Journal of the Indian Institute of Science
A Multidisciplinary Reviews Journal
ISSN: 0970-4140 Coden-JIISAD

REVIEWS

Regionalization of Precipitation in India—A Review

V.V. Srinivas

Abstract | Regionalization of precipitation refers to delineation of rain gauges in an area into homogeneous groups (clusters or regions). Various regionalization procedures are employed by researchers in hydrometeorology for addressing a wide spectrum of problems. This paper provides an overview of underlying concepts as well as advantages and limitations of procedures that have been developed over the past six decades. Emphasis is given to studies that have been carried out in India. Following this, gaps where more research needs to be focussed are highlighted, and challenges for regionalization in a climate change scenario are discussed.

Keywords: *regionalization, homogeneous rainfall regions, precipitation, India.*

1 Introduction

Regionalization of precipitation is necessary for various applications that include (i) meteorological drought analysis and agricultural planning to cope with water shortages that are likely due to low rainfall, (ii) forecasting and downscaling of precipitation (iii) design of water control (e.g., barrages, dams, levees) and conveyance structures (e.g., culverts, storm sewers, spillways) to mitigate damages that are likely due to floods triggered by extreme precipitation, and (iv) land-use planning and management.

Regionalization becomes a challenging task when spatial and/or temporal variability of rainfall is quite complex. Regions are often delineated for specific purposes by analyzing records of relevant rainfall patterns (e.g., extreme, long-term cumulative) compiled at relevant temporal scale(s) (e.g., sub-daily, daily, monsoon, annual). Therefore homogeneous regions delineated for different purposes need not be identical, and the definition of regional homogeneity in not unique.

The remainder of this paper is organized as follows. Section 2 provides a historical perspective of regionalization of precipitation in India, and presents an overview of underlying concepts as well as advantages and limitations of various procedures that have been developed to aid in the task of regionalization. Following this, conclusions are provided in Section 3. In the following write-up the words 'site, gauge and station' and the words 'group, clusters and region' are used interchangeably.

2 Overview of Approaches to Regionalization of Precipitation

The conventional practise was to delineate regions as geographically contiguous areas based on physiography and/or political/administrative boundaries. As such regions may not have any definite relationship to causal/explanatory variables influencing rainfall, regions based on those factors need not be homogeneous in rainfall. Based on this argument India was delineated into 21 rainfall provinces using rainfall records of stations spread all over the country.[1] In spite of these developments, several studies continued to consider the entire India as a homogeneous rainfall region for the sake of investigating association of monsoon rainfall with large-scale circulation. Even the India Meteorological Department (IMD) traditionally considers Meteorological Subdivisions (MSDs) for reporting precipitation related forecasts and other estimates at various temporal scales (e.g., weakly, seasonal). Those subdivisions are based on political boundaries and a well distributed network of 306 rain gauge stations over India. Their number keeps increasing as new states emerge in India, as evident in the recent past after formation of the states Uttaranchal, Jharkhand and Chhattisgarh in the northern part of India. As the number of MSDs is fairly large (36), the general notion is that

Department of Civil Engineering, Indian Institute of Science, Bangalore 560 012, India.
vvs@civil.iisc.ernet.in

they are homogeneous. However, several of them are too extensive to be considered homogeneous as the possibility of entire subdivision getting similarly affected by excess or deficit of rainfall is rare. Various studies revealed that (i) India is too large to be considered as a homogeneous region, as very small or even negative correlations were found between All-India rainfall and rainfall of various MSDs, and between some MSDs,[2] (ii) many of the MSDs are not homogeneous in frequency distribution of rainfall,[3] and (iii) inter-site correlations of rainfall for various pairs of gauges within certain MSDs are insignificant or even negative.[4–6] Therefore rainfall estimates for those MSDs that are generally reported by IMD might not be meaningful. Some studies even suggested pooling of certain MSDs owing to similarity in their rainfall characteristics.[7] A brief review of various studies that death with such investigations can be found in literature.[5,6] In the backdrop of some of these significant findings, India was delineated into five homogeneous Summer Monsoon Rainfall (SMR) regions by grouping MSDs on the basis of the following criteria:[8] (i) Contiguity of area, (ii) contribution of monsoon seasonal rainfall to the annual amount, (iii) inter-correlations of sub-divisional and All-India monsoon rainfall, and (iv) relationships between sub-divisional monsoon rainfall and regional/global circulation parameters. The delineated five regions are referred to as, Northwest India, West Central India, Central Northeast India, Northeast India and Peninsular India.[8] They were recently found to be heterogeneous in frequency distribution of summer monsoon rainfall.[9] A group of 14 MSDs in Northwest and West Central India regions was referred to as Homogeneous Indian Monsoon (HIM) region,[6] which was recently found to be spectrally heterogeneous.[10]

To delineate effective homogeneous regions a variety of regionalization approaches have been developed over the past six decades. They include those based on (i) correlation analysis (e.g., elementary linkage analysis and its variation), (ii) Principal Component Analysis (PCA) also referred to as eigenvector analysis or the related Empirical Orthogonal Function (EOF) analysis, (iii) common factor analysis, (iv) spectral analysis, (v) cluster analysis, and PCA in association with cluster analysis, (vi) hierarchical approach, and (vii) Region of Influence. Of these, PCA, cluster analysis and the last two approaches found use for regionalization of extreme rainfall.

