.58	0.73	1.76	2.82	0.51	0.75	1.65	3.05	0.60	0.74	1.72	2.80
WC3	0.69	0.79	3.42	6.76	0.66	0.82	2.92	6.23	0.70	0.81	3.06	5.89
WC4	0.27	0.51	1.49	2.09	0.33	0.66	1.03	1.67	0.28	0.64	1.08	1.76
(b)												
CN1	0.48	0.64	2.31	3.33	0.44	0.71	1.87	3.32	0.48	0.66	2.16	3.41
CN2	0.37	0.58	2.46	3.16	0.36	0.71	1.72	3.20	0.36	0.62	2.28	3.24
CN3	0.34	0.56	2.84	3.30	0.64	0.76	1.53	2.40	0.34	0.59	2.59	3.40
CN4	0.57	0.71	2.23	3.32	0.53	0.73	2.05	3.27	0.55	0.70	2.31	3.52
H11	0.02	-0.11	4.32	3.21	0.08	0.40	2.30	2.41	0.01	0.43	2.20	2.70
H12	0.15	0.46	3.71	4.44	0.51	0.65	2.41	3.56	0.15	0.54	3.19	4.47
H13	0.01	-0.03	3.93	3.10	0.03	0.58	1.60	2.69	0.00	0.52	1.82	2.79
H14	0.24	0.12	4.66	3.17	0.40	0.62	1.98	2.79	0.22	0.55	2.35	3.21
H21	0.40	0.59	4.06	4.72	0.56	0.63	3.61	4.57	0.40	0.54	4.57	4.70
H22	0.59	0.65	4.49	5.85	0.62	0.74	3.38	4.68	0.61	0.68	4.12	5.19
H23	0.57	0.68	3.08	4.60	0.55	0.70	2.93	4.18	0.58	0.68	3.11	4.54
H24	0.38	0.60	4.32	5.63	0.54	0.64	3.89	5.00	0.39	0.53	5.05	5.45

(Continues)



TABLE 1 (Continued)

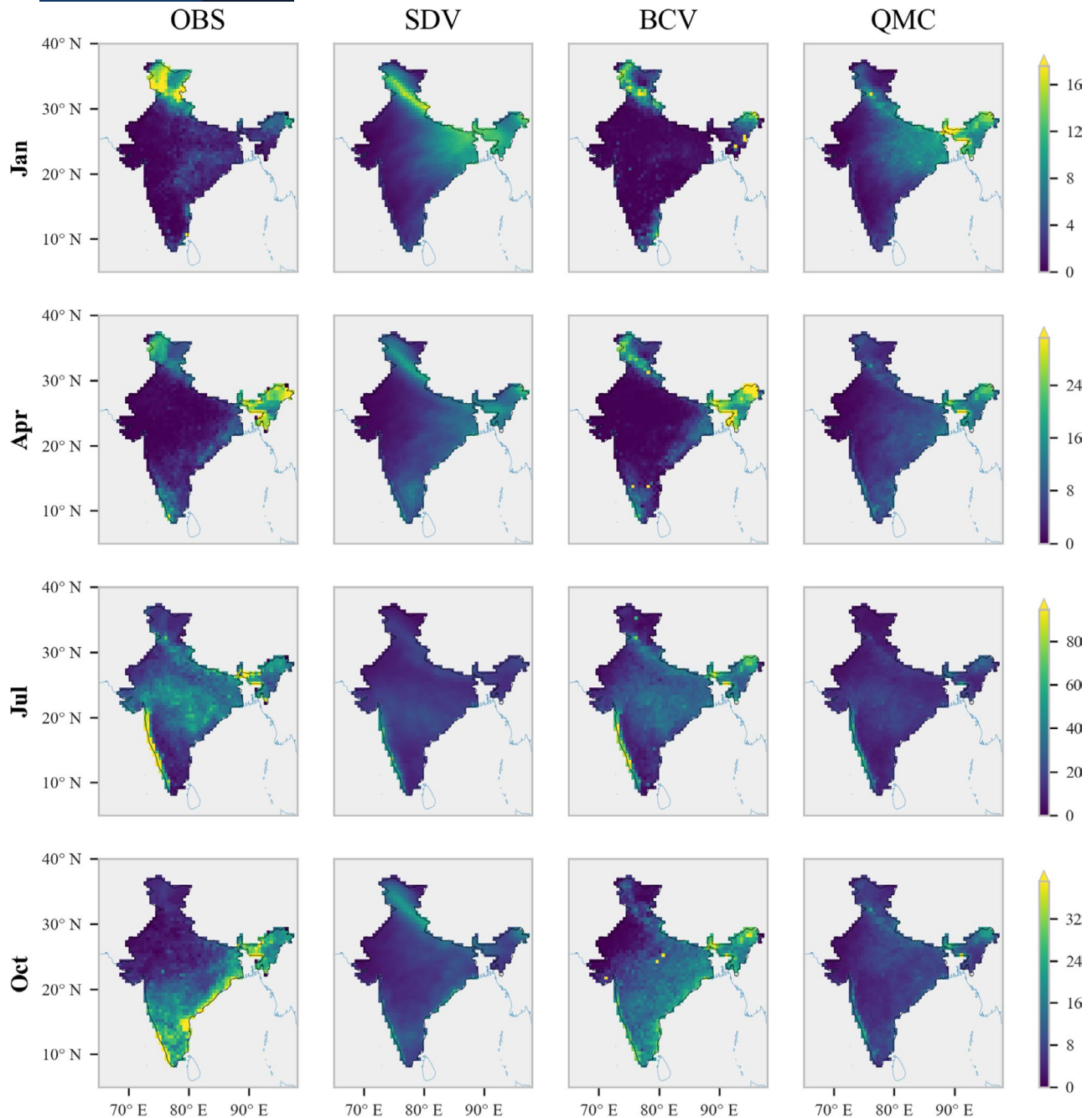
Location code	SDV				BCV				QMC			
	$R^2$	Dr	MAE	uRMSE	$R^2$	Dr	MAE	uRMSE	$R^2$	Dr	MAE	uRMSE
NE1	0.30	0.62	3.31	4.41	0.54	0.72	2.42	3.56	0.37	0.63	3.21	4.29
NE2	0.16	0.57	3.15	4.29	0.41	0.67	2.41	3.59	0.23	0.59	3.01	4.10
NE3	0.49	0.66	3.07	3.86	0.60	0.75	2.29	3.35	0.50	0.66	3.05	3.96
NE4	0.65	0.68	2.95	3.55	0.71	0.72	2.54	3.31	0.67	0.67	3.01	3.50
NW1	0.21	0.41	1.05	1.31	0.26	0.61	0.69	1.21	0.19	0.50	0.89	1.25
NW2	0.49	0.62	0.92	1.24	0.38	0.67	0.81	1.38	0.49	0.66	0.83	1.28
NW3	0.29	0.59	0.79	1.33	0.28	0.66	0.65	1.37	0.33	0.69	0.60	1.23
NW4	0.48	0.70	1.17	1.99	0.50	0.73	1.04	2.03	0.45	0.74	1.02	1.94
PE1	0.68	0.75	4.58	7.89	0.76	0.79	3.75	6.30	0.65	0.74	4.70	6.88
PE2	0.18	0.53	1.91	2.55	0.34	0.66	1.40	2.22	0.17	0.60	1.65	2.51
PE3	0.01	0.47	2.98	4.13	0.26	0.64	2.01	3.38	0.01	0.49	2.84	4.12
PE4	0.52	0.67	5.06	6.68	0.62	0.74	4.08	5.77	0.54	0.71	4.51	6.25
WC1	0.74	0.72	2.44	3.61	0.62	0.78	1.94	3.44	0.74	0.72	2.44	3.50
WC2	0.51	0.71	1.75	2.70	0.55	0.77	1.38	2.55	0.51	0.72	1.71	2.67
WC3	0.70	0.78	4.41	8.48	0.71	0.80	3.93	7.57	0.73	0.81	3.86	7.31
WC4	0.21	0.42	1.87	2.53	0.35	0.67	1.04	1.75	0.23	0.61	1.26	1.99

