



Non-stationary modelling framework for regionalization of extreme precipitation using non-uniform lagged teleconnections over monsoon Asia

Meghana Nagaraj¹ · Roshan Srivastav²

Accepted: 10 March 2022 / Published online: 31 March 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Global warming has increased the spatio-temporal variations of Extreme Precipitation (EP), causing floods, in turn leading to losses of life and economic damage across the globe. It is found that EP variability strongly correlates with large-scale climate teleconnection resulting from ocean–atmosphere oscillations. In this study, the Non-Stationary Generalized Extreme Value (NSGEV) framework is used to model EP for high resolution daily gridded (0.5° latitude $\times 0.5^\circ$ longitude) APHRODITE dataset over Monsoon Asian Region (MAR) using climate indices as covariates. The proposed framework has three major components (i) Selection of non-uniform time-lag climate indices as covariates, (ii) Regionalization of NSGEV model parameters, and (iii) Estimation of zone-wise EP changes. According to Akaike Information Criterion (AICc), results reveal that the NSGEV model is prevalent in 92% of the grid locations across MAR compared to Stationary(S) GEV models. The Gaussian Mixture Model (GMM) clustering algorithm has identified six zones for MAR. It is observed that the derived zonal parameters of NSGEV model is able to mimic the EP characteristics. Further, zone-wise estimation of EP changes for selected return periods shows that the relative percentage change in intensity ranges between 4 and 11% across the six zones. The change in EP is significantly higher in the monsoonal windward and coastal regions when compared to the other parts of MAR. Overall, the intensities of the EP across MAR are increasing, and return periods are decreasing, which can majorly impact on planning, design and operations of the water infrastructure in the region.

Keywords Extreme precipitation · Non-stationary · Teleconnections · Monsoon Asia regions

1 Introduction

Recent studies indicate that global warming and its consequences substantially impact the world's socio-economic development. It is reported that the available moisture in the atmosphere will increase at a rate of $\sim 7\%$ per kelvin of temperature rise, potentially increasing the magnitude and frequencies of extreme precipitation (EP) events (Trenberth et al. 2003; Donat et al. 2013; O’Gorman 2015; Trenberth 2012). Most of the water infrastructure designs

are based on the assumption that EP over time remains stationary. However, due to the effects of the warming climate, the stationarity assumption is not valid. Therefore, non-stationary modelling approaches are developed to address the changes in EP over time. Coles (2001) proposed the non-stationary generalized extreme value (NSGEV) modelling framework based on the Extreme Value Theory (EVT). It provides a robust mathematical framework and has been widely used for modelling of EP (Cheng et al. 2014; Zhang et al. 2010).

The NSGEV models are fitted by assuming the distribution parameters as a function of dependent variables (also known as covariates) (Coles 2001). Most of the studies considered the location or scale or both the parameters of the NSGEV model as a function of time (Cheng and Aghakouchak 2014; El Adlouni et al. 2007; Sarhadi and Soulis 2017; Yilmaz and Perera 2014). However, large-scale climate teleconnections resulting from the variations in the atmosphere–ocean oscillations show

✉ Roshan Srivastav
roshan@iittp.ac.in

¹ Department of Civil Engineering, Indian Institute of Technology Madras, Chennai 600036, India

² Department of Civil and Environmental Engineering, Indian Institute of Technology Tirupati, Yerpedu, Tirupati 517619, India

strong correlations with EP and have improved the performance of the NSGEV model. Zhang et al. (2010) used covariates such as Pacific Decadal Oscillation (PDO), Southern Oscillation Index (SOI), and North Atlantic Oscillation (NAO) to model the EP over North America. The results showed that the SOI and PDO strongly influence EP. Agilan and Umamahesh (2016) developed the NSGEV model by selecting five covariates, i.e., land-use changes, local and global temperature (annual mean), El Niño Southern Oscillation (ENSO) (averaged Nov-Mar), Indian Ocean Dipole (IOD) (averaged Jun-Nov), and time. They found that the teleconnections as a covariate are appropriate for the study area. Thiombiano et al. (2018) found that Arctic Oscillation (AO) and Pacific North American (PNA) patterns significantly influence the extreme daily precipitation over South-Eastern Canada. Ouarda et al. (2019) modeled NSGEV for three stations in Canada and California, USA with AMO, Western Hemisphere Warm Pool, SOI, and PDO covariates and showed that the NSGEV is a better fit than the stationary. Jha et al. (2021) assessed the influence of ENSO, IOD, and Atlantic Multidecadal Oscillation (AMO) indices (averaged over Jun-Sep) on EP using the NSGEV model for 24 river basins of India, assuming a simultaneous relationship between changes in SST and precipitation and found that NSGEV models perform better than the SGEV models.

The climate index series considered are usually averaged over a few months/seasons or annually, depending on the spatio-temporal effects of teleconnections on the EP. These uniform time-lag climate index series were used as covariates in the non-stationary models (Gao et al. 2016; Jha et al. 2021; Zhang et al. 2010). However, the variations in teleconnection patterns and their effect on extreme precipitation occur non-uniformly in space and time. He and Guan (2013) observed that the time-lags between precipitation and teleconnections better interpret precipitation's temporal variability. Kim et al. (2018) showed that the non-uniform climate indices could account for the delayed effect of changes in atmospheric circulations on precipitation. It was observed that the ENSO indices with four- and ten-month lags strongly represented periodic oscillations in monthly precipitation. Armal et al. (2018) investigated the frequency of precipitation across the United States, and they suggested examining the time-lagged dependence of the climate variables and precipitation spatially. Most of these works focused on using uniform time-lag covariates with hydroclimatic series but not for non-stationary modelling of EP. Therefore, it is envisaged that non-uniform time lag climate indices as covariates may improve the performance of the NSGEV models.

Many studies were carried out to understand the effect of teleconnections on EP over a large-scale region. Mondal and Mujumdar (2015) modelled non-stationary extreme

rainfall over India using covariates such as Niño 3.4 SST anomalies, global temperature, and local temperature. Significant spatial variability was observed in EP characteristics with the teleconnections. Tan and Gan (2017) considered teleconnections ENSO, NAO, PDO, North Pacific Oscillations (NPO) for analyzing the frequency and intensity of EP under non-stationary conditions over Canada. They found that the ENSO's effect on EP was consistent at all stations across the region. Also, the PDO, NAO, and NPO significantly influence EP variability at a few stations across Canada. Vu and Mishra (2019) studied non-stationary EP events over the United States using time, an annual average maximum, and mean temperature and ENSO as covariates. It was observed that the different combinations of covariates influence EP variability over the US under non-stationary conditions. One of the possible ways to overcome this variability in large-scale study areas is to cluster the regions of EP. The identification of zones helps in reducing the prediction uncertainties. The regionalization of precipitation in a stationary framework is well studied (Irwin et al. 2017; Roushangar et al. 2018; Roushangar and Alizadeh 2018). However, the regionalization in a non-stationary framework accounting for the influence of teleconnection indices on EP has not been well explored, especially for Monsoon Asia Region (MAR).

The proposed study addresses three key issues: i) Selection of non-uniform time-lag climate indices: most of the studies in the NSGEV framework have used uniform time-lagged covariates, i.e., irrespective of their EP month of occurrence, the covariate is averaged over particular months (uniform). However, the maximum correlation between EP and climate indices will vary in space and time and hence the covariates series should be non-uniformly lagged. It is envisaged in this study that identifying the non-uniform lag structure between precipitation and the multiple climatic indices would improve the prediction of EP and contribute to a better understanding/representation of the mechanics underlying these large-scale EP responses to teleconnection patterns; ii) Regionalization of NSGEV model parameters: identification of coherent regional spatial patterns and the uncertainty associated with non-stationary extreme precipitation across large study domains need to be addressed. This is achieved by regionalizing NSGEV model parameters by grouping similar EP sites into coherent zones. The zones will reduce the uncertainties in the prediction of EP. The performance of the regional NSGEV model needs to be evaluated to the individual sites within the zone; and (iii) Estimation of Zonal EPs': the regional NSGEV model is used to estimate the intensity of extreme precipitation for the selected return periods in each zone, and compared with the estimates from the SGEV model. The EP intensity for various return periods is investigated and presented for each cluster. In

addition, the efficacy of the proposed modelling framework is also compared with traditional NSGEV models using corrected Akaike Information Criteria (AICc) for MAR.

The paper comprises the following sections: Sect. 2 describes the MAR study area, precipitation data, and teleconnections. Following this, Sect. 3 presents a detailed methodology explaining (i) NSGEV model framework, (ii) Selection of non-uniform climate indices, (iii) Clustering of EP zones. In Sect. 4, the results and discussions showing the efficacy of the non-uniform climate index time series and regional parameters are presented. Further, the section presents the variations and changes in EP characteristics for stationary and non-stationary. Finally, the conclusion and future scope of the proposed study are presented in Sect. 5.

2 Case study

In the present study, the Asian Precipitation-Highly Resolved Observational Data Integration Towards Evaluation (APHRODITE) gridded precipitation dataset for 1951–2007 with a 0.5×0.5 -degree spatial resolution for the Monsoon Asia Region (MAR) is used as illustrated in Fig. 1 (Yatagai et al. 2009, 2012) (<http://www.chikyu.ac.jp/precip/index.html>). It is developed by the Research Institute for Humanity and Nature (RIHN) and the JMA (Japan Meteorological Agency). The region includes mainly South Asia (India, Nepal, Pakistan, Bangladesh),

South-East Asia (SEA), and Northern Australia. The 180×140 grid locations cover $60^\circ - 150^\circ\text{E}$, $15^\circ - 55^\circ\text{N}$ region. There are 25,200 grid locations, of which 13,672 and 11,528 are ocean and land grids. The Annual Maximum Precipitation (AMP) series is extracted for each grid location and used in the study, referred to as EP. Aphrodite data are extensively used for various applications such as (i) Study of seasonal variations of the summer monsoon events (Befort et al. 2016); (ii) Climate change's effect on non-monsoonal precipitation in the Western Himalayas (Krishnan et al. 2019); (iii) Simulating the response of East Asia's rainfall to ENSO (Ng et al. 2019); (iv) Understanding dry spell characteristics over India (Sushama et al. 2014); (v) Understand the El Nino and Non-El Nino effects on droughts (Varikoden et al. 2015).

2.1 Climate indices

Climate patterns and circulation processes are influenced by global atmospheric circulation at multiple scales. The low-frequency oscillations that reflect large-scale atmospheric circulation patterns are called “teleconnections patterns” and arise from internal atmospheric dynamics. Many planetary-scale teleconnection patterns have wide-scale effects, cover large areas, and contribute to changes in the earth's temperature, precipitation, and storm tracks. These are responsible for severe weather conditions, including extreme precipitation and flooding (Barnston and Livezey 1987; Mo and Livezey 1986). The climate centers

