

Characterization of Precipitation in the Subdivisions of the Mahanadi River Basin, India

Ramgopal T Sahu*, Mani Kant Verma and Ishtiyahq Ahmad

*Civil Engineering Department, National Institute of Technology - Raipur, Chhattisgarh, India****Corresponding Author:** Ramgopal T Sahu, Civil Engineering Department, National Institute of Technology - Raipur, Chhattisgarh, India.**Received:** August 31, 2021**Published:** November 30, 2021© All rights are reserved by **Ramgopal T Sahu, et al.****Abstract****Purpose:** To study the homogeneity and precipitation characterization of the subdivisions of the Mahanadi River basin, India.**Methodology/Approach:** Precipitation characterization was studied using spatial pattern analysis (Eigen-based technique), approaching principal component analysis and keeping three equal timeslot segments for better understanding change. Homogeneity characterization of the subdivisions is accomplished using L-moment and probability weighted moments (PWMs) based heterogeneity measures. Divisive hierarchical cluster analysis for effective partitioning of the subdivisions.**Findings:** Heterogeneity (H1) assessment; Upper subdivision U1 (42sites) with H1 - 1.74 (possibly heterogeneous), Middle subdivision M1 (76sites) with H1 - 0.80 (acceptably homogeneous), Two splits of lower subdivision (1) L1 (45sites) with H1 - 0.68 (acceptably homogeneous), (2) L2 (34sites) with H1 - 1.02 (acceptably homogeneous), All the subdivisions were heterogeneous to each other during all timeslot segments with p-value < 0.021, the lower subdivision observed $\cong 35\%$ similarity during time slot 1940-1978 and 1979-2017 with p-value - 0.349. The iqr width of the middle subdivision does not change significantly but has a dropping trend suggesting a decrease in precipitation magnitude, and the measure of spread has significantly reduced with some outliers suggesting non-uniformity during segment 3. The lower subdivision with longest rainy months May-October, while November month observed a decreasing magnitude with time.**Implications:** The characterized precipitation of the subdivisions identified has separate applications and uses. In areas with known precipitation variability, future precipitation forecasting can be abridged for water resource management. Scheduling different crops, agriculture planning, rain-fed and dry farming, and farming calendars are various activities for known precipitation regimes.**Importance:** The findings are applicable to water resource planning and management in both a practical and scientific manner.**Keywords:** Characterization; DIANA; Divisive Analysis; Eigenvalue; Precipitation; Spatial Pattern**Introduction**

Process complexity occurs when precipitation variability is complex. Moreover, it is assessed as a specific purpose when a long-term or extreme property is temporally assembled for the purpose of delineating the regions. Therefore, homogenous regions delineated for different purposes do not necessarily have to be identical, and the definition of regional homogeneity is not

unique [49]. As a result, spatial events must be grouped spatially, where spatial analysis is an important technique [52]. In a general process of the hydrological cycle, water evaporates into the atmosphere and the amount increases with the increase in temperature. As a result, the spatial and temporal form of changes appears in the patterns. Global warming has indisputably accelerated the water cycle and had a significant impact on water resources, result-

ing in more frequent and unexpected climate and weather events (heat waves, drought, and heavy precipitation, etc.). To discover the characterization of precipitation magnitude and variability in the subdivisions of the Mahanadi River basin, eigen-based spatial pattern analysis for identifying patterns of precipitation variability was employed.

The motivating subject for this study is phenological instabilities (e.g., uncertain, sudden, increases, and decreases) in rainfall that appear from season to season. The increasing impact of rainfall events on health and safety is affecting not only humans, but the ecosystem as well. Changing trends have altered rainfall frequency and intensity around the world. Uncertainty and erratic precipitation patterns have a negative impact on the agricultural production of dependent farmers and other stakeholders who are directly or indirectly involved. Assessing the risks of flooding and drought, as well as water resources planning, and management are the scientific and practical importance of the study.

The measure of heterogeneity is based on L-moment and probability-weighted moments (PWM) to evaluate the homogeneity of the subdivisions with a view to effective regionalization [18,21,22,38]. Over the past few decades, several regionalization techniques cited by Srinivas [49] together with Principal Component Analysis [17,53], Elementary Linkage Analysis [27], PCA with Cluster Analysis [2], Spectral Analysis [4], Hierarchical Cluster Analysis [3,14], Region of Influence [7,8], and Common Factor Analysis [10] have developed. The meteorological subdivisions of India (currently 36) were investigated using various regression and classification techniques and found heterogeneous and negative correlations for most MSDs [4,14-16,29-31,35,36,44,45,47]. Bharath and Srinivas [6] studied the Indian region for effective regionalization using a GFCM-based RFA approach and a gridded rainfall dataset of resolution $0.5^\circ \times 0.5^\circ$. The effectiveness was assessed using the ROI approach, and the quantiles were found to be in line with the ROI approach. Later, high resolution of $0.25^\circ \times 0.25^\circ$ gridded rainfall data was assessed by Sahu, *et al.* [41] for the Mahanadi River basin to discover the efficacy of the maximum loading value approach (MLV) for precipitation regionalization.

