Received: 16 July 2019 | Revised: 6 March 2020 | Accepted: 9 March 2020 | Published on: 30 March 2020

DOI: 10.1002/joc.6556

RESEARCH ARTICLE

International Journal of Climatology RMetS

Spatial variation in long-lead predictability of summer monsoon rainfall using a time-varying model and global climatic indices

Riya Dutta | **Rajib Maity**

Department of Civil Engineering, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal, India

Correspondence
Rajib Maity, Department of Civil Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721302, West Bengal, India.
Email: rajib@civil.iitkgp.ac.in, rajibmaity@gmail.com

Funding information
Department of Science and Technology, Climate Change Programme (SPLICE), Government of India, Grant/Award Number: DST/CCP/CoE/79/2017(G)

Abstract

Long-lead prediction of summer monsoon rainfall in India is a challenging task, especially at finer spatial scale. The spatial variability in the long-lead (one or two season in advance) prediction plays a vital role in planning of hydrological and agricultural aspects of the society. One of the major issues in this field is the climate-induced time-varying characteristics (non-stationarity) that lead to a deteriorating model performance as the time passes by after the model calibration. This study proposes the use of time-varying approaches in order to check such deteriorating performance over time. Considering the time-varying association between Indian summer monsoon rainfall and large-scale climatic indices (e.g., El Niño-Southern Oscillation, Equatorial Indian Ocean Oscillation, North Atlantic Oscillation, Pacific Decadal Oscillation, and El Niño Modoki Index), a time-varying approach based on hybrid graphical modelling (GM) and vine copula (GM-Copula) is demonstrated for rainfall prediction over five homogeneous monsoon regions (HMRs) in India. The time-varying characteristic is imparted in the GM-Copula approach by recursively updating the model inputs and the corresponding model parameters at a regular time interval (τ) through recalibration. In the time-varying framework, the parameter τ is referred to as optimum recurrence interval of model recalibration and it is identified as 5 years for the regions with moderate rainfall and 3 years for regions with above and below moderate rainfall. The developed time-varying approach is able to yield reasonably good prediction performance (mean absolute percentage error being within 4–10% across HMRs) with a prediction lead time of 5 months. HMR-wise seasonal rainfall predictions with such quality and lead time are expected to be highly useful.

KEYWORDS

climatic indices, graphical modelling, homogeneous monsoon regions (HMRs), Indian summer monsoon rainfall (ISMR), long-lead prediction, spatial variability, vine copula

1 | INTRODUCTION

In India, annual rainfall is largely accounted from the southwest monsoon season, occurring during the months of June, July, August and September, also referred to as summer monsoon rainfall (Wang *et al.*, 2009; Singh *et al.*, 2012). The spatiotemporal variability in the Indian Summer Monsoon Rainfall (ISMR) is linked with the atmospheric circulation patterns through hydroclimatic teleconnection (Kahya and Dracup, 1993; Ashok *et al.*, 2001, 2004; Maity and Nagesh, 2008). The important teleconnection patterns known to influence the variability of summer monsoon rainfall are El Niño-Southern Oscillation (ENSO), Equatorial Indian Ocean Oscillation (EQUINOO), North Atlantic Oscillation (NAO), Pacific Decadal Oscillation (PDO), El Niño Modoki Index (EMI) to name a few (Nair *et al.*, 2018). These large-scale indices are associated with ISMR at lead times of months to seasons and hence, used for long-range prediction of ISMR. However, the nature of association between different large-scale climatic indices and monsoon rainfall varies with both space and time. Therefore, information on the temporal evolution of large-scale indices and their impact may provide a better understanding of the spatiotemporal variability in the summer monsoon rainfall. This is the focus of this study.

Several observational and modelling studies established the teleconnection pattern between the large-scale indices and the summer monsoon rainfall (Pant and Parthasarathy, 1981; Rasmusson and Carpenter, 1983; Ju and Slingo, 1995; Meehl and Arblaster, 1998; Kumar *et al.*, 1999). ENSO mode, for instance, is identified as the third largest component of Asian summer monsoon variations. Regarding the teleconnection mechanism between ISMR and ENSO events, Rasmusson and Carpenter (1983) concluded that 'episodes of above normal sea surface temperatures (SSTs) over the Eastern and Central Equatorial Pacific are associated with a low Southern Oscillation Index, that is, negative pressure anomalies in the Southeast Pacific and positive anomalies over the Indian Ocean region, weaker than normal southwest monsoon over the Arabian Sea, and below normal rainfall over India'. The inverse relationship between ISMR and ENSO is well documented (Kumar *et al.*, 1999); however, the relationship is modulated on decadal timescale due to the influence of other causative climate forcing like Atlantic Multidecadal Oscillation (Lu *et al.*, 2006; Kucharski *et al.*, 2009; Chen *et al.*, 2010; Ault *et al.*, 2013; Lewis and Legrande, 2015; Brown *et al.*, 2016) and zonal shifts in ENSO's centre from Eastern Pacific to Central Pacific (Kumar *et al.*, 2006; Fan *et al.*, 2017). Temporal shifts in ENSO have also been studied across three periodicity bands, 2–3 years (near biennial), 3–8 years (classical ENSO) and 8–25 years (decadal), and the former shows increased variability in Niño3.4 index for the recent years (Hope *et al.*, 2017). The change in ENSO teleconnections is largely attributable to ENSO variance itself (Chowdary *et al.*, 2012), which shows a change in the 2–3 years' band.

Another driver of rainfall variability in this region is EQUINOO, which is the atmospheric component of Indian Ocean Dipole (IOD) mode (Kumar *et al.*, 2007; Rajeevan *et al.*, 2007; Francis and Gadgil, 2010; Charlotte and Mathew, 2012). During the summer monsoon season, the convection over the eastern part of the Equatorial Indian Ocean is negatively correlated to that over the western part of the Equatorial Indian Ocean. When the convection is enhanced (suppressed) over the western part of the Equatorial Indian Ocean, the anomalous surface pressure gradient, high to low, is towards the west (east) so that the anomalous surface wind along the equator becomes easterly (westerly). The oscillation between these two states is considered as the EQUINOO index. Recent meteorological observations indicate a strong link between ISMR and EQUINOO due to the association of large-scale monsoon rainfall over the Indian region with the northward propagation of convective system generated over the Indian Ocean region (Gadgil *et al.*, 2004; Gadgil and Gadgil, 2006). Furthermore, the two dominant modes ENSO and IOD are correlated and the complex evolution of the IOD–ENSO relationship is majorly controlled by their variability at three dominant timescales of 1.5, 3 and 24 years (Sang *et al.*, 2018). Moreover, it is also observed that ENSO–ISMR relationship is modified by the influence of IOD which is in-turn associated with ENSO.

Another climatic index associated with ISMR is NAO, which is a temporal fluctuation of the zonal wind strength across the Atlantic Ocean due to pressure variations in the subtropical anticyclone belt and in the subpolar low near Iceland. Dugam *et al.* (1997) gave details on studies on the association between Northern Hemispheric pressure anomalies and Indian summer monsoon. Strong negative (positive) NAO events, through hemispheric change in winds and storm tracks, lead to tropospheric temperature anomalies over Eurasia. These anomalies decrease (increase) meridional gradient of tropospheric temperature, resulting in below (above) normal summer monsoon rainfall in India (Goswami *et al.*, 2006). Temporal changes are found in the means and variability of the NAO index. For instance, there has been a sustained significant decrease in the summer NAO since the 1990s and, a striking increase in variability of the winter NAO (Hanna *et al.*, 2015).

PDO is another large-scale, strongly associated climatic index that modulates the ISMR–ENSO relation.

The mechanism by which the PDO could affect the monsoon was hypothesized by Krishnamurthy and Krishnamurthy (2014). According to them, the pressure change associated with the SST footprint of warm phase of PDO affects the equatorial trade winds, which is the part of the equatorial Walker circulation. As a result, in the Central Pacific, an enhanced ascending motion and over the maritime continent a descending motion is observed. This is sustained by the ascending motion over the Equatorial Indian Ocean and leads to drought condition over the Indian region through the descending Hadley branch. Thereby, PDO index exhibits significant negative correlation with ISMR, similar to the relation between ISMR and ENSO. The warm (cold) phases of PDO are related to droughts (floods) over India. Studies have also identified a change in the zonal propagation of ENSO-related SST anomaly from the end of 1970s which coincides with the late 1970s switch from cold to warm phase in the PDO states (Mantua *et al*., 1997; Mantua and Hare, 2002). It is also observed that ENSO-related SST anomaly preferably evolves to the east (west) during the warm (cold) phase of the PDO (Antico and Barros, 2017).

Lastly, EMI, warming in the Central Pacific (~Nino4 region) flanked by colder SST anomalies to the west and east, is considered to modulate the variability of ISMR as well. A double-cell pattern during El Niño Modoki shows a marked difference from the single-cell pattern in the typical El Niño case. The impacts of the El Niño Modoki on the climate of the surrounding subtropical regions may be attributed to the Rossby waves generated by the diabatic heating in the central tropical Pacific. The impact of ENSO events on India is seen to be limited and confined to Eastern Central India. In comparison, the impact from El Niño Modoki is seen over a larger area in Southern India (Nair *et al*., 2018). Thereby, it is evident that the large-scale climatic indices show temporal variability with a space and time-varying nature of interaction among the large-scale indices, which vastly impacts the rainfall pattern in the Indian region.