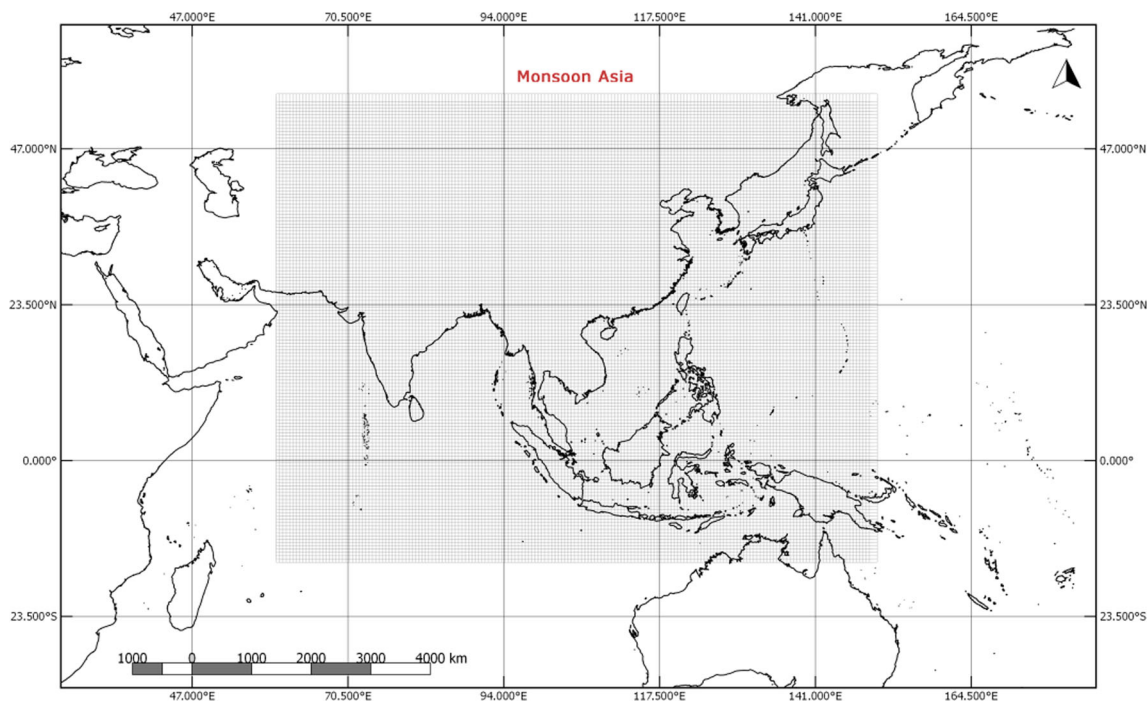


Fig. 1 Map with APHRODITE grids representing Monsoon Asia Region (MAR)

worldwide monitor and tabulate the teleconnection patterns, also known as “climate indices,” based on geopotential height, sea level pressure, and sea surface temperature. Several observations indicate that extreme precipitation variability has a strong correlation with teleconnections. For example, ENSO is reported as the most significant climate index affecting global extreme precipitation variability (Dai and Wigley 2000; Krishnaswamy et al. 2015; Sun et al. 2015; Wang 2002). Krishnaswamy et al. (2015) showed ENSO and IOD as significant drivers for Indian Monsoon variability and increased extreme precipitation events. Wei et al. (2017) found that extreme precipitation changes are mainly induced by teleconnections like ENSO, PDO, and, East Asian summer monsoon over China. Few studies over Australia reported that the ENSO phases (Verdon et al. 2004), SAM (Meneghini et al. 2007), and IOD (Ashok et al. 2003) have a high impact on precipitation changes. Ng et al. (2019) observed that the extreme precipitation varies during ENSO phases (El Nino and La Nino) over East Asia. Duan et al. (2015) found that climate indices like SOI have significant associations with the increasing extreme precipitation over Japan. Rathinasamy et al. (2019) confirmed a strong correlation between the IOD, PDO, and NAO and precipitation extremes across India. In the present study, 15 climate indices are used for the NSGEV model as covariates to analyze the extreme precipitation over MAR. Table 1 shows the list of climate indices with a brief description.

3 Methodology

The proposed methodology has five main modules: (1) Selection of uniform and non-uniform time-lag climate indices; (2) NSGEV model; (3) Model selection using AICc; (4) Zone formation using clustering algorithm; (5) Performance evaluation of the model. The proposed methodology is schematically represented in Fig. 2. The details for each of the modules are explained in the following subsections.

3.1 Selection of uniform and non-uniform time-lag climate indices

This study identifies the time-lags between precipitation and climate indices based on the cross-correlation technique. The lag corresponding to the maximum correlation coefficient is the offset or strong dependence between precipitation and the corresponding climate index. The advantage of cross-correlation is that it is simple to compute and identify the covarying pattern between precipitation and climate indices in space and time instantaneously. It is to be noted that the cross-correlation

analysis addresses: (i) The identification of the time-lags with each climate indices at each grid location and (ii) Extraction of the non-uniform climate index series corresponding to Annual Maximum Precipitation (AMP)/EP at each grid location. The hypothetical example of extracting the uniform and non-uniform climate index series corresponding to a grid location in MAR is shown in Fig. 3 and are presented below:

- (i) **Extraction of Month Index:** Consider AMP series of ‘n’ years at a grid location. The shaded boxes in Fig. 3a represent the month-index of EP. For example, in year-1, EP is observed in the month-7, i.e., the month index is 7 for year 1; for year-2, AMP’s month-index- 8 and so on.
- (ii) **Uniform time-lag climate index series** is selected based on the average over a season or few months irrespective of EP occurrence in any month/time of a year. Figure 3b shows that the climate index series selection is constant. In this hypothetical example, the index value is obtained by taking an average over the three-month window as shown in Fig. 3b, i.e., months 6–8, and remains uniform throughout the study region.
- (iii) **Non-uniform time-lag climate index series**

- Step 1: The cross-correlation between monthly precipitation and climate index is carried out. A time lag corresponding to the maximum correlation coefficient is selected at each grid location. In this example, a time lag of 5 months is assumed.
- Step 2: The climate index series is selected corresponding to the time-lag for each year’s EP series, as shown in Figure 3c. From Figure 3a, c, the AMP of year-1 occurs in the month-7, the climate index value averaged over three months windows (i.e., month-2 with time-lag of 5 as the center of the time-window) for year-1 is obtained as shown in Figure 3c. Similarly, for the subsequent years, the non-uniform time-lag climate index series will be derived for MAR.

3.2 NSGEV model

The GEV model is extensively used for analyzing EP based on Extreme Value Theory (EVT) (Coles 2001). The location (μ), scale (σ), and shape (ξ) are the three parameters of the GEV distribution. Consider $x = x_1, x_2, \dots, x_n$ to be independent and identically distributed AMP series. The Probability Density Function (PDF), f is given by Eq. 1 and the Cumulative Distribution Function (CDF), F of the GEV is shown in Eq. 2

Table 1 Selected climate indices and their description

Sl. No.	Climate indices	Description	Download link
1	North Atlantic oscillation (NAO)	The sea level pressure difference between the stations at Azores and Iceland in the Atlantic Ocean	https://psl.noaa.gov/gcos_wgsp/Timeseries/NAO/
2	Southern oscillation index (SOI)	The normalized sea level pressure difference between the stations at Tahiti and Darwin in the Pacific Ocean	https://psl.noaa.gov/gcos_wgsp/Timeseries/SOI/
3	Arctic oscillation (AO)	Average of 10–1000 Mb geopotential height anomalies (30 N poleward)	https://psl.noaa.gov/gcos_wgsp/Timeseries/AO/
4	Pacific Decadal oscillation (PDO)	sea surface temperature anomalies calculated from empirical orthogonal function analyses poleward of 20 N in the Pacific Ocean	https://psl.noaa.gov/gcos_wgsp/Timeseries/PDO/
5	Southern annular mode (SAM)	The difference of mean sea level pressure measured between 40 and 65°S. it is also known as Antarctic Oscillation	https://psl.noaa.gov/data/20thC_Rean/timeseries/monthly/SAM/sam.20crv2c.long.data
6	Trans polar index (TPI)	The normalized sea level pressure difference between Hobart and Stanley	https://psl.noaa.gov/gcos_wgsp/Timeseries/TPI/
7	Dipole mode index (DMI)	sea surface temperature anomalies between the western equatorial Indian Ocean and the southeastern Indian Ocean. It represents the Indian Ocean Dipole (IOD)	https://psl.noaa.gov/gcos_wgsp/Timeseries/DMI/
8	Pacific North American pattern (PNA)	The low-frequency variability modes in the Northern Hemisphere extratropic reflect a quadrupole pattern of 500 millibar height anomalies	https://www.ncdc.noaa.gov/teleconnections/pna/
9	North Pacific (NP)	The sea level pressure from 30 N-65 N to 160E-140 W (area-weighted)	https://psl.noaa.gov/gcos_wgsp/Timeseries/NP/
10	Nino 4	The area-averaged sea surface temperature over the area 5S-5 N and 160E-150 W in the Pacific Ocean	https://psl.noaa.gov/gcos_wgsp/Timeseries/Nino4/
11	Nino 1 + 2	The area-averaged sea surface temperature over the area 0–10S and 90–80 W in the Pacific Ocean	https://psl.noaa.gov/gcos_wgsp/Timeseries/Nino12/
12	Atlantic multidecadal oscillations (AMO)	The sea surface temperature measured in the North Atlantic Ocean (0–70 N)	https://psl.noaa.gov/gcos_wgsp/Timeseries/AMO/
13	Global Average Temperature & SST anomalies (GLBT)	Average surface and sea surface temperature measured over the global scale	https://psl.noaa.gov/gcos_wgsp/Timeseries/GLBTSST/
14	Multivariate ENSO indices (MEI)	The first principal component was obtained from five variables: pressure, the surface wind (zonal and meridional), sea surface temperature, and outgoing longwave radiation	https://psl.noaa.gov/enso/mei/data/meiv2.data
15	East Pacific-North Pacific (EPNP)	The pattern is measured over three anomaly centers located over Alaska/ Western Canada, central North Pacific, and eastern North America	https://www.cpc.ncep.noaa.gov/data/teledoc/ep.shtml

The PDF, $f : f(x)$

$$= \frac{1}{\sigma} \left\{ 1 + \frac{\xi(x - \mu)}{\sigma} \right\}^{(-\frac{1}{\xi})-1} \exp\left[-\left\{ 1 + \frac{\xi(x - \mu)}{\sigma} \right\}^{-\frac{1}{\xi}}\right] \tag{1}$$

The CDF, $F : F(x)$

$$= \exp\left\{-\left[1 + \frac{\xi(x - \mu)}{\sigma} \right]^{-\frac{1}{\xi}}\right\}, 1 + \frac{\xi(x - \mu)}{\sigma} > 0 \tag{2}$$

covariates (Coles 2001; Katz 2002). In the present study, covariates are introduced as a function of only location parameters (Eq. 3) and substituted in CDF of GEV (Eq. 4). The scale parameter is not considered for computational simplicity (Ouarda and Charron 2018), and the shape parameter is kept constant due to its uncertainty and difficulty to estimate with precision (Coles 2001).

$$\mu(C) = \mu_0 + \mu_1 C \tag{3}$$

$$\sigma(C) = \sigma \tag{4}$$

$$\xi(C) = \xi \tag{5}$$

Non-stationarity is included by considering one or more GEV distribution parameters as a function of time-varying

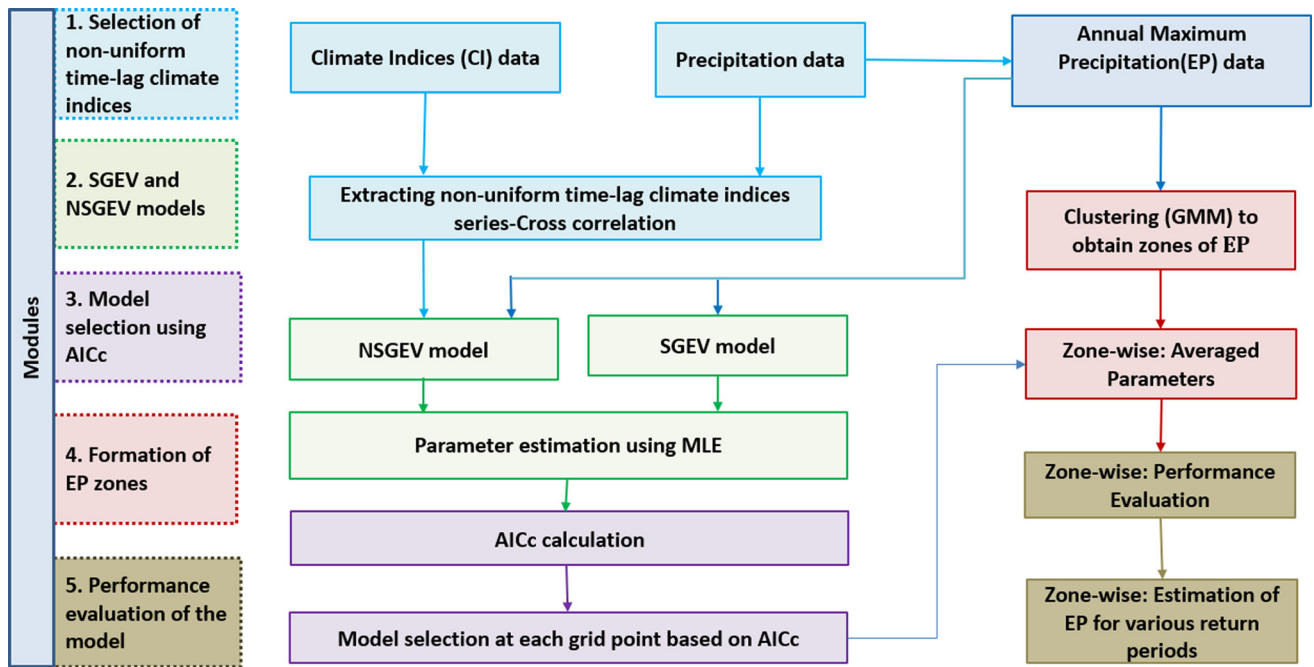


Fig. 2 Schematic representation of the proposed methodology for the NSGEV framework

