Climate Network Based Index Discovery for Prediction of Indian Monsoon

Moumita Saha$^{(\boxtimes)}$ and Pabitra Mitra

Department of Computer Science and Engineering,
Indian Institute of Technology Kharagpur,
Kharagpur, India
{moumitasaha.2012,pabitra}@gmail.com

Abstract. Identification of climatic indices are vital in essence of their ability to characterize different climatic events. We focus on discovery of climatic indices important for Indian summer monsoon from climatic parameters surface pressure and zonal wind velocity. We use climatic network based community detection approach for discovery of climatic indices. New indices depict better correlation with monsoon than existing indices. Regression and non-linear models are designed using newly discovered climatic indices for prediction of Indian summer monsoon. Models show superior accuracy to existing state of art models.

Keywords: Climatic network · Community detection · Climatic indices · Indian monsoon prediction

1   Introduction

Mechanism behind climatic process is complex. Identification and analysis of different patterns in global climatic system is vital in understanding its intricate nature. The state and dynamics of climatic process are explained by different climatic indices. Climatic indices are based on climatic parameters like sea surface temperature (SST), sea level pressure (SLP), wind velocity, surface pressure (SP), that elucidate specific climatic change. Climatic indices are important for their ability to predict different climatic events. Prediction of Indian summer monsoon rainfall ($ISMR$) is challenging due to its dynamic nature. It is important for economic development of agricultural land like India.

Building and analysis of climatic networks in Earth Sciences is one of the emerging topic with immense future scopes. Complex networks have been widely used in building climatic networks and finding out interesting patterns and interconnections present in the climatic system [1]. Steinhaeuser et al. [2] have proposed use of complex networks in descriptive analysis and predictive modelling of climatic events. Donges et al. [3] have revealed the important internal structure present in the climatic network build upon surface air temperature data and uncover a pattern related to global surface ocean currents. Steinhaeuser et al. [4] have detected community in climatic system, given a climatological

interpretation of the communities and applied the model for discovery of new climatic indices.

Climatic index discovery assists in visualizing different aspects of climatic system. Clustering approaches are used in discovery of climatic indices. Sap and Awan [5] have used kernel k-means algorithm with spatial constraint to identify the spatio-temporal patterns in the system. Similar nearest neighbours-based clustering approach is used for detection of novel climatic indices, which are validated against known climatic indices and are shown to overcome limitations of *PCA* and *SVD* approaches [6].

The purpose of our work is two folds– (i) discovery of new climatic indices using climatic network based community detection approach from climatic parameters surface pressure and zonal wind velocity, (ii) utilization of discovered climatic indices as predictors for forecasting Indian summer monsoon rainfall, which acts as validation of our proposed index discovery approach. In our work, climatic networks are formed considering each spatial grid point as a node in the network with time series of climatic parameter in the grid. We use normalized euclidean distance to create weighted edges between the nodes. Three important community detection algorithms are applied for invention of different climatic regions that are significant. Community detection performs better than the traditional clustering method as unlike clustering approach, it also focusses on the structure of the network along with the node attributes. Correlation value between time series of node and Indian monsoon is also included as a node attribute to assists in detecting communities important for prediction of monsoon. The communities found after proper thresholding are shown to be good predictors of Indian monsoon. The discovered climatic indices are compared with established climatic indices of Indian monsoon for validation and they are shown to be more correlated to Indian monsoon than the present climatic indices. Finally, different linear and non-linear models are designed with the newly invented climatic indices as input parameters to predict monsoon. The discovered climatic indices show their imprint and ascertain their superiority in prediction of Indian summer monsoon rainfall.

2 Climatic Network Formation

Climatic networks are built based on two different climatic parameters, namely, surface pressure and zonal wind velocity. Each spatial grid points over the world is considered as a node in the network. Our network consist of *10,512* nodes. Each node is characterized by its corresponding latitude, longitude, climatic parameter time series values over the temporal scale, and scalar correlation value between the climatic parameter time series and Indian monsoon time series at best lead month. Weighted edges are added studying the strength of bonding between each pair of nodes in the network with normalized euclidean distance measure. Top one percent and five percent edges are considered for networks built for climatic parameters surface pressure (*NET_SP*), and zonal wind velocity (*NET_ZW*), respectively. Finally, isolated nodes are removed from the networks to obtain

connected networks. *NET_SP* has *1,999* nodes and *23,326* edges, and *NET_ZW* has *4,922* nodes and *6,851* edges.