The Mahanadi River basin is a point of interest and has been studied by numerous researchers using different datasets and methods. Singh, *et al.* [48] discovered significant spatiotemporal variation to understand precipitation characteristics and patterns

in the agro-climatic zone of the middle Mahanadi River basin on a regional scale. Sahu, *et al.* [43] discovered significance among the ENSO and ENSO modoki during (June, July, and August) and (September, October, and November) on river boundaries while studying the combined effect of the Indian monsoon and Indo-Pacific climatic modes. Azharuddin, *et al.* [5] conducted similar research for Seonath River basin. The Seonath (upper subdivision) basin is the largest tributaries of the Mahanadi River basin and is investigated for precipitation variability and trends of climate data (precipitation and temperature) [13,50,51]. Later, Sahu, *et al.* [39] studied density-based spatial clustering of applications with noise clustering for precipitation regionalization, and the effectiveness was assessed using unsupervised hierarchical cluster analysis. Study apart from the Mahanadi River basin includes, the upper Godavari River basin [1], Poyang Lake basin [55], Huai River basin [54], Hongshui River basin [23], and Iran [37].

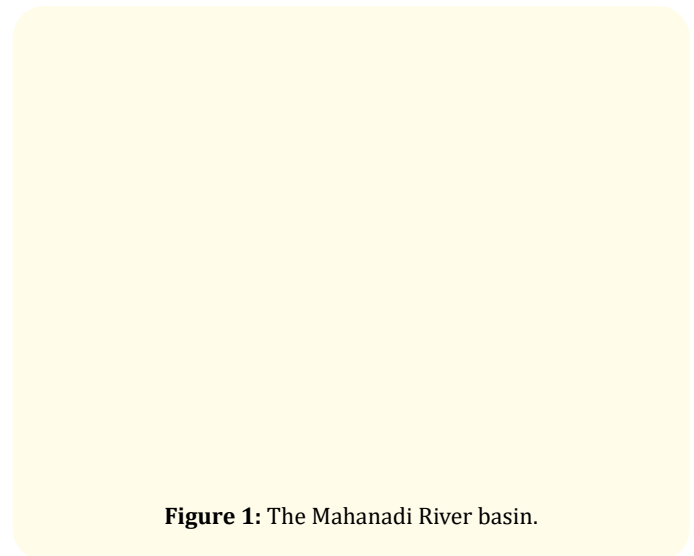


Figure 1: The Mahanadi River basin.

Study area and data

The Mahanadi is a major east-flowing river in India (see Figure 1), located east-centrally having a geographical coordinate $19^\circ 20' N$ to $23^\circ 35' N$ latitude and $80^\circ 30' E$ to $86^\circ 50' E$ longitude. The thalweg is 858 kilometres long from origin to false point, with a catchment area of 141,600 km². More than 82% of the annual rainfall encounters during the summer monsoon. Typical hydrology of the Mahanadi River as it is an eastern-current peninsular river with the river length extending to 858km from the origin to its false point. The primary tributaries covering the maximum catchment area are the

Seonath River, the Ib River, and the Tel River, which is approx. 47%. High resolution of 0.25° x 0.25° gridded rainfall data obtained from the Indian Meteorological Department, Pune. The data is prepared using records of 2140 selected stations from 6329 available stations in India based on interpolation schemes [33,34,46]. The data for DIANA is an assembled dataset includes, mean annual, monthly max, monthly mean, location attributes, distance from western sea coast at (0°, 15°, 30°, 45° from west to south), monthly proportion to annual, summer monsoon (JJA) proportion to annual, winter monsoon (DJF) proportion to annual, a ratio of the summer monsoon (JJA) to winter monsoon (DJF), warmer season (AMJJAS) proportion to annual and colder season (ONDJFM) proportion to annual, and a ratio of the warmer season (AMJJAS) to Colder season (ONDJFM) for 230 station points in [N x M] matrix where M = 230 refers to station points and N = 54 refers to the observations of variables.

SRTM - DEM, R - programming language, and ArcGIS are other tools and data used for studying and analysing the stated purpose of the concern objective.

Methodology

Effective clustering and regionalization

Unsupervised machine learning is used to cluster grid points using the DIANA - Divisive Analysis (Divisive hierarchical algorithm) proposed by Kaufman and Rousseeuw [24] based on the splitting concept proposed by Macnaughton-Smith, *et al.* [26]. DIANA works on identifying an object (in any cluster *c*), that has the maximum average distance and accumulates all objects with the closest similarity to that object. There exist (2ⁿ⁻¹-1) possibilities to split data into two clusters with O(n²) time complexity (exhaustive search) and this process is repeated until the last object is separated.

$$\text{(Maximum Average Distance)} \bar{d}_i = \frac{1}{n-1} \sum_{j=1}^n d(x_i - x_j) \text{----(1)}$$

Where,

d(x_i-x_j) denotes similarity distance between the object x_i and x_j.

Divisive coefficient: A measure to estimate cluster quality (strength of the cluster structure) is denoted as DC and defined as

$$DC = \frac{\sum_{i=1}^n \Delta(x_i)}{n} \text{----- (2)}$$

Here Δ(x) denotes the diameter of the least cluster to which it belongs.

Statistical test: L-moments [21] and probability-weighted moments [18] based homogeneity assessment. The H-statistics [22] based heterogeneity measure and the following are the criteria for determining whether a region is homogeneous or heterogeneous: “H₁ < 1 for acceptable homogeneous, 1 < H₁ < 2 for possible heterogeneous/homogeneous, and H₁ > 2 for acceptable heterogeneous.” It is a primary indicator of whether to accept or reject a region as homogenous and is expressed in equation (3).

$$H = \frac{v-\mu}{\sigma} \text{----- (3)}$$

Measure of discordance: this is a secondary indicator to accept or eliminate a site whose L-moment ratio is different from its L-moment grouped ratios. The parameter u_i and \bar{u} represent vectors of sample l-moment i.e., t⁽ⁱ⁾, t₃⁽ⁱ⁾ and t₄⁽ⁱ⁾ and unweighted regional average, respectively see equation (4).