In regionalization based on elementary linkage analysis,[11] formation of homogeneous precipitation regions is based on correlations between precipitation time series corresponding to all possible pairs of rain gauges in the study area. In the first step, a pair of gauges having the highest correlation in precipitation is identified to form a tentative group. In the second step, a gauge whose precipitation record has significant correlation (greater than a specified correlation-threshold) with that of each of the gauges in the tentative group is identified from the remaining gauges and then assigned to the tentative group. Following this, the second step is repeated till no more gauges can be assigned to the tentative group, which is then declared as a homogeneous region. Subsequently the foregoing procedure of forming homogeneous regions is repeated on gauges that are not already assigned to region(s). Thus each of the regions delineated by elementary linkage analysis has high inter-site correlation in precipitation. This approach to regionalization found application in countries such as Sierra-Leone, West Africa,[12] Tanzania,[13] United Kingdom[14] and Nigeria.[15]

An alternate correlation analysis based regionalization approach[4,5] begins with identification of two gauges having least or weakest correlation in their precipitation data to form central stations or seed points. To form a coherent rainfall zone (region) around a central station, gauges whose precipitation records have positive significant correlation (greater than a specified correlation-threshold) with that of the central station are identified and then ranked in descending order of the correlation. Geographical proximity could be the basis for ranking of gauges with a comparable correlation value. The central station is first grouped with the gauge that has the highest value of correlation with it in terms of rainfall to form a zone. Following this, other gauges are considered one at a time, in order of their rank, and assigned to the zone if the gauge's precipitation time series has positive significant correlation with precipitation corresponding to each of the gauges available in the zone at that stage. The subsequent central station is chosen from the gauges that are not already assigned to the existing zone(s), such that its corresponding precipitation is as poorly correlated as possible to that of the existing central stations, and the foregoing procedure is followed to form a coherent rainfall zone around it. This process is continued until an adequate network of central stations representing the entire gamut of rainfall variation in the study region is obtained. This approach was applied to delineate Karnataka state of India into 2 to 9 coherent rainfall zones.[4] Intersite correlation of annual rainfall was considered to identify central stations and the coherence within each delineated zone was examined in terms of intersite correlation on seasonal and monthly scales. In a subsequently study, the approach was applied to delineate India into 31 zones that were coherent with respect to the variations of the SMR that extends over the period June–September.[5] Monthly rainfall records of 200 IMD gauges having at least 25 years record (during 1901–1987) and situated at locations having elevation within 1 km above sea-level were considered for the study. Seventeen of those regions were found to be definitely heterogeneous and another four were found to be possibly heterogeneous in terms of frequency distribution of SMR based on 1° × 1° resolution IMD gridded rainfall data.[16] Very recently the approach was adapted to delineate India into 26 homogeneous rainfall zones using annual as well as southwest monsoon rainfall, and 20 zones using annual as well as northeast monsoon rainfall.[17] Fifty-one years long daily rainfall records collated during the period 1951 to 2001 at 1025 rain gauges spread all over India were considered for the study.

An issue with regionalization by the aforementioned correlation analysis based methods is that the delineated regions are sensitive to the choice of correlation threshold value, but there is no established procedure to arrive at an optimal value of the same.

In regionalization by Principal Component Analysis (PCA), which is often referred to as eigen vector analysis or empirical orthogonal function analysis, Principal Components (PCs) that are orthogonal to each other are derived from inter-station correlation and/or covariance matrix of precipitation in the study area. If the first few leading PCs account for significant percent of the total variance, their spatial patterns are analysed to form homogeneous precipitation regions. This approach could involve either plotting the unrotated and/or rotated PC loadings on the map of the study area, or representing stations as points in two-dimensional space of leading PCs. In the context of regionalization of extreme rainfall, regional frequency (growth) curves of precipitation extremes are constructed for each of the delineated regions using pooled information from the region,[18] or by fitting regional regression relationship between PCs and parameters of the distribution.[19] The developed curves or relationships are useful to arrive at quantile estimates at sites that are ungauged or have inadequate data. The PCA approach has been widely applied to precipitation data from various parts of the world including southern Africa,[20] United Kingdom,[21] United States[22] and its border regions with Mexico,[23] European Mediterranean Countries,[24] Austria,[25] Yugoslavia,[26] India,[27,28] Switzerland,[29] Island of Tahiti,[19] Italy,[30] and Spain.[18,31] Delineation of homogeneous rainfall zones based on this approach becomes cumbersome when a large number of PCs collectively account for a significant percent of the total variance, and/or if the number of gauges in the study area is large. In the Indian context, prior studies[5,27,28,32] indicated that the first few leading PCs are unable to account for significant percent of the total variance in summer monsoon season during which the country receives copious rainfall, possibly due to complexity in spatial variation of the rainfall. Hence PCA approach is not the best possible choice to delineate homogeneous Summer Monsoon Rainfall (SMR) regions. In a study confined to sub-Himalayan region and Gangetic plains of India, four homogeneous rainfall regions for both seasonal and monthly time scales were delineated based on analysis of seasonal and monthly summer monsoon rainfall of 90 stations considered over the period 1871 to 1984.[28]

To address lack of resolution power of conventional PCA with progression to the higher components (PCs), PCA based on sequential sieving of stations was proposed for regionalization.[27] The procedure involves identification of the first major sequential region by grouping sites that are highly correlated with the first significant PC. At this stage, those grouped sites are omitted and the reduced dataset is once again subjected to PCA. If the new first PC is significant then second sequential region is formed by grouping stations that are highly correlated with it. The process of eliminating already grouped sites and locating first new PC using data of the remaining sites to delineate a new sequential region is repeated till the residual stations (unallocated to regions) naturally group themselves into well separated geographically contiguous regions. Ten sequentially coherent homogeneous rainfall regions were identified in India by applying this procedure to SMR totals of 200 stations over the period 1901–1980. The regions accounted for 91% of the study area.[27] Six of those regions were found to be definitely heterogeneous and another three were found to be possibly heterogeneous when tested for homogeneity in SMR frequency distribution based on 1° × 1° resolution IMD gridded rainfall data.[16]