3   Community Detection and Index Discovery

Three important community detection algorithms, namely, *infomap (Info)*, *walktrap (Wlktrp)*, and *fastgreedy (Fstgrdy)* are applied on the climatic network to detect communities over the world which will correspond to discovery of novel climatic indices important for prediction of *ISMR*. We have chosen these algorithms guided by requirements as following– (i) ability to utilize edge weights, (ii) suitability for dense networks, (iii) overall computational efficiency, and (iv) inclusion of node weights (in case of *info-map* community detection method).

Info-map Community Detection (Info): The algorithm is based on an information theoretic approach, which use the probability flow of random walks on a network and decompose the network into modules by compressing a description of the probability flow [7]. It discovers community structure in weighted and directed networks, taking into account the node values, weighted edges, and network structure.

Walk-trap Community Detection (Wlktrp): The algorithm employs the concept of random walks through the network for community detection. A node similarity measure based on short walks is used for community detection via hierarchical agglomeration, considering the edge weight and structure of the network. It is efficient in terms of time and space complexity [8].

Fast-greedy Community Detection (Fstgrdy): It is a hierarchical agglomeration algorithm for detecting community structure based on modularity optimization method [9]. It follows greedy optimization in which, starting with each vertex being the sole member of a community of one, two communities are repeatedly join together, whose amalgamation produces the largest increase in modularity value.

The communities found by the above three approaches are evaluated by measure of modularity defined in Sect. 4.2. These communities are utilized for discovery of new climatic indices. We select top few communities by thresholding based on number of nodes present in the community, density of community, correlation of time-series of community with Indian monsoon. Communities filtered out are the representative for new climatic index. We average the time series values over all the nodes present in a specific community and the resulting time-series represents the new climatic index. The correlation of discovered indices with Indian monsoon is studied and compared with correlation of present Indian Meteorological Department's (*IMD*) predictors with Indian monsoon.

Discovered indices show higher correlation than present predictor indices of monsoon. A study of correlation between discovered climatic indices and existing climatic predictors of monsoon is performed as a validation of our discovered
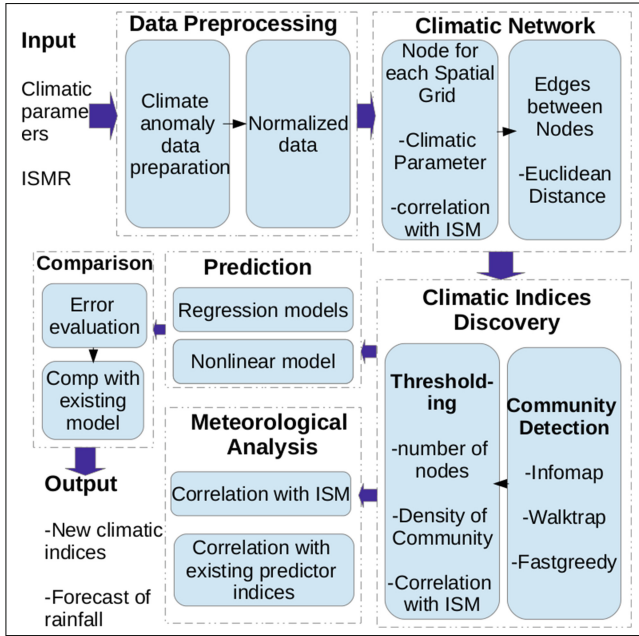
Fig. 1. Block diagram of proposed approach discovery of climatic indices important for Indian summer monsoon rainfall

indices. Finally, discovered climatic indices having high correlation with Indian monsoon are used for prediction of monsoon. Regression and non-linear models are designed with discovered climatic indices as predictors for forecasting annual Indian summer monsoon rainfall. The block diagram of proposed approach of discovery of climatic indices important for Indian monsoon and utilization of indices in forecasting monsoon are shown in Fig. 1.

4   Experimental Evaluation

4.1   Data Sets

Surface pressure and zonal wind velocity are collected from *NCEP* reanalysis data provided by the NOAA/OAR/ESRL (www.esrl.noaa.gov/psd/) [10] at spatial resolution of $2.5° \times 2.5°$ with coverage of $90°N$–$90°S$ and $0°E$–$358°E$. There are *73* latitude and *144* longitude grids, which give *10,512* nodes (73×144) in the network. Annual Indian summer monsoon rainfall (*ISMR*), occurring in months of June, July, August, and September is acquired from Indian Institute of Tropical Meteorology (www.imdpune.gov.in/research/ncc/longrange/data/data.html) [11]. *ISMR* is expressed as percentage of long period average (*LPA*) value of rainfall, which is *878.1 mm* for our period of study *1948–2013*.