The spatial variation of observed monthly mean and extreme precipitation, and corresponding bias in ensemble mean CORDEX precipitation for four selected months (January, April, July and October; each representing one season) during calibration period are shown in Figure 2. For monthly mean precipitation, the ensemble mean CORDEX mostly overestimates for the dry season and underestimates during the wet season. For instance, in the month of April (a comparatively dry month), the ensemble mean CORDEX overestimates the monthly mean precipitation in most of India excluding north-east. However, during July (a monsoon month), the ensemble mean CORDEX is found to underestimate the precipitation in most parts of the north and north-east India. Additionally, the bias is high in areas having higher monthly mean precipitation. In the Himalayan regions of north India, the ensemble mean CORDEX shows a predominantly wet bias, *i.e.*, ensemble mean CORDEX overestimates precipitation magnitude. Similar but relatively intensified patterns (in both bias magnitude and its spatial extent) are observed for monthly extreme precipitation. During July, the Western Ghats is showing the highest dry bias (underestimation of precipitation by CORDEX models) for extreme precipitation; however, during October, most of the regions showing the highest dry bias for extreme precipitation are situated in the Eastern Ghats. The spatio-temporal variation of the bias is different for different CORDEX models (Figures S3–S5 in the Supporting Information show the spatio-temporal distribution of bias for three CORDEX models: REMO2009 RCM forced by MPI-ESM-LR GCM, RegCM4 forced by

CSIRO-Mk3.6.0 and GFDL-ESM2M GCMs), which can be attributed to design/modelling differences between them.

Before feeding to the model, observed precipitation at each point is checked for outliers during the calibration period. Daily precipitation magnitudes showing a deviation of more than five standard deviation unit from the mean are regarded as outlier. These are not extreme events and may have resulted from erroneous observation records. Such instances are less than 0.33% for the 99.5% of Indian mainland for the entire calibration period. The OBS and corresponding SDV for those time steps are not used while calibrating the model. Additionally, the IMD observation stations are not distributed uniformly over Indian landmass, resulting in non-uniform relative error in precipitation records (discussed in details in Section S1 in the Supporting Information). This information is kept in mind while analysing the quality of bias-corrected precipitation dataset.