Another important aspect is that the summer monsoon rainfall at regional scale over the Indian domain responds to the above-mentioned large-scale climatic indices in complex ways. Dutta and Maity (2018) studied the time-varying association among two large-scale climatic indices and ISMR and concluded the need to carry out a detailed analysis at finer spatial scale with a larger pool of large-scale climatic indices. Several other studies have been performed on Indian rainfall; however, it is difficult to treat India as a single unit for interpreting the association with the large-scale indices, as the association have seasonal and regional differences (Maity and Nagesh, 2006; Vathsala and Koolagudi, 2017). For instance, highest correlation, with respect to ENSO and EQUINOO (atmospheric component of IOD), was observed for Central Northeast India and West Central India followed by northeast and northwest regions of India and least for Peninsular India (Kashid and Maity, 2012). The influence of ENSO is distributed more widely as compared to IOD, which mainly influences the mean position of the monsoon trough over India. Most of these regions are significantly associated with IOD; however, the effects are opposite in nature from region to region (Ashok and Saji, 2007). It is further observed that interannual variability in the summer monsoon rainfall in peninsular and northwest regions of India is better explained by NAO index (Dugam *et al*., 1997). Nair *et al*. (2018) showed that the summer monsoon rainfall in Northeast Indian region is controlled by climatic indices like NAO, and the nature of contribution of the different climatic indices is in general opposite as compared to the other regions. Moreover, the northeast (heavy rainfall region) and west central part of India show a strong association with PDO index; however, summer monsoon rainfall over Central Northeast India shows weak associated (Sen, 2011). Different statistical, machine learning and dynamic models have been used to make such inferences. Methods like step-wise regression, canonical correlation, artificial neural network and genetic programming have also been utilized to develop prediction models at finer spatial scale (Parthasarathy *et al*., 1993; Kane, 2006; Ashok and Saji, 2007; Phatak *et al*., 2011; Gadgil and Srinivasan, 2012; Kashid and Maity, 2012; Singh *et al*., 2012; Guhathakurta *et al*., 2015; Pattanaik *et al*., 2019). However, recent findings clearly state that ISMR at regional scale is influenced by the combined effect of large number of climatic indices (Nair *et al*., 2018). Thereby, it is vital to identify the complex association of the large-scale climatic indices and summer monsoon rainfall. However, limitations of the existing prediction models mostly lies in their inability to consider the conditional independence structure among the predictors and predictand that helps to identify and ignore the redundant variables. Furthermore, these models consider time-invariant inputs and model parameters, not taking into consideration the time variability in the association of the large-scale climatic indices and summer monsoon rainfall at regional scale. Addressing these issues related to (a) complex temporal association of the large-scale indices and summer monsoon rainfall and (b) spatial variation in association and predictability of summer monsoon rainfall forms the motivation of this study.

The objective of this study is to analyse the spatial variation in long-lead predictability of summer monsoon rainfall by identifying the time-varying association between the large-scale climatic indices and rainfall over homogeneous monsoon regions (HMRs) of India. A time-varying approach based on hybrid graphical modelling (GM) and vine copula (GM-Copula) (Dutta and Maity, 2018), hereafter referred to as time-varying GM-Copula model, has been utilized in this study. GM approach is used to identify the complex association among the variables as it provides the complete conditional independence structure among the variables and the parameters of the prediction model are obtained using vine copula, which effectively considers the nonlinear association among the variables. The model development period is considered as a moving window, with step size of τ and the model is recalibrated (in terms of model input and parameters) in regular intervals to capture the time-varying association among the large-scale indices and ISMR. Furthermore, the results of the proposed time-varying model are compared to its time-invariant counterpart.

The prediction models are developed for each HMR of India in order to assess the region-wise variation in association with the large-scale climatic indices and to improve the prediction performance at regional scale. These regions are divided based on the similarity in rainfall characteristics and association of subdivisional monsoon rainfall with regional/global circulation parameters (Parthasarathy *et al.*, 1993), as per the specifications of Indian Institute of Tropical Meteorology (IITM). Further details are provided in Section 3.

2 | METHODOLOGY

The large-scale climatic indices used as the input variables for the long-lead prediction of summer monsoon rainfall are ENSO, EQUINOO, NAO, PDO and EMI. Initially, all lags up to 15 months for each of the climatic index have been considered and the best lag combination is ascertained. Lag indicates the gap (number of months) between input climatic indices and the starting month of summer monsoon rainfall, that is, June. For example, for summer monsoon rainfall (June to September) in 1998, inputs with Lag 1 are from the month of May 1998 and inputs with Lag 2 are from the month of April 1998 and so on. Since a lag up to 15 months is sufficient to pick out the best lags for the input variables, no further lags have been considered. The lags beyond 15 months show insignificant association with the summer monsoon rainfall.

2.1 | Model development

The methodological framework for development of a time-varying GM-copula model includes two major aspects: (a) development of the parsimonious (with optimum number of inputs and parameters without compromising the prediction performance) prediction model using GM-Copula approach and (b) imparting the time-varying characteristic by updating the inputs and parameters of the model.

Addressing the first aspect, the prediction model is developed by investigating the association among the predictors (large-scale climatic indices with several lags) and predictand (summer monsoon rainfall) using the GM approach. This approach provides a well-defined conditional independence structure among the variables that helps to assess the dependent (directly influencing), independent (not influencing) and conditionally independent (indirectly influencing) variables (Jordan, 2004; Ihler *et al.*, 2007; Whittaker, 2009; Dutta and Maity, 2018). Thereby, the GM approach provides the dependencies between the different lags of the large-scale climatic indices (input variables) and summer monsoon rainfall (target variable) as well as the dependencies among the input variables. Next, discarding all the independent and conditionally independent input variables with respect to the target variable, the prediction model is developed using C-Vine copula approach (Aas *et al.*, 2009; Bauer *et al.*, 2012; Brechmann and Schepsmeier, 2013; Gómez *et al.*, 2017). Mathematical formulations used for development of the model are provided in Appendix.

2.2 | Spatiotemporal variability

Addressing the second aspect, that is, in order to develop the time-varying prediction model, the model needs to be updated after regular time intervals. To ensure best possible prediction results, this time interval needs to be optimized such that it should not be too long that misses the temporal variation in association and too short that leads to frequent updating (Dutta and Maity, 2018). For the development of the proposed time-varying model, the model development period is considered as a moving window of 30 years and the immediately following n years is considered as the testing period. For instance, the first model development period is considered from 1901 to 1930 and the model testing period is from 1931 to $1931 + (n - 1)$. Thereby, the model developed considering the data for the first 30 years is validated considering the data for the following n years. As the model is updated after n years the next model development period is shifted by n years. Thereby, the second model development period is considered from $1901 + n$ to $1930 + n$ and the model testing period is from $(1930 + n) + 1$ to $(1930 + n) + n$. To identify the optimum recurrence interval (ORI) of model recalibration (τ), that is, the optimum value of n or the step size of the moving development period, the procedure is repeated for n = 1, 2...,10 years. The model performance during the entire contiguous model testing periods, which is essentially the entire period of 1931 to 2010, is evaluated for different values of n to identify the value of τ. The value of n, for which best performance is achieved, is selected as τ.

In order to study the spatial variation in predictability using the time-varying GM-Copula model, the analysis is carried out for five HMRs in India. These are northwest (NW), central northeast (CNE), northeast (NE), west central (WC), and peninsular (PE) (Figure 1). The time-varying GM-Copula model is developed for each of these HMRs. The values of τ for each HMR are ascertained separately.

2.3 | Benefit of considering multiple large-scale climatic indices and time-varying characteristics

Another time-varying GM-Copula model is developed using a subset of the large-scale climatic indices, namely ENSO and EQUINOO, two most widely associated climatic indices to Indian summer monsoon rainfall. Comparing the prediction performance of the two models (developed using two different sets of input variables) helps to establish the contribution of other climatic indices, that is, EMI, NAO and PDO, which may vary from region to region. The model output of both the models is compared with its time-invariant counterpart (henceforth referred to as the time-invariant model) for each region. The procedure explained above remains same but only one model is developed using 30-year data, and the developed prediction model is used for entire testing period without imparting the time-varying characteristics, explained before.

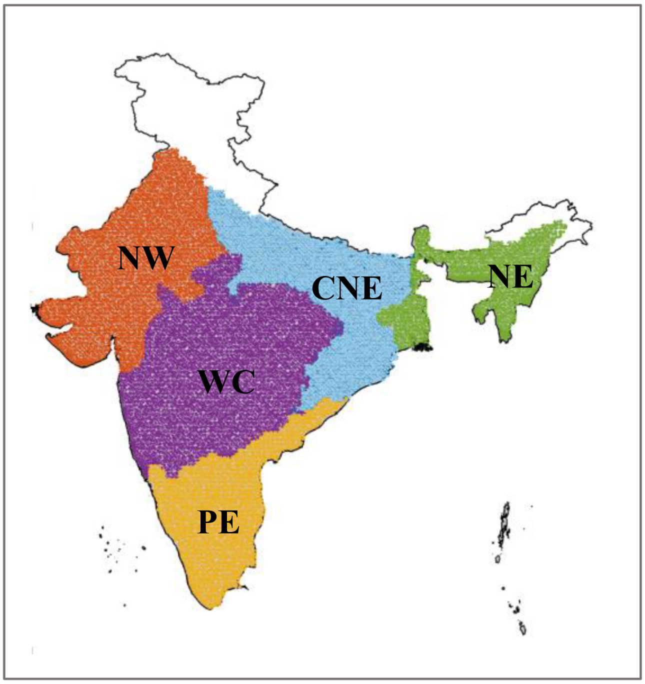

FIGURE 1 Map showing HMRs in India. These regions are the grouped contiguous sub-divisions based on the rainfall characteristics and association with global/regional circulation parameters to form the HMRs [Colour figure can be viewed at wileyonlinelibrary.com]

3 | DATA USED

Summer monsoon rainfall (June–September) data for the aforementioned HMRs (Figure 1) are obtained from IITM (https://www.tropmet.res.in/, accessed in July 2019). The summer monsoon rainfall is known to have considerable spatial variability across India and there are multiple meteorological subdivisions (total 36). These subdivisions are proposed based on the local distribution characteristics of seasonal rainfall. Some of the contiguous subdivisions are grouped based on the rainfall characteristics and association with global/regional circulation parameters to form the HMRs (Parthasarathy *et al.*, 1993). The monthly area weighted rainfall series for each meteorological subdivision have been prepared by assigning the district area as the weight for each rain-gauge station in that subdivision. Similarly assigning the subdivision area as the weight to each of the subdivisions in the region, area weighted monthly rainfall series are prepared for the HMRs of India. The details on the evaluation of the monthly time-series for the HMRs are available at https://tropmet.res.in/Data%20Archival-51-Page (accessed in July 2019). The cumulative seasonal rainfall (June–September) is obtained by summing up the monthly values for each of the HMRs.