$$u_i = [t^{(i)}, t_3^{(i)}, t_4^{(i)}] \text{ and } \bar{u} = N^{-1} \sum_{i=1}^n u_i \text{----- (4)}$$

D_i denotes critical value that describes a criticality line for any site to remain within the group. According to Hosking and Wallis [21], the standard criticality is Di < 3.

$$D_i = \frac{1}{3} (u_i - \bar{u})^T S^{-1} (u_i - \bar{u}) \text{-----(5)}$$

Here S can be stated as S = $\sum_{i=1}^N (u_i - \bar{u})(u_i - \bar{u})^T$

Castellarin, *et al.* [11] proposed another measure (if necessary) to improve the homogeneity or heterogeneity of the proposed region: the measure of heterogeneity adjustment and is assessed by cross-correlation among the sites of the proposed group and expressed as:

$$H_{adj} = H + 0.122 \times \overline{\rho^2} (N_c - 1) \text{-----(6)}$$

$\overline{\rho^2}$ denotes “Mean of the concurrent precipitation cross-correlation squares for cluster sites”

N_c denotes ‘Number of sites in the region’.

Spatial pattern analysis (Eigen-based technique) for precipitation time variability

Time-variability patterns of precipitation across the basin for a time series of monthly relative precipitation subjected to spatial pattern analysis (eigen-based technique) [20,32]. The perception of the method is to distinguish stations with analogous annual pre-

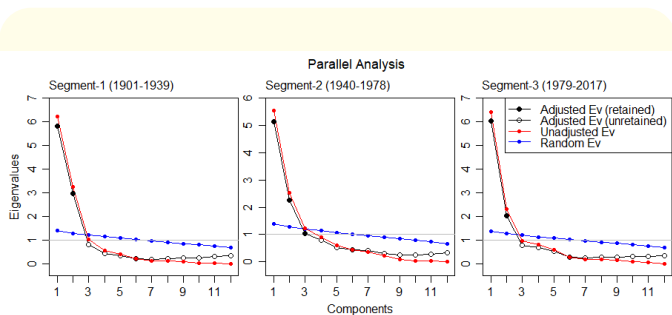


Figure 2: Eigenvalue decomposition of the correlation matrix for the three timeslots with the application of Horn's parallel analysis for component retention.

precipitation features, irrespective of their magnitude. The spectacle used in the method is a kind of regulation to find out the overweighing of the high precipitation values observed at different locations. A partitioning approach is needed to understand the spectacle of rainfall variability across the basin. The length of monthly relative precipitation is partitioned into 3 segments of 39 years, including 1901-1939, 1940-1978, and 1979-2017, respectively for the 1st, 2nd, and 3rd segment. The eigenvalue decomposition signifies the same number of components as variables in the transformed dataset. Furthermore, further analysis is required to decide which components cover the maximum amount of variance. To determine those components, the scree plot [9] and parallel analysis [19] are used. The sampling error for assessed eigenvalues is intended using North's thumb rule [28]. To obtain more localised statistics useful for pattern interpretation, the retained components of the spatial pattern analysis were rotated orthogonally using the varimax criteria. The components that result are the leading rotated loadings on a few variables and the non-zero loadings on others.

In terms of PCA, a scree plot is a graphical representation of the eigenvalue to the eigenvector, whose optimism lies in its elbow. Interpretation is the major disadvantage of the scree plot, which is spontaneous and random, including two or three components. Hence, the feasibility of deciding the optimum number of components cannot be accomplished using a single criterion because the decomposed eigenvalues are generally biased and contain errors. A parallel analysis involves adjusting the bias and errors as an enhancement to the scree plot for better interpretation and accounting for the feasibility of the eigenvalue. "The rationale for the parallel analysis is to figure out factors that account for maximum

variance, whose probabilistic existence is not by chance" [19]. "If the sampling error of an eigenvalue has a magnitude comparable to or greater than the linear spacing relative to its nearest eigenvalue, then those two eigenvectors are mixed" [28]. Lawley's formula [25] is suggested in the following equation - 7 for estimating the standard error.

$$\Delta\lambda_i = \left(\frac{2}{n^*}\right)^{1/2} \lambda_i \text{ -----(7)}$$

Where,

n* denotes effective sample size,

λ_i denotes eigenvalue for ith order.

Equation - 7 suggests, "the error associated with each eigenvalue depends on the eigenvalue itself and the effective sample size, which means the error is directly proportional to the eigenvalue and inversely proportional to effective sample sizes" [25]. Furthermore, the error limit associated with each eigenvalue at 95 percent CI is referred to as an Error= $\lambda_i \pm 1.96\Delta\lambda_i$.

Re- gion	Re- gion Size	Sub- Region	Heterogeneity			Nature
			H ₁	H ₂	H ₃	
1	(53)	Upper	5.22	-1.24	-1.59	Acceptably Heterogeneous
2	(82)	Middle	2.19	-0.25	-1.06	Acceptably Heterogeneous
3	(95)	Lower	11.52	1.36	0.18	Acceptably Heterogeneous

Table 1: Initial assessment of subdivisions for heterogeneity, and nature of measure for 230 station point (grid) in the Mahanadi River basin (MRB).

Results and Discussion

Statistics of cluster partition and Heterogeneity assessment

Initial assessment of subdivisions with 53, 82, and 95 station points, respectively, for upper, middle, and lower subdivisions. The heterogeneity assessment of the same is presented in table 1. The implementation results are acceptably heterogeneous for all the subdivisions of the Mahanadi River basin. The upper and middle subdivisions were assessed as possibly heterogeneous and acceptably homogeneous, whereas they were acceptably heterogeneous

for the lower subdivisions after the heterogeneity adjustment and discordancy measure. The lower subdivision is then applied to DI-ANA (DIvisive ANALysis) Divisive hierarchical analysis, which yield-

ed two subclusters, the latter of which was found to be acceptably homogeneous after discordancy and heterogeneity assessments (see Table 2 and Figure 9).