As stated earlier, all the OBS-SDV pairs are separated into three groups for each CORDEX model (Figure 1). For the first group of OBS-SDV pairs where both are non-zero, a suitable marginal distribution is fitted individually to OBS and SDV for converting them to their reduced variates (non-exceeding probability). The spatial and seasonal distribution of the selected best-fit marginal distribution for OBS and SDV for REMO2009 RCM model forced by MPI-ESM-LR GCM is shown in Figure 3. The spatial and seasonal distribution of the best-fit marginal distribution for two other CORDEX models are shown in Figure S6 of the Supporting Information. The spatio-temporal distribution of the best-fit marginal distribution varies with different CORDEX models. Additionally, the



**FIGURE 6** Spatio-temporal distribution of monthly extreme OBS, SDV, BCV and QMC during validation period for four representative months. The SDV signifies data from ensemble mean CORDEX precipitation

statistical characteristics of pairs of non-zero OBS and SDV change spatially and seasonally for a given CORDEX model. The kernel-based empirical marginal distribution is selected as best-fit distribution at most locations for non-zero OBS precipitation. Using selected marginal distributions, pairs of non-zero OBS and SDV are converted to their respective reduced variates. Four different copula functions are used to model the inter-relationship between non-zero OBS and SDV using their reduced variates. The best copula function is selected based on the smallest Cramér-von Mises statistics. The spatio-temporal distribution of the selected copula functions

for REMO2009 RCM model forced by MPI-ESM-LR GCM for different seasons is shown in Figure 4. Similar figures for two other CORDEX models are shown in Figure S7 of the Supporting Information.

The conditional distribution of non-zero OBS, given non-zero SDV, is obtained from the selected copula function. From the second group of OBS-SDV pairs (with zero OBS), the entire range of SDV is divided into different classes, and the frequency of SDV values having corresponding zero OBS in those classes is ascertained. The frequency of SDV having zero OBS is expected to decrease with increasing SDV.

Hence, an exponentially decaying curve of form Equation (2) is used to model the frequency of SDV corresponding to zero OBS given the class mark (midpoint of class) of SDV as shown in Figure S8a of the Supporting Information. The assumption of decreasing frequency of SDV having zero OBS with increasing SDV holds well for most locations and season combinations; but some cases are found in which the frequency is found to increase with increasing SDV, which means that the RCM is not able to capture the seasonality well. Additionally, it should be noted that the uncertainty in the OBS is not uniform across India due to the non-uniform distribution of raingauge stations as discussed in Section S1 of the Supporting Information. It may be noted that the performance of RCM may not be spatially uniform, given the climatological, topological or other spatially varying factors. Hence, the correspondence between SDV and OBS itself may be poor at some locations. As any bias-correction method heavily depends upon the quality of output of driving climate model and observed data, so the bias-correction methods are expected to perform poor at those locations (Maraun et al., 2017).

The parametric conditional distribution obtained from the best-fit copula is then modified to include the probability mass of SDV corresponding to zero OBS, resulting in a set of mixed probability distribution of OBS for different positive values of SDV. One such set of conditional distribution for location NE2 is shown in Figure S8b of the Supporting Information. These simulation curves are for non-zero SDV from REMO2009 RCM forced by MPI-ESM-LR GCM during the autumn season. Furthermore, the probability of OBS given zero SDV is estimated from the third set of OBS-SDV pairs with zero SDV. The BCV is then obtained by using (a) the simulation curves for  $SDV > 0$  and (b) probability of OBS given zero SDV, as shown in Figure 1.

The calibrated model is then used to estimate the mean daily BCV. The optimal quantile for the mean daily BCV is obtained by comparing the monthly mean OBS with BCV at different quantiles during calibration period. The spatio-temporal distribution of the mean monthly BCV along with OBS, ensemble mean CORDEX precipitation (SDV) and corresponding QMC for four representative months during the validation period is shown in Figure 5. The spatio-temporal distribution of BCV matches better to OBS when compared with SDV. Despite the spatio-temporal variation of bias and change in its characteristics between calibration and validation period, the satisfactory quality of bias-corrected precipitation bolsters the hypothesis that the model is able to capture and reduce the spatial and seasonal variation of bias. Location-wise correspondence of BCV and SDV with OBS is presented in Table 1, and the quality of bias-corrected precipitation is inferred from grouped Taylor diagrams provided in Figures S9 and S10 of the Supporting Information. From the figures, the correlation and variability pattern of

BCV are better than SDV when compared to OBS in most of the hydro-meteorological homogenous regions. Additionally, from Table 1, the BCV is found to have lower MAE, and better Dr as compared to SDV. In some locations, the correspondence between SDV and OBS is not satisfactory, for example, the locations falling inside H1 and H2 regions, NE2, NW3, PE2 and PE3. The reason for the low correspondence for locations PE2 and PE3 might be due to topological and climatic factors. These regions are in Eastern Ghats and unlike most of India receive rain for longer time (July–November). For the locations NE2, NW3 and locations falling inside regions H1 and H2, the small number of stations used for recording observed precipitation (hence, resulting in higher uncertainty; Section S1 of the Supporting Information) might be the reason for the low correspondence. At these locations, the BCV is found to have better correspondence to OBS compared to SDV as indicated by higher  $R^2$  and Dr, and lower values of MAE and uRMSE, which shows the efficacy of the RMPH model. However, the BCV in such locations should be used with caution, as a bias-correction method cannot be taken as a substitute for inadequate modelling of seasonality by climate models (Maraun et al., 2017). Even in cases where the correspondence between SDV and OBS changed between calibration and validation periods (hence, the characteristics of bias changed; e.g., all the points falling in H2 region), the RMPH model is found to perform satisfactorily.