The time period of the study is from 1901 to 2010 and the data for the aforementioned large-scale climate indices are obtained for the given time period. SST, Sea Level Pressure (SLP) and Zonal Wind data are obtained from ERA-20C, a reanalysis product. ERA-20C is the first atmospheric reanalysis of the 20th century (1900–2010) of the European Centre for Medium-Range Weather Forecast (www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-20c). It is a coupled atmosphere/land surface/ocean waves model used to reanalyze the weather, by assimilating observations of surface pressure and surface marine winds. ERA20C is the latest reanalysis product spanning over a long time period, in comparison to the other reanalysis products. Such a long time frame is beneficial in the proposed time-varying framework to investigate time-varying characteristics of the relationship as mentioned before.

The SLP and Zonal Wind data are utilized to derive the large-scale climatic indices (hereafter, referred to as D1). ENSO index is evaluated as the SST anomaly over Niño3.4 region (120°–170°W, 5°S–5°N). EMI is derived from the difference in area average SST anomalies in the regions of 10°S–10°N and 165°E–140°W; 15°S–5°N and 110°–70°W; and 10°S–20°N and 125°–145°E (Ashok *et al*., 2007). PDO (Deser et al., 2016) is derived by evaluating the leading Empirical Orthogonal Function (EOF) of SST anomalies in the North Pacific basin (polewards of 20°N). NAO (Hurrel et al., 2018) is derived by evaluating the leading EOF of SLP anomalies over the Atlantic sector, 20°–80°N, 90°W–40°E. Lastly, EQUINOO (Gadgil *et al*., 2004) index is computed as the negative of the zonal wind anomaly at surface in the Equatorial Indian Ocean region (60°–90°E, 2.5°S–2.5°N). All the series are checked for trend and, if present, it is detrended. For instance, the rainfall series for the regions of northeast, west central and peninsular regions exhibit significant trend at 5% significance level, which are detrended before analysis.

4 | RESULTS AND DISCUSSION

4.1 | Identification of the ORI of model recalibration

In order to identify the ORI of model recalibration (τ), the analysis is carried out for the values of n starting from 1 to 10 years and the predicted rainfall obtained using the different values of n are compared. The range of 1 to 10 years is an initial guess assuming the value of τ will be well within this range. Performance statistics used for comparison of model performance are correlation coefficient (R), root mean square error (RMSE), Nash–Sutcliffe model efficiency (NSE) coefficient, degree of agreement (Dr) and coefficient of determination (R^2). The above-mentioned steps are repeated for each HMR as the values of τ may vary from region to region based on the temporal association of rainfall with large-scale climatic indices. Results are shown in Figure 2. It is noticed that at least four out of five performance statistics, if not all of them, converge to the best performance for a particular value of n in all the regions. The values of τ, identified for each region, are also shown in Figure 2 by a rectangle. These are identified as 5 years for CNE, PE and WC

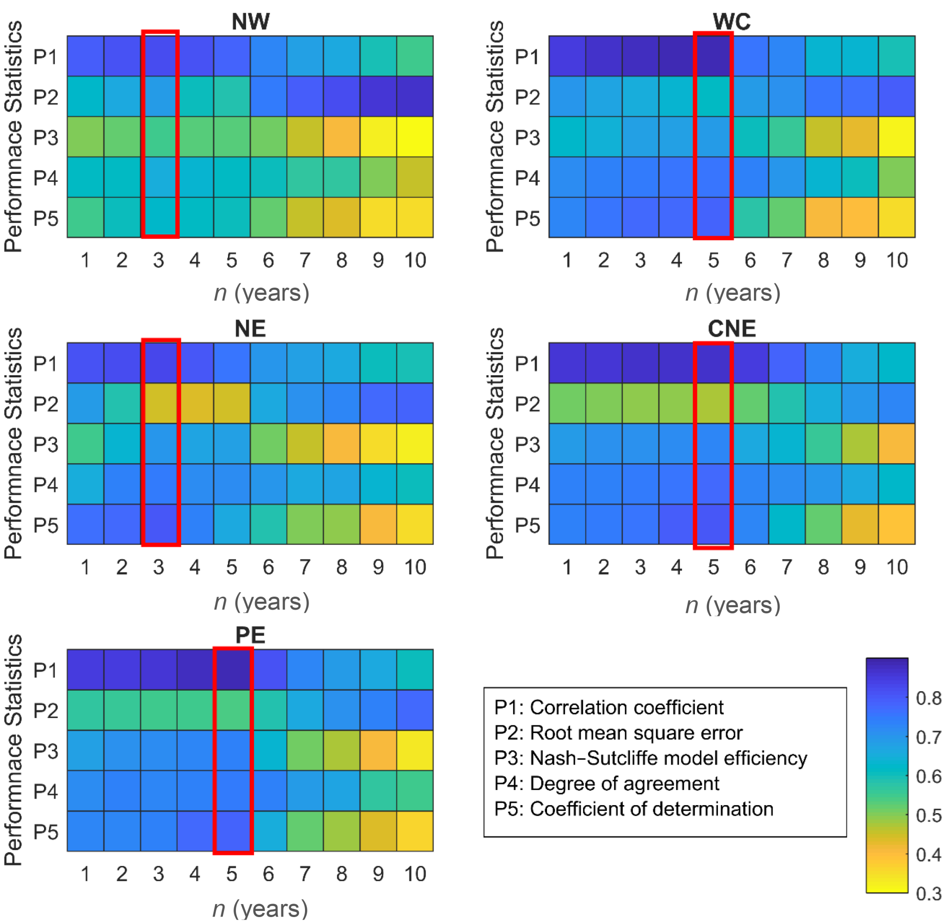

FIGURE 2 Comparison of the results obtained using different values of n to select its optimum value for all the HMRs. The value of n that yields best performance, reflected through most of the performance statistics, is identified as ORI of model recalibration (τ) and highlighted by a rectangle. The value of τ is 3 years for NE and NW and 5 years for CNE, PE and WC [Colour figure can be viewed at wileyonlinelibrary.com]

regions, and 3 years for NW and NE regions. Thus, initial guess of 1–10 years was sufficient to capture the value of τ. It is noted that the regions receiving either higher or lower than average rainfall with high rainfall variability have smaller values of τ. As the rainfall variability is high, smaller values of τ are more effective to capture the temporal variation in the association among the large-scale indices and rainfall. Thereby, the prediction model needs to be updated after every 3 or 5 years, depending on the region, to appropriately capture the time variability of association between the summer monsoon rainfall and individual lags of the climatic indices.

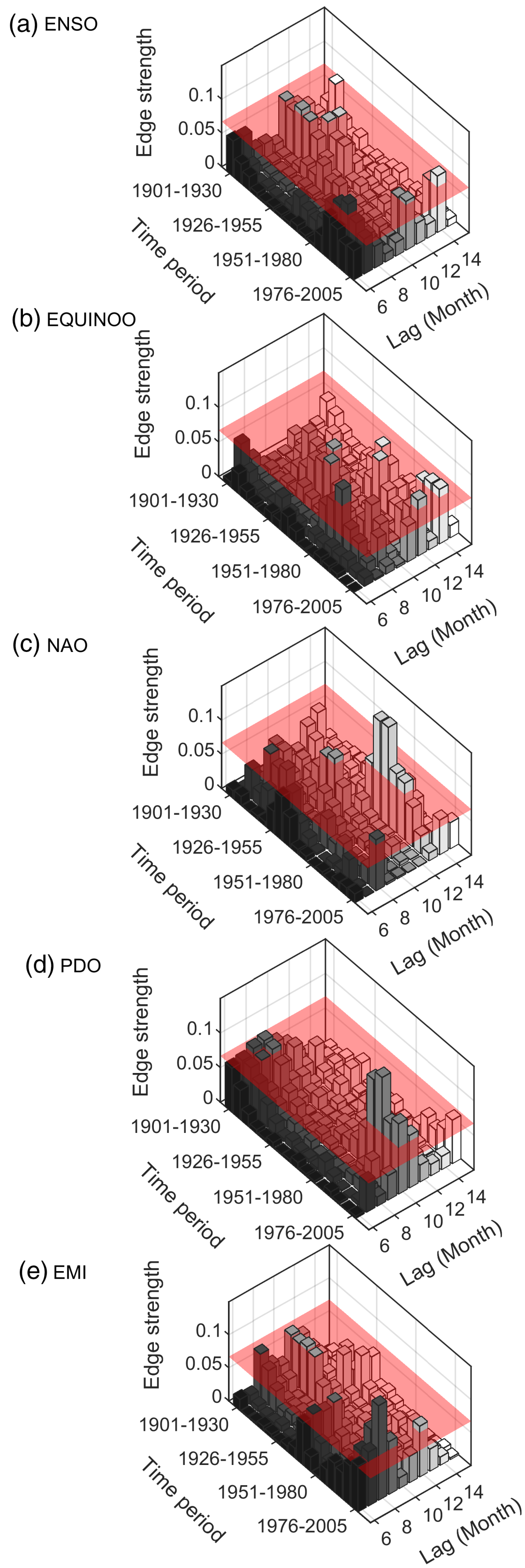

FIGURE 3 Time-varying association between summer monsoon rainfall (target variable) and different large-scale climatic indices with different lags (input variables) for CNE region. The value of τ identified for the region is 5 years based on which the development periods are shown in the x-axis. The threshold value of the edge strength is demarcated using a translucent horizontal surface [Colour figure can be viewed at wileyonlinelibrary.com]

4.2 | Temporal evolution of association among climatic indices and rainfall

In case of 5 years ORI of model recalibration (regions mentioned in Section 4.1), successive model development (testing) periods are 1901–1930 (1931–1935), 1906–1935 (1936–1960), …, 1976–2005 (2006–2010), and in case of 3 years ORI of model recalibration (regions mentioned in Section 4.1), the model development (testing) periods are 1901–1930 (1931–1933), 1904–1933 (1934–1936), …, 1979–2008 (2009–2010). The conditional independence structure obtained for each development period (moving window) provides the time-varying association between the different lags of the climatic indices and summer monsoon rainfall. The degree of association evaluated using the edge strength can be used to examine the degree of association between the variables. The temporal variation in association of the large-scale climatic indices namely ENSO, EQUINOO, EMI, NAO and PDO considering the 6th (December of previous year) to 15th (March of previous year) lags for the CNE region is shown as a typical plot (Figure 3) along with the significance threshold. The association at other lags (1–5 months) is found insignificant for all the regions, thereby excluded from the illustration.