Region	Region Size	Sub-Region	HOM Region	D _{critical}	Heterogeneity			Discor-dant sites	Nature
					H ₁	H ₂	H ₃		
1	(53)	Upper	42	3.000	1.74	-2.32	-3.17	11	Possibly Heterogeneous
2	(82)	Middle	76	3.000	0.80	-1.10	-2.32	6	Acceptably Homogeneous
3	(95)	Lower	48	3.000	0.68	-1.80	-3.39	6	Acceptably Homogeneous
			31	3.000	1.02	0.11	1.21	10	Acceptably Homogeneous

Table 2: Results of heterogeneity, discordancy, and nature of homogeneity for 230 station point (grid) in the Mahanadi River basin (MRB).

Statistics of spatial pattern analysis (Eigen-based technique)

The monthly relative time series is subjected to spatial pattern analysis to investigate the patterns of precipitation variability in different timeslots across the basin. Table 3 shows the results of eigenvalue decomposition, parallel analysis, and varimax rotated loadings, which show that segments 1 (1901-1939), 2 (1940-1978), and 3 (1979-2017) account for 79.00%, 77.57%, and 72.60% of to-

tal variance, respectively. In addition, component retention analysis (parallel analysis) discovered 2, 3, and 2 components, respectively, for segments 1, 2, and 3 (see Figure 2). The spatial distribution in figure 3 depicts the rotated loadings of retained PC scores, indicating that two components from segments 1 and 3, and three from segment 2, are relatively sufficient for explaining the monthly relative precipitation. The results obtained in this section were comparatively similar to those reported by Sahu, *et al.* [40].

Segment	Component (1)	Eigenvalue (2)	Sampling Error		Parallel Analysis		Un-rotated		Varimax Rotated	
			Low-Eigen (3)	High-Eigen (4)	Adjusted Eigenvalue (5)	Estimated Bias (6)	Variance (7)	Cumulative Variance (8)	Variance (9)	Cumulative Variance (10)
1901-1939	1	6.219	5.639	6.799	5.829	0.390	51.831	51.831	42.930	42.930
	2	3.259	2.955	3.563	2.976	0.283	27.166	78.997	36.067	78.997
1940-1978	1	5.542	5.025	6.059	5.158	0.383	46.187	46.187	30.351	30.351
	2	2.533	2.297	2.769	2.253	0.280	21.111	67.298	31.359	61.710
	3	1.232	1.117	1.347	1.024	0.208	10.274	77.573	15.863	77.573
1979-2017	1	6.420	5.821	7.019	6.026	0.393	53.503	53.503	44.510	44.510
	2	2.292	2.078	2.505	2.013	0.278	19.101	72.605	28.095	72.605

Columns 2-4 and 7-8 are the result of eigenvalue decomposition

Table 3: Retained components from all timeslots were investigated for precipitation variability. Statistics including sampling error, parallel analysis, un-rotated, and varimax-rotated were examined for interpreting different timeslot variability.

According to the result statistics in segment 1, the spatial pattern variability of spring (Apr and May), autumn (Sept, Oct and Nov), and early winter (Dec) has negative loadings, whereas summer (Jul and Aug) has positive loadings on component 1, accounting for 42.93 percent of the total variance. The Mahanadi’s upper

(U1), middle (M1), and lower (L1) subdivisions have negative loading, indicating reasonable precipitation during autumn (Sep, Oct, and Nov) and spring (Apr and May), as well as representative rainfall during early winter (Dec). The lower (L2) subdivision of Mahanadi has high positive loadings, indicating reasonable summer (Jul

and Aug) precipitation. For component 2, the spatial pattern variability of winter (Jan and Feb) and early spring (Mar) features positive loadings, whereas early summer (June) with negative loadings accounts for 36.07% of the total variance. The north Mahanadi, in-

cluding the north of U1 and M1 feature with positives loadings, receives reasonable precipitation during winter (Jan and Feb). During early summer (Jun) and spring (Mar), the rest of the Mahanadi (L1, L2, and south of U1 and M1) with negative loading, receives representative precipitation (see Tables 3 and 4).



Figure 3: The spatial interpolation of rotated loadings of PC scores, obtained by implementing monthly relative precipitation to spatial pattern analysis.

Similarly, for segment 2, the statistics indicate that component 1 accounts for 30.35 percent of the total variance, implying spatial pattern variability with negative loadings in the mid-autumn (Oct) and spring (Apr and May). Summer spatial pattern variability (Jul and Aug) has positive loadings. The lower Mahanadi (L2) has positive loading and receives practical rainfall during summer (Jul

and Aug), whereas the rest (U1, M1, and L1) have negative loading and receive reasonable precipitation during mid-autumn (Oct) and spring (Apr and May). Component 2 accounts for 31.36% of the total variance, suggesting the spatial pattern variability of early autumn (Sept) with negative loading adds representative precipitation to the annual total. Winter (Jan and Feb) and early spring

(Mar) with positive loading implying reasonable precipitation. The north-east and south-east Mahanadi (L2, and north of U1, M1, and L1) receive practical rainfall during early spring (Mar) and winter (Jan and Feb) with positive loading. The south-west Mahanadi (south of U1, M1, and L1) witnesses representative precipitation during early autumn (Sep) with negative loadings (see Figures 3 and 9). For component 3, a combo spatial pattern variability of early summer (Jun), late-autumn (Nov) and early-winter (Dec) accounts for 15.86% of the total variance, suggesting representative precipitation during summer (Jun) and a retreating monsoon during late-autumn (Nov) and early-winter (Dec). The middle and upper subdivisions (U1, M1, and north L1) feature negative loading, suggesting reasonable precipitation during the retreating monsoon (i.e., Nov and Dec), while the lower subdivision (L2), excluding the coastal reach around Chilika Lake, has positive loadings, signifying representative precipitation during early summer (Jun), see tables 3 and 4.