As a preprocessing step, data is converted to monthly anomaly data by subtracting the monthly mean from corresponding data. Pearson correlation of climatic parameter and Indian rainfall for best lead month, considering lead of zero to six months is taken as a node attribute, which assists our search of climatic indices which will act as good predictor of Indian summer monsoon rainfall.

$$Climatic\ anomaly_m = X_m - mean(X_m),$$

where, X_m denotes climatic parameter value for month m and mean (X_m) is the average of the parameter values over all the years under study for month m.

4.2   Evaluation Methodology

Modularity. The goodness of communities detected are evaluated in terms of modularity measure. It is defined as the fraction of the edges that fall within the given communities minus the expected such fraction if edges were distributed at random. Higher value corresponds to good community detection. It is shown by Eq. 1.

$$Q = \frac{1}{2e} \sum_{vw} \left[A_{vw} - \frac{k_v k_w}{2e} \right] \delta\left(c_v, c_w\right), \tag{1}$$

where, e represents the number of edges in the graph, v and w are the nodes, $A_{vw} = 1$, if edges present between nodes v and w, 0 otherwise, k_v, k_w are the degree of nodes v and w, $\delta(c_v, c_w) = 1$, if both nodes belong to same community, otherwise 0.

Modularity and number of communities formed by three community detection algorithms for *NET_SP* and *NET_ZW* are shown in Tables 1 and 2, respectively. Communities detected have high modularity measure of *0.93* for surface pressure, and *0.97* for zonal wind velocity by *Fstgrdy* community detection method.

Table 1. Modularity and number of communities detected for network built for surface pressure (*NET_SP*)

Algorithm	Modularity	Number of communities
Info	0.890	512
Wlktrp	0.925	197
Fstgrdy	**0.930**	400

Table 2. Modularity and number of communities detected for network built for zonal wind velocity (*NET_ZW*)

Algorithm	Modularity	Number of communities
Info	0.913	680
Wlktrp	0.977	351
Fstgrdy	**0.978**	358

Selecting Top Communities. Few predictive communities are selected from the obtained communities by thresholding. Three measures are taken as baseline, namely, (i) number of nodes, (ii) density of communities, (iii) communities having correlation with Indian monsoon greater than threshold correlation.

The threshold correlation is ascertained by plotting a histogram of correlation of random *1000* climatic parameter series and Indian monsoon. The result for climatic parameter surface pressure is shown in Fig. 2. It is observed that most of the correlation lies below *0.1*, so we have taken our threshold as *0.13* for surface pressure and similarly *0.15* for zonal wind velocity. The selected predictive communities of both surface pressure and zonal wind velocity are considered as the new discovered climatic indices important for prediction of Indian monsoon.

4.3   Correlation Studies

Discovered climatic indices are evaluated by estimating their correlation with Indian monsoon. Number of selected discovered climatic indices and best correlation of indices with monsoon for all three community detection algorithms are elaborated in Tables 3 and 4 for surface pressure and zonal wind velocity, respectively. Correlation of *0.34* is observed for discovered climatic indices from surface pressure parameter and *0.35* is obtained for zonal wind velocity parameter. Pearson correlation of discovered climatic indices for *NET_SP* by *info-map* community detection method is shown in Fig. 3.

Table 3. Number of discovered climatic indices and their best correlation with Indian monsoon for surface pressure (*NET_SP*)

Algorithm	Number of selected communities	Best correlation
Info	11	0.32
Wlktrp	11	0.32
Fstgrdy	12	**0.34**

Table 4. Number of discovered climatic indices and their best correlation with Indian monsoon for wind velocity (*NET_ZW*)

Algorithm	Number of selected communities	Best correlation
Info	12	**0.35**
Wlktrp	12	0.28
Fstgrdy	14	0.28