4.2.1 | CNE region

For the CNE region (Figures 3), the 11th, 12th and 13th lags show increasing association starting from the first model development period and shows the highest association around the 1920s. November ENSO (from previous year, that is, 7th lag) shows an increasing strength of association starting from 1950s, till before 1970s. Gradually, the association becomes weaker around the 1970s and 11th and 14th lags appear (edge strength above the threshold value) with very strong association and the edge strength keeps increasing till the last development period. All the other lags of ENSO either show very low or insignificant association for the entire study period. Very similar to ENSO, in case of EQUINOO also the October ENSO (single lag) of the previous year shows significant association during the 1950s. The 10th, 11th, 13th and 14th lags appear and disappear during 1930s to 1970s. Moreover, during the 1970s, the June and April (2 lags) EQUINOO of the previous year appear and the association becomes stronger and stronger. In case of NAO, the 13th lag shows very strong association during the 1950s and gradually disappears (edge strength below the threshold value). The 11th lag shows significant association during the 1930s. Later, during the 1980s, the October NAO of the previous year appears with comparatively weaker association. Considering PDO, the 7th, 8th, 9th and 10th lags show significant association for initial few development periods. The 10th lag again appears during the 1950s and shows a significant association with gradual increase and decrease in the strength. All the other lags of PDO do not show any significant association with the summer monsoon rainfall in the CNE region. Lastly, in case of EMI, the 6th lag shows significant association from 1900s to 1940s and gradually weakens in strength. Again, after 1970s, the 7th, 8th and 10th lags appear and gradually the association becomes strong specifically for the 8th lag. In general, it can be observed that considering this region, almost all the climatic indices show sharp change in the association mostly in terms of the lag associated with summer monsoon rainfall in and around the 1920s, 1950s and 1970s.

4.2.2 | PE region

For the PE region (Figure S1), the June and May ENSO show significant association during the 1910s and gradually disappears. Later, during the 1950s to 1970s, the 14th and 15th lags of ENSO show significant association whereas, during the 1980s the 10th lag appears and the edge strength gradually increases. Considering EQUINOO, the November EQUINOO shows significant association during the 1900s and gradually weakens in strength and reappears during the early 1960s and gradually the strength of association increases. Near the end, the December EQUINOO of the previous year also appears. All the rest of the lags are insignificant. Considering NAO, the 9th, 12th and 15th lags show significant association from the beginning. However, the 9th and 12th lags gradually disappear and the 15th lag shows slight dip in strength during the 1970s and again increasing towards the last development period. In case of PDO, April and May show strong association during the initial time period. Furthermore, the 10th to 13th lags show very strong association during the initial periods and then completely disappear. In case of EMI, the March, April and May show very strong association between 1900s and 1940s. Moreover, similar to EQUINOO, the 6th and 7th lags show very strong association since the 1970s. It is interesting to note that ENSO, EQUINOO and EMI show significant changes around the 1970s. It can also be observed that, ENSO, NAO and PDO show significant changes around the 1950s.

4.2.3 | NW region

For the NW region (Figure S2), most of the lags of ENSO show significant association during the 1960s. All these lags, except the April ENSO of the previous year gradually disappear by the early 1970s. Also, the 9th lag gradually appears with a consistent strength of association. In case of EQUINOO, the April, June and September show appearing and disappearing association during the 1920s to 1960s. The 14th lag appears in the middle of the time period and gradually disappears. By the end, the 11th and 15th lags show strong association. Considering NAO, the 6th lag shows significant association during the 1920s and the 12th lag shows significant association during the 1930s and 1940s. Furthermore, NAO does not show very strong association during the 1950s; however, the edge strength for the 13th lag gradually increases towards the last development period. Lastly considering PDO, the 6th and 13th lags show consistent association before and during the 1960s, gradually disappears and again reappears in the late 1970s along with the 11th lag. Similar to ENSO, few lags of EMI show significant association during the 1960s; however, completely disappear for the rest of the years. It may be observed that for the NW region, major changes can be observed around the 1950s and 1970s.

4.2.4 | WC region

For the WC region (Figure S3), the April ENSO of the previous year appears in 1960s and the edge strength gradually increases till the last development period. The 11th and 12th lags also show significant association in the 1950s and early 1960s. In case of EQUINOO, the 10th, 12th and 15th lags appear and disappear during the period of 1900s to 1950s. The September EQUINOO appears in the late 1960s and gradually increases in strength. Considering NAO, the 9th lag shows significant association 1900s to 1940s and gradually disappears. The November NAO shows a consistent association till the 1980s. After this period the 6th, 8th and 11th lags appear with the highest edge strength of the 11th lag during the last development period. PDO shows stronger association during the initial development periods. Gradually the strength of association decreases and all the lags disappear; however, the 6th and 8th lags appear during the late 1970s. It can be observed that most of the changes occur during the late 1960s. EMI can be considered to have very weak association with the summer monsoon rainfall considering all the lags. The September and April EMI of the previous year show significant but weak association around the late 1970s. Later, in and around the last development period, certain lags of ENSO, EQUINOO and NAO show very strong association with the summer monsoon rainfall unlike the PE region.

4.2.5 | NE region

For the NE region (Figure S4), the 10th lag shows increasing and decreasing association for the time period of 1900s to 1950s. The December and November ENSO consistently shows strong association starting from 1960s with the 6th lag gradually reducing in strength and the 7th lag gradually increasing in strength towards the last development period. EQUINOO, on the other hand shows a strong demarcation, in terms of significant lags, before and after the late 1960s. Before this time period, the 7th, 12th, 13th and 14th lags show significant association, however, after this time period the 10th and 11th lags show strong association. Similar to EQUINOO, the 12th and 14th lags show strong association in the initial time periods and gradually disappear. The 6th lag of NAO shows significant association during the 1900s to 1940s. The March and May NAO (two lags) shows significant change in association around the 1950s. Gradually at the end, the December and August NAO of the previous year show significant association. Considering PDO, the 7th, 8th and 10th lags show very strong association till early 1960s and gradually disappear. In the recent years none of the lags of PDO show any association the summer monsoon rainfall. Considering EMI, none of the lags except the 14th and 15th lags show insignificant with the summer monsoon rainfall, through out the time period of the study. It is interesting to note that for NE region also (similar to WC region), certain lags of ENSO, EQUINOO and NAO show very strong association during the recent years.

In summary, the influence of ENSO, EQUINOO, EMI, NAO and PDO with a particular lag on summer monsoon rainfall varies from one HMR to another. Moreover, time period for which a certain lag shows strong association continuous evolves, emphasizing on the need to study the temporal association among the climatic indices and rainfall.

4.3 | Performance of the time varying GM-Copula model

Based on the identified value of τ, the region-wise predicted rainfall obtained using the time-varying GM-Copula approach is compared with the observed rainfall at the actual scale, for the entire testing period (Figure 4). It is clearly observed that the time-varying model is able to capture the nature of variation and actual magnitude of observed rainfall for all the regions. Specifically, for the regions that receive moderate rainfall during the summer monsoon season namely, CNE, PE and WC, the time-varying model appropriately captures the peaks in the rainfall, for example, above normal rainfall in the year 2008 for the CNE region, above normal rainfall in the year 2007 for the PE region and below normal rainfall in the year 2009 for the WC region, etc. Even for the NE and NW regions the modulation in the rainfall series is well represented by the predicted rainfall. Figure 4 also shows the comparison between mean, range and outliers of the observed (red) and predicted using time-varying model (blue) rainfall for each region. It is observed that the proposed model precisely reproduces the range and mean of actual rainfall for all the regions. HMR wise the mean absolute percentage error between the predicted and observed rainfall are as follows: NW – 10%, WC – 5.5%, NE – 6.2%, CNE – 4.9% and PE – 4%. Thus, the error margins lie within 10% across HMRs within India.