The two components of segment 3 account for 44.509% and 28.089% of the total variance, respectively, for which the spatial pattern variability of component 1 includes spring (Mar, Apr, and May) and autumn (Oct and Nov) with high negative loadings, suggesting feasible rainfall. Whereas summer (Jun, Jul, and Aug) features high positive loadings, implying reasonable rainfall. The upper (U1), middle (M1), and lower (L1) subdivisions of Mahanadi experience reasonable precipitation during autumn (Oct and Nov) and spring (Mar, Apr, and May) with high negative loadings. The lower (L2) portion experiences relatively high precipitation with high positive loadings during the summer (Jun, Jul, and Aug). Early-autumn (Sept) and winter (Dec, Jan, and Feb) explain the spatial pattern variability of component 2 and account for 28.089% of the total variance. The northern Mahanadi, including north of the upper (U1) and middle (M1) subdivisions, features negative loading, indicating representative precipitation adds to the annual total during early autumn (Sep). The lower (L2) and south of upper (U1) and middle (M1) subdivisions of Mahanadi experience practical rainfall with positive loadings during winter (Dec, Jan, and Feb). Refer to tables 3 and 4, and figures 3 and 9.

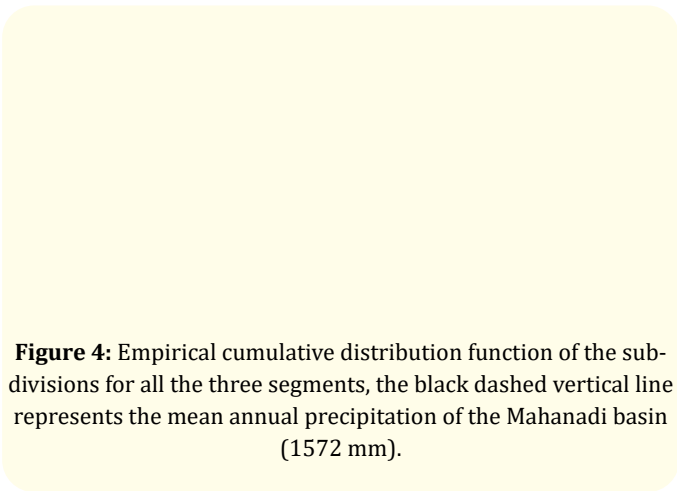


Figure 4: Empirical cumulative distribution function of the subdivisions for all the three segments, the black dashed vertical line represents the mean annual precipitation of the Mahanadi basin (1572 mm).

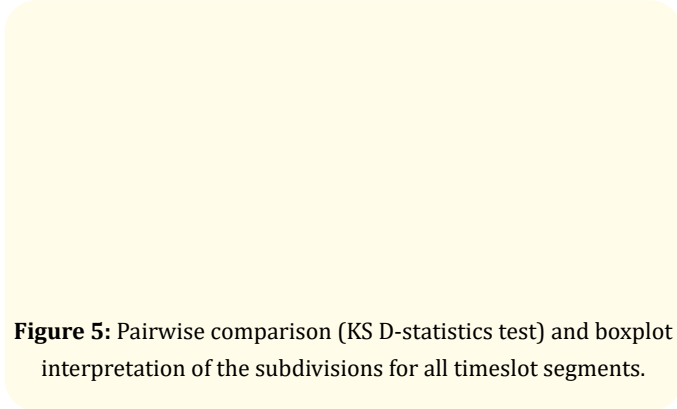


Figure 5: Pairwise comparison (KS D-statistics test) and boxplot interpretation of the subdivisions for all timeslot segments.

Month	Segment - 1 (1901-1939)		Segment - 2 (1940-1978)			Segment - 3 (1979-2017)	
	Comp.1	Comp.2	Comp.1	Comp.2	Comp.3	Comp.1	Comp.2
January	0.11	0.49		0.54	-0.12	0.15	0.56
February		0.51	-0.14	0.57	0.12	-0.17	0.49
March	-0.10	0.42	-0.24	0.28	0.22	-0.30	0.18
April	-0.28	-0.16	-0.41		0.33	-0.34	
May	-0.36		-0.43			-0.38	
June		-0.42			0.64	0.23	-0.18
July	0.38	0.11	0.39		0.11	0.38	0.10
August	0.39		0.41			0.31	-0.11
September	-0.31	-0.20	-0.15	-0.49	0.12	-0.16	-0.20
October	-0.39		-0.34		-0.30	-0.38	
November	-0.38		-0.31		-0.38	-0.36	
December	-0.28	0.25		0.20	-0.36		0.54

Table 4: Statistics of orthogonally rotated loadings of retained PC score obtained by implementing monthly relative precipitation to spatial pattern analysis (Bold value describes dominating component).