The RMPH model is able to reduce the bias for extreme values too. For comparing extreme precipitation, 95<sup>th</sup> percentile of daily OBS and SDV is compared with the mean of 95<sup>th</sup> quantile daily BCV. The spatio-temporal distributions of extreme BCV as compared to extreme OBS and SDV (ensemble mean CORDEX) for four representative months are shown in Figure 6. The spatio-temporal distribution of monthly extreme BCV is found to match better with monthly extreme OBS when compared to monthly extreme SDV. This suggests that the RMPH model suitably reduces the bias in the extreme precipitation, even for locations where the bias characteristics of extreme precipitation are different than bias characteristics of monthly mean precipitation. This benefit of the RMPH model is due to the provision of different simulation curves used for different values of SDV; hence, it captures the varying nature of bias at different quantiles. Therefore, for the analysis of extreme events affected by extreme precipitation, BCV is better suited for hydrological simulation as compared to SDV (*i.e.*, CORDEX simulation outputs). Furthermore, location-wise correspondence between OBS and SDV, and OBS and BCV for extreme precipitation are shown in Table 2. The quality of bias-corrected product can be inferred from the grouped Taylor diagrams presented in Figures S11 and S12 of the Supporting Information. Figures S11 and S12 show reduced root-mean-square distance between BCV and OBS when compared to SDV and OBS, barring a few locations with inadequate stations for OBS (Figure S1). Comparing Tables 1 and 2, the correspondence

between monthly extreme OBS and SDV is found to be inferior compared to the case of monthly mean. Hence, the RCMs used in CORDEX do not have same skill while predicting mean and extreme precipitation, resulting in different spatio-temporal characteristics of bias in both cases as shown in Figure 2. The correspondence between monthly extreme BCV and OBS is better compared to the correspondence between monthly extreme SDV and OBS, which is revealed by lower values of MAE and uRMSE and higher values of  $R^2$  and Dr (Table 2). Even in the case of locations with very low correspondence between SDV and OBS, such as H12, H14, NE2, PE2 and PE3, the RMPH model is found to perform satisfactorily.

## 4.2 | Comparison with quantile mapping

The BCV is also compared with the bias-corrected precipitation obtained using the QM method (QMC) in Figures 5 and 6, Tables 1 and 2, and Figures S9–S12 of the Supporting Information. The comparison helps in highlighting the quality of bias-corrected precipitation obtained from RMPH compared to that from QM method. Figures 5 and 6 show that the spatio-temporal variation of BCV (from RMPH model) matches better with OBS as compared to QMC. This difference in the quality is due to the assumption of QM method itself: the QM and other methods based on QM method bias correct the simulated values by matching the quantile with OBS. This assumption might not be correct as the bias is affected by multiple factors as outlined before and can also vary spatially and seasonally (Maity et al., 2019; Mao et al., 2015). In the case of the RMPH model, the relationship between OBS and SDV is modelled using the generated simulation curves, conditional on the simulated value, and these curves vary spatially. Depending upon the SDV, the conditional distribution of OBS changes, which helps in reducing the bias in different quantiles. Further, the provision of spatially varying most optimal quantile as a calibrated model parameter also helps in capturing spatially varying bias. Hence, the RMPH model is more flexible, and it is supposed to be more effective in reducing biases (Maity et al., 2019) compared to QM. This is also evident in the case of location-wise comparison as provided in Tables 1 and 2. From Tables 1 and 2, the correspondence of monthly mean or extreme BCV with OBS is found to be better compared to the correspondence of QMC with OBS as evident by lower MAE and uRMSE and higher values of  $R^2$  and Dr. Additionally, it is found that during the wet period of the year, QMC is under-predicting the extreme precipitation; however, it is over-predicting during the dry period of year. Hence, the generated bias-corrected precipitation (BCV; bias-corrected using the RMPH model) is a better choice for hydrological simulations as compared to either of CORDEX precipitation data (SDV) or CORDEX precipitation data corrected by the QM method (QMC).

Finally, assumption of stationary bias in the data is another important point to mention. In general, if the bias is non-stationary, bias-corrected products will be more reliable in the near future than in the far future. Most of the existing bias-correction methods including RMPH model inherently assume stationarity in the bias. Thus, the limitation with respect to model capability in the far future is undeniable. Incorporating the time-varying characteristics in the existing bias-correction models may be kept as a future scope of this study.

## 5 | SUMMARY AND CONCLUSIONS

In this study, a bias-corrected CORDEX precipitation dataset (bias corrected using the newly proposed RMPH model based on copula functions) is evaluated for tropical region like Indian mainland. Additionally, the bias-correction performance is compared with quantile mapping. The observed data provided by the India Meteorological Department suffer from non-uniform relative error due to non-uniform distribution of raingauge station across the country. Despite this, the quality of bias-corrected product from RMPH model is found to be satisfactory across the country compared to quantile mapping in case of mean precipitation.