In general, the variation in association suggests that the time-invariant set of predictors may suffer from consistency in performance. To investigate this fact, the performance of the time-varying models and time-invariant model are compared. In case of the time-invariant GM-Copula based model, one model is developed using the first 30-years data (1901–1930) and the developed prediction model is used for the entire testing period (1931–2010). Figure 5 compares the predicted rainfall at standardized anomaly scale obtained using the time-varying model and its time-invariant counterpart with the observed rainfall. Comparative statistics (in actual scale) are shown in Table 1 (Input set #1). The time-varying models are found to yield more accurate results, as compared to the time-invariant model, rightly capturing the nature/behaviour of the recorded anomalous rainfall

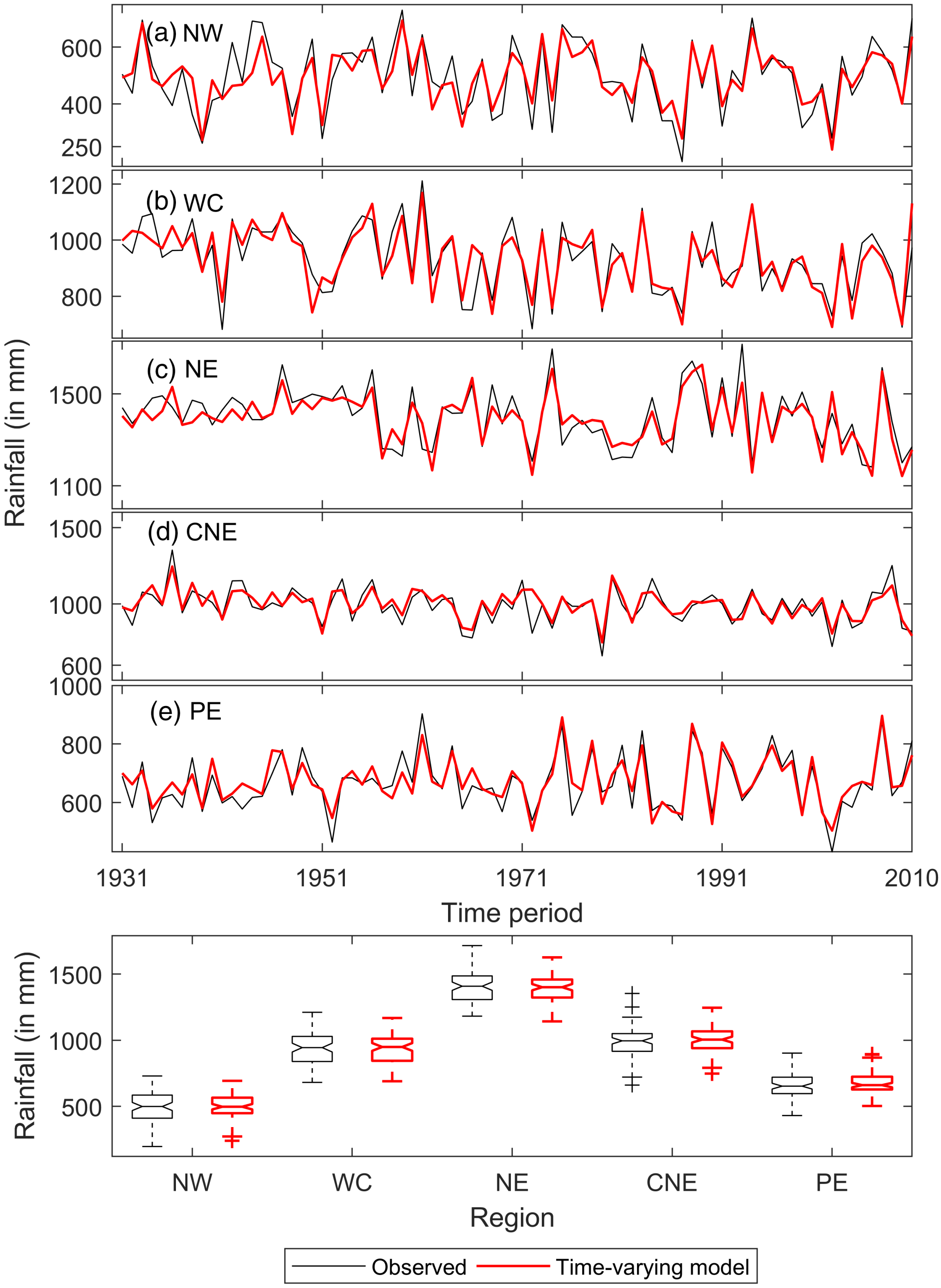

FIGURE 4 Performance of the time-varying GM-Copula model by comparing the observed and predicted region-wise summer monsoon rainfall for the entire testing period (1931–2015). The lowermost panel shows the mean and range of the observed and predicted rainfall through the boxplots for each HMR [Colour figure can be viewed at wileyonlinelibrary.com]

values. For WC, PE, NE, CNE and NW regions, in the years 1981 to 2015 that is, the entire model testing period, the time varying model performs better than the time invariant model. It can be observed that the years of positive as well as negative rainfall anomalies reasonably match in model testing period. It can be reiterated that the ability of the time varying model rely on the consideration of evolving association between the climatic indices and summer monsoon rainfall, which is identified as one of the major issues of long-lead prediction models at regional scale.

Next, it is also worthwhile to quantify the additional contribution of multiple indices, apart from ENSO and EQUINOO only, in the time-varying and time-invariant framework. To investigate this, the performance statistics obtained using the both the approaches with two different sets of input variables: (a) all inputs (Input set #1) and (b) only ENSO and EQUINOO (Input set #2) are given in Table 1. For both the inputs sets, performance of time-varying approach is better than its time-invariant counterparts. In case of time-varying approach, considering Input set #1, the values of correlation coefficient lies between 0.827 and 0.890 and that for Input set #2, the values of correlation coefficient lie between 0.790 and 0.859. Considering other performance statistics also, it can be clearly observed that the Input set #1 provides superior performance as compared to Input set #2, based on the values of the performance statistics. The explained

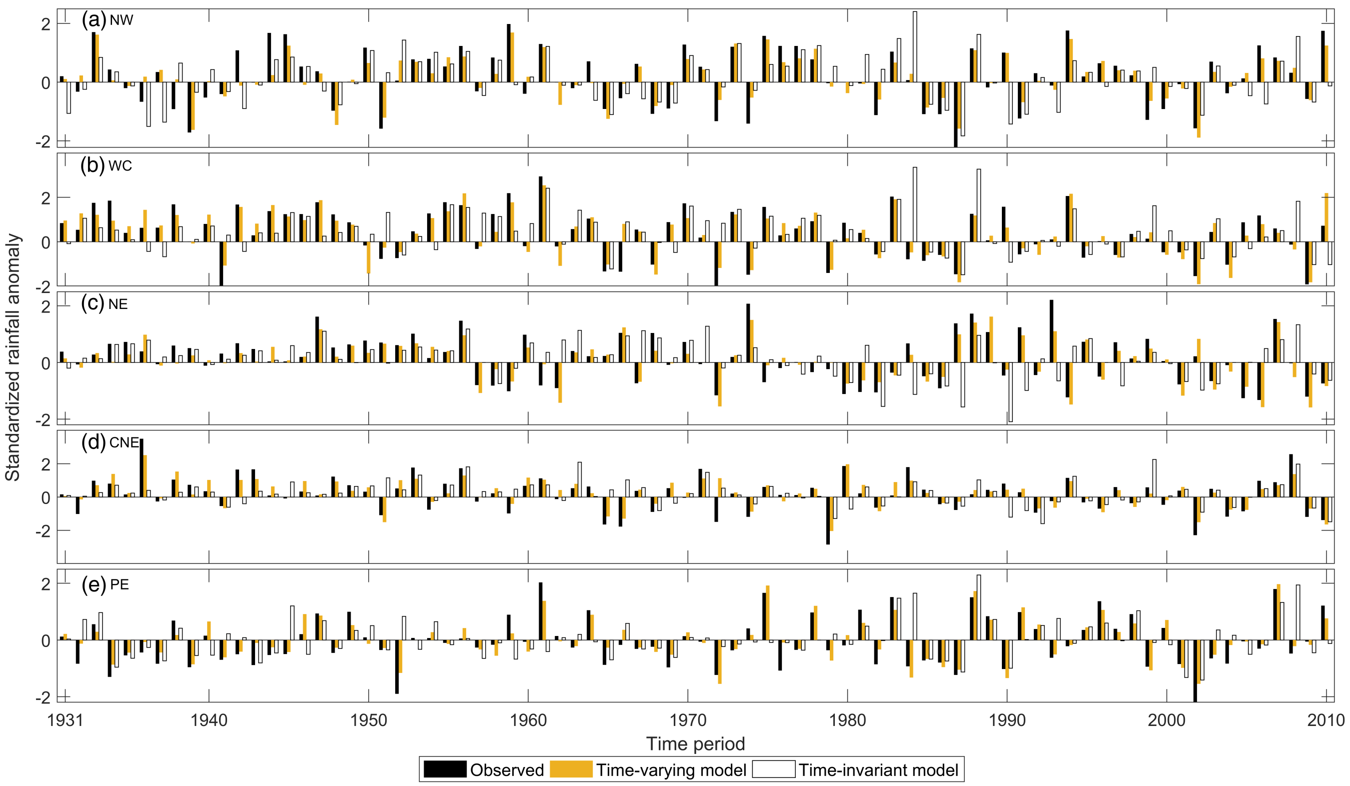

FIGURE 5 Comparison of the observed and predicted rainfall anomaly obtained using the time-varying GM-Copula model and its time-invariant counterpart for different HMRs [Colour figure can be viewed at wileyonlinelibrary.com]

variability (R^2 expressed in percentage) ranges between 62.5% (NW) and 80.6% (CNE) while considering Input set #1, whereas it is between 59.4% (NW) and 74.3% (WC) while considering Input set #2. Specifically, for the CNE, PE and WC regions the performance with Input set #2 deteriorates as the other indices have strong influence on the variability of rainfall in these regions and removing them from the input set is depriving the model of important information. For the WC and NE regions also, though the impact of EMI, NAO and PDO is low, still for certain time periods they provide valuable information, thereby improving the overall performance of the model during the entire testing period. Thereby, spatial variation of predictability is high and inclusion or exclusion of a particular/set of large-scale climatic indices affects the performance of the prediction model to varying degrees based on their association with summer monsoon rainfall.