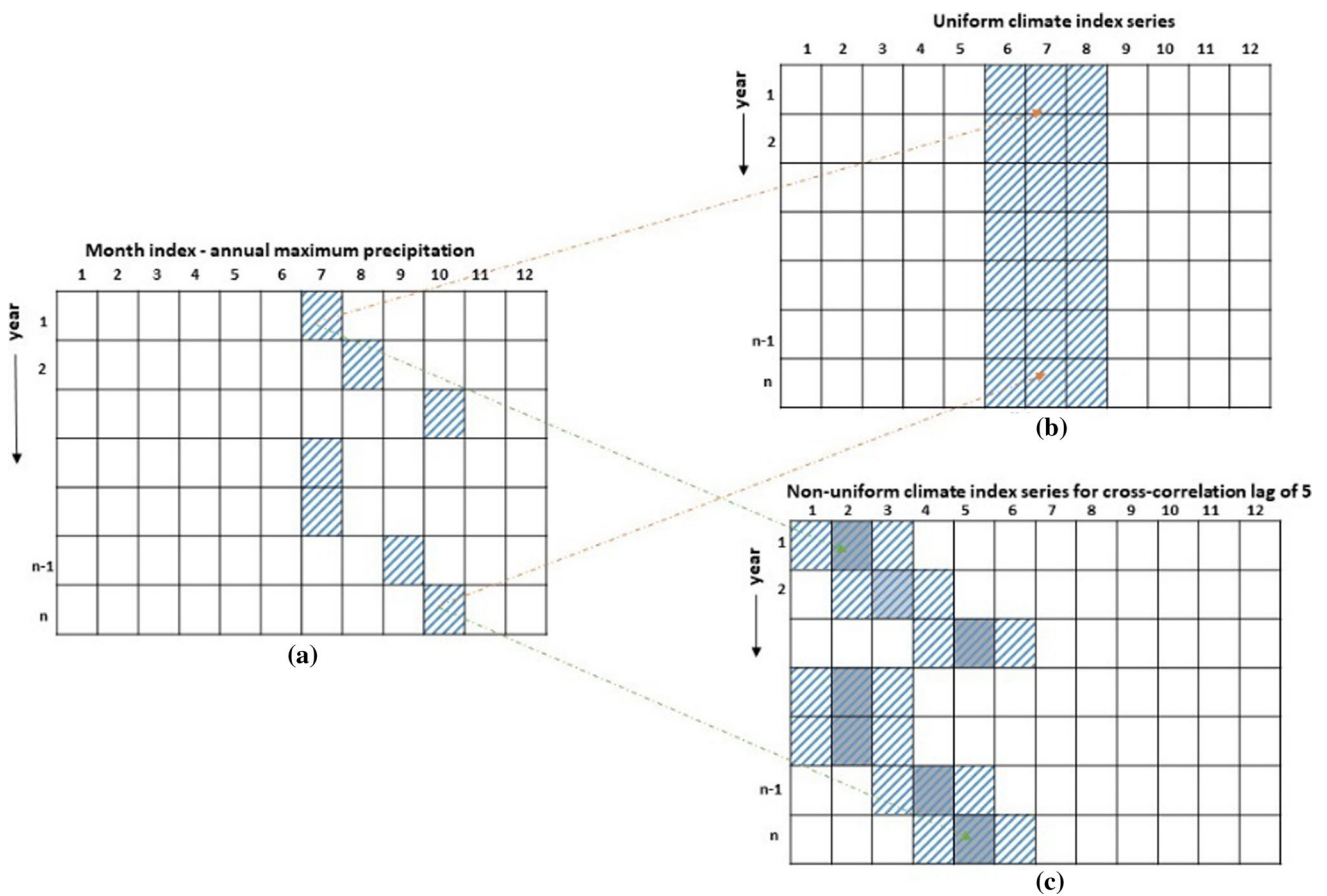


Fig. 3 Hypothetical example of extracting the climate index series at a grid location **a** annual maximum precipitation series **b** uniform climate index series **c** non-uniform climate index series assuming cross-correlation lag of 5

The climate Indices described in Table 1 are considered covariates for the NSGEV model and are represented as “C”. There will be 15 NSGEV models corresponding to 15 climate indices and an SGEV model at each grid location over MAR. The maximum likelihood estimation is used to estimate the model parameters. The log-likelihood for the cumulative distribution function for both stationary (Eq. 6) and non-stationary (Eq. 7) GEV model is given below:

$$\log L(\mu, \sigma, \xi|X) = -n \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum \log \left[1 + \frac{\xi(x - \mu)}{\sigma}\right]^{-\frac{1}{\xi}} - \sum \left[1 + \frac{\xi(x - \mu)}{\sigma}\right]^{-\frac{1}{\xi}} \quad (6)$$

$$\log L(\mu(C), \sigma, \xi|X) = -n \log \sigma \left(1 + \frac{1}{\xi}\right) \sum \log \left[1 + \frac{\xi(x - \mu(C))}{\sigma}\right]^{-\frac{1}{\xi}} - \sum \left[1 + \frac{\xi(x - \mu(C))}{\sigma}\right]^{-\frac{1}{\xi}} \quad (7)$$

The parameters for the SGEV and NSGEV model are obtained by maximizing the negative log-likelihood (Eqs. 6 and 7). The ‘ismev’ R package estimates the models’ parameters (Coles 2001; Heffernan 2018).

The extreme precipitation is estimated for various return periods under stationary and non-stationary conditions using the low-risk approach (Coles 2001). In this study, the ninety-fifth percentile of the location parameter values (Eq. 8) is estimated and used to calculate the precipitation intensity for selected return periods for the stationary (Z_{rs}) and non-stationary (Z_{ms}) conditions are given below (Eqs. 9 and 10):

$$\mu_{95} = Q_{95}(\mu_{t1}, \mu_{t2}, \dots, \dots, \dots, \mu_{tm}) \quad (8)$$

$$Z_{rs} = \mu + \frac{\sigma}{\xi} \left[(-\log(1 - p))^{\xi} - 1\right] \quad (9)$$

$$Z_{ms} = (\mu_{95}) + \frac{\sigma}{\xi} \left[(-\log(1 - p))^{\xi} - 1\right] \quad (10)$$

where $p = 1/T$, the probability of occurrence and T is the return period. μ_{95} is the 95th percentile of the location parameter. The Relative Percentage Change (RPC) in extreme precipitation intensity between non-stationary and stationary are computed using Eq. 11

$$RPC = \frac{Z_{ms} - Z_{rs}}{Z_{rs}} \times 100(\%) \quad (11)$$

where Z_{ms} and Z_{rs} are precipitation intensities for non-stationary and stationary models.

3.3 Selection of the model

In this study, the corrected Akaike Information Criterion (AICc) is adopted to choose a model from a set of alternative models arising from the 15 covariates. The negative log-likelihood function ($-\log L$) is penalized according to the number of estimated parameters (p) for each model. The AICc helps prevent the bias of over-fitting the data and is recommended when $n/p > 40$, where n is the sample size and p is the number of parameters (Burnham and Anderson 2004; Hurvich and Tsai 1989). The AICc for the fitted model is expressed as (Eq. 12)

$$AICc = -2\log L + 2p + \frac{2p(p + 1)}{n - p - 1} \quad (12)$$

where $-\log L$ is the negative log-likelihood value, the first two terms of Eq. 12 represent the classic AIC. Second-order bias correction is added in AICc by the last term. It is worth noting that as n increases, AICc converges to AIC. Each model is evaluated for its AICc, and the one with the lowest AICc is chosen as the best model.

3.4 Clustering using gaussian mixture model (GMM)

The GMM based clustering algorithm is applied to identify an adequate number of EP zones over MAR. This method is very flexible and has advantages over traditional heuristic methods such as k-means and hierarchical clustering, in which the grouping of coherent regions is based on measured distances (Euclidean distance). The number of clusters and initial cluster center must specify a priori. These limitations can be overcome using model-based methods such as GMM in a probability framework involving multiple mixtures (clusters), each defined by a Gaussian distribution. The Bayesian Information Criterion (BIC) is used as an objective metric to determine the best model and an optimal number of clusters. It uses soft clustering, where probability is assigned to each data location in a cluster. The probability value indicates the degree of association between the cluster and the data locations. The mathematical expression of the Gaussian model representing each cluster is given in Eq. 13

$$\begin{aligned} \emptyset_k(z|\mu_k, \Sigma_k) &= (2\pi)^{-\frac{1}{2}} \times |\Sigma_k|^{-\frac{1}{2}} \\ &\times \exp\left\{-\frac{1}{2}(z_i - \mu_k)^T \times \Sigma_k^{-1}(z_i - \mu_k)\right\} \end{aligned} \quad (13)$$

where z_i is the attribute data for i th location, μ_k is the centre of the distribution for k th cluster, and Σ_k is the variance matrix. The Expectation–Maximization (EM) algorithm is used to estimate the maximum likelihood for the

parameters (mean and variance) of the Gaussian clusters initialized by hierarchical model-based agglomerative clustering. EM proceeds with an 'E'-step in which a matrix z is computed such that z_{ik} is a conditional probability estimate that i th location is in k th cluster for the parameters μ_k and Σ_k . Following 'M-step' estimates the parameters, μ and Σ converges to the maximum log-likelihood values (Eq. 14) of the model (Dempster et al. 1977).

$$\log L(z|\mu_k, \Sigma_k) = \sum_{i=1}^n \log \left[\sum_{k=1}^K \pi_k N(x_i; \mu_k, \Sigma_k) \right] \quad (14)$$

The BIC determines the adequate number of clusters. It penalizes log-likelihood with increment in sample size by introducing a second component for the parameter size. The BIC is typically used in association with GMMs and takes the following form (Eq. 15)

$$\text{BIC}_{M,K} = 2 \times \log L_{M,G} - p \times \log(n) \quad (15)$$

where $\log L_{M,G}$ is the log-likelihood for model M , K is the number of clusters, n is the sample size, and p is the number of parameters. The paired $\{M, K\}$ which optimizes $\text{BIC}_{M,K}$ is found by varying K between a range of clusters. In this study, K is varied between 1 and 20 to find the optimal clusters. The GMM clustering analysis is carried out using the 'mclust' R package (Scrucca et al. 2016).

3.5 Performance evaluation of models

The quality of the fitted NSGEV models must be evaluated to ensure the selected model fits the data well. The Quantile–Quantile (QQ) plots are used to check the quality of the selected model based on model and empirical quantiles. The AMP data of n - years (Z_i) used to fit the NSGEV model is not identically distributed and has to be transformed into residuals (ε) (Eq. 16) (Coles 2001; Katz 2002). Model quantiles (M_i) are the ordered values of residuals and Empirical quantile (E_i) is given by Eq. 17. In this study, the Root Mean Square Error (RMSE) based on quantile plots are calculated using the model and empirical quantiles given in Eq. 18. It measures the model's efficacy while quantifying its quality.

$$\varepsilon_i = \frac{1}{\xi_t} \log \left\{ 1 + \xi_t \left(\frac{Z_i - \mu_t}{\sigma_t} \right) \right\} \quad (16)$$

$$E_i = -\log \left(-\log \left(\frac{i}{n+1} \right) \right), i = 1, \dots, n \quad (17)$$

$$\text{RMSE} = \left(\frac{1}{n} \sum_{i=1}^n (E_i - M_i)^2 \right)^{0.5} \quad (18)$$

4 Results and discussions

4.1 Extreme precipitation

The EP (annual daily maximum precipitation) is extracted from daily precipitation data for each grid location from 1951 to 2007. The statistics of EP such as mean and standard deviation are extracted for each grid location and are presented in Fig. 4a, b. It is observed that the mean and standard deviation vary spatially across MAR. The range of mean EP varies from 3 to 189 mm, and the standard deviation is between 0.92 mm and 86 mm. The variations show high EPs in the coastal and monsoon windward regions. The Western Ghats, Himalayan foothills, a few parts of Southeast Asia, and most eastern coastal areas show the highest EP. One of the primary reasons for high EP in these regions could be the elevated terrain, windward parts of mountains, and ocean proximity (Saini et al. 2020; Shige et al. 2017; Khouakhi and Villarini 2017; Zhang and Zhou 2019). On the other hand, the EP gradually decreases to inland areas where there is not much moisture left for precipitation and are far from large water bodies. In addition, it is worth to be noted that the number of EPs (i.e., frequency) is more prevalent in coastal regions and generally decreasing towards inland areas. The number of grid locations having mean and the standard deviation of EPs is shown in Fig. 4c, d, respectively. It is observed that the maximum number of grids locations are of EP less than 20 mm and 10 mm, respectively.