4.4   Prediction Performance

Discovered climatic indices are evaluated in terms of their predictability of Indian summer monsoon rainfall. We use climatic indices which have high correlation with Indian monsoon as predictors. The predictor climatic indices obtained from networks built for surface pressure and zonal wind velocity are listed in Tables 5 and 6, respectively. Regression models, namely linear regression, ridge regression model with cross validation, bayesian regression and non-linear model, namely generalized regression neural network (*GRNN*) are built with discovered climatic indices as predictors for forecasting annual Indian summer monsoon rainfall. Test period of twenty years from *1994* to *2013* is considered for evaluation. Mean absolute errors in terms of percentage of long period average value (*LPA*)
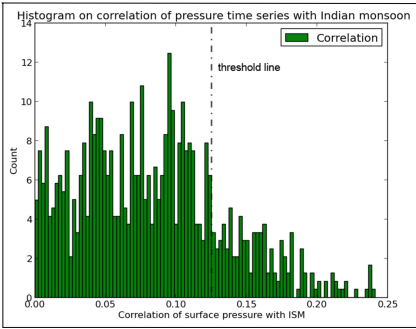
Fig. 2. Histogram for finding baseline *threshold* correlation with Indian monsoon for *NET_SP*
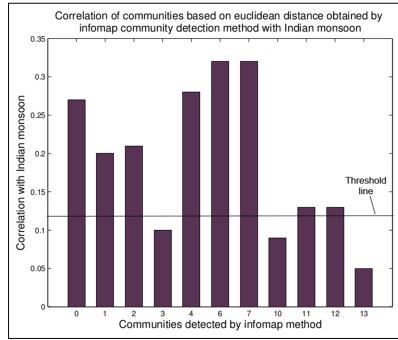
Fig. 3. Correlation of communities with Indian monsoon detected by *Info-map* method for *NET_SP*

of rainfall is presented for regression and non-linear models in Tables 7 and 8 for *NET_SP* and *NET_ZW*, respectively. Climatic indices discovered by *info-map* method give best performance with mean absolute errors of *5.5%* and *5.4%* for *NET_SP* and *NET_ZW*, respectively. This verifies the inclusion of correlation of parameter with Indian monsoon as node weight, which is considered by *info-map* technique for discovery of climatic indices.

Table 5. Number of predictors and discovered climatic indices with community id for surface pressure (*NET_SP*)

Algorithm	Number of predictors	Community ids
Info	4	0,4,6,7
Wlktrp	6	1,15,93,103,109,136
Fstgrdy	4	182,186,217,237

Table 6. Number of predictors and discovered climatic indices with community id for wind velocity (*NET_ZW*)

Algorithm	Number of predictors	Community ids
Info	4	1,4,5,8
Wlktrp	4	56,66,67,224
Fstgrdy	6	34,35,56,66,78,184

4.5   Comparisons with Existing Models

The predictability of climatic indices in forecasting Indian monsoon are compared with present Indian Meteorological Department's (*IMD*) models. Models built with indices discovered from network based on surface pressure by all the three community detection methods give better performance than existing *16-* parameter power regression model [12] and *8* and *10*-parameter *IMD* models [13]. Proposed models built with discovered predictor climatic indices by *Info*, *Wlktrp*, and *Fstgrdy* methods give root mean square errors of *4.8%*, *5.6%*, and *6.2%*, respectively, outperforming all three *IMD* models giving *10.8%*, *6.4%*, and *7.6%* errors for period *1996–2002*. Models built from predictor climatic

Table 7. Mean absolute errors (%) for prediction of Indian monsoon by discovered climatic indices from *NET_SP* for test period *1994–2013*

Models	Info	Wlktrp	Fstgrdy
Linear	*5.5*	6.5	5.8
RidgeCV	6.0	**5.7**	**5.7**
Bayesian ridge	6.0	**5.7**	6.0
GRNN	**6.0**	6.3	6.3

Table 8. Mean absolute errors (%) for prediction of Indian monsoon by discovered climatic indices from *NET_ZW* for test period *1994–2013*

Models	Info	Wlktrp	Fstgrdy
Linear	**5.8**	6.5	6.1
RidgeCV	**5.8**	6.6	6.2
Bayesian ridge	*5.4*	6.5	6.2
GRNN	**6.4**	**6.4**	**6.4**