5 | CONCLUSIONS

The problem of spatial variation in predictability of the summer monsoon rainfall for different HMRs in India has been addressed, through a time-varying GM-Copula model, which uses lagged large-scale climatic indices as the input variables. The conditional independence structure (as obtained using the GM approach) is employed to identify the complex association among the different lags of climatic indices and summer monsoon rainfall. In order to capture the time-variability in the association, the model is recalibrated at regular time intervals. The optimum value of this time interval (ORI of model recalibration) is identified as 3 years for the NW and NE regions and 5 years for the remaining 3 regions. The index of ENSO used for this study (Nino3.4 index) shows temporal shifts in 2–3 and 3–8 years. Moreover, the ENSO–ISMR relationship is also varied in 1.5–3 years. The timescale of the temporal variability in the dependence of the large-scale climatic indices explains the need to update the prediction model after every 3–5 years. It is interesting to note that the input variables and corresponding model structure varies over both space and time on updating the prediction model after 3–5 years' time period. Studying the time-varying association of the 15 lags of the five large-scale climatic indices provides some interesting insight regarding the time periods with sharp changes in the dependencies. For instance, considering the CNE region all the indices show a change in the lag of the climatic index having direct influence with the summer monsoon rainfall around the 1950s and 1970s. Certain indices like EMI, emerge as a very strong contender for the recent years considering

TABLE 1 Performance statistics between the observed and predicted rainfall for the entire model testing period using all inputs (Input set #1) and only ENSO and EQUINOO (Input set #2)

Region	Input set #	Performance statistics R	RMSE (mm)	Dr	NSE	R^2
CNE	1	0.871 (0.414)	47.105 (73.462)	0.721 (0.579)	0.782 (0.344)	0.806 (0.182)
	2	0.815 (0.268)	62.323 (99.146)	0.644 (0.425)	0.613 (0.128)	0.672 (0.189)
PE	1	0.890 (0.401)	41.327 (113. 231)	0.720 (0.379)	0.701 (0.174)	0.782 (0.164)
	2	0.827 (0.365)	87.271 (112.351)	0.546 (0.312)	0.542 (0.042)	0.608 (0.121)
NW	1	0.825 (0.359)	124.586 (139.231)	0.653 (0.467)	0.564 (0.214)	0.625 (0.159)
	2	0.790 (0.210)	110.248 (145.254)	0.547 (0.479)	0.536 (0.146)	0.594 (0.144)
WC	1	0.885 (0.373)	65.608 (82.168)	0.708 (0.497)	0.692 (0.129)	0.779 (0.103)
	2	0.859 (0.295)	65.739 (93.146)	0.612 (0.452)	0.664 (0.220)	0.743 (0.134)
NE	1	0.827 (0.356)	63.802 (96.344)	0.743 (0.510)	0.702 (0.161)	0.802 (0.128)
	2	0.798 (0.276)	72.27 (115.251)	0.771 (0.426)	0.672 (0.205)	0.645 (0.127)

Note: For the both the cases, performance of both time-varying and time-invariant (within parentheses) approaches are shown. Input set #1, that is, all inputs, consist of ENSO, EQUINOO, EMI, NAO and PDO.

the PE region. Similarly, for the NW region NAO shows very strong association in the recent years. For both WC and NE regions, NAO, ENSO and EQUINOO show very strong association in the 2000s.

The results indicate that the time varying GM-Copula model very well captures the positive and negative anomalies in the observed rainfall for all the regions. The mean and range in the region-wise rainfall is also well captured. The mean absolute percentage errors of prediction using the time-varying GM-Copula model are within 4–10% across HMRs. Moreover, when compared with its time-invariant counterpart, it shows superior performance due to the inability of the time-invariant model to capture the dynamic association. Based on the lags of the climatic indices identified as the potential predictors for each model development period, the long-lead prediction can be made 5 months in advance for all the regions. ENSO and EQUINOO are the dominant indices considering all the regions, as considering only these two indices also provides satisfactory prediction performance. However, considering EMI, NAO and PDO strongly improves the performance of the prediction model for the CNE, PE and NW regions. The time-varying model developed at finer spatial scale successfully captures the temporal variability in association and the spatial variation in predictability of the summer monsoon rainfall. The long-lead prediction using the time-varying model at finer spatial scale enables the hydrologists and policy makers to establish effective management measures. Such prediction for the HMRs is especially effective for the agricultural community.

ACKNOWLEDGEMENT

This work is partially 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.

ORCID

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

REFERENCES

Aas K, Czado C, Frigessi A, Bakken H. 2009. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*. Elsevier B.V., 44(2): 182–198. https://doi.org/10.1016/j.insmatheco.2007.02.001.

Antico, P.L. and Barros, V.R. (2017) Changes in the zonal propagation of El Niño-related SST anomalies: A possible link to the PDO. *Theoretical and Applied Climatology*, 129(1–2), 171–176. https://doi.org/10.1007/s00704-016-1766-1.

Ashok, K., Behera, S.K., Rao, S.A., Weng, H. and Yamagata, T. (2007) El Niño Modoki and its possible teleconnection. *Journal* *of Geophysical Research: Oceans*, 112(11), 1–27. https://doi.org/10.1029/2006JC003798.

Ashok, K., Guan, Z., Saji, N.H. and Yamagata, T. (2004) Individual and combined influences of ENSO and the Indian Ocean dipole on the Indian summer monsoon. *Journal of Climate*, 17(16), 3141–3155. https://doi.org/10.1175/1520-0442(2004)017<3141:IACIOE>2.0.CO;2.

Ashok, K., Guan, Z. and Yamagata, T. (2001) Impact of the Indian Ocean dipole on the relationship between the Indian Monsoon Rainfall and ENSO. *Geophysical Research Letters*, 28(23), 4499–4502. https://doi.org/10.1029/2001GL013294.

Ashok, K. and Saji, N.H. (2007) On the impacts of ENSO and Indian Ocean dipole events on sub-regional Indian summer monsoon rainfall. *Natural Hazards*, 42(2), 273–285. https://doi.org/10.1007/s11069-006-9091-0.

Ault, T.R., Cole, J.E., Overpeck, J.T., Pederson, G.T., George, S.S., Otto-Bliesner, B., Woodhouse, C.A. and Deser, C. (2013) The continuum of hydroclimate variability in Western North America during the last millennium. *Journal of Climate*, 26(16), 5863–5878. https://doi.org/10.1175/JCLI-D-11-00732.1.

Bauer, A., Czado, C., Klein, T., Bauer, A., Czado, C. and Klein, T. (2012) Pair-copula constructions for non-Gaussian DAG models. *The Canadian Journal of Statistics*, 40(1), 86–109. https://doi.org/10.1002/cjs.10131.

Box, G.E. and Cox, D.R. (1964) An analysis of transformations. *Journal of the Royal Statistical Society*, 26(2), 211–252.

Brechmann, E. and Schepsmeier, U. (2013) Modeling Dependence with C-and D-Vine Copulas: The R Package CDVine. *Journal of Statistical Software*, 52(3), 1–27.

Brown, J.R., Hope, P., Gergis, J. and Henley, B.J. (2016) ENSO teleconnections with Australian rainfall in coupled model simulations of the last millennium. *Climate Dynamics*, 47(1–2), 79–93. https://doi.org/10.1007/s00382-015-2824-6.

Charlotte, B.V. and Mathew, B. (2012) EQUINOO: The entity and validity of this oscillation to Indian monsoon, 1(11), 45–54.

Chen, W., Dong, B. and Lu, R. (2010) Impact of the Atlantic Ocean on the multidecadal fluctuation of El Nio-Southern Oscillation–South Asian monsoon relationship in a coupled general circulation model. *Journal of Geophysical Research Atmospheres*, 115(17), 1–12. https://doi.org/10.1029/2009JD013596.

Chowdary, J.S., Xie, S.P., Tokinaga, H., Okumura, Y.M., Kubota, H., Johnson, N. and Zheng, X.T. (2012) Interdecadal variations in ENSO teleconnection to the Indo-Western Pacific for 1870–2007. *Journal of Climate*, 25(5), 1722–1744. https://doi.org/10.1175/JCLI-D-11-00070.1.

Dalla Valle, L., De Giuli, M.E., Tarantola, C. and Manelli, C. (2016) Default probability estimation via pair copula constructions. *European Journal of Operational Research*, 249(1), 298–311. https://doi.org/10.1016/j.ejor.2015.08.026.

Deser, C., and Trenberth, K., (Eds). (2016). National Center for Atmospheric Research Staff. The Climate Data Guide: Pacific Decadal Oscillation (PDO): Definition and Indices. Retrieved from https://climatedataguide.ucar.edu/climate-data/pacific-decadal-oscillation-pdo-definition-and-indices.

Dugam, S.S., Kakade, S.B. and Verma, R.K. (1997) Interannual and long-term variability in the North Atlantic Oscillation and Indian Summer monsoon rainfall. *Theoretical and Applied Climatology*, 58(1–2), 21–29. https://doi.org/10.1007/BF00867429.

Dutta, R. and Maity, R. (2018) Temporal evolution of hydroclimatic teleconnection and a time-varying model for long-lead prediction of Indian summer monsoon rainfall. *Scientific Reports*, 8(1), 10778. https://doi.org/10.1038/s41598-018-28972-z.

Fan, F., Dong, X., Fang, X., Xue, F., Zheng, F. and Zhu, J. (2017) Revisiting the relationship between the South Asian summer monsoon drought and El Niño warming pattern. *Atmospheric Science Letters*, 18(4), 175–182. https://doi.org/10.1002/asl.740.

Francis, P.A. and Gadgil, S. (2010) Towards understanding the unusual Indian monsoon in 2009. *Journal of Earth System Science*, 119(4), 397–415. https://doi.org/10.1007/s12040-010-0033-6.