The percentage of EP's per month occurring over the 57 years for MAR is shown in Fig. 5. It is observed that the maximum number of EP's in the MAR occur during summer monsoon or the Southwest monsoon months (i.e., June–September), i.e., about 68% of grids (June—13.72%, July—23.3%, August—20.64%, and September—10.48). Following this, about 9% of EP occurs during the Northeast monsoon months (October–December), mainly in the Bay of Bengal coastal (Southern India), a few East Asian countries, and Sri Lanka. It is worth to be noted that although a few months have a higher occurrence of EP, there is a considerable variation within the months of the year and spatially across MAR. It is evident from the results that the selection of constant lag time series of covariates for the NSGEV model may not represent the plausible delayed-effect of the teleconnections. Thus the effect of variations in the EP occurrence has to be accounted for extracting the covariates.

To understand the temporal variations in the EP, 50 grid locations across MAR is selected (10 regions containing five grids each, as shown in Figure S1). It is observed from Fig. 6 that sites closer to each other have similar patterns of occurrences. Two significant observations are: (i) the grid

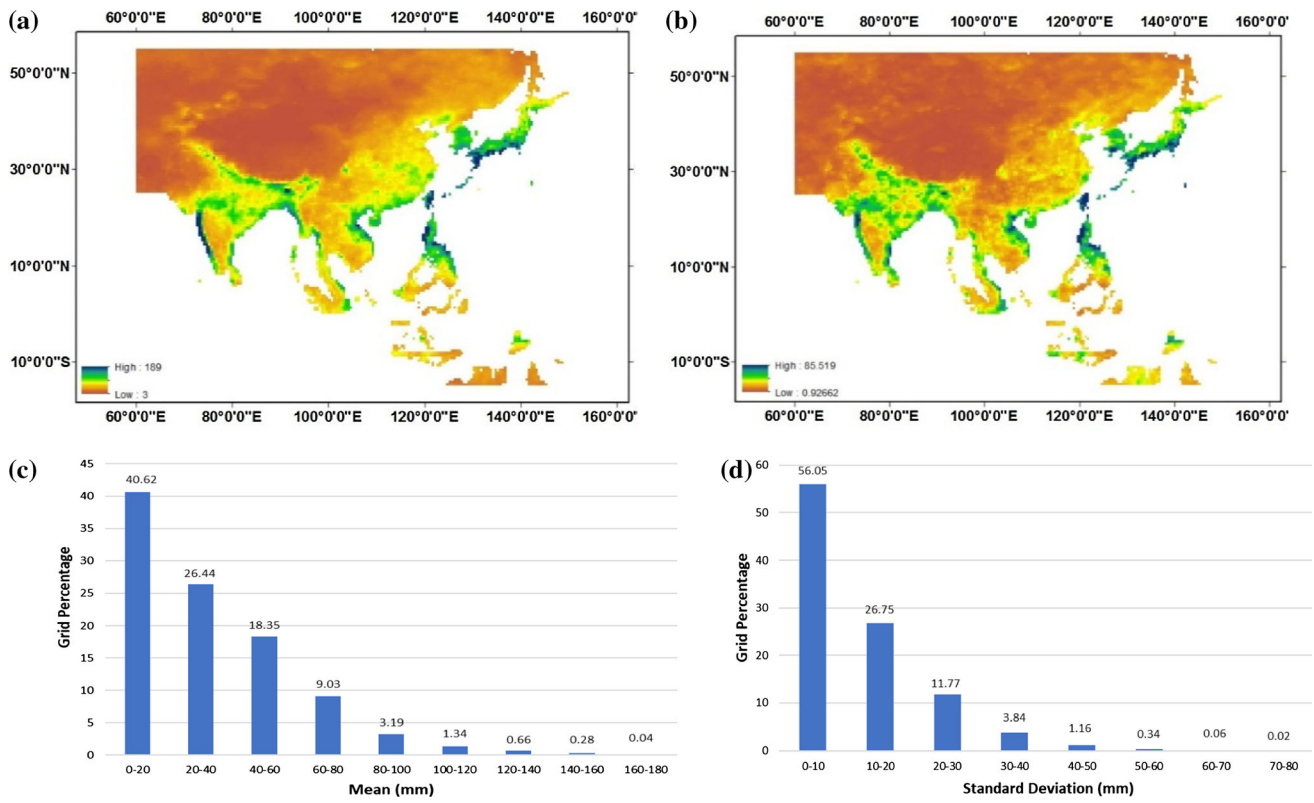


Fig. 4 Map illustrating the spatial distribution **a** mean and **b** standard deviation of EP; Distribution of grid locations over MAR **c** mean and **d** standard deviations, i.e., how many grids are in each bin (precipitation interval)

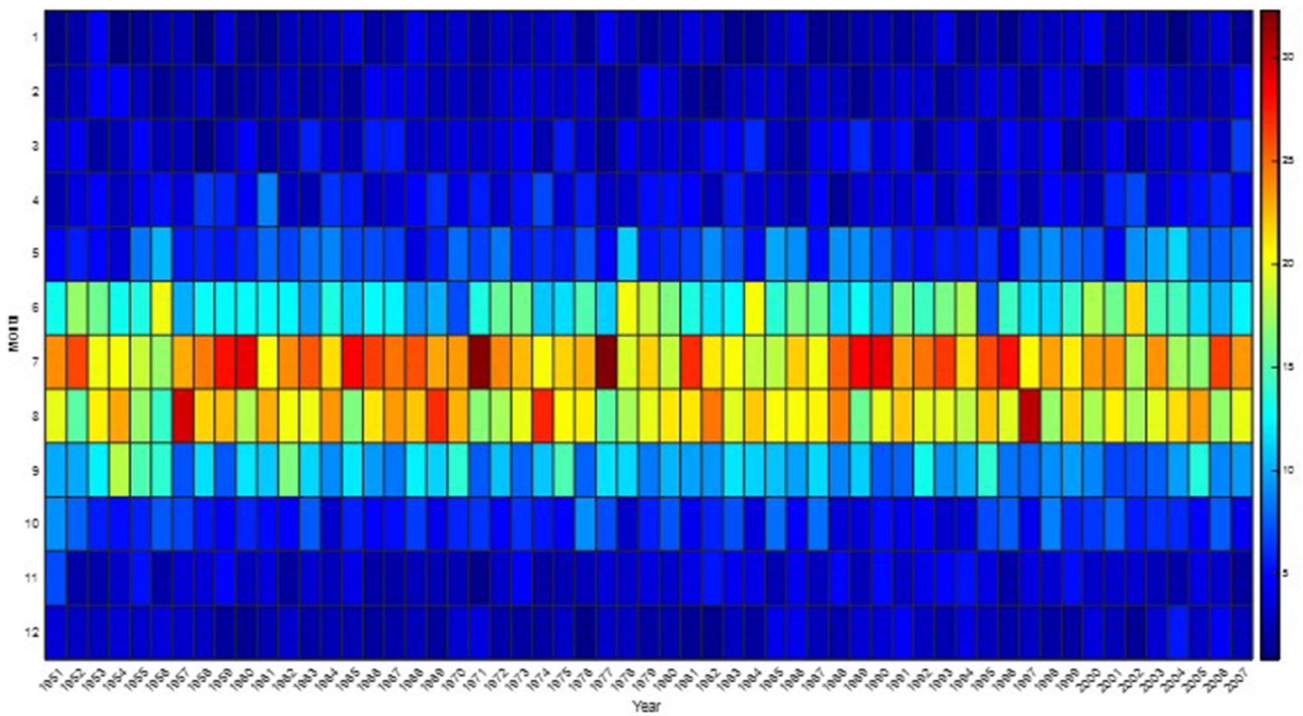


Fig. 5 Month-wise number of EP occurrence (in %) over MAR (color bar represents the number of EP occurrences in percentage)

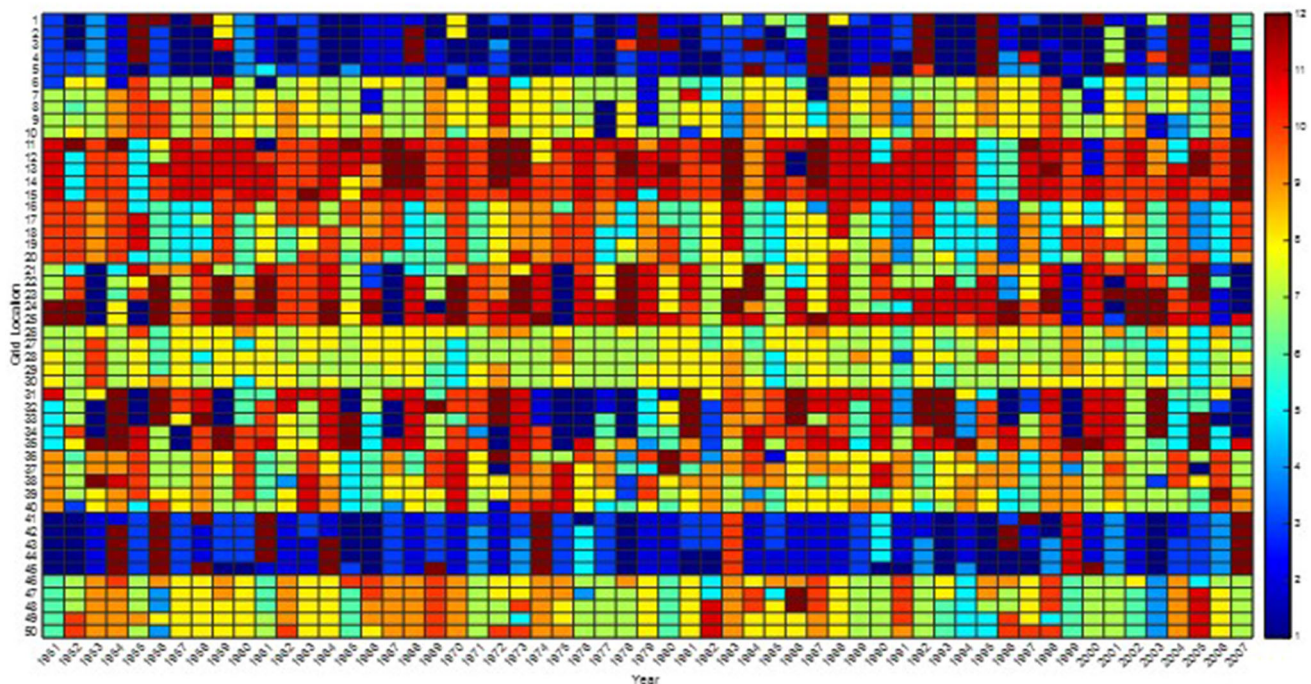


Fig. 6 Spatial variation of month index of EPs at selected 50 grid locations over MAR (color bar represents the month index)

location sampled from a similar EP characteristics region shows relatively similar patterns; and (ii) EP varies in space and time and is not uniform throughout the MAR. Therefore, since the month index of EP varies over the study domain, the corresponding effect of teleconnection time-lag should also vary in space and time. Consequently, it is anticipated that using non-uniform time-lagged climate indices will capture the dynamics of EP variability better than the uniform/constant lagged climate indices.

4.2 Time-lagged climate indices

The uniform time-lagged (UTL) series for a covariate is selected based on the MAR region's averaged months. For example, ENSO is averaged over Nov–Mar (Agilan and Umamahesh 2016; Mondal and Mujumdar 2015), DMI is averaged over Jun–Nov (Agilan and Umamahesh 2016), PNA is averaged over Dec–Mar, and AO is averaged over Jan–Dec (Thiombiano et al. 2018), AMO is averaged over Jun–Aug (Ouarda and Charron 2018), PDO is averaged over Nov–Mar (Gao et al. 2016), global temperature is the annual mean (Mondal and Mujumdar 2015) and NAO is averaged over Dec–Feb (Hao et al. 2019). It is to be noted that the UTL series for a climate index will be identical irrespective of the grid location over MAR. On the other hand, the non-uniform time-lagged (NUTL) climate indices are selected based on the maximum cross-correlation between monthly precipitation and climate index at each grid location. For brevity, NAO results only are presented

in this section (and for other indices, refer to supplementary material—Figure S2, S3 and S4). For the NUTL series, the time-lag corresponding to the maximum correlation for the NAO climate index of all grid locations is shown in Fig. 7. It is observed that the time-lags obtained are spatially non-uniform over the region. These lags represent the possible time-delayed relationship between precipitation and NAO teleconnection. It is also worth to be noted that the NUTL are randomly clustered into smaller regions, which might have similar localized or geographic conditions. NUTL-NAO series are extracted at each grid location based on these identified time-lags.