The better quality of bias correction in the dataset, bias-corrected using the RMPH method, can be explained by different design considerations in the model. For instance, the RMPH model is found to capture the regional variation in climatic condition better than quantile mapping. This observed better spatial transferability of the model might result from provision of the location-specific most optimal percentile and different simulation curves for different values of RCM precipitation. The provision of different simulation curve for different values of RCM precipitation is based on hypothesis that characteristic of bias changes with magnitude of RCM simulated precipitation. These flexibilities result in better modelling of bias for different quantiles of daily precipitation, which in turn result in better quality of bias-corrected precipitation dataset from the model as compared to other method like quantile mapping.

Similar to the case of mean precipitation, the bias-corrected precipitation from RMPH model is better than the one from quantile mapping in reducing the bias in extreme daily precipitation. The correspondence between monthly extreme observed precipitation and monthly extreme RCM precipitation is found to be inferior as compared to the case of monthly mean. This suggests that climate models used in CORDEX do not have the same skill to predict extreme precipitation when compared to mean precipitation, resulting in different spatio-temporal variation of bias in extreme precipitation compared to bias in monthly mean precipitation. Even in this case, the RMPH model performed better due to the provision of different simulation curves

**TABLE 2** Correspondence of monthly extreme BCV, QMC and ensemble mean CORDEX (SDV) with respect to monthly extreme OBS precipitation during the (a) calibration period, and (b) validation period.

Location code	SDV				BCV				QMC			
	$R^2$	Dr	MAE	uRMSE	$R^2$	Dr	MAE	uRMSE	$R^2$	Dr	MAE	uRMSE
(a)												
CN1	0.39	0.63	8.8	13.3	0.47	0.71	7.1	11.6	0.40	0.63	9.0	13.7
CN2	0.21	0.60	11.2	17.0	0.44	0.72	7.6	13.8	0.21	0.60	11.2	17.1
CN3	0.13	0.56	12.0	16.3	0.46	0.71	8.0	12.7	0.16	0.58	11.5	16.3
CN4	0.33	0.62	11.4	16.7	0.46	0.70	9.0	14.2	0.34	0.60	12.0	17.1
H11	0.00	0.39	9.3	11.1	0.03	0.54	7.1	10.5	0.00	0.51	7.5	10.9
H12	0.04	0.54	13.5	20.6	0.40	0.64	10.6	16.3	0.04	0.54	13.5	20.6
H13	0.00	0.35	7.4	8.8	0.00	0.57	4.9	8.4	0.00	0.54	5.3	8.7
H14	0.06	0.46	11.1	13.3	0.23	0.60	8.2	11.9	0.07	0.54	9.4	13.2
H21	0.21	0.53	24.4	32.1	0.35	0.63	19.2	27.9	0.27	0.57	22.5	31.3
H22	0.42	0.50	17.7	20.1	0.36	0.63	13.0	18.2	0.40	0.58	14.9	19.5
H23	0.34	0.58	11.4	14.7	0.45	0.68	8.7	12.8	0.37	0.58	11.5	15.0
H24	0.25	0.51	27.7	34.3	0.37	0.63	21.2	29.5	0.29	0.57	24.6	33.3
NE1	0.18	0.58	11.4	15.7	0.45	0.70	8.1	12.8	0.25	0.56	12.0	15.8
NE2	0.08	0.52	11.3	14.5	0.40	0.65	8.1	11.7	0.12	0.51	11.4	14.2
NE3	0.20	0.58	13.8	19.9	0.41	0.68	10.3	16.6	0.21	0.57	14.2	20.1
NE4	0.33	0.60	15.6	21.2	0.49	0.71	11.4	17.3	0.34	0.59	16.0	21.6
NW1	0.17	0.60	5.8	9.5	0.30	0.67	4.8	8.7	0.19	0.60	5.8	9.5
NW2	0.16	0.59	5.0	8.3	0.33	0.69	3.8	7.4	0.17	0.61	4.8	8.3
NW3	0.25	0.63	3.6	6.6	0.29	0.71	2.8	6.5	0.24	0.65	3.4	6.7
NW4	0.30	0.72	5.4	13.6	0.34	0.73	5.2	13.1	0.31	0.71	5.6	13.9
PE1	0.49	0.64	23.8	39.0	0.64	0.75	16.5	27.6	0.50	0.69	20.6	35.8
PE2	0.17	0.61	7.8	11.5	0.41	0.69	6.2	9.8	0.18	0.60	7.9	11.6
PE3	0.05	0.55	11.9	17.5	0.28	0.63	9.8	15.4	0.06	0.55	11.8	17.5
PE4	0.30	0.55	14.1	17.4	0.48	0.67	10.3	14.7	0.35	0.59	12.8	17.0
WC1	0.46	0.64	10.6	15.3	0.57	0.72	8.1	12.2	0.48	0.63	10.9	15.6
WC2	0.41	0.68	8.2	13.6	0.47	0.73	6.9	12.4	0.48	0.67	8.6	13.9
WC3	0.61	0.73	15.0	25.6	0.69	0.80	10.9	19.3	0.68	0.74	14.5	24.2
WC4	0.21	0.62	5.5	8.7	0.33	0.65	5.0	7.9	0.26	0.64	5.3	8.5
(b)												
CN1	0.34	0.62	11.1	17.5	0.43	0.70	8.8	15.2	0.36	0.63	11.0	17.8
CN2	0.19	0.59	11.1	17.0	0.38	0.70	8.1	14.4	0.20	0.59	11.0	17.0
CN3	0.14	0.55	11.4	14.6	0.60	0.72	7.0	9.9	0.17	0.57	10.9	14.6
CN4	0.31	0.61	11.9	17.2	0.47	0.72	8.6	14.2	0.29	0.60	12.3	17.6
H11	0.02	0.47	10.1	13.2	0.10	0.45	10.5	11.6	0.02	0.35	12.3	12.6
H12	0.02	0.52	12.3	17.7	0.40	0.61	10.1	13.9	0.03	0.53	12.0	17.6
H13	0.01	0.48	10.0	13.5	0.02	0.58	8.1	13.0	0.00	0.56	8.5	13.2
H14	0.10	0.43	10.2	10.8	0.27	0.59	7.3	9.7	0.08	0.53	8.3	10.9
H21	0.36	0.60	12.1	15.7	0.46	0.53	14.1	15.0	0.32	0.60	12.1	15.8
H22	0.49	0.53	18.3	19.5	0.50	0.68	12.4	16.3	0.50	0.60	15.6	18.7
H23	0.37	0.56	13.3	18.2	0.40	0.66	10.3	16.3	0.38	0.56	13.5	18.6
H24	0.26	0.59	13.4	18.9	0.44	0.43	18.5	17.0	0.29	0.61	12.9	18.2