Common Factor Analysis (CFA) approach to regionalization involves application of factor analysis to inter-site correlation matrix to arrive at common factors and specific factors. The analysis assumes an underlying basic model for the data and it allows accounting for specific variance, which is not possible with PCA. The specific variance is related to local forcing that influences rainfall variability at individual gauges in the study area, and is different from the forcing that is common to a group of gauges.[33,34] In PCA, the specific variance is distributed among all the loadings, indicating that the sum of localized effects is spread over all the modes. In CFA, the specific variance is not manifested on any of the loadings. The number of common factors is optimized to maximize the variance shared by stations (communality) and minimize the specific variance. Each of the common factors represents a region comprising of stations that share the communality variance. If reasonable number of common factors do not account for significant percent of the variance shared by stations, CFA may not be a good choice for regionalization. This approach found applications on data from countries such as Kenya[35] and Puerto Rico.[33,34]

Cluster analysis procedures gained recognition for use in delineation of homogeneous rainfall groups owing to their effectiveness in interpreting patterns in data of explanatory variables influencing rainfall, or in leading PCs resulting from PCA. Each gauge is represented by a feature vector that comprises of explanatory variables or PCs, which are referred to as attributes. Conventional practise is to consider attributes as statistics computed from precipitation records, besides location indicators of gauges (e.g., latitude, longitude, elevation/altitude, distance to the sea). The statistics include mean annual/seasonal/monthly/daily precipitation, mean number of wet days, ratio of minimum average two-month precipitation to maximum average two-month precipitation, parameters of hydrologic distributions fitted to precipitation data. When such statistics (attributes) form basis for regionalization, adequate number of sites with sufficiently long period of contemporaneous observations is necessary to form meaningful regions, and regions cannot be delineated in ungauged areas. The cluster analysis approach facilitates formation of regions that are not necessarily contiguous in geographical space. It has been extensively applied throughout the world, including countries like United States,[36–38] Spain,[39,40] Australia,[41] Iran,[42,43] Austria,[44] South Africa and Lesotho,[45] South America,[46] India[3,9,47–51] and China.[52]

In a study covering peninsular India, PCA in conjunction with cluster analysis was applied to the mean pentad (5-day) rainfall data of 53 stations pertaining to the period 1901–1950 for regionalization.[47] The stations were represented as points in the coordinate space of the amplitudes of the two leading eigenvectors and nearest centroid sorting (clustering) method was used to delineate them into eight geographically contiguous areas or clusters. Further, in a study confined to Western Ghat region of India, Ward's hierarchical clustering method was applied to gridded mean annual rainfall data to form three zones (regions): Coastal zone, Transition zone and Malanad zone.[50] The gridded data was based on records of 93 rain gauge stations in the region (whose lengths varied from 10 to 50 years), and the grid was designed to have resolution of 7.5' × 7.5' on windward side, and 15' × 15' on leeward side of Western Ghats. Recently, in two studies covering India, partitional clustering procedures were considered for delineation of homogeneous summer monsoon and annual rainfall regions.[3,9] Attributes that facilitate regionalization in ungauged as well as sparsely gauged areas were suggested. Finer details of those studies are provided in latter part of this review. More recently a cluster ensemble algorithm was applied to 1° × 1° gridded annual rainfall data of IMD over the period 1951–2003 to arrive at 9 regions in India that are homogeneous in annual rainfall.[51]

In regionalization based on spectral analysis, gauges that show significantly low spectral variability are grouped together to form a region. Recently a Monte-Carlo procedure to assess spectral homogeneity of a region was proposed and its utility in delineating a spectrally homogeneous region was demonstrated.[10] The procedure involves examining whether spectral variability of the region differs significantly from that expected in realizations of white noise process whose number is equal to that of sites in the region. Herein, the spectral variability refers to variation among Power Spectral Densities (PSDs) of standardized anomaly precipitation time series of stations in the region, and it is computed with respect to PSD of the ensemble average rainfall over the region. Application of the spectral homogeneity test to rainfall data over the period 1871–1990 and analysis of periodicities in rainfall revealed that a group of 14 MSDs in the central and northwestern parts of India, which was declared to be a homogeneous Indian monsoon (HIM) region in a previous study,[6] is not spectrally homogeneous.[10] One can explore forming homogeneous rainfall sub-regions in India by grouping MSDs whose spectral variability is less than that expected in realizations of white noise process at specified confidence level. An alternative option is to form groups (regions) through application of agglomerative hierarchical clustering procedures (single, complete and average linkage) to rainfall spectra, by considering metric for estimating the distance between any two gauges to be the deviation of the cross-correlation coefficient between their rainfall spectra from unity.[10]

Hierarchical regionalization approach explicitly accounts for spatial variability in moments of predictand (extreme rainfall) data.[53,54] It is based on the hypothesis that higher-order moments (e.g., skewness and kurtosis) of extreme rainfall data do not display significant spatial variability over a larger area than relatively lower order moment (coefficient of variation), which in turn is assumed to vary more slowly over space than the first-order moment. Therefore for the sake of frequency analysis at a target site, information from more (less) sites is used to estimate distribution parameters controlling the higher (lower) order moments. This approach is not appropriate for direct application to large geographical areas like Indian subcontinent with diverse climatic conditions.