Characterization of precipitation patterns

The task of assessing dispersions in annual rainfall of the identified patterns (for all timeslots) observed at different locations (stations) by employing certain attributes, e.g., similarity among the members of a group or between groups. Also refer to Verma, *et al.* [51] for a long term-trend assessment of climate data (precipitation and temperature) for Chhattisgarh state. The median relative distance (see Figures 5 and 7) is playing around the centre, implying a symmetric distribution for all subdivisions in different timeslots. The middle and upper subdivisions, respectively, of segments 1 and 2 tend to have negative skewness, whereas the lower subdivision of segment 1 experiences positive skewness. The naturally defined subdivisions of the basin for different timeslot segments exhibited decent compactness, implying homogeneous precipitation variability amongst subdivisions of the basin. Figures 7b and 7c depict the middle and lower subdivisions with a wider inter-quartile range, indicating a higher measure of spread and surface area. The iqr width also suggests that the middle subdivision has a declining trend in range of precipitation magnitude as well as cumulative precipitation (Figure 7b), whereas the upper subdivision had a sudden fall in cumulative precipitation (Figure 7a). Furthermore, the measure of spread and cumulative precipitation magnitude has both decreased significantly, with some outliers indicating non-uniformity for the middle subdivision during segment 3 (Figure 5).

Distinctness authentication of the identified subdivisions using empirical cumulative distribution functions is illustrated in figures 4 and 6, whereas figures 5 and 7 depict a pairwise comparison of CDFs with the application of the Kolmogorov-Smirnov D-statistics test [12]. Visually, the CDF plots of subdivisions seem alike and close, so the KS D-statistics test for pairwise comparison of CDFs indicates that the subdivisions are distinct with a p-value less than 0.05 (see Figure 5). Different timeslot comparisons of subdivisions as illustrated in figure 7 indicate high heterogeneity among the segment pairs (except for 1 pair). The exempted segment pair is 2-3 for the lower subdivision with a p-value of 0.349 (34.9%).

Figure 8 shows the boxplot interpretation of the patterns of the subdivisions for the dispersions of the monthly precipitation. The interpretation suggests medium to low inter-subdivision variations of the identified precipitation patterns, implying possible to acceptable heterogeneity. The boxplot analysis of the subdivisions for different timeslot segments depicted in figure 8 recommends

Figure 6: Comparing eCDFs of subdivisions to their different segments. The black dashed vertical line represents the mean annual precipitation of the Mahanadi basin (1572 mm).

Figure 7: Pairwise comparison (KS D-statistics test) and boxplot interpretation of timeslot segments of each subdivision.

that the regimes bounding different topographical regions of 230 stations are distinct, well-defined, and compact. The median line in some months has diverged from its mean line, suggesting variation in precipitation magnitude observed at different locations, while the rest are compact and symmetrical.

The rainfall marching months for upper (U1), middle (M1), and lower (L1) subdivisions are June, July, August, and September for all timeslots. The boxplot analysis also suggests a symmetrical distribution of rainfall in the direction of marching month and trailed by a dry season from November to May. The lower (L2) subdivision has the widest inter-quartile range for all timeslots, indicating rapid changes in magnitude and a large measure of spread. In the lower (L2) subdivision, intense precipitation is observed in conjunction with large rainfall marching months that begin in May and end in October. For the southernmost or lower (L2) subdivision of the Mahanadi basin, May to October are the wettest months, indicative

Figure 8: Boxplot of relative monthly precipitation for interpreting monthly march of the regionalized homogeneous patterns.

of the two main rainy seasons, including summer and autumn, with late-spring (May) adding representative rainfall to the annual total. Further, followed by a dry season including late autumn (Nov), winter (Dec, Jan, and Feb), and spring (Mar and Apr). Key findings in Sahu, *et al.* [42] included similar and identical non-uniformities.

Figure 9: Characterized subdivisions of the Mahanadi basin.

Conclusion

Characterization of precipitation and homogeneity of subdivisions of the Mahanadi River basin using Eigen-based spatial pattern analysis and DIANA. The accumulation of changes for a long term of 117 years of high-resolution (0.25° x 0.25°) gridded precipitation data was investigated in different timeslot segments. A regional scale investigation of precipitation variability to assess the characteristics of accumulated changes and the DIANA-based regional analysis for regionalizing the heterogeneous subdivision. The rationale for regionalization is to figure out the spatial dependency and distribution of the grid points. Findings from the investigation are as below:

- DIANA (Divisive Analysis) divided the lower subdivision into two homogeneous groups: (L1) 48 station points with eight discordant sites, and (L2) 31 station points with ten discordant sites.
- Spatial pattern variability in summer (Jun - Aug) is not observed in the middle (M1) subdivision.
- The retreating monsoon (Oct - Dec) was more dominant and observed effectively in the middle (M1) subdivision.

- Lower (L1 and L2) subdivisions with large measures of spread suggest a large surface area and a rapid change in precipitation magnitude. The division also contains outliers, which are an indication of non-uniformities prevailing in precipitation patterns.
- In the lower (L2) subdivision, intense precipitation is observed in conjunction with large rainfall marching months that begin in May and end in October.
- The iqr width also suggests that the upper subdivision has a consistent range of precipitation magnitude in all timeslots but has a sudden fall in cumulative precipitation.
- Spatial pattern variability of summer (Jun - Aug) was not observed and significantly influenced mid-autumn (Oct) in the upper (U1) subdivision for the 3rd timeslot interval.

The three subdivisions are the spontaneous divisions based on natural slopes and terrain formed due to the effect of local topography, long-term climatology, and changing hydrology. The homogeneity of the subdivisions and regions resulting from the DIANA technique is feasible because of the spatial dependency and distribution of grid points. The spatial and temporal form of changes in the patterns of precipitation is generally due to the movement of moist currents and the summer monsoon from the southwest (Indian ocean) to the northeast (Himalayas). The evidence can be visualized in Figures 8a, b, and c, which precisely intimate the variability in magnitude observed at different locations.