(Continues)

TABLE 2 (Continued)

Location code	SDV				BCV				QMC			
	$R^2$	Dr	MAE	uRMSE	$R^2$	Dr	MAE	uRMSE	$R^2$	Dr	MAE	uRMSE
NE1	0.11	0.59	11.7	18.8	0.34	0.69	8.9	16.2	0.11	0.56	12.6	19.1
NE2	0.07	0.51	12.4	15.2	0.31	0.64	9.0	13.1	0.11	0.49	12.8	15.0
NE3	0.24	0.57	15.5	19.3	0.54	0.71	10.5	14.6	0.29	0.56	15.9	19.5
NE4	0.29	0.60	11.2	14.1	0.56	0.69	8.6	11.3	0.33	0.59	11.4	14.2
NW1	0.24	0.57	4.2	6.0	0.30	0.62	3.7	6.0	0.23	0.58	4.1	6.1
NW2	0.34	0.61	5.1	7.9	0.33	0.68	4.2	7.7	0.35	0.62	5.0	8.2
NW3	0.27	0.66	3.3	6.4	0.34	0.71	2.7	6.2	0.36	0.69	3.0	6.3
NW4	0.41	0.72	5.6	10.1	0.51	0.74	5.1	9.5	0.45	0.71	5.8	10.5
PE1	0.53	0.66	18.8	28.1	0.72	0.76	13.6	20.1	0.53	0.72	15.9	25.4
PE2	0.08	0.61	8.4	14.3	0.31	0.67	7.1	12.4	0.10	0.61	8.3	14.2
PE3	0.02	0.56	12.3	19.3	0.22	0.61	11.2	17.3	0.02	0.56	12.5	19.3
PE4	0.35	0.54	21.8	24.6	0.56	0.69	15.0	20.2	0.42	0.58	20.3	24.0
WC1	0.52	0.64	11.4	16.7	0.58	0.75	8.0	13.2	0.53	0.63	11.8	16.9
WC2	0.29	0.67	9.4	15.9	0.46	0.74	7.3	13.8	0.29	0.65	9.9	16.3
WC3	0.59	0.73	18.9	33.5	0.71	0.80	13.8	25.0	0.63	0.72	18.9	32.9
WC4	0.19	0.62	6.7	10.7	0.33	0.66	6.0	9.7	0.24	0.66	6.0	10.4

for different quantiles of precipitation as mentioned above. Compared to the RMPH model, quantile mapping is found to generally over-predict extreme precipitation during dry months and to under-predict precipitation during wet months. Hence, the bias-corrected precipitation dataset (obtained from RMPH model) is better than either CORDEX outputs or CORDEX outputs bias-corrected using quantile mapping.

## 6 | POTENTIAL DATA SET USE AND USAGE NOTES

The bias-corrected daily CORDEX precipitation dataset from RMPH model is being released for the research community. As the BCV is found to be better than both SDV and QMC, the developed dataset is expected to be particularly suited for hydrological simulations, formulating extreme event mitigation strategies, and climate change adaptation strategies over India. The dataset (Suman et al., 2019) is provided in the form of self-documented NetCDF files (\*.nc). The file names are formatted as follows:

```
pr_<domain>_<GCM_name>_<scenario>_
r1ilp1_<RCM_name>_day_<corr_method>.nc
```

Different parts of the file name are described as:

1. **domain**: Domain name as per CORDEX project. It can be either 'WAS-44' or 'WAS-44i' for Indian mainland.

2. **GCM\_name**: Name of driving GCM as given in Table S2.
3. **scenario**: RCP Scenario. It can be among 'historical' (for historical CORDEX simulations), 'rcp26' (for RCP 2.6; if available), 'rcp45' (RCP4.5) and 'rcp85' (RCP8.5).
4. **RCM\_name**: Name of RCM used as given in Table S2.
5. **corr\_method**: Identifier for bias-correction method used. It can be among 'bias\_corrected\_expected', 'bias\_corrected\_extreme' or 'quantile\_mapping'. File name ending with 'bias\_corrected\_expected' has mean BCV from RMPH model. File name ending with 'bias\_corrected\_extreme' has 95<sup>th</sup> percentile daily BCV, hence, an estimate of extreme precipitation daily. File name ending with 'quantile\_mapping' has QMC.