Region of influence approach considers each target site to have its own region consisting of those sites whose distance to the target site in a weighted multi-dimensional space of predictors does not exceed a chosen threshold distance value.[55,56] Pooled information from all the sites in a region forms basis for constructing a regional frequency curve, and in the process each site could be weighed according to its proximity to the target site. This approach contradicts with the notion of forming geographically contiguous regions and has subjectivity in the selection of weights and threshold value.

Regions delineated using the conventional regionalization procedures may not be reliable when the study area is sparsely gauged and the period of contemporaneous records for the gauges considered is inadequate to account for sampling variability. Further it is not possible to delineate regions in ungauged areas using those procedures. To address these issues an approach is proposed recently, which recommends choosing attributes for cluster analysis as location indicators of gauges and PCs extracted from Large Scale Atmospheric Predictor Variables (LSAPVs) influencing rainfall at those gauges.[9] This enables delineation of regions even in ungauged areas because gridded data on LSAPVs are available for the entire globe (earth) from various reanalysis databases, irrespective of the available number of rain gauges in the study area. Another advantage of choosing such attributes is creation of an opportunity to validate the delineated regions for homogeneity using statistics of the observed precipitation. Such validation is not possible with conventional approaches which delineate regions based on statistics estimated using at-site rainfall data (e.g., correlations, summary statistics), because use of the same or closely related statistics for regionalization as well as validation of regions is inappropriate. Regions formed in ungauged areas using the suggested approach could be validated in future, when rainfall data becomes available. Besides this, the approach provides scope to arrive at future projections of homogeneous precipitation regions in a study area, as information on LSAPVs is available even for future periods from a large number of General Circulation Models (GCMs). Such projections might be useful to prepare plans for water resources development and management in the area. Based on this approach India was delineated into 17 homogeneous summer monsoon rainfall (SMR) regions (Figure 1) through analysis of results obtained from application of K-means partitional clustering procedure to feature vectors corresponding to fifty-two $2.5° \times 2.5°$ grid boxes portraying grid of NCEP (National Centers for Environmental Prediction) reanalysis data covering the country.[9] Monthly mean values of 15 LSAPVs (listed in Table 1) over a spatial domain of 16 NCEP grid points surrounding each grid box were subjected to PCA to extract five PCs that preserved more than 97% of the variance. Each feature vector comprised of the PCs and three standardized location attributes: latitude, longitude, and average elevation of terrain in each of the NCEP grid boxes.

Besides this, the conventional regionalization procedures yield hard clusters or groups of rain gauges, which means that each gauge can utmost belong to one region. This is based on the assumption that rainfall of a gauge fully resembles that at other gauges in the region (to which it is assigned) in terms of explanatory/causal variables influencing the same, and it does not have any resemblance to rainfall at sites in other regions. This assumption is not valid in real world scenario and hence formation of regions with soft boundaries would be more meaningful. In this perspective, the foregoing approach was extended to fuzzy framework, which allows a gauge to belong to more than one region simultaneously.[3] India was delineated into 25 homogeneous annual rainfall regions (Figure 2) through application of Fuzzy *c*-means cluster analysis to feature vectors corresponding to two hundred and ninety four $1° \times 1°$ grid boxes depicting grid of IMD rainfall data covering the country. The NCEP reanalysis data of 15 LSAPVs (listed in Table 1) were re-gridded from $2.5° \times 2.5°$ to $1° \times 1°$ resolution and twelve monthly mean values of all the LSAPVs over a spatial domain of 25 grid points surrounding each grid box were subjected to PCA to extract four PCs that preserved more than 95%

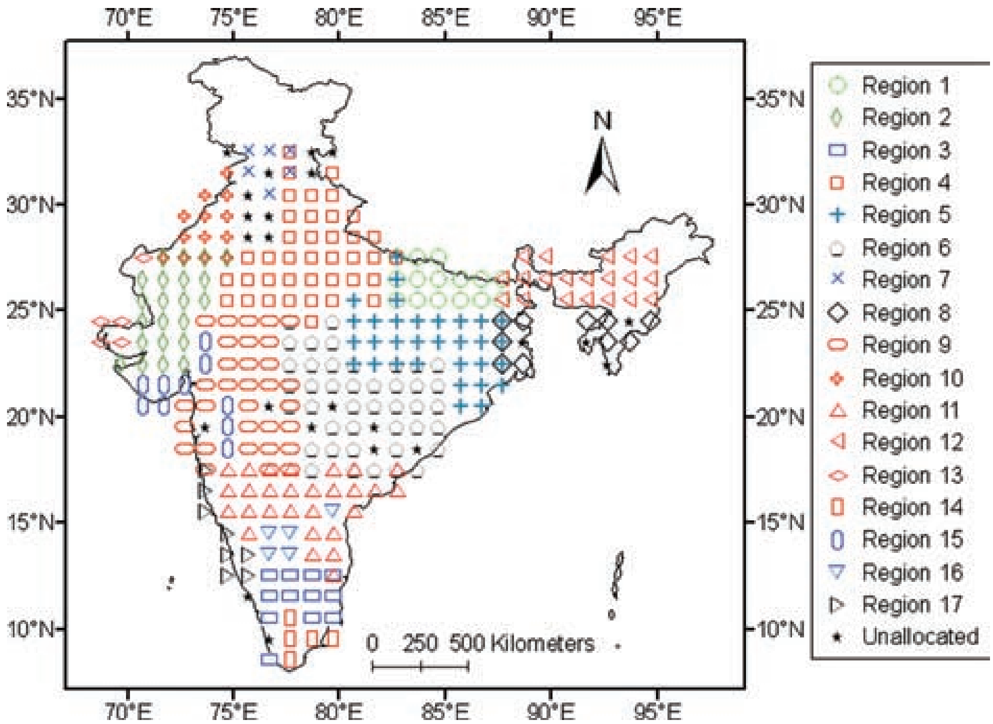

Figure 1: Homogeneous summer monsoon rainfall regions delineated in India using large scale atmospheric predictor variables [Reprinted from Satyanarayana and Srinivas (2008),[9] Copyright 2008 by the American Geophysical Union, with permission from John Wiley & Sons Inc].