Gadgil, S. and Gadgil, S. (2006) The Indian monsoon, GDP and agriculture. *Economic & Political Weekly*, 41, 4887–4895. https://doi.org/10.2307/4418949.

Gadgil, S. and Srinivasan, J. (2012) Monsoon prediction: are dynamical models getting better than statistical models? *Current Science*, 103(3), 257–259.

Gadgil, S., Vinayachandran, P.N., Francis, P.A. and Gadgil, S. (2004) Extremes of the Indian summer monsoon rainfall, ENSO and Equatorial Indian Ocean oscillation. *Geophysical Research Letters*, 31(12), 2–5. https://doi.org/10.1029/2004GL019733.

Gómez, M., Concepción Ausín, M. and Carmen, D.M. (2017) Seasonal copula models for the analysis of glacier discharge at King George Island, Antarctica. *Stochastic Environmental Research and Risk Assessment*, 31(5), 1107–1121. https://doi.org/10.1007/s00477-016-1217-7.

Goswami, B.N., Madhusoodanan, M.S., Neema, C.P. and Sengupta, D. (2006) A physical mechanism for North Atlantic SST influence on the Indian summer monsoon. *Geophysical Research Letters*, 33(2), 1–4. https://doi.org/10.1029/2005GL024803.

Guhathakurta, P., Rajeevan, M., Sikka, D.R. and Tyagi, A. (2015) Observed changes in southwest monsoon rainfall over India during 1901–2011. *International Journal of Climatology*, 35(8), 1881–1898. https://doi.org/10.1002/joc.4095.

Hanna, E., Cropper, T.E., Jones, P.D., Scaife, A.A. and Allan, R. (2015) Recent seasonal asymmetric changes in the NAO (a marked summer decline and increased winter variability) and associated changes in the AO and Greenland Blocking Index. *International Journal of Climatology*, 35(9), 2540–2554. https://doi.org/10.1002/joc.4157.

Hope, P., Henley, B.J., Gergis, J., Brown, J. and Ye, H. (2017) Time-varying spectral characteristics of ENSO over the last millennium. *Climate Dynamics*, 49(5–6), 1705–1727. https://doi.org/10.1007/s00382-016-3393-z.

Hurrell, J. (Eds). (2018) National Center for Atmospheric Research Staff. The Climate Data Guide: Hurrell North Atlantic Oscillation (NAO) Index (station-based). Retrieved from https://climatedataguide.ucar.edu/climate-data/hurrell-north-atlantic-oscillation-nao-index-station-based.

Ihler, A.T., Kirshner, S., Ghil, M., Robertson, A.W. and Smyth, P. (2007) Graphical models for statistical inference and data assimilation. *Physica D: Nonlinear Phenomena*, 230(1–2), 72–87. https://doi.org/10.1016/j.physd.2006.08.023.

Jordan, M.I. (2004) Graphical models. *Statistical Science*, 19(1), 140–155. https://doi.org/10.1214/088342304000000026.

Ju, J. and Slingo, J. (1995) The Asian summer monsoon and ENSO. *Quarterly Journal of the Royal Meteorological Society*, 121(525), 1133–1168. https://doi.org/10.1002/qj.49712152509.

Kahya E, Dracup J a. 1993. U.S. streamflow patterns in relation to the E1 Niño/Southern Oscillation, 29(8): 2491–2503.

Kane, R.P. (2006) Unstable ENSO relationship with Indian regional rainfall. *International Journal of Climatology*, 26(6), 771–783. https://doi.org/10.1002/joc.1281.

Kashid, S.S. and Maity, R. (2012) Prediction of monthly rainfall on homogeneous monsoon regions of India based on large scale circulation patterns using genetic programming. *Journal of Hydrology*, 454–455, 26–41. https://doi.org/10.1016/j.jhydrol.2012.05.033.

Krishnamurthy, L. and Krishnamurthy, V. (2014) Influence of PDO on South Asian summer monsoon and monsoon-ENSO relation. *Climate Dynamics*, 42(9–10), 2397–2410. https://doi.org/10.1007/s00382-013-1856-z.

Kucharski, F., Bracco, A., Yoo, J.H., Tompkins, A.M., Feudale, L., Ruti, P. and Dell'Aquila, A. (2009) A Gill-Matsuno-type mechanism explains the tropical Atlantic influence on African and Indian monsoon rainfall. *Quarterly Journal of the Royal Meteorological Society*, 135(640), 569–579. https://doi.org/10.1002/qj.406.

Kumar, D.N., Reddy, M.J. and Maity, R. (2007) Regional rainfall forecasting using large scale climate teleconnections and artificial intelligence techniques. *Journal of Intelligent Systems*, 16 (4), 307–322. https://doi.org/10.1515/JISYS.2007.16.4.307.

Kumar, K.K., Rajagopalan, B. and Cane, M.A. (1999) On the weakening relationship between the Indian monsoon and ENSO. *Science*, 284(5423), 2156–2159. https://doi.org/10.1126/science.284.5423.2156.

Kumar, K.K., Rajagopalan, B., Hoerling, M., Bates, G. and Cane, M. (2006) Unraveling the mystery of Indian monsoon failure during El Nino. *Science*, 314(5796), 115–119. https://doi.org/10.1126/science.1131152.

Lewis, S.C. and Legrande, A.N. (2015) Stability of ENSO and its tropical Pacific teleconnections over the last millennium. *Climate of the Past*, 11(10), 1347–1360. https://doi.org/10.5194/cp-11-1347-2015.

Liu, Z., Zhou, P., Chen, X. and Guan, Y. (2015) A multivariate conditional model for streamflow prediction and spatial precipitation refinement. *Journal of Geophysical Research: Atmospheres*, 120(19), 10,116–10,129. https://doi.org/10.1002/2015JD023787.

Lu, R., Dong, B. and Ding, H. (2006) Impact of the Atlantic Multidecadal Oscillation on the Asian summer monsoon. *Geophysical Research Letters*, 33(24), 1–5. https://doi.org/10.1029/2006GL027655.

Maity, R. and Nagesh, K.D. (2006) Bayesian dynamic modeling for monthly Indian summer monsoon rainfall using El Niño–Southern Oscillation (ENSO) and Equatorial Indian Ocean Oscillation (EQUINOO). *Journal of Geophysical Research*, 111 (D7), D07104. https://doi.org/10.1029/2005JD006539.

Maity, R. and Nagesh, K.D. (2008) Probabilistic prediction of hydroclimatic variables with nonparametric quantification of uncertainty. *Journal of Geophysical Research Atmospheres*, 113(14), 1–12. https://doi.org/10.1029/2008JD009856.

Mantua, N.J. and Hare, S.R. (2002) The Pacific decadal oscillation. *Journal of Oceanography*, 58, 35–44. https://doi.org/10.1023/A:1015820616384.

Mantua, N.J., Hare, S.R., Zhang, Y., Wallace, J.M. and Francis, R.C. (1997) A Pacific interdecadal climate oscillation with impacts on Salmon production. *Bulletin of the American Meteorological Society*, 78(6), 1069–1079. https://doi.org/10.1175/1520-0477(1997)078<1069:apicow>2.0.co;2.

Meehl, G.A. and Arblaster, J.M. (1998) The Asian-Australian monsoon and El Nino-Southern Oscillation in the NCAR climate system model. *Journal of Climate*, 11(6), 1356–1385. https://doi.org/10.1175/1520-0442(1998)011<1356:TAAMAE>2.0.CO;2.

Nair, P.J., Chakraborty, A., Varikoden, H., Francis, P.A. and Kuttippurath, J. (2018) The local and global climate forcings induced inhomogeneity of Indian rainfall. *Scientific Reports*, 8 (1), 1–12. https://doi.org/10.1038/s41598-018-24021-x.

Pant, G.B. and Parthasarathy, S.B. (1981) Some aspects of an association between the southern oscillation and Indian summer monsoon. *Archives for Meteorology, Geophysics, and Bioclimatology Series B*, 29(3), 245–252. https://doi.org/10.1007/BF02263246.

Parthasarathy, B., Rupa Kumar, K. and Munot, A.A. (1993) Homogeneous Indian monsoon rainfall: variability and prediction. *Proceedings of the Indian Academy of Sciences - Earth and Planetary Sciences*, 102(1), 121–155. https://doi.org/10.1007/BF02839187.

Pattanaik, D.R., Sahai, A.K., Mandal, R., Phani Muralikrishna, R., Dey, A., Chattopadhyay, R., Joseph, S., Tiwari, A.D. and Mishra, V. (2019) Evolution of operational extended range forecast system of IMD: Prospects of its applications in different sectors. *Mausam*, 70(2), 233–264.

Phatak, A., Bates, B.C. and Charles, S.P. (2011) Statistical downscaling of rainfall data using sparse variable selection methods. *Environmental Modelling and Software*, 26(11), 1363–1371. https://doi.org/10.1016/j.envsoft.2011.05.007.

Rajeevan, M., Pai, D.S., Anil Kumar, R. and Lal, B. (2007) New statistical models for long-range forecasting of southwest monsoon rainfall over India. *Climate Dynamics*, 28(7–8), 813–828. https://doi.org/10.1007/s00382-006-0197-6.

Rasmusson, E.M. and Carpenter, T.H. (1983) The relationship between eastern equatorial Pacific Sea surface temperatures and rainfall over India and Sri Lanka. *Monthly Weather Review*, 111(3), 517–528. https://doi.org/10.1175/1520-0493(1983)111<0517:TRBEEP>2.0.CO;2.

Righi, M.B., Schlender, S.G. and Ceretta, P.S. (2015) Pair copula constructions to determine the dependence structure of Treasury bond yields. *IIMB Management Review*, 27(4), 216–227. https://doi.org/10.1016/j.iimb.2015.10.008.