Files with name ending with 'bias\_corrected\_expected' should be used for hydrological simulation as it contains most expected daily BCV. All the files have spatial extent of 6.75°N, 66.75°E to 38.25°N, 99.75°E with a spatial resolution of 0.5° (latitude) × 0.5° (longitude) and temporal resolution as daily. In the dataset, the duration of historical period and future period for different RCPs are 1961–2005 and 2006–2100 (like CORDEX). It should be noted that files with names ending with 'bias\_corrected\_extreme' have estimates of precipitation with the assumption that extreme condition prevails every day. Hence, these files cannot be used for daily simulation, rather they can be used to assess mean extreme precipitation for weeks, months or years by averaging their daily values.

## ACKNOWLEDGEMENTS

Corresponding author, Dr. Rajib Maity, is a recipient of the Humboldt Fellowship for Experienced Researchers from the Alexander von Humboldt (AvH) Foundation, Germany. Support from AvH is duly acknowledged. Funding from Helmholtz-Association of German Research Centres (HGF) and its Programme "Changing Earth -Sustaining our Future", Topic "The Atmosphere in Global Change " (12.11.33) is gratefully acknowledged. The work was also partially supported by the sponsored projects supported by Department of Science and Technology, Climate Change Programme (SPLICE), Government of India (Ref No. DST/CCP/CoE/79/2017(G)) through a sponsored project.

## OPEN PRACTICES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at <http://doi.org/10.6084/m9.figshare.9686648.v2> Learn more about the Open Practices badges from the Center for OpenScience: <https://osf.io/tvyxz/wiki>.

## ORCID

Mayank Suman  <https://orcid.org/0000-0002-1402-8982>

Rajib Maity  <https://orcid.org/0000-0001-5631-9553>

Harald Kunstmann  <https://orcid.org/0000-0001-9573-1743>

## REFERENCES

- Cho, C., Li, R., Wang, S.Y., Yoon, J.H. & Gillies, R.R. (2016) Anthropogenic footprint of climate change in the June 2013 northern India flood. *Climate Dynamics*, 46(3–4), 797–805.
- Christensen, J.H., Boberg, F., Christensen, O.B. & Lucas-Picher, P. (2008) On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophysical Research Letters*, 35(20), L20709. <https://doi.org/10.1029/2008GL035694>
- Cinquini, L., Crichton, D., Mattmann, C., Harney, J., Shipman, G., Wang, F. et al. (2014) The Earth system grid federation: an open infrastructure for access to distributed geospatial data. *Future Generation Computer Systems*, 36, 400–417.
- de Elía Ramón, Laprise René, Biner Sébastien, Merleau James (2017). Synchrony between reanalysis-driven RCM simulations and observations: variation with time scale. *Climate Dynamics*, 48, (7–8), 2597–2610. <http://dx.doi.org/10.1007/s00382-016-3226-0>
- De Souza, K., Kituyi, E., Harvey, B., Leone, M., Murali, K.S. & Ford, J.D. (2015) Vulnerability to climate change in three hot spots in Africa and Asia: key issues for policy-relevant adaptation and resilience-building research. *Regional Environmental Change*, 15(5), 747–753.
- Fowler, H.J., Ekström, M., Blenkinsop, S. & Smith, A.P. (2007) Estimating change in extreme European precipitation using a multimodel ensemble. *Journal of Geophysical Research: Atmospheres*, 112(D18), D18104. <https://doi.org/10.1029/2007JD008619>
- Genest, C., Rémillard, B. & Beaudoin, D. (2009) Goodness-of-fit tests for copulas: a review and a power study. *Insurance: Mathematics and economics*, 44(2), 199–213.
- Ghosh, S., Das, D., Kao, S.C. & Ganguly, A.R. (2012) Lack of uniform trends but increasing spatial variability in observed Indian rainfall extremes. *Nature Climate Change*, 2(2), 86–91.
- Giorgi, F. & Gutowski, W.J. Jr (2015) Regional dynamical downscaling and the CORDEX initiative. *Annual Review of Environment and Resources*, 40, 467–490.
- Houze, R.A. Jr, McMurdie, L.A., Rasmussen, K.L., Kumar, A. & Chaplin, M.M. (2017) Multiscale aspects of the storm producing the June 2013 flooding in Uttarakhand, India. *Monthly Weather Review*, 145(11), 4447–4466.
- Kitoh, A., Hirokazu Endo, K., Kumar, K., Cavalcanti, I.F.A., Goswami, P. & Zhou, T. (2013) Monsoons in a changing world: a regional perspective in a global context. *Journal of Geophysical Research: Atmospheres*, 118(8), 3053–3065.
- Kothawale, D.R. & Rajeevan, M. (2017) Monthly, seasonal and annual rainfall time series for all-India, homogeneous regions and meteorological subdivisions: 1871–2016. Contribution from IITM Research Report No. RR-138. Available at: <https://www.tropmet.res.in/~lip/Publication/RR-pdf/RR-138.pdf>
- Krishnan, R., Sabin, T.P., Vellore, R., Mujumdar, M., Sanjay, J., Goswami, B.N. et al. (2016) Deciphering the desiccation trend of the South Asian monsoon hydroclimate in a warming world. *Climate Dynamics*, 47(3–4), 1007–1027.
- Kumar, K.N., Rajeevan, M., Pai, D.S., Srivastava, A.K. & Preethi, B. (2013) On the observed variability of monsoon droughts over India. *Weather and Climate Extremes*, 1, 42–50.
- Lafon, T., Dadson, S., Buys, G. & Prudhomme, C. (2013) Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. *International Journal of Climatology*, 33(6), 1367–1381.
- Maity, R., Suman, M., Laux, P. & Kunstmann, H. (2019) Bias correction of zero-inflated RCM precipitation fields: a copula-based scheme for both mean and extreme conditions. *Journal of Hydrometeorology*, 20(4), 595–611.
- Maity, R., Suman, M. & Verma, N.K. (2016) Drought prediction using a wavelet based approach to model the temporal consequences of different types of droughts. *Journal of Hydrology*, 539, 417–428.
- Mallya, G., Mishra, V., Niyogi, D., Tripathi, S. & Govindaraju, R.S. (2016) Trends and variability of droughts over the Indian monsoon region. *Weather and Climate Extremes*, 12, 43–68.
- Mao, G., Vogl, S., Laux, P., Wagner, S. & Kunstmann, H. (2015) Stochastic bias correction of dynamically downscaled precipitation fields for Germany through Copula-based integration of gridded observation data. *Hydrology and Earth System Sciences*, 19(4), 1787–1806. <https://doi.org/10.5194/hess-19-1787-2015>
- Maraun, D., Shepherd, T.G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J.M. et al. (2017) Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, 7(11), 764–773.
- Mohan, T.S. & Rajeevan, M. (2017) Past and future trends of hydroclimatic intensity over the Indian monsoon region. *Journal of Geophysical Research: Atmospheres*, 122(2), 896–909.
- NDMA (2008) *National Disaster Management Guidelines: Management of Floods*. Available at: <https://ndma.gov.in/images/guidelines/flood.pdf>
- Pai, D.S., Sridhar, L., Rajeevan, M., Sreejith, O.P., Satbhai, N.S. & 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.