The limitation of the investigation is specific to precipitation regionalization and is different for different approaches as well as the size of the characterized region (e.g., spatial resolution or period size). As a result, sensitivity analysis is a must for the statistics resulting from implementation, characterized by the number of grid points and different spatial resolutions over the entire basin.

Acknowledgments

The precipitation data used in this study was provided by the Indian Meteorological Department (IMD), Pune and is highly appreciated. Suggestions and comments from reviewers are greatly acknowledged.

Conflict of Interest

No conflict of interest declared by the authors.

Bibliography

1. Aher Sainath., *et al.* "Spatio-temporal analysis and estimation of rainfall variability in and around upper Godavari River basin, India". *Arabian Journal of Geosciences* 12.22 (2019): 1-16.
2. Ahuja Sangeeta and C T Dhanya. "Regionalization of rainfall using RCDA cluster ensemble algorithm in India". *Journal of Software Engineering and Applications* 5.8 (2012): 568-573.
3. Alila Younes. "A hierarchical approach for the regionalization of precipitation annual maxima in Canada". *Journal of Geophysical Research: Atmospheres* 104.D24 (1999): 31645-31655.
4. Azad Sarita., *et al.* "Periodicities in Indian monsoon rainfall over spectrally homogeneous regions". *International Journal of Climatology* 30.15 (2010): 2289-2298.
5. Azharuddin Mohammed., *et al.* "A Synoptic-Scale Assessment of Flood Events and ENSO—Streamflow Variability in Sheonath River Basin, India". In: Rao CM, Patra KC, Jhajharia D, and Kumari S (eds) *Advanced Modelling and Innovations in Water Resources Engineering*. Springer, Singapore, 176 (2022): 93-104.
6. Bharath R and V V Srinivas. "Regionalization of extreme rainfall in India". *International Journal of Climatology* 35.6 (2015): 1142-1156.
7. Burn Donald H. "Evaluation of regional flood frequency analysis with a region of influence approach". *Water Resources Research* 26.10 (1990): 2257-2265.
8. Burn Donald H. "An appraisal of the "region of influence" approach to flood frequency analysis". *Hydrological Sciences Journal* 35.2 (1990): 149-165.
9. Cattell Raymond B. "The scree test for the number of factors". *Multivariate Behavioral Research* 1.2 (1966): 245-276.
10. Carter M M and J B Elsner. "A statistical method for forecasting rainfall over Puerto Rico". *Weather and Forecasting* 12.3 (1997): 515-525.
11. Castellarin A., *et al.* "Homogeneity testing: How homogeneous do heterogeneous cross-correlated regions seem?". *Journal of Hydrology* 360.1-4 (2008): 67-76.
12. Conover William Jay. *Practical nonparametric statistics*. Vol-350. John Wiley and Sons (1999).

13. Dhiwar Bhupendra K., *et al.* "Identification of Flood Vulnerable Area for Kharun River Basin by GIS Techniques". In: Rao CM, Patra KC, Jhahharia D, and Kumari S (eds) *Advanced Modelling and Innovations in Water Resources Engineering*. Springer, Singapore, 176 (2022): 385-408.
14. Gabriele Salvatore and Nigel Arnell. "A hierarchical approach to regional flood frequency analysis". *Water Resources Research* 27.6 (1991): 1281-1289.
15. Gadgil Sulochana and R Gowri. "Coherent rainfall zones: Case study for Karnataka". *Proceedings of the Indian Academy of Sciences-Earth and Planetary Sciences* 97.1 (1988): 63-79.
16. Gadgil Sulochana and N V Joshi. "Coherent rainfall zones of the Indian region". *International journal of climatology* 13.5 (1993): 547-566.
17. García-Marín A P, *et al.* "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.
18. Greenwood J Arthur, *et al.* "Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form". *Water Resources Research* 15.5 (1979): 1049-1054.
19. Horn John L. "A rationale and test for the number of factors in factor analysis". *Psychometrika* 30.2 (1965): 179-185.
20. Horn R A and C R Johnson. "Matrix Analysis, Cambridge Univ". Press. MR0832183 (1985).
21. Hosking J R M and J R Wallis. "Some statistics useful in regional frequency analysis". *Water Resources Research* 29.2 (1993): 271-281.
22. Hosking J R M and J R Wallis. "Regional Frequency Analysis: An Approach Based on L-Moments". Cambridge University Press. New York, NY (1997).
23. Huang Ya, *et al.* "Spatial and temporal variability in the precipitation concentration in the upper reaches of the Hongshui River basin, southwestern China". *Advances in Meteorology* 2018 (2018).
24. Kaufman Leonard and Peter J Rousseeuw. "Finding groups in data: an introduction to cluster analysis". John Wiley and Sons 344 (2009).
25. Lawley D N. "Tests of significance for the latent roots of covariance and correlation matrices". *Biometrika* 43.1/2 (1956): 128-136
26. Macnaughton-Smith P, *et al.* "Dissimilarity analysis: a new technique of hierarchical subdivision". *Nature* 202.4936 (1964): 1034-1035.
27. McQuitty Louis L. "Elementary linkage analysis for isolating orthogonal and oblique types and typical relevancies". *Educational and Psychological Measurement* 17.2 (1957): 207-229.
28. North Gerald R, *et al.* "Sampling errors in the estimation of empirical orthogonal functions". *Monthly Weather Review* 110.7 (1982): 699-706.
29. 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.
30. Parthasarathy B, *et al.* "Homogeneous Indian monsoon rainfall: variability and prediction". *Proceedings of the Indian Academy of Sciences-Earth and Planetary Sciences* 102.1 (1993): 121-155.
31. Parthasarathy B, *et al.* "Monthly and Seasonal Rainfall Series for All-India Homogeneous Regions and Meteorological Subdivision: 1871-1994". Indian Institute of Tropical Meteorology, Pune Research Report No. RR-065.
32. Pearson Karl. "LIII. On lines and planes of closest fit to systems of points in space". *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.11 (1901): 559-572.
33. Rajeevan M, *et al.* "Development of a high resolution daily gridded rainfall data for the Indian region". *Met. Monograph Climatology* 22 (2005): 2005.
34. Rajeevan M, *et al.* "High resolution daily gridded rainfall data for the Indian region: Analysis of break and active monsoon spells". *Current Science* (2006): 296-306.
35. Rao A Ramachandra and V V Srinivas. "Regionalization of watersheds by hybrid-cluster analysis". *Journal of Hydrology* 318.1-4 (2006): 37-56.
36. Rao A Ramachandra and V V Srinivas. "Regionalization of watersheds by fuzzy cluster analysis". *Journal of Hydrology* 318.1-4 (2006): 57-79.