The performance comparison of UTL- and NUTL based NSGEV models is presented in Fig. 8a. It is observed that in about 94.7% of grid locations, the performance of the NUTL based NSGEV model is better (lower AICc values) when compared to UTL. For brevity, the difference in the AICc values at three grid locations are highlighted in Fig. 8a. In addition, the boxplot (Fig. 8b) of the deviance statistics (difference in AICc) shows that NUTL has the lower AICc for all the grid locations. It is also observed that the deviance statistics for most of the grid locations is greater than two, i.e., the selected NUTL has shown significant improvements in the NSGEV model performance compared to UTL.

Fig. 7 Time lag corresponding to the maximum correlation for NAO at each grid over MAR (color bar represents the time-lag in months)

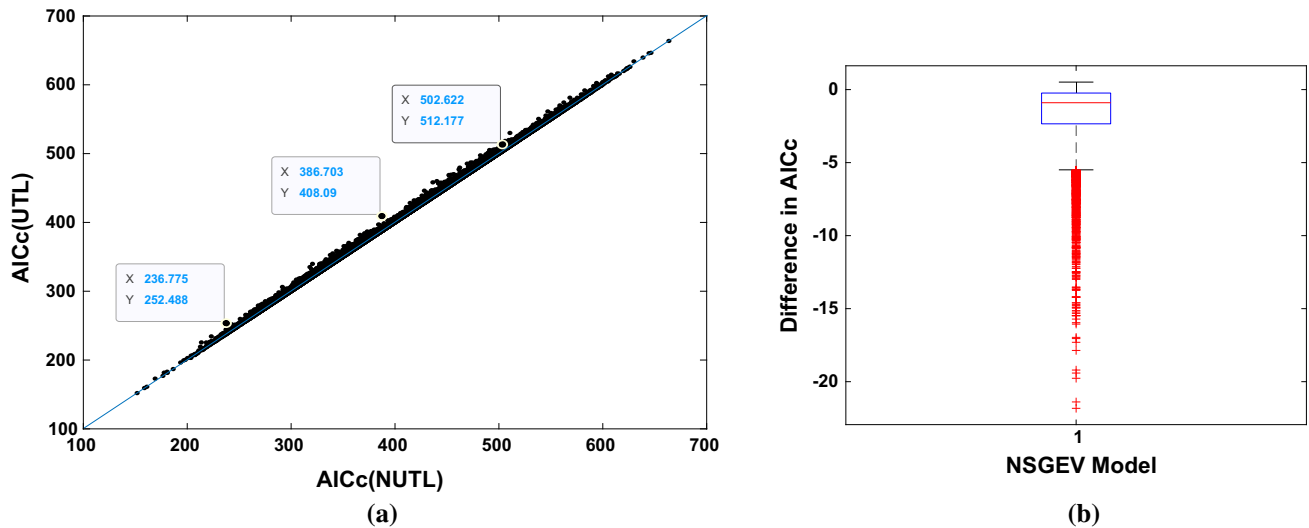
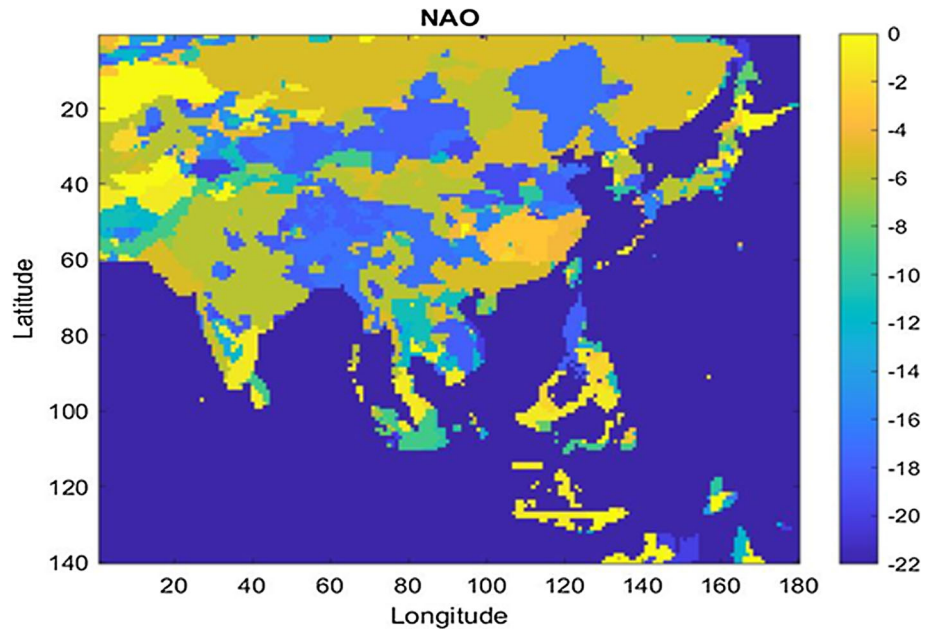


Fig. 8 Comparison of Uniform (UTL) and Non-Uniform Time-Lagged (NUTL) NSGEV models **a** AICc (X and Y represent data label corresponding to the x-axis and y-axis respectively at three

randomly selected points (lower the AICc value better the model)); **b** deviance statistics (difference in AICc between UTL and NUTL NSGEV)

4.3 NSGEV versus SGEV models

The total number of models at each grid location is 16, i.e., 15 NSGEV models and 1 SGEV model. The spatial distribution of selected best models based on AICc over MAR is shown in Fig. 9. It is observed that about 92% of grid locations over MAR, the NSGEV models perform better than SGEV models. However, the spatial pattern of grid-wise selected models (Figs. 10 and 11) shows that no single/group of climate index/covariate-based NSGEV models are chosen over the entire MAR. Similar results were observed by Mondal and Mujumdar (2015) over

India; they found that the significant covariates differed from one grid location to another showing a non-uniform pattern. The results indicate that among all the indices, the GLBT and AMO are majorly influencing MAR. The reason could be that the GLBT represents the global warming phenomenon and directly affects the EP changes through the warming of oceans, transporting moist air to land and into the weather systems (Groisman et al. 2005; IPCC 2007; Res and Trenberth 2011). Similarly, the SSTs in the tropical Atlantic (AMO) strongly influence Monsoon precipitation (Azad and Rajeevan 2016; Kucharski et al. 2007). Overall, the use of NULT based climate indices for

Fig. 9 Performance of SGEV and NSGEV (includes 15 covariates) models at each grid over MAR

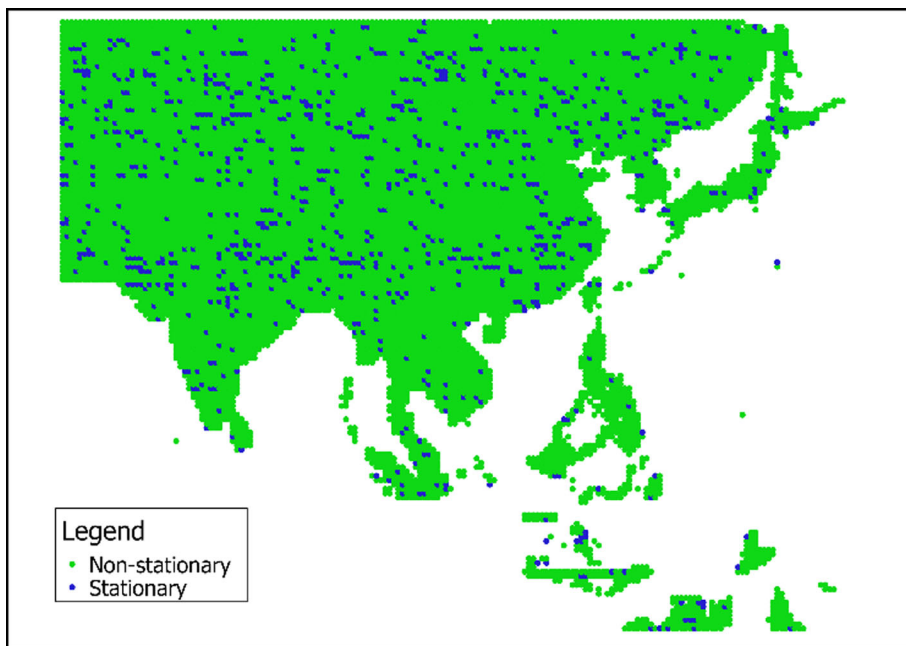
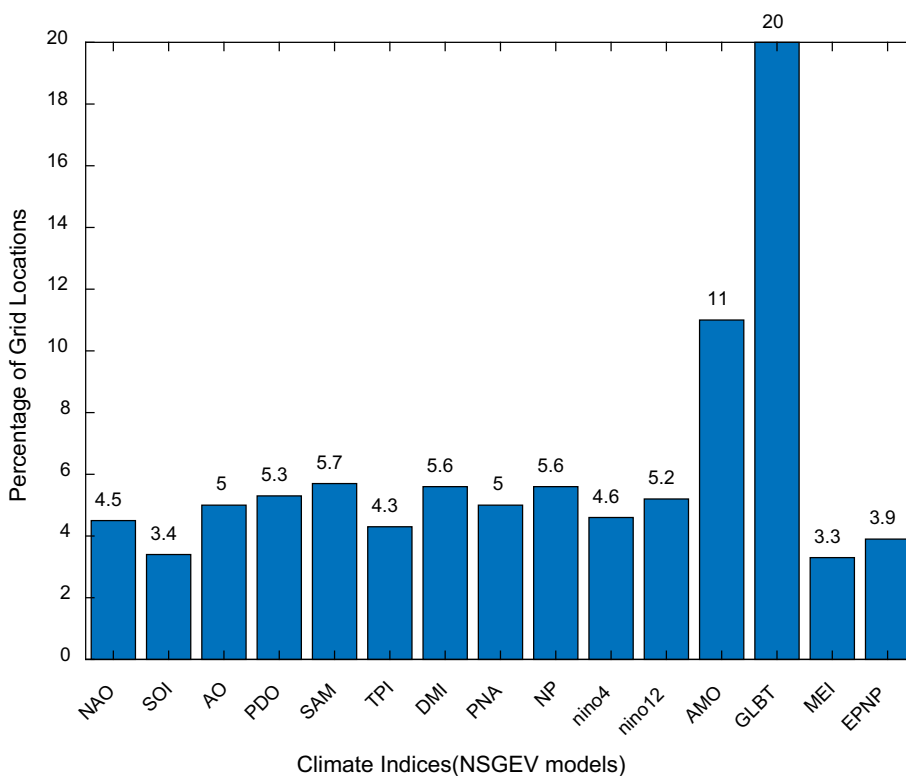