Table 1: Atmospheric variables considered for regionalization of Indian rainfall.

Variable name	Pressure levels (in kPa)
Air temperature	92.5, 70, 50, 20
Geopotential height	92.5, 50, 20
Meridional wind	92.5, 20
Precipitable water	–
Specific humidity	92.5, 85
Surface pressure	–
Zonal wind	92.5, 20

of the variance. A feature vector representing each grid box comprised of the corresponding PCs, seasonality measure (Julian day depicting middle day of a 30-day maximum rainfall period in each year), and three standardized location attributes: latitude, longitude, and average elevation of terrain in the grid box. The seasonality of rainfall was chosen as an attribute because the gauges which receive maximum rainfall during the same time of a year could form a homogeneous rainfall region, owing to similarity in causal factors influencing rainfall. Information on seasonality can be reliably obtained from local inhabitants even at ungauged sites.[3]

Each of the SMR (annual) regions delineated in the foregoing studies[3,9] was validated (tested) for homogeneity in *L*-moment framework[57] using *L*-statistics of gridded SMR (annual rainfall) data for the period 1951–2004. The gridded data were prepared from daily records at 2140 stations (having a minimum of 90% of data during 1951–2004) chosen from the available 6329 stations in India by Interpolation scheme of Shepard.[58–60] Density of considered stations was high over southern Peninsula and was low over northern plains and eastern parts of India. In homogeneity test a region was said to be homogeneous if precipitation records of all the sites in it could be considered as samples drawn from a particular frequency distribution (population). An advantage with regional homogeneity test[57] is that it avoids *a priori* assumptions about the form of the probability distribution of

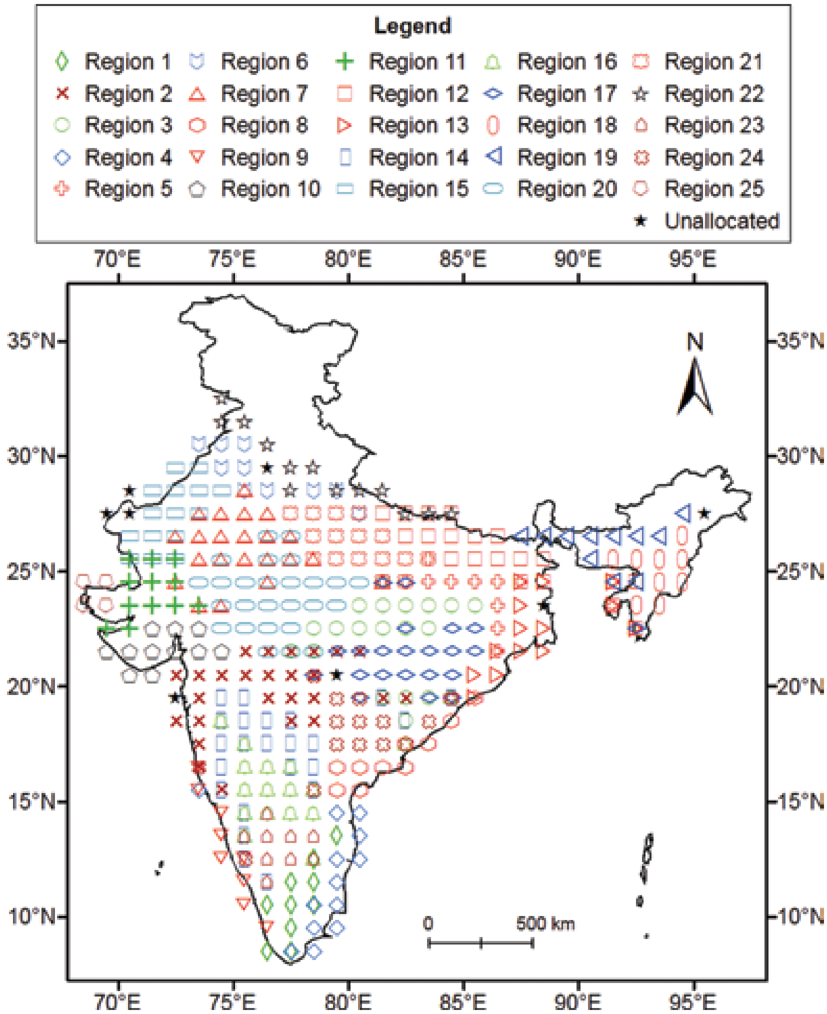

Figure 2: Homogeneous annual rainfall regions delineated in India using large scale atmospheric predictor variables [Reprinted from Satyanarayana and Srinivas (2011),[3] Copyright 2011, with permission from Elsevier].

precipitation records. Besides this, it is worth noting that the studies considered LSAPVs based on reanalysis data as surrogate for those in observed data to arrive at attributes, because latter data were unavailable for target locations over the study area. The reanalysis data are produced by integration of historical observations (wherever available) and model simulations using data assimilation techniques. Typical examples of such data that are available for use from public domains include NCEP/NCAR reanalysis-1,[61] NCEP-DEO AMIP-II reanalysis,[62] ECMWF 15 year re-analysis (ERA-15),[63] ECMWF 40 year re-analysis (ERA-40),[64] and Japanese 25-year Reanalysis (JRA-25).[65] As reanalysis data based on different sources are not the same, the number and composition of delineated homogeneous rainfall regions might differ with change in the data. Further, as the available at-site rainfall records are not sufficiently large, there is uncertainty in arriving at a precise estimate for homogeneity of each of delineated regions. In view of these issues, further investigations are necessary to quantify various types of uncertainties associated with the delineated regions. Furthermore, the regions could be revalidated for homogeneity when more rain gauge data becomes available.