indices discovered from network based on zonal wind velocity by *Info*, *Wlktrp*, and *Fstgrdy* methods give root mean square errors of *7.3 %*, *7.0 %*, and *7.5 %*, respectively, which outrun *IMD*'s *16* and *8*-parameter model, but is greater than *IMD*'s *10*-parameter model having *6.4 %* error. Discovered climatic indices for network based on surface pressure serve as better predictor of Indian monsoon. Therefore, it can be ascertained that surface pressure has more important role than wind velocity for climatic event of monsoon. Comparisons of predictability of models built with discovered climatic indices from *NET_SP* and *IMD* models are shown in Fig. 4.
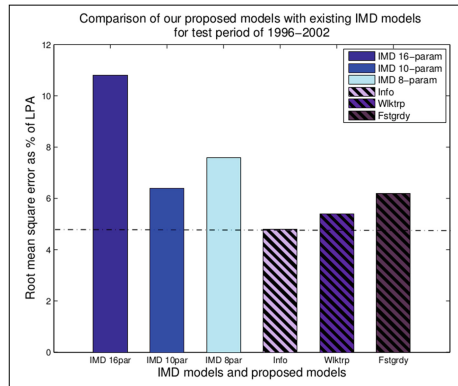
Fig. 4. Comparison of root mean square errors in prediction of Indian monsoon by proposed models based on climatic indices discovered from *NET_SP* and *IMD's 16* [12], *10*, *8*-parameter [13] models for period *1996–2002*

5   Meteorological Significance

5.1   Analysis Based on Correlation with ISMR

The Pearson correlation (μ) of discovered climatic indices with Indian monsoon are compared to correlation of existing predictor climatic indices with Indian

monsoon [14]. Important predictor of monsoon, as considered by *IMD*, namely, North Atlantic SST (*NA_SST*), Equatorial South Eastern Indian Ocean SST (*ESE_IO_SST*), East Asia surface pressure (*EA_SP*), North Atlantic surface pressure (*NA_SP*), North Central Pacific Ocean zonal wind anomaly (*NC_PO_ zonal_wnd*), and North West Europe surface pressure (*NW_Eu_SP*) are considered for validation of the discovered climatic indices. Newly discovered climatic indices are shown to be having higher correlation than *IMD*'s predictor indices. The result for climatic indices discovered for *NET_SP* and *NET_ZW* are shown in Figs. 5 and 6, respectively. High correlation of *0.34* and *0.35* are observed for indices discovered for climatic parameters surface pressure and zonal wind velocity, respectively, which show superior behaviour.
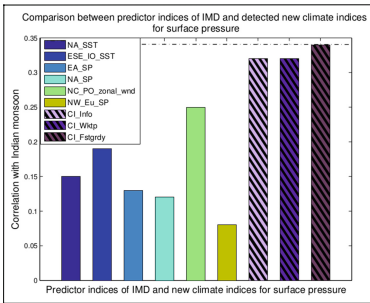
Fig. 5. Comparison of correlation with *ISMR* for *IMD* predictors and discovered climatic indices for *NET_SP*
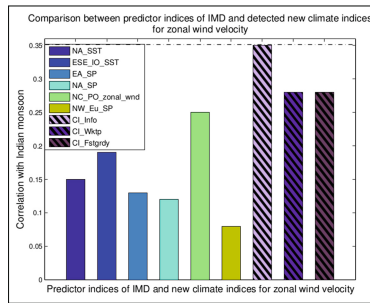
Fig. 6. Comparison of correlation with *ISMR* for *IMD* predictors and discovered climatic indices for *NET_ZW*

5.2   Validation of Discovered Climatic Indices

New climatic indices (*CI*) are validated by correlation study of the newly discovered indices and *IMD* predictors. Tables 9 and 10 show the best correlation of climatic indices discovered by *Info*, *Wlktrp*, and *Fstgreedy* methods with existing *IMD* predictors as discussed earlier for *NET_SP* and *NET_ZV*, respectively. High correlation value (≥ 0.5) validates the proposed approach of climatic index discovery by inventing the existing indices (highlighted in bold). Medium correlation value ($0.2 \leq \mu < 0.5$) represents invention of new indices, which are related to existing indices, but may act as good predictor than the existing ones (normal font). Low correlation value (< 0.2) represents newly discovered indices different from known indices (highlighted in italics). Discovered climatic index for *NET_SP* shows high correlation with *EA_SP* and *NA_SP*, validate our approach by re-invention of existing predictor indices.