Sang, Y.F., Singh, V.P. and Xu, K. (2018) Evolution of IOD-ENSO relationship at multiple time scales. *Theoretical and Applied Climatology*, 136, 1303–1309. https://doi.org/10.1007/s00704-018-2557-7.

Schepsmeier, U., Stoeber, J., Christian, E., Graeler, B., Nagler, T., Erhardt, T., Almeida, C., Min, A., Czado, C., Hofmann, M., Killiches, M. and Joe, H. (2017) *Package 'VineCopula' R topics documented: Version: 2.1.3*.

Sen, R.S. (2011) Identification of periodicity in the relationship between PDO, El Niño and peak monsoon rainfall in India using S-transform analysis. *International Journal of Climatology*, 31(10), 1507–1517. https://doi.org/10.1002/joc.2172.

Singh, A., Kulkarni, M.A., Mohanty, U.C., Kar, S.C., Robertson, A.W. and Mishra, G. (2012) Prediction of Indian summer monsoon rainfall (ISMR) using canonical correlation analysis of global circulation model products. *Meteorological Applications*, 19(2), 179–188. https://doi.org/10.1002/met.1333.

Vathsala, H. and Koolagudi, S.G. (2017) Long-range prediction of Indian summer monsoon rainfall using data mining and statistical approaches. *Theoretical and Applied Climatology*, 130 (1–2), 19–33. https://doi.org/10.1007/s00704-016-1862-2.

Wang, B., Ding, Q. and Joseph, P.V. (2009) Objective definition of the Indian summer monsoon onset. *Journal of Climate*, 22(12), 3303–3316. https://doi.org/10.1175/2008JCLI2675.1.

Whittaker, J. (2009) *Graphical Models in Applied Multivariate Statistics*. Hoboken, NJ: Wiley Publishing.

Xiao, H. (2011) Pair-copula construction for non-Gaussian graphical models. *The Canadian Journal of Statistics*, 40(1), 86–109.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Dutta R, Maity R. Spatial variation in long-lead predictability of summer monsoon rainfall using a time-varying model and global climatic indices. *Int J Climatol*. 2020;40:5925–5940. https://doi.org/10.1002/joc.6556

APPENDIX: | MATHEMATICAL DETAILS FOR DEVELOPMENT OF GM-COPULA MODEL

Development of the GM-Copula model involves two important aspects. The first one is the selection of the potential predictors from the large pool of influencing variables (15 lags of the five large-scale indices) using GM approach. GM helps to identify a complete conditional independence structure among all the variables (predictors and predictand). The second one is development of the prediction model using copula based approach. The variables that are directly influencing the target variables (parent variables), as identified by the conditional independence structure, are considered for development of the prediction model using vine copula. Copulas help to obtain the joint distribution and then conditional distribution of the target variables (summer monsoon rainfall), conditioned on the parent variables. These models are updated after regular intervals to capture the time-varying association among the large-scale indices and summer monsoon rainfall. Mathematical details are explained in the following subsections.

Selection of predictor using GM

The conditional independence among the input and target variables is revealed through a graph – a mathematical object, denoted by $G = (V, E)$, where V is a set of vertices or nodes (representing the variables) and E is a set of edges (representing the association among the variables). The identification of the conditional independence structure among the input variables and target variable is determined using the maximum likelihood approach (Whittaker, 2009). For application of this approach the data should follow normal distribution, else it can be transformed using some transformation methodology (e.g., Box and Cox transformation; Box and Cox, 1964). In the maximum likelihood approach, initially a fully interconnected graph structure, also referred to as a saturated model, is considered where all pairs nodes are connected. Next, the edge exclusion deviance (EED) is used to test if an edge can be eliminated from the saturated model (Whittaker, 2009) depending on its statistical significance. EED is computed as follows:

$$\mathrm{EED} = -N\log\left(1-\mathrm{corr}_N^2\left(X_i,X_j|\mathrm{rest}\right)\right), \qquad (1)$$

where N is the size of the sample and $\mathrm{corr}_N^2\left(X_i,X_j|\mathrm{rest}\right)$ is the partial correlation coefficient between any two random variables X_i and X_j given the rest. The statistic EED follows a chi-squared distribution with one degree of freedom as one edge is removed at a time. Thus, the threshold value of EED is 3.84 at 5% significance level and the edge with the lowest EED is removed if it is less than this threshold value. This is an iterative process which continues till all possible edges (EED less than the threshold value) are removed to obtain the final graph structure. To check the acceptability of the obtained graph structure at a particular confidence level, deviance of the obtained graph structure is evaluated. The generalized likelihood ratio test statistics, evaluated based on the observed sample variance and the estimated variance obtained from the independence structure is known as the deviance of the model. The deviance (Dv) is evaluated as follows (Whittaker, 2009):

$$\mathrm{Dv} = N\left\{tr\left(S\hat{V}^{-1}\right)-\log\det\left(S\hat{V}^{-1}\right)-K\right\}, \qquad (2)$$

where S is the variance matrix, $\hat{V}$ is the estimated variance matrix evaluated based on the number of edges removed from the model, K is the total number of variables and N is as stated before. The deviance (Dv) follows an approximate chi-squared distribution with d degrees of freedom (where, d is the number of edges excluded from the saturated graph). Thus, p value of the test statistics can be computed as $P\left(\chi_p^2>\mathrm{Dv}\right)$. For this study the acceptable significance level is fixed at .05, that is, the obtained conditional independence structure is acceptable if the p value is higher than .05. In case the structure fails to meet the acceptability criteria, structure is to be modified with re-adding edges one by one that were previously removed. The final graph structure is synonymously referred to as conditional independence structure also.

Although the conditional independence structure helps to identify the potential predictors, time variability of their potential is another important aspect. This is quantified through a statistical measure known as edge strength. Thus, the surviving edges of the conditional independence structure are investigated for their strength of association, also known as edge strength. The edge strength between two nodes (for a surviving edge) in the conditional independence structure can be calculated as follows (Whittaker, 2009):

$$\mathrm{Inf}\left(X_i \coprod X_j|\mathrm{rest}\right) = -\frac{1}{2}\log\left(1-\mathrm{corr}_N{}^2\left(X_i,X_j|\mathrm{rest}\right)\right), \quad (3)$$

where $\mathrm{Inf}(X_i \coprod X_j|\mathrm{rest})$ is the edge strength between X_i and X_j given rest. This is also known as divergence against conditional independence (Whittaker, 2009). This information on edge strength is used to investigate the temporal evolution of association of a particular input with the summer monsoon rainfall.

Development of the prediction model using copula

The probabilistic model is developed using the selected predictors identified through the conditional independence structure. Even after obtaining the structure, there could be multiple predictors directly associated with the target variable (summer monsoon rainfall). Multivariate copulas, like nested copula or vine copula are the best choice to develop a multivariate probabilistic model. Among different alternatives in vine copulas, canonical vine (C-Vine) is used in this study to develop the probabilistic model. C-Vine copulas are used for prediction in many studies of a variable by a sequence of trees (Xiao, 2011; Bauer *et al.*, 2012; Liu *et al.*, 2015; Righi *et al.*, 2015; Dalla Valle *et al.*, 2016). These trees are referred as C-Vines and the corresponding multivariate distribution is called C-Vine distribution. For a D-dimensional C-Vine (considering $D-1$ number of predictors are selected based on the conditional independence structure), the first tree identifies $(D-1)$ pairs of variables whose distribution is modelled directly, utilizing the random variables. The second tree identifies $(D-2)$ pairs of variables whose distribution is conditional on a single variable evaluated by pair copula. This tree uses transformed variables based on the structure of the preceding tree. Proceeding, in this manner the final tree, determines a single pair of variables conditional on the remaining variables. The analysis using C-Vine includes identifying the trees, its pair copula families and estimating their parameters.

Selection of each tree is based on a maximum spanning tree algorithm, where edge weights are chosen to reflect the dependencies. In this case, the absolute value of the empirical Kendall's tau ($\hat{\tau}_{i,j}$) (evaluated for two adjoining variables of the tree X_i and X_j) is utilized as the edge weight and optimization problem is solved ($\max \sum_{\mathrm{edges}\, e_{ij} \in \mathrm{in\ spanning\ tree}} \left|\hat{\tau}_{ij}\right|$, where a spanning tree is a tree on all nodes) for each tree (Schepsmeier *et al.*, 2017). Evaluation of the transformed variables (for selection of the subsequent trees after the first tree) requires estimation of the pair copula families and parameter estimation based on the conditioning variables. Considering X_D as the target variable and $X_1, X_2, ..., X_{D-1}$ as the conditioning variables (predictors), the conditional distribution can be developed for a $D-1$ dimensional vector $\boldsymbol{V} = (X_1, X_2, ..., X_{D-1})$ by applying the following recursive relationship:

$$F(X_D/\boldsymbol{V}) = \frac{\partial C_{X,\boldsymbol{V}_j/\boldsymbol{V}_{-j}}\left\{F\left(X_D/\boldsymbol{V}_{-j}\right), F\left(\boldsymbol{V}_j/\boldsymbol{V}_{-j}\right)\right\}}{\partial F\left(\boldsymbol{V}_j/\boldsymbol{V}_{-j}\right)}, \quad (4)$$

where $\boldsymbol{V}_j (j = 1, 2, ..., D-1)$ is an arbitrary component of $\boldsymbol{V}$, and $\boldsymbol{V}_{-j} = (X_1, X_2, ..., X_{j-1}, X_{j+1}, ..., X_{D-1})$ denotes the vector $\boldsymbol{V}$ excluding element $\boldsymbol{V}_j$. The bivariate copula function is specified by $C_{X,\boldsymbol{V}_j/\boldsymbol{V}_{-j}}$. The final tree can be utilized to evaluate the conditional dependence for the prediction of the target variable given the input variables using the above equation.