Another important issue is that the conventional regionalization approaches consider rainfall (predictand) to be a stationary stochastic process, and assume that the relationships between rainfall and its explanatory/causal (predictor) variables are time invariant. In the recent decades there is increase in evidence of climate change and non-stationarity in rainfall is suspected in various parts of the world. In this backdrop, it is necessary to verify stationarity of datasets at time scales of relevance before proceeding to regionalization and develop effective regionalization approaches for use with non-stationary data. In Indian context, it is worth noting that a study covering the period 1871–1990 revealed large interannual and decadal variations, but no systematic trend in the all India rainfall.[6] Further, a study covering the period 1951–2000 did not find significant trend in the mean monsoon seasonal rainfall over central India.[66] Nevertheless, non-stationarity is suspected to be more in data at finer time scales such as daily and hourly, and it cannot be ignored while performing regionalization of extreme rainfall. Future research should explore these issues.

3 Conclusions

The following conclusions can be drawn from the foregoing discussion:

1. India is too large to be considered as a homogeneous region. Further, many of the meteorological subdivisions that are being considered by IMD for reporting precipitation related forecasts and other estimates are statistically heterogeneous in precipitation.
2. Principal component analysis approach is not the best possible choice to delineate homogeneous Summer Monsoon Rainfall (SMR) regions in India, because the first few leading PCs (derived from inter-station correlation and/or covariance matrix) do not collectively account for a significant percent of the total variance shared by stations.
3. Common factor analysis approach has capability to account for specific variance that is related to local forcing which influences rainfall variability at individual gauges in the study area.
4. Cluster analysis based regionalization approach appears to be appealing owing to its effectiveness in interpreting patterns in data of explanatory variables influencing rainfall, or in leading PCs resulting from PCA.
5. Regions delineated using conventional regionalization procedures (e.g., correlation analysis, principal component analysis; common factor analysis) may not be reliable when the study area is sparsely gauged and the period of contemporaneous records for the gauges considered is inadequate to account for sampling variability.
6. Regions must be validated for homogeneity. The validation could be done using rainfall statistics that are not closely related to the statistics which form basis to arrive at the regions. Delineation of regions using large scale atmospheric (climate) variables and validating the regions using rainfall statistics could be one such strategy.
7. Regions need not be geographically contiguous.
8. Formation of homogeneous rainfall regions with soft (fuzzy) boundaries would be meaningful.

Received 14 May 2013.

References

1. Blanford H F, *India Meteorological Memoirs*, Government of India, 1884.
2. Parthasarathy B, Interannual and long-term variability of Indian summer monsoon rainfall, *Proceedings of the Indian Academy of Sciences* (*Earth and Planetary Sciences*), 93(4), (1984) 371–385.
3. Satyanarayana P and Srinivas V V, Regionalization of precipitation in data sparse areas using large scale atmospheric variables—A fuzzy clustering approach, *Journal of Hydrology*, 405(3–4), (2011) 462–473.
4. Gadgil S, Gowri R and Yadumani, Coherent rainfall zones: case study for Karnataka, *Proceedings of the Indian Academy of Sciences* (*Earth and Planetary Sciences*), 97(1), (1988), 63–79.
5. Gadgil S, Yadumani and Joshi N V, Coherent rainfall zones of the Indian region. *International Journal of Climatology*, 13, (1993) 547–566.
6. Parthasarathy B, Kumar R K, Munot A A, Homogeneous Indian monsoon rainfall: variability and prediction. *Proceedings of the Indian Academy of Sciences* (*Earth and Planetary Sciences*), 102(1), (1993) 121–155.
7. Shukla J, Interannual variability of monsoons; In: Monsoons (eds) J S Fein and P L Stephens, Wiley and sons, New York, Chapter 14, 399–464, 1987.
8. Parthasarathy B, Munot A A, Kothwale D R, Monthly and Seasonal Rainfall Series for All-India Homogeneous Regions and Meteorological Subdivision: 1871–1994, *Research Report No. RR-065*, Indian Institute of Tropical Meterology, Pune, 1995.
9. Satyanarayana P and Srinivas V V, Regional frequency analysis of precipitation using large-scale atmospheric variables, *Journal of Geophysical Research—Atmospheres*, 113, (2008) D24110, doi:10.1029/2008JD010412.
10. Azad S, Vignesh T S and Narasimha R, Periodicities in Indian monsoon rainfall over spectrally homogeneous regions, *International Journal of Climatology*, 30(15), (2010) 2289–2298.
11. McQuitty L L, Elementary linkage analysis for isolating orthogonal and oblique types and typal relevances, *Educational and Psychological Measurement*, 17(2), (1957) 207–229.
12. Gregory S, Rainfall over Sierra-Leone, *Research Paper No. 2*, Department of Geography, University of Liverpool, 1965, 55 pp.
13. Jackson I J, The spatial correlation of fluctuations in rainfall over Tanzania: A preliminary analysis. Archives for Meteorology, Geophysics, and Bioclimatology, Series B, 20(2), (1972) 167–178.
14. Gregory S, On the delimitation of regional patterns of recent climatic fluctuations. *Weather*, 30(9), (1975) 276–287.