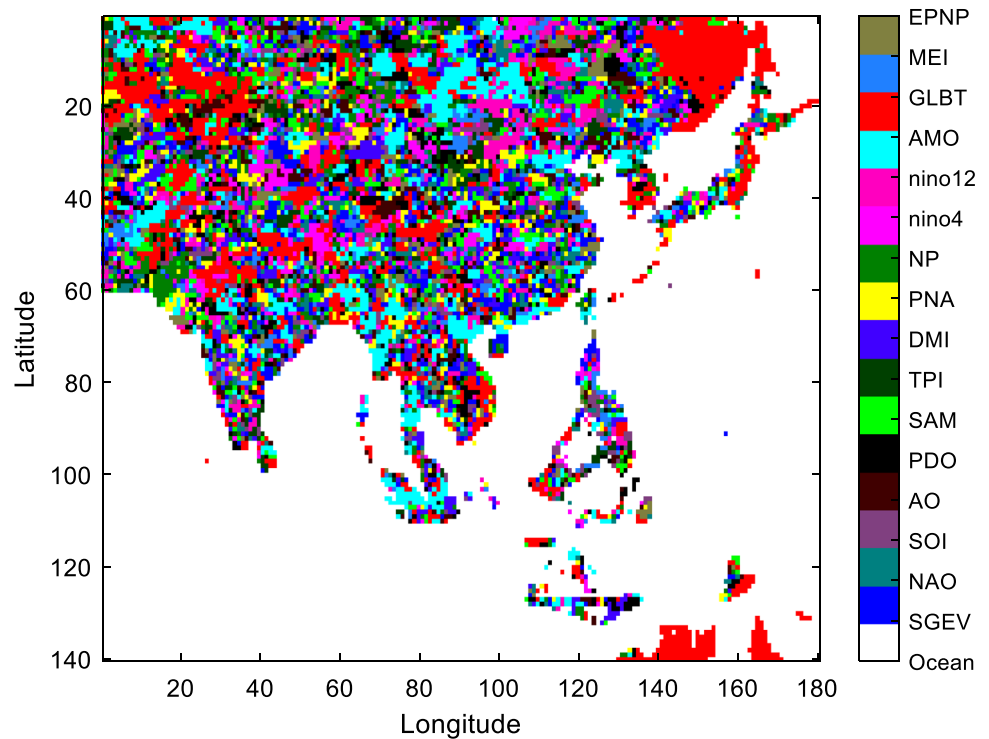
Fig. 10 Percentage distribution of grid locations over MAR for selected NSGEV models (X-axis represents the 15 climate indices corresponding to the NSGEV models, and Y-axis represents the number of grid locations in percentage)



NSGEV has significantly improved the modelling of EP compared to SGEV. However, it is evident from Fig. 11 that there is a considerable variation in the selected NSGEV model, covariates and NULT depending on the effect of the teleconnections over the MAR region. The uncertainty in the prediction of the EP will vary significantly over MAR. Therefore, it is proposed to investigate

the variability of NSGEV model parameters by grouping the grid locations having similar magnitude EP into coherent zones. In a large-scale domain, the grid locations closer to each other are expected to have identical EP behaviour, as discussed earlier.

Fig. 11 Selected best covariate based NSGEV and SGEV models at each grid over MAR



4.4 Clustering of EP zones

The analysis over MAR suggests that the within-year distribution of the EPs (Fig. 6) for the grids in a region are almost similar, indicating the NSGEV model parameters can be grouped over an area. Regionalization of model parameters reduces the model and parametric uncertainties in simulating the EP. In the present study, the EP zones are derived based on the magnitude of AMPs using the GMM probability-based clustering algorithm. An optimal number of six EP zones are identified over MAR, as shown in

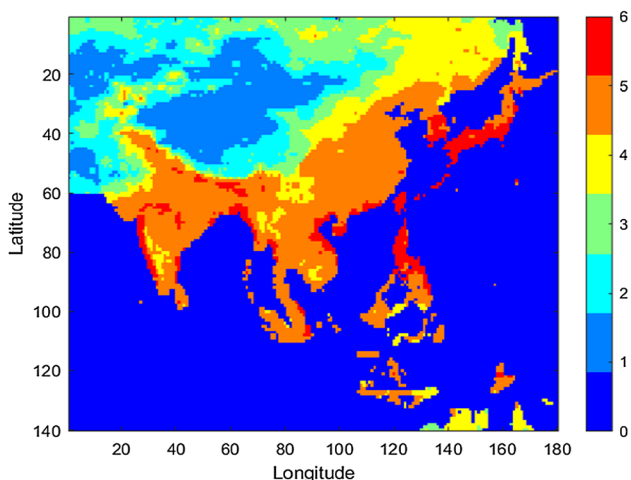


Fig. 12 Six EP zones obtained over MAR from GMM based clustering (color bar represents the zone number)

Fig. 12. The percentage of grid locations covered by each zone and their descriptive statistics are presented in Table 2. Zone-5 is the largest zone enclosing 30.95% of grid locations, mainly representing precipitation from the Southwest monsoon, which has an EP range from 35.95 to 78.74 mm and mean of 53.05 mm with a variation of 11.07 mm. Zone-6 is the smallest zone with a 5.95 percentage of grid locations having EP range between 78.76 to 189.99 mm with a mean of 101.06 mm and variation of 20.98 mm and lies on the windward side of the Western ghats (wettest peninsular India) and grids closer to coast-lines such as parts of Tamil Nadu, Vietnam, Japan, and receives Northeast and East Asian monsoons. The remaining zone-1, zone-2, zone-3, and zone-4 each cover 14–16 percentage of the study area with EP mean and variation in the range of 3–35 mm and 1.6–3.5 mm, respectively. The zones (1–3) cover the lee-ward portions of the Himalayas or Hindu-Kush, and the Arakan Mountains, Tibetan Plateau, and Tarim Basins, primarily arid and semi-arid regions receiving low precipitation, and zone-4 is mainly confined to the Northern Australia, Southeast countries, and leeward sides of Western ghats.

The zone-wise distribution of the number of grid locations for the selected NSGEV model (each climate index) is represented as a stacked column in Fig. 13a. It is observed that the GLBT and AMO are dominant among all climate indices in each of the zone, which is similar to the earlier results. The zone-wise observations show that the NSGEV models are the best alternative in all the EP zones

Table 2 Percentage of grid locations and descriptive statistics of AMP for identified zones over MAR

Zone no.	Percentage of grid locations	Minimum (mm)	Maximum (mm)	Mean (mm)	Standard deviation (mm)
1	16.52	3.88	10.65	8.05	1.66
2	16.05	10.65	16.49	13.57	1.66
3	14.48	16.50	23.59	19.77	1.98
4	16.05	23.60	35.91	29.75	3.45
5	30.95	35.95	78.74	53.05	11.07
6	5.95	78.76	189.99	101.06	20.98

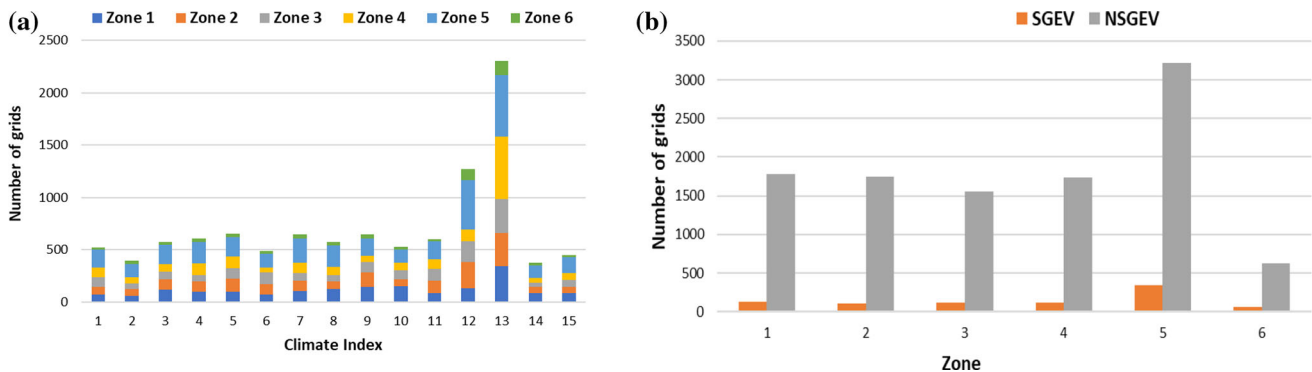


Fig. 13 a Selected covariate based NSGEV models at a number of grids within each zone; b Performance comparison of the NSGEV and SGEV models across the zones

compared to the SGEV model (Fig. 13b). However, the grids within each zone have different covariate-based NSGEV models. The selected best NSGEV models’ parameters at each grid location within each zone are pooled up to obtain the regional NSGEV parameters. Table 3 provides descriptive statistics of Zone-wise (Z-) NSGEV parameters across MAR. It is observed that the location parameter is lower for zone-1 ranging between 3.24 and 12.44, and a mean of 7.64, and parameter values increase across the zones, with zone-6 having the highest values in a range of 59.95–191.17 and mean of 98.64. The following sections present the evaluation of Z-NSGEV models’ performance and zone-wise EP changes for various return periods compared to SGEV.

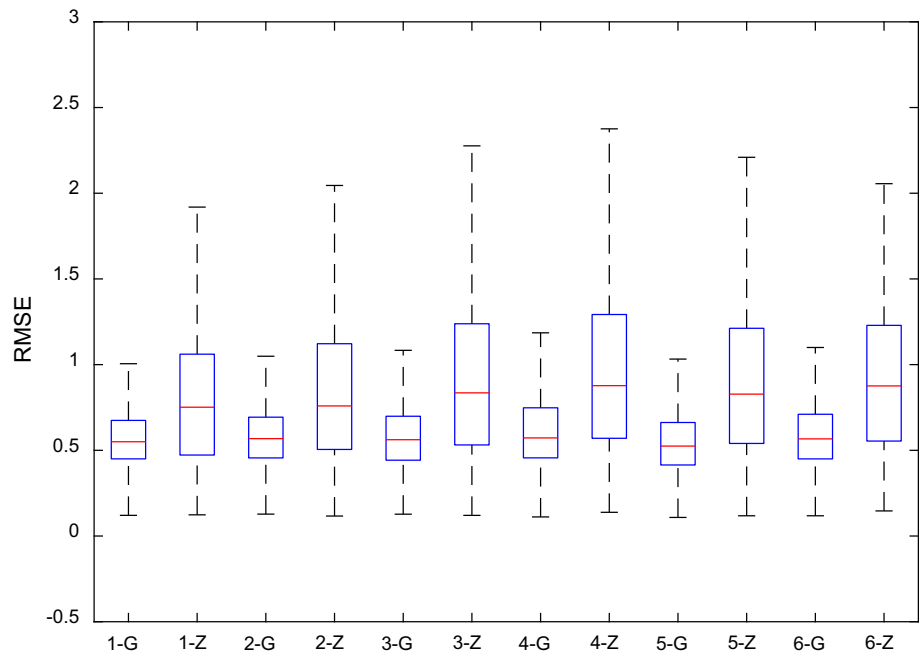
4.5 Z-NSGEV model performance

The efficacy of the model’s performance is brought out in two steps. In step one, an intra-comparison of the Z-NSGEV model with an NSGEV model within a zone is carried out. The Z-NSGEV models within the area are compared to the individual grid NSGEV models. Figure 14 shows the intra-comparison of models in each zone (grid location (G) and zone-wise (Z)). It is observed that the RMSE for Z-NSGEV models is comparable to the grid NSGEV models in the respective zones over MAR, indicating that the former can capture the EP dynamics of the respective zones. In step two, an inter-comparison of Z-NSGEV models across the zones is carried out to validate the performance of the regional models. It is to be

Table 3 Descriptive statistics of the zone-wise parameters of NSGEV models for six zones

Zone	Location parameter			Scale parameter			Shape parameter		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
1	3.24	12.44	7.64	0.73	4.05	2.18	- 0.57	0.72	0.075
2	7.27	19.45	13.14	1.26	6.49	3.29	- 0.60	0.75	0.038
3	12.15	34.47	19.31	1.74	11.04	4.65	- 0.73	0.58	0.026
4	14.32	52.71	29.15	3.31	14.73	6.81	- 0.54	0.66	0.03
5	27.67	89.77	50.50	4.80	36.33	12.08	- 0.64	0.71	0.094
6	59.95	191.17	98.54	12.93	58.51	25.32	- 0.56	0.57	0.046