- Pierce, D.W., Cayan, D.R., Maurer, E.P., Abatzoglou, J.T. & Hegewisch, K.C. (2015) Improved bias correction techniques for hydrological simulations of climate change. *Journal of Hydrometeorology*, 16(6), 2421–2442.
- Prasad, A.K. & Singh, R.P. (2005) Extreme rainfall event of July 25–27, 2005 over Mumbai, West Coast, India. *Journal of the Indian Society of Remote Sensing*, 33(3), 365–370.
- Qiu, J. (2008) China: the third pole. *Nature*, 454(7203), 393–396.
- Schmidli, J., Frei, C. & Vidale, P.L. (2006) Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 26(5), 679–689.
- Schwarz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464.
- Seenirajan, M., Natarajan, M., Thangaraj, R. & Bagyaraj, M. (2017) Study and analysis of Chennai flood 2015 using GIS and multicriteria technique. *Journal of Geographic Information System*, 9(02), 126.
- Sharma, S. & Mujumdar, P. (2017) Increasing frequency and spatial extent of concurrent meteorological droughts and heatwaves in India. *Scientific Reports*, 7(1), 1–9.
- Singh, S., Ghosh, S., Sahana, A.S., Vittal, H. & Karmakar, S. (2017) Do dynamic regional models add value to the global model projections of Indian monsoon? *Climate Dynamics*, 48(3), 1375–1397.
- Subramanya, K. (2013) *Engineering Hydrology*, 4th edition. New Delhi, India: Tata McGraw-Hill Education.
- Suman, M., Maity, R. & Kunstmann, H. (2019) Bias Corrected daily most expected and extreme CORDEX precipitation data set for India. <https://doi.org/10.6084/m9.figshare.9686648.v2>
- Taylor, K.E. (2001) Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7), 7183–7192.
- Teutschbein, C. & Seibert, J. (2010) Regional climate models for hydrological impact studies at the catchment scale: a review of recent modeling strategies. *Geography Compass*, 4(7), 834–860.
- Thayyen, R.J., Dimri, A.P., Kumar, P. & Agnihotri, G. (2013) Study of cloudburst and flash floods around Leh, India, during august 4–6, 2010. *Natural Hazards*, 65(3), 2175–2204.
- Vishnu, C.L., Sajinkumar, K.S., Thomas Oommen, R.A., Coffman, K.P.T., Rani, V.R. & Keerthy, S. (2019) Satellite-based assessment of the august 2018 flood in parts of Kerala, India. *Geomatics, Natural Hazards and Risk*, 10(1), 758–767.
- Wit, E., Heuvel, E.V.D. & Romeijn, J.W. (2012) ‘All models are wrong..’: an introduction to model uncertainty. *Statistica Neerlandica*, 66(3), 217–236.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Suman M, Maity R, Kunstmann H. Precipitation of Mainland India: Copula-based bias-corrected daily CORDEX climate data for both mean and extreme values. *Geosci Data J.* 2021;00:1–17. <https://doi.org/10.1002/gdj3.118>