15. Adelekan I O, Spatio-temporal variations in thunderstorm rainfall over Nigeria, *International Journal of Climatology*, 18(11), (1998) 1273–1284.
16. Satyanarayana P, Regional frequency analysis of hydrometeorological events—An approach based on climate information, PhD Thesis, Indian Institute of Science, Bangalore, 2009.
17. Saikranthi K, Rao T N, Rajeevan M, and Rao S V B, Identification and validation of homogeneous rainfall zones in India using correlation analysis, *Journal of Hydrometeorology*, 14(1), (2013) 304–317, DOI: 10.1175/JHM-D-12–071.1.
18. García-Marín A P, Ayuso-Munoz J L, Taguas-Ruiz E V and Estevez J, Regional analysis of the annual maximum daily rainfall in the province of Malaga (southern Spain) using the principal component analysis, *Water and Environment Journal*, 25(4), (2011) 522–531.
19. Wotling G, Bouvier Ch., Danloux J, and Fritsch J-M, Regionalization of extreme precipitation distribution using the principal components of the topographical environment, *Journal of Hydrology*, 233(1–4), (2000) 86–101.
20. Dyer T G J, The assignment of rainfall stations into homogeneous groups: An application of principal component analysis, *Quarterly Journal of the Royal Meteorological Society*, 101(430), (1975) 1005–1013.
21. Gregory J M, Jones P D, and Wigley T M L, Precipitation in Britain: an analysis of area-average data updated to 1989, *International Journal of Climatology*, 11(3), (1991) 331–345.
22. Richman M B and Lamb P J, Climatic pattern analysis of three- and seven-day summer rainfall in the central United States: Some methodological considerations and a regionalization, *Journal of Climate and Applied Meteorology*, 24(12), (1985) 1325–1343.
23. Comrie A C and Glenn E C, Principal components-based regionalization of precipitation regimes across the southwest United States and northern Mexico, with an application to monsoon precipitation variability, *Climate Research*, 10, (1998), 201–215.
24. Goossens C, Principal component analysis of Mediterranean rainfall. *Journal of Climatology*, 5(4), (1985) 379–388.
25. Ehrendorfer M, A regionalization of Austria's precipitation climate using principal component analysis, *Journal of Climatology*, 7(1), (1987) 71–89.
26. Pandzic K, Principal component analysis of precipitation in the Adriatic-Pannonian area or Yugoslavia, *Journal of Climatology*, 8(4), (1988) 357–370.
27. Iyengar R N and Basak P, Regionalization of Indian monsoon rainfall and long-term variability signals, *International Journal of Climatology*, 14(10), (1994) 1095–1114.
28. Singh K K and Singh S V, Space time variation and regionalization of seasonal and monthly summer monsoon rainfall of the sub-Himalayan region and Gangetic plains of India, *Climate Research*, 6(3), (1996) 251–262.
29. Baeriswyl P A and Rebetez M, Regionalization of precipitation in Switzerland by means of principal component analysis. *Theoretical and Applied Climatology*, 58(1–2), (1997) 31–41.
30. Brunetti M, Maugeri M, Nanni T and Navarra A, Droughts and extreme events in regional daily Italian precipitation series, *International Journal of Climatology*, 22(5), (2002) 543–558.
31. Llasat M C, Ceperuelo M and Rigo T, Rainfall regionalization on the basis of the precipitation convective features using a raingauge network and weather radar observations, *Atmospheric Research*, 83(2–4), (2007) 415–426.
32. Bedi H S, Bindra M M S, Principal components of monsoon rainfall. Tellus 32, (1980) 296–298.
33. Carter M M, Convective rainfall regions in Puerto Rico, *MS Thesis*, *The Florida State University*, 1995.
34. Carter M M and Elsner J B, A statistical method for forecasting rainfall over Puerto Rico, *Weather and forecasting*, 12(3), (1997) 515–525.
35. Bärring, L. Spatial patterns of daily rainfall in central Kenya: application of principal component analysis and spatial correlation. *Journal of Climatology*, 7(3), (1987) 267–289.
36. Easterling D R, Regionalization of thunderstorm rainfall in the contiguous United States. *International Journal of Climatology*, 9(6), (1989) 567–579.
37. Guttman N B, The use of *L*-moments in the determination of regional precipitation climates, *Journal of Climate*, 6(12), (1993) 2309–2325.
38. Trefry C M, Watkins Jr. D W, Johnson D, Regional rainfall frequency analysis for the state of Michigan, *Journal of Hydrologic Engineering*, 10(6), (2005) 437–449.
39. Periago M C, Lana X, Serra C, Mills G F, Precipitation regionalization: An application using a meteorological network in Catalonia (NE Spain), *International Journal of Climatology*, 11(5), (1991) 529–543.
40. Mills G F, Principal component analysis of precipitation and rainfall regionalization in Spain, *Theoretical Applied Climatology*, 50(3–4), (1995) 169–183.
41. Bonell M and Sumner G, Autumn and winter daily precipitation areas in Wales, 1982–1983 to 1986–1987, *International Journal of Climatology*, 12(1), (1992) 77–102.
42. Domroes M, Kaviani M and Schaefer D, An analysis of regional and intra-annual precipitation variability over Iran using multivariate statistical Methods, *Theoretical and Applied Climatology*, 61(3–4), (1998) 151–159.
43. Dinpashoh Y, Fakheri-Fard A, Moghaddam M, Jahanbakhsh S and Mirnia M, Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods, *Journal of Hydrology*, 297(1–4), (2004) 109–123.
44. Holawe F and Dutter R, Geostatistical study of precipitation series in Austria: time and space, *Journal of Hydrology*, 219(1–2), (1999) 70–82.
45. Mimmack G M, Mason S J and Galpin J S, Choice of distance matrices in cluster analysis: Defining regions, *Journal of Climate*, 14(12), (2001) 2790–2797.