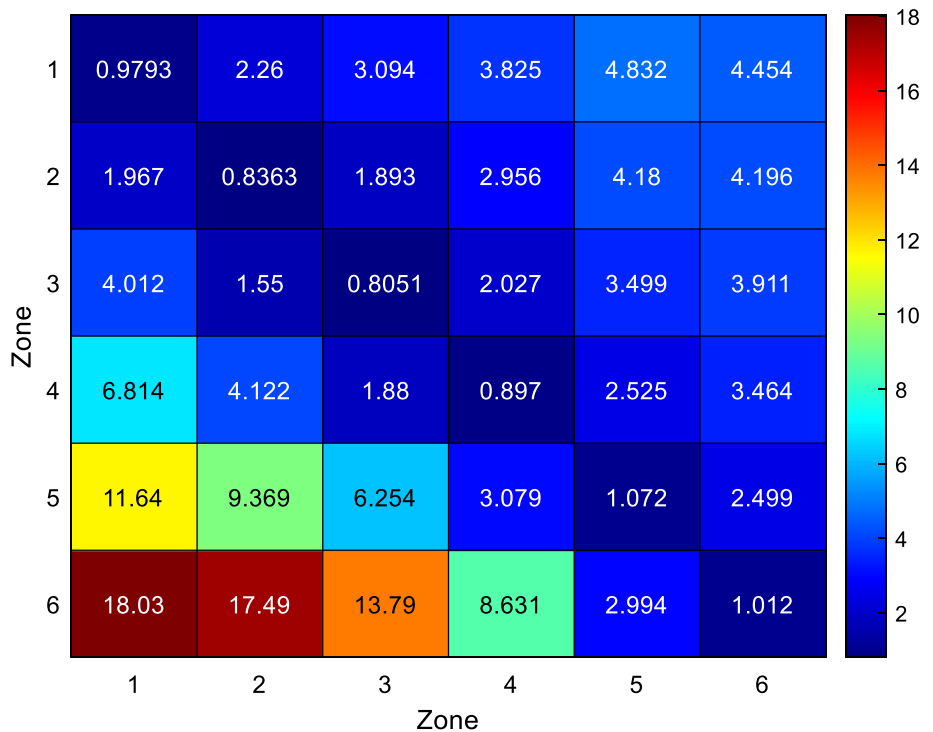
Fig. 14 Intra-comparison of the Z-NSGEV model with an NSGEV model at all the grids within a zone (G represents the grids within the zone and Z represents the zone-wise model i.e., 1-G and 1-Z refer to zone-1)



noted that since EP is used for the formation of the region, it cannot be used for homogeneity/validation of model performance. Alternatively, a simplified approach is proposed to validate the zone model’s performance by assuming that the model derived for a particular zone will underperform in other zones. A heat map representing the performance of the Z-NSGEV models across the zones is shown in Fig. 15. The non-diagonal RMSE values show

the zone model parameters used to predict EP in other zones, whereas the diagonal values show the zone-wise model performance. All the non-diagonal RMSE values are higher than the diagonal, indicating that the estimated regional parameters for all the six zones can capture the dynamics of the EP within the cluster grids better when compared to the other zones.

Fig. 15 Validation of Z-NSGEV models across the zones (Diagonal elements represent the within-zone performance and non-diagonal elements represent the model performance with other zones)



4.6 Zone-wise: estimation of EP

The zone-wise EP is estimated for various return periods of 5-, 10-, 25- 50-, and 100- years for SGEV and Z-NSGEV models. The relative percentage change in intensity of EP (NSGEV vs. SGEV) is shown in Fig. 16. It is observed that the percentage change ranges between 7 -10%, 5 -8%, 4-8%, and 4-10% for 5-, 10-, 50- and 100- year return periods, respectively, across the six zones. For each zone, the percentage change ranges between 6.8-9.8%, 7.11-9.24%, 6.4-8.9%, 4.4-7.6%, 8.4-10.85% and 7.4-10.5% for various return periods. Percentage change is observed higher in zone-5 and -6, which lie in monsoonal windward and coastal regions (prone to high tropical cyclones). The EP in other zones (1-4), mainly arid and semi-arid regions (inland of MAR), show a moderate increase in the EP intensities. It is observed that for all the six zones of MAR, the intensities of the EP are increasing, and return periods are decreasing, i.e., high-intensity EP will occur more frequently. The revised return periods based on the SGEV and Z-NSGEV models are provided in Table 4. The results show that the return periods have reduced on an average from 5- to 3.4-year, 10- to 6.7-year, 25- to 16.6-year, 50- to 32.2-year, and 100- to 61.64- year in six zones, respectively. In addition, the relative percentage change in zone-wise return periods (Fig. 17) shows that the percentage change in return periods ranges from 26 to 47% across all the zones. Overall, the results indicate an increase in extreme precipitation and reduction in the return periods across all zones over MAR compared to SGEV models.

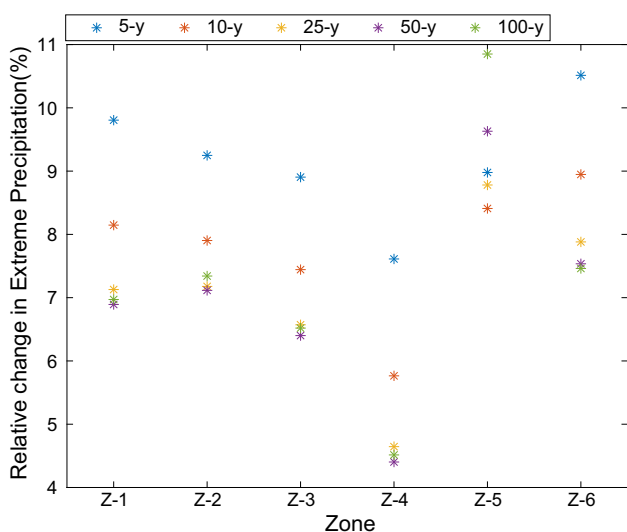


Fig. 16 Zone-wise relative percentage change in EP (NSGEV vs. SGEV) for various return periods

Table 4 Revised return periods based on the SGEV and Z-NSGEV models across the zones

Return period (year)	Return period (year)				
	5	10	25	50	100
Z-1	3.48	6.95	17.24	33.83	65.55
Z-2	3.40	6.69	16.27	31.39	59.65
Z-3	3.38	6.72	16.51	32.10	61.46
Z-4	3.54	7.27	18.48	36.55	70.77
Z-5	3.46	6.69	15.60	28.89	52.44
Z-6	3.27	6.43	15.81	30.97	60.02
Average	3.42	6.79	16.65	32.29	61.65

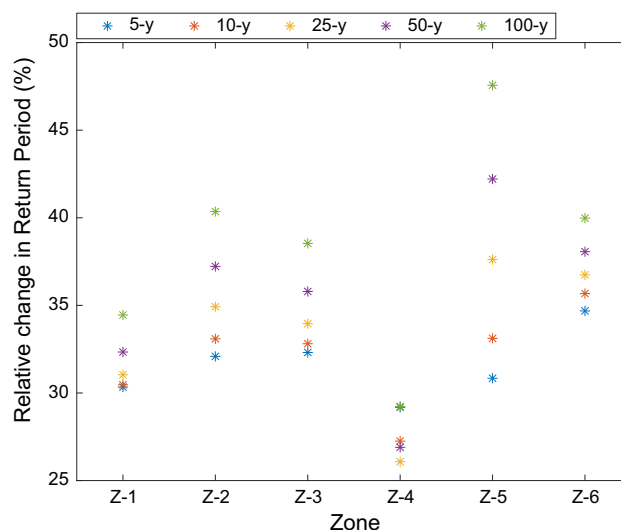


Fig. 17 Zone-wise relative percentage change in various Return Period (RP) across the six zones over MAR

5 Summary and conclusions

This study presents an NSGEV framework for modelling EP over Monsoon Asia Region (MAR). The non-uniform time-lag climate indices are used as covariates for the NSGEV models and have improved the model’s performance compared to uniform time-lag climate indices. It is observed that the NSGEV models selected at about 92% of the grid locations over MAR outperform the SGEV model. The MAR region is divided into six zones using the GMM clustering algorithm. The Z-NSGEV models are derived by regionalizing the model parameters for analyzing the EPs. The results show that the Z-NSGEV models can mimic the EP characteristics in all the zones. Finally, the zone-wise EPs are investigated for various return periods for Z-NSGEV and SGEV models. It is observed that there is an increase in intensity and reduction in return periods across all zones, especially in zone 5 and 6 there is a relatively

higher increase in intensity, which lies in monsoonal windward regions and coastal, respectively. Due to climate variability, these regions are highly receptive to EP and are expected to intensify (in both magnitude and frequency). The Z-NSGEV model outcomes will help in the reliable design and rehabilitation of water infrastructure, risk assessment, and to develop adaptation and mitigation strategies over MAR.

5.1 Limitations and extensions

1. In the NSGEV model, only the location parameter as a function of covariate is varied to address the non-stationarity. The effect of scale and shape parameters is not considered in this study. The work can be extended to understand other parameters' role on the NSGEV model's performance in MAR.
2. The Peak Over Threshold (POT) series can be considered instead of the AMP series in the proposed framework. POT has advantages over AMP as more maximums can be extracted for a set threshold value (Cheng and Aghakouchak 2014).
3. In this study, the location parameter for the NSGEV model is only a function of a single climate index (covariate). However, the association between EP and one teleconnection pattern is most likely modulated by other teleconnections or may include other effects of teleconnections. The framework can be extended to include the distribution parameters as a function of various combinations of covariates. Further, the study analyzes only the influence of teleconnections representing the global processes on EP variability over MAR. Other covariates, such as local temperature, topography, and land cover changes, can also be included to understand their effect on EP.
4. The NSGEV models consider linear functions to relate covariates and parameters. However, the linear assumption may not be able to capture the complex and nonlinear EP processes. The work is in progress to develop a nonlinear functional relationship to capture/mimic the processes between teleconnections and EP.
5. The regional homogeneity of the clusters requires the formation of the zones based on independent hydro-climatic datasets. However, in this study, EP was directly used as an attribute to cluster the MAR region, and hence homogeneity of the region is not carried out. Instead, a simplified approach is adopted for validating the zones, i.e., inter-comparison Z-NSGEV models. Alternatively, the homogeneous zones of EP can be formed using other hydro-climatic data influencing EP such as temperature, pressure, location parameters, geopotential height derived using reanalysis datasets (NCAR–NCEP) as attributes to cluster the zones. In

addition, geographic characteristics such as elevation, topographic features, and location variables can also be considered.

Appendix

Cross-correlation

Cross-correlation analysis determines the degree of similarity between two time-series over a time-lag range. The cross-correlation coefficient (r_k) indicates the strength of the association between two different time series. A strong negative or positive association is determined by the correlation coefficient's proximity to -1 or 1 . Equation (19) defines the correlation coefficient between two-time series Y_1 and Y_2 , with a time lag (K). (Coles 2001)

$$r_k = \frac{\sum_{t=1}^{T-K} (Y_1(t) - \bar{Y}_1(t))(Y_2(t+K) - \bar{Y}_2(t))}{\sqrt{\sum_{t=1}^T (Y_1(t) - \bar{Y}_1(t))^2} \sqrt{\sum_{t=1}^T (Y_2(t) - \bar{Y}_2(t))^2}} \quad (19)$$

where $\bar{Y}_1(t)$ and $\bar{Y}_2(t)$ are the mean for each time series.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00477-022-02211-4>.

Acknowledgements The authors are grateful to the two anonymous reviewers, associate editor and editor for their helpful comments and suggestions that substantially improved this work. We also thank Jency. M. Sojan, IIT Tirupati for helping with the preliminary analysis using gaussian mixture model clustering.

Funding Science and Engineering Research Board, SRG/2019/001251, Roshan Karan Srivastav

Declarations

Conflict of interest The authors have not disclosed any competing interests.