Table 9. Correlation of discovered climatic indices (*CI*) for *NET_SP* with *IMD* predictors for Indian monsoon

Existing Predictor CI	Info CI	Wlktrp CI	Fstgrdy CI
NA_SST	*0.09*	0.21	*0.16*
ESE_IO_SST	0.37	0.37	0.39
EA_SP	0.29	0.32	**0.52**
NA_SP	0.32	**0.56**	0.32
NC_PO_ zonal_wnd	0.23	0.32	0.23
NW_Eu_SP	0.24	0.30	0.50

Table 10. Correlation of discovered climatic indices (*CI*) for *NET_ZW* with *IMD* predictors for Indian monsoon

Existing Predictor CI	Info CI	Wlktrp CI	Fstgrdy CI
NA_SST	0.27	0.24	0.24
ESE_IO_SST	0.26	*0.16*	0.21
EA_SP	*0.19*	0.21	0.21
NA_SP	*0.19*	0.21	0.21
NC_PO_ zonal_wnd	0.27	*0.18*	0.20
NW_Eu_SP	0.22	0.29	*0.19*

6   Conclusions

New climatic indices important for Indian summer monsoon rainfall are discovered using algorithms of community detection for climatic parameters surface pressure and zonal wind velocity. Indices discovered are shown to have high correlation with Indian monsoon. Their correlation are even better than that of the known predictor indices used by *IMD* for predicting monsoon. Different regression and non-linear models are designed with discovered climatic indices as predictors. Mean absolute error of *5.4 %* is achieved, which is appreciable for forecasting complex phenomenon of Indian monsoon. Prediction of monsoon by discovered indices of surface pressure is superior to *IMD*'s existing models. Finally, a study of correlation between discovered indices and predictor indices of Indian monsoon is performed as meteorological validation of our approach.

In future, other climatic parameters can be explored and new climatic indices can be discovered from combination of different climatic parameters which may be highly correlated and act as a better estimator of Indian monsoon.

References

1. Donges, J.F., Zou, Y., Marwan, N., Kurths, J.: Complex networks in climate dynamics. Eur. Phys. J.-Special Topics **174**(1), 157–179 (2009)
2. Steinhaeuser, K., Chawla, N.V., Ganguly, A.R.: Complex networks as a unified framework for descriptive analysis and predictive modeling in climate science. Stat. Anal. Data Min. **4**(5), 497–511 (2011)
3. Donges, J.F., Zou, Y., Marwan, N., Kurths, J.: The backbone of the climate network. Europhys. Lett. (EPL) **87**(4), 48007 (2009)
4. Steinhaeuser, K., Chawla, N.V., Ganguly, A.R.: An exploration of climate data using complex networks. ACM SIGKDD Explor. Newsl. **12**(1), 25–32 (2010)
5. Noor Md Sap, M., Awan, A.M.: Finding spatio-temporal patterns in climate data using clustering. In: Proceeding 2005 International Conference on Cyberworlds, pp. 8–15. IEEE (2005)
6. Steinbach, M., Tan, P.N., Kumar, V., Klooster, S., Potter, C.: Discovery of climate indices using clustering. In: Proceeding of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 446–455. ACM (2003)

7. Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. Proc. Natl. Acad. Sci. **105**(4), 1118–1123 (2008)
8. Pons, P., Latapy, M.: Computing communities in large networks using random walks. In: Yolum, I., Güngör, T., Gürgen, F., Özturan, C. (eds.) ISCIS 2005. LNCS, vol. 3733, pp. 284–293. Springer, Heidelberg (2005)
9. Clauset, A., Newman, M.E., Moore, C.: Finding community structure in very large networks. Phys. Rev. E **70**(6), 066111 (2004)
10. Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Jenne, R., Joseph, D.: The NCEP/NCAR 40-year reanalysis project. Bull. Am. Meteorol. Soc. **77**(3), 437–471 (1996)
11. Parthasarathy, B., Munot, A.A., Kothawale, D.R.: Monthly and seasonal rainfall series for All-India homogeneous regions and meteorological subdivisions, 1871–1994. Research Report No. RR-065, Indian Institute of Tropical Meteorology (1995)
12. Gowariker, V., Thapliyal, V., Kulshrestha, S.M., Mandal, G.S., Sen Roy, N., Sikka, D.R.: A power regression model for long range forecast of southwest monsoon rainfall over India. Mausam **42**(2), 125–130 (1991)
13. Rajeevan, M., Pai, D.S., Dikshit, S.K., Kelkar, R.R.: IMD's new operational models for long-range forecast of southwest monsoon rainfall over India and their verification for 2003. Curr. Sci. **86**(3), 422–431 (2004)
14. Rajeevan, M., Pai, D.S., Kumar, R.A., Lal, B.: New statistical models for long-range forecasting of southwest monsoon rainfall over India. Clim. Dyn. **28**(7–8), 813–828 (2007)