37. Raziei Tayeb. "A precipitation regionalization and regime for Iran based on multivariate analysis". *Theoretical and Applied Climatology* 131.3 (2018): 1429-1448.
38. Sahu Ramgopal T, *et al.* "Regional Frequency Analysis Using L-Moment Methodology—A Review". In: Pathak KK, Bandara JMSJ and Agrawal R (eds) Recent Trends in Civil Engineering. Lecture Notes in Civil Engineering. Springer, Singapore, 77 (2021): 811-832.
39. Sahu Ramgopal T, *et al.* "Density-Based Spatial Clustering of Application with Noise approach for regionalization and its effect on Hierarchical CA". [Manuscript submitted for publication]. Civil Engineering Department, National Institute of Technology, Raipur (2021).
40. Sahu Ramgopal T, *et al.* "Segmental variability of precipitation in the Mahanadi River basin during 1901-2017". 24 August 2021, PREPRINT (Version 1) available at Research Square (2021).
41. Sahu Ramgopal T, *et al.* "Maximum loading value-based Regionalization for the Mahanadi River basin, a multivariate approach". [Manuscript submitted for publication]. Civil Engineering Department, National Institute of Technology, Raipur (2021).
42. Sahu, Ramgopal T, *et al.* "Some non-uniformity patterns spread over the lower Mahanadi River basin, India". *Geocarto International* (2021).
43. Sahu Netrananda, *et al.* "Impact of Indo-Pacific Climate Variability on High Streamflow Events in Mahanadi River Basin, India". *Water* 12.7 (2020): 1952.
44. Satyanarayana P and V V Srinivas. "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.
45. Satyanarayana P and V V Srinivas. "Regional frequency analysis of precipitation using large-scale atmospheric variables". *Journal of Geophysical Research: Atmospheres* 113.D24 (2008).
46. Shepard Donald. "A two-dimensional interpolation function for irregularly-spaced data". Proceedings of the 1968 23rd ACM national conference, August 27 - 29, 1968: Association for Computing Machinery, New York, US (1968) 517-524.
47. Shukla, J. "Interannual variability of monsoons". Monsoons, edited by Fein, Jay S., and Stephens Pamela L., John Wiley and sons, Chapter 14. (1987): 399-464.
48. Singh Gurjeet, *et al.* "Regional scale trend and variability of rainfall pattern over agro-climatic zones in the mid-Mahanadi River basin of eastern India". *Journal of Hydro-environment Research* 29 (2020): 5-19.
49. Srinivas V V. "Regionalization of precipitation in India-a review". *Journal of the Indian Institute of Science* 93.2 (2013): 153-162.
50. Verma Shashikant, *et al.* "Trend Analysis and Rainfall Variability of Monthly Rainfall in Sheonath River Basin, Chhattisgarh". In: Pathak KK, Bandara JMSJ and Agrawal R (eds) Recent Trends in Civil Engineering. Lecture Notes in Civil Engineering. Springer, Singapore, 77 (2021): 777-790.
51. Verma Shashikant, *et al.* "Trends of Rainfall and Temperature over Chhattisgarh During 1901-2010". In: Rao CM, Patra KC, Jhajharia D, and Kumari S (eds) Advanced Modelling and Innovations in Water Resources Engineering. Springer, Singapore, 176 (2022): 3-19.
52. Waller Lance A. "Detection of clustering in spatial data". The SAGE handbook of spatial analysis (2009): 299-320. edited by Fotheringham, A. S., and Rogerson, P. A., SAGE Publications: Thousand Oaks, CA, USA, (2009): 299-320.
53. Wotling G., *et al.* "Regionalization of extreme precipitation distribution using the principal components of the topographical environment". *Journal of hydrology* 233.1-4 (2000): 86-101.
54. Yin Yixing, *et al.* "Spatiotemporal Changes and Frequency Analysis of Multiday Extreme Precipitation in the Huai River Basin during 1960 to 2014". *Advances in Meteorology* 2019 (2019).
55. Yuan Lifeng, *et al.* "Spatio-temporal variation analysis of precipitation during 1960-2008 in the Poyang Lake basin, China". *Open Journal of Modern Hydrology* 6.2 (2016): 115-127.

Volume 5 Issue 12 December 2021

© All rights are reserved by Ramgopal T Sahu, *et al.*