46. Obregón G O and Nobre C A, Rainfall regionalization on the Amazon basin. Proceedings of 8th International Conference on Southern Hemisphere Meteorology and Oceanography (ICSHMO), Foz do Iguaçu, Brazil, INPE, (2006) 1149–1152.
47. Gadgil S and Iyengar R N, Cluster analysis of rainfall stations of the Indian peninsula, *Quarterly Journal of the Royal Meteorological Society*, 106(450), (1980) 873–886.
48. Kulkarni A, Kripalani R H, Singh S V, Classification of summer monsoon rainfall patterns over India. *International Journal of Climatology*, 12, (1992) 269–280.
49. Kulkarni B S and Reddy D D, The cluster analysis approach for classification of Andhra Pradesh on the basis of rainfall, *Mausam*, 45(4), (1994) 325–332.
50. Venkatesh B and Jose M K, Identification of homogeneous rainfall regimes in parts of Western Ghats region of Karnataka, *Journal of Earth System Science*, 116(4), (2007) 321–329.
51. Ahuja S and Dhanya C T, Regionalization of Rainfall Using RCDA Cluster Ensemble Algorithm in India, *Journal of Software Engineering and Applications*, 5, (2012) 568–573.
52. Yang T, Shao Q, Hao Z-C, Chen X, Zhang Z, Xu C-Y, Sun L, Regional frequency analysis and spatio-temporal pattern characterization of rainfall extremes in the Pearl River Basin, China. *Journal of Hydrology*, 380(3–4), (2010) 386–405.
53. Gabriele S and Arnell N, A hierarchical approach to regional flood frequency analysis, *Water Resources Research*, 27(6), (1991) 1281–1289.
54. Alila Y, A hierarchical approach for the regionalization of precipitation annual maxima in Canada, *Journal of Geophysical Research*, 104(D24), (1999) 31, 645–31, 655.
55. Burn D H, Evaluation of regional flood frequency analysis with a region of influence approach, *Water Resources Research*, 26(10), (1990) 2257–2265.
56. Gaál L, Kyselý J, Szolgay J, Region-of-influence approach to a frequency analysis of heavy precipitation in Slovakia, *Hydrology and Earth System Sciences*, 12(3), (2008) 825–839.
57. Hosking J R M, Wallis J R, Regional frequency analysis: an approach based on *L*-moments. Cambridge University Press, New York, USA, 1997.
58. Rajeevan M, Bhate J, Kale J D, Lal B, Development of a high resolution daily gridded rainfall data for the Indian Region (Version 2). India Meteorological Department, India, Met. Monograph Climatology No. 22/2005, 2005.
59. Rajeevan M, Bhate J, Kale J D, Lal B, High resolution daily gridded rainfall data for the Indian region: Analysis of break and active monsoon spells. *Current Science*, 91(3), (2006) 296–306.
60. Shepard D, A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of ACM National Conference*, (1968) 517–524.
61. Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo K C, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D, The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3), (1996) 437–471.
62. Kanamitsu M, Ebisuzaki, W. J., Woollen, S -K, Yang, Hnilo J J, Fiorino M, and Potter G L, NCEP-DEO AMIP-II Reanalysis (R-2), *Bulletin of the American Meteorological Society*, 83(11), (2002) 1631–1643.
63. Gibson J K, Kållberg P, Uppala S, Hernandez A, Nomura A, and Serrano E, ERA-15 Description (Version 2), In ECMWF Re-Analysis Project Report Series, European Centre for Medium-Range Weather Forecasts, Shinfield, Reading, UK, 1999.
64. Uppala S M, Kållberg P W, Simmons A J, Andrae U, da Costa Bechtold V, Fiorino M, Gibson J K, Haseler J, Hernandez A, Kelly G A, Li X, Onogi K, Saarinen S, Sokka N, Allan R P, Andersson E, Arpe K, Balmaseda M A, Beljaars A C M, van de Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Hólm E, Hoskins B J, Isaksen L, Janssen P A E M, Jenne R, McNally A P, Mahfouf J -F, Morcrette J -J, Rayner N A, Saunders R W, Simon P, Sterl A, Trenberth, K E, Untch A, Vasiljevic D, Viterbo P and Woollen J, The ERA-40 re-analysis. *Quarterly Journal of The Royal Meteorological Society*, 131(612), (2005) 2961–3012.
65. Onogi K, Tsutsui J, Koide H, Sakamoto M, Kobayashi S, Hatsushika H, Matsumoto T, Yamazaki N, Kamahori H, Takahashi K, Kadokura S, Wada K, Kato K, Oyama R, Ose T, Mannoji N and Taira R, The JRA-25 Reanalysis, *J. Meteor. Soc. Japan*, 85, (2007) 369–432.
66. Goswami B N, Venugopal V, Sengupta D, Madhusoodanan M S, Xavier P K, Increasing trend of extreme rain events over India in a warming environment. *Science*, 314, (2006) 1442–1445.

Dr. V.V. Srinivas is Associate Professor in the Department of Civil Engineering, Indian Institute of Science, Bangalore. His research interests include Stochastic Surface Water Hydrology, Climate Hydrology, and Regionalization of Hydro-meteorological variables. He has co-authored a book titled "Regionalization of Watersheds—An Approach Based on Cluster Analysis".