References

- Agilan V, Umamahesh NV (2016) Is the covariate based non-stationary rainfall IDF curve capable of encompassing future rainfall changes? *J Hydrol.* <https://doi.org/10.1016/j.jhydrol.2016.08.052>
- Armal S, Devineni N, Khanbilvardi R (2018) Trends in extreme rainfall frequency in the contiguous United States: attribution to climate change and climate variability modes. *J Clim* 31:369–385. <https://doi.org/10.1175/JCLI-D-17-0106.1>
- Ashok K, Guan Z, Yamagata T (2003) Influence of the Indian ocean dipole on the Australian winter rainfall. *Geophys Res Lett* 30:3–6. <https://doi.org/10.1029/2003GL017926>

- Azad S, Rajeevan M (2016) Possible shift in the ENSO-Indian monsoon rainfall relationship under future global warming. *Sci Rep* 6:3–8. <https://doi.org/10.1038/srep20145>
- Barnston AG, Livezey RE (1987) Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Acta Univ Agric Silvic Mendel Brun* 53:1689–1699
- Befort DJ, Leckebusch GC, Cubasch U (2016) Intraseasonal variability of the Indian summer monsoon: wet and dry events in COSMO-CLM. *Clim Dyn* 47:2635–2651. <https://doi.org/10.1007/s00382-016-2989-7>
- Burnham KP, Anderson DR (2004) Multimodel inference: understanding AIC and BIC in model selection. *Sociol Methods Res* 33:261–304. <https://doi.org/10.1177/0049124104268644>
- Trenberth KE, Dai A, Rasmussen RM, Parsons DB (2003) The changing character of precipitation. *Bull Am Meteorol Soc* 84:1205–1217. <https://doi.org/10.1175/BAMS-84-9-1205>
- Cheng L, Aghakouchak A (2014) Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. *Sci Rep* 4:1–7. <https://doi.org/10.1038/srep07093>
- Cheng L, Aghakouchak A, Gilleland E, Katz RW (2014) Nonstationary extreme value analysis in a changing climate. *Clim Change* 127:353–369. <https://doi.org/10.1007/s10584-014-1254-5>
- Coles S (2001) An introduction to statistical modelling of extreme values. Springer
- Dai A, Wigley TML (2000) Precipitation precip . Anomaly (mm) for typical El ninos. 27: 1283–1286.
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J Royal Stat Soc Ser B (methodol)*. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>
- Donat MG, Alexander LV, Yang H, Durre I, Vose R, Dunn RJH, Willett KM, Aguilar E, Brunet M, Caesar J, Hewitson B, Jack C, Klein Tank AMG, Kruger AC, Marengo J, Peterson TC, Renom M, Oria Rojas C, Rusticucci M, Salinger J, Elrayah AS, Sekele SS, Srivastava AK, Trewin B, Villarreal C, Vincent LA, Zhai P, Zhang X, Kitching S (2013) Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: the HadEX2 dataset. *J Geophys Res Atmos* 118:2098–2118. <https://doi.org/10.1002/jgrd.50150>
- Duan W, He B, Takara K, Luo P, Hu M, Alias NE, Nover D (2015) Changes of precipitation amounts and extremes over Japan between 1901 and 2012 and their connection to climate indices. *Clim Dyn* 45:2273–2292. <https://doi.org/10.1007/s00382-015-2778-8>
- El Adlouni S, Ouarda TBMJ, Zhang X, Roy R, Bobée B (2007) Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resour Res* 43:1–13. <https://doi.org/10.1029/2005WR004545>
- Gao M, Mo D, Wu X (2016) Nonstationary modelling of extreme precipitation in China. *Atmos Res* 182:1–9. <https://doi.org/10.1016/j.atmosres.2016.07.014>
- Groisman PY, Knight RW, Easterling DR, Karl TR, Hegerl GC, Razuvaev VN (2005) Trends in intense precipitation in the climate record. *J Clim* 18:1326–1350. <https://doi.org/10.1175/JCLI3339.1>
- Hao W, Shao Q, Hao Z, Ju Q, Baima W, Zhang D (2019) Nonstationary modelling of extreme precipitation by climate indices during rainy season in Hanjiang River Basin. *China Int J Climatol* 39:4154–4169. <https://doi.org/10.1002/joc.6065>
- He X, Guan H (2013) Multiresolution analysis of precipitation teleconnections with large-scale climate signals: a case study in South Australia. *Water Resour Res* 49:6995–7008. <https://doi.org/10.1002/wrcr.20560>
- Heffernan JE (2018) ISMEV: an introduction to statistical modeling of extreme values. R package version 1.42
- Hurvich MC, Tsai C-L (1989) Regression and time series model selection in small samples. *Biometrika*, Oxford University Press 76:297–307. URL: <https://www.jstor.org/stable/2336663>.
- Irwin S, Srivastav RK, Simonovic SP, Burn DH (2017) Delineation of precipitation regions using location and atmospheric variables in two Canadian climate regions: the role of attribute selection. *Hydrol Sci J* 62:191–204. <https://doi.org/10.1080/02626667.2016.1183776>
- IPCC (2007) Climate change 2007: synthesis report
- Jha S, Das J, Goyal MK (2021) Low frequency global-scale modes and its influence on rainfall extremes over India: nonstationary and uncertainty analysis. *Int J Climatol* 41:1873–1888. <https://doi.org/10.1002/joc.6935>
- Katz R (2002) Statistics of Extremes in climatology and hydrology. *Adv Water Resour* 25:1287–1304
- Khouakhi A, Villarini G (2017) Contribution of tropical cyclones to rainfall at the global scale. *J Clim*. <https://doi.org/10.1175/JCLI-D-16-0298.1>
- Kim T, Shin JY, Kim S, Heo JH (2018) Identification of relationships between climate indices and long-term precipitation in South Korea using ensemble empirical mode decomposition. *J Hydrol* 557:726–739. <https://doi.org/10.1016/j.jhydrol.2017.12.069>
- Krishnan R, Sabin TP, Madhura RK, Vellore RK, Mujumdar M, Sanjay J, Nayak S, Rajeevan M (2019) Non-monsoonal precipitation response over the Western Himalayas to climate change. *Clim Dyn* 52:4091–4109. <https://doi.org/10.1007/s00382-018-4357-2>
- Krishnaswamy J, Vaidyanathan S, Rajagopalan B, Bonell M, Sankaran M, Bhalla RS, Badiger S (2015) Non-stationary and nonlinear influence of ENSO and Indian Ocean Dipole on the variability of Indian monsoon rainfall and extreme rain events. *Clim Dyn* 45:175–184. <https://doi.org/10.1007/s00382-014-2288-0>
- Kucharski F, Bracco A, Yoo JH, Molteni F (2007) Low-frequency variability of the Indian monsoon-ENSO relationship and the tropical Atlantic: the “weakening” of the 1980s and 1990s. *J Clim* 20:4255–4266. <https://doi.org/10.1175/JCLI4254.1>
- Meneghini B, Simmonds I, Smith NI (2007) Association between Australian rainfall and the Southern annular mode. *Int J Climatol* 2029:2011–2029. <https://doi.org/10.1002/joc>
- Mo K, Livezey RE (1986) Tropical-extratropical geopotential height teleconnections during the north hemisphere winter 1–27
- Mondal A, Mujumdar PP (2015) Modelling non-stationarity in intensity, duration and frequency of extreme rainfall over India. *J Hydrol* 521:217–231. <https://doi.org/10.1016/j.jhydrol.2014.11.071>
- Ng CHJ, Vecchi GA, Muñoz ÁG, Murakami H (2019) An asymmetric rainfall response to ENSO in East Asia. *Clim Dyn* 52:2303–2318. <https://doi.org/10.1007/s00382-018-4253-9>
- O’Gorman PA (2015) Precipitation extremes under climate change. *Curr Clim Change Rep* 1:49–59. <https://doi.org/10.1007/s40641-015-0009-3>
- Ouarda TBMJ, Charron C (2018) Nonstationary temperature-duration-frequency curves. *Sci Rep* 8:1–9. <https://doi.org/10.1038/s41598-018-33974-y>
- Ouarda TBMJ, Yousef LA, Charron C (2019) Non-stationary intensity-duration-frequency curves integrating information concerning teleconnections and climate change. *Int J Climatol* 39:2306–2323. <https://doi.org/10.1002/joc.5953>
- Rathinasamy M, Agarwal A, Sivakumar B, Marwan N, Kurths J (2019) Wavelet analysis of precipitation extremes over India and teleconnections to climate indices. *Stoch Environ Res Risk Assess* 33:2053–2069. <https://doi.org/10.1007/s00477-019-01738-3>

- Res C, Trenberth KE (2011) Changes in precipitation with climate change 47: 1–18. <https://doi.org/10.3390/atmos8050083>
- Roushangar K, Alizadeh F (2018) Identifying complexity of annual precipitation variation in Iran during 1960–2010 based on information theory and discrete wavelet transform. *Stoch Environ Res Risk Assess* 32:1205–1223. <https://doi.org/10.1007/s00477-017-1430-z>
- Roushangar K, Alizadeh F, Adamowski J (2018) Exploring the effects of climatic variables on monthly precipitation variation using a continuous wavelet-based multiscale entropy approach. *Environ Res* 165:176–192. <https://doi.org/10.1016/j.envres.2018.04.017>
- Saini A, Sahu N, Kumar P, Nayak S, Duan W, Avtar R, Behera S (2020) Advanced rainfall trend analysis of 117 years over west coast plain and hill agro-climatic region of India. *Atmosphere (basel)* 11:1–25. <https://doi.org/10.3390/atmos11111225>
- Sarhadi A, Soulis ED (2017) Time-varying extreme rainfall intensity-duration-frequency curves in a changing climate. *Geophys Res Lett* 44:2454–2463. <https://doi.org/10.1002/2016GL072201>
- Scrucca L, Fop M, Murphy TB, Raftery AE (2016) Mclust 5 clustering, classification and density estimation using gaussian finite mixture models. *R J* 8:289–317
- Shige S, Nakano Y, Yamamoto MK (2017) Role of orography, diurnal cycle, and intraseasonal oscillation in summer monsoon rainfall over the western ghats and myanmar coast. *J Clim* 30:9365–9381. <https://doi.org/10.1175/JCLI-D-16-0858.1>
- Sun X, Renard B, Thyer M, Westra S, Lang M (2015) A global analysis of the asymmetric effect of ENSO on extreme precipitation. *J Hydrol* 530:51–65. <https://doi.org/10.1016/j.jhydrol.2015.09.016>
- Sushama L, Ben Said S, Khaliq MN, Nagesh Kumar D, Laprise R (2014) Dry spell characteristics over India based on IMD and APHRODITE datasets. *Clim Dyn* 43:3419–3437. <https://doi.org/10.1007/s00382-014-2113-9>
- Tan X, Gan TY (2017) Non-stationary analysis of the frequency and intensity of heavy precipitation over Canada and their relations to large-scale climate patterns. *Clim Dyn* 48:2983–3001. <https://doi.org/10.1007/s00382-016-3246-9>
- Thiombiano AN, St-Hilaire A, El Adlouni SE, Ouarda TBMJ (2018) Nonlinear response of precipitation to climate indices using a non-stationary poisson-generalized pareto model: case study of southeastern Canada. *Int J Climatol* 38:e875–e888. <https://doi.org/10.1002/joc.5415>
- Trenberth KE (2012) Framing the way to relate climate extremes to climate change. *Clim Change* 115:283–290. <https://doi.org/10.1007/s10584-012-0441-5>
- Varikoden H, Revadekar JV, Choudhary Y, Preethi B (2015) Droughts of Indian summer monsoon associated with El Niño and Non-El Niño years. *Int J Climatol* 35:1916–1925. <https://doi.org/10.1002/joc.4097>
- Verdon DC, Wyatt AM, Kiem AS, Franks SW (2004) Multidecadal variability of rainfall and streamflow: Eastern Australia. *Water Resour Res* 40:1–8. <https://doi.org/10.1029/2004WR003234>
- Vu TM, Mishra AK (2019) Nonstationary frequency analysis of the recent extreme precipitation events in the United States. *J Hydrol* 575:999–1010. <https://doi.org/10.1016/j.jhydrol.2019.05.090>
- Wang H (2002) The instability of the East Asian summer monsoon-ENSO relations. *Adv Atmos Sci* 19:1–11
- Wei W, Shi Z, Yang X, Wei Z, Liu Y, Zhang Z, Ge G, Zhang X, Guo H, Zhang K, Wang B (2017) Recent trends of extreme precipitation and their teleconnection with atmospheric circulation in the Beijing-Tianjin sand source region, China, 1960–2014. *Atmosphere (basel)* 8:83. <https://doi.org/10.3390/atmos8050083>
- Yatagai A, Arakawa O, Kamiguchi K, Kawamoto H, Nodzu MI, Hamada A (2009) A 44-year daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Sci Online Lett Atmos* 5:137–140. <https://doi.org/10.2151/sola.2009-035>
- Yatagai A, Kamiguchi K, Arakawa O, Hamada A, Yasutomi N, Kitoh A (2012) Aphrodite constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bull Am Meteorol Soc* 93:1401–1415. <https://doi.org/10.1175/BAMS-D-11-00122.1>
- Yilmaz AG, Perera BJC (2014) Extreme Rainfall nonstationarity investigation and intensity–frequency–duration relationship. *J Hydrol Eng* 19:1160–1172. [https://doi.org/10.1061/\(asce\)he.1943-5584.0000878](https://doi.org/10.1061/(asce)he.1943-5584.0000878)
- Zhang W, Zhou T (2019) Significant increases in extreme precipitation and the associations with global warming over the global land monsoon regions. *J Clim* 32:8465–8488. <https://doi.org/10.1175/JCLI-D-18-0662.1>
- Zhang X, Wang J, Zwiers FW, Groisman PY (2010) The influence of large-scale climate variability on winter maximum daily precipitation over North America. *J Clim* 23:2902–2915. <https://doi.org/10.1175/2010JCLI3249.1>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.