Segmental Variability of Precipitation In The Mahanadi River Basin During 1901-2017

Ramgopal Tilakram Sahu (📧 ramgopal_sahu@yahoo.com)
National Institute of Technology Raipur  https://orcid.org/0000-0003-2564-3419

Mani Kant Verma
National Institute of Technology-Raipur

Ishtiyaq Ahmad
National Institute of Technology-Raipur

Research Article

**Keywords:** Spatial, Temporal, Loadings, Multivariate, parallel analysis, Mahanadi, Segment, Regionalization

**Posted Date:** August 24th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-542786/v1

**License:** ☑️ ⏩ This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Eigen-based sequential spatial pattern analysis an application of PCA is presented here. The analysis examines the spatial distribution of precipitation over the Mahanadi river basin. The Spatial(S)-mode of sequential spatial pattern analysis with the application of Maximum loading value of rotated retained principal component (also referred to as Maximum loading value approach) to assess a gridded monthly rainfall data of resolution 0.25° x 0.25° having a record length of 117 years (1901-2017). The meteorological records have a sequential spatial field for Spatial and Temporal-mode, which can be used for recognizing the area for precipitation variability and regime. The identified patterns of the different timeslot segments were then analyzed for their dispersions of the annual precipitation observed at different station points using similarities and dissimilar characteristics of inter-cluster and between clusters respectively. Validation of the regionalized pattern for distinctness and a pairwise comparison of CDFs using the Kolmogorov-Smirnov ‘D’ statistic test.

Introduction

The undisputed fact of warming climate against the precipitation variability over time and an IPCC's 5th assessment report suggesting 0.85°C rise during 1880–2012 [IPCC, 2013]. To ascertain this phenomenon of precipitation variability in the Mahanadi river basin, a segmental approach was employed. The monthly precipitation record length was segmented into 3 layers of 39 years each, including the 1st layer (1901–1939), 2nd layer (1940–1978), and 3rd layer (1979–2017). In a general hydrological cycle process, water evaporates and is stored in the atmosphere as vapour. The amount of water vapour increases as the temperature increases and, ultimately, the spatial and temporal form of the precipitation pattern changes. “The first law of geography states that closer spatial entities are more strongly related to each other than distant ones” [24]. So spatial events therefore need to be spatially clustered for which spatial analysis is an important technique [23]. Such analysis is important and works by dividing the data point/object into groups or clusters based on similarities with respect to attributes and location aspects. Regional and global precipitation patterns are likely to change in spatial and temporal terms once the atmospheric water retention capacity is exceeded. The warming climate has intensified precipitation characteristics such as precipitation intensities and the number of wet days. Changes in the intensity, quantity, and pattern of precipitation can result in more frequent extreme events, such as droughts and floods. The stakeholders how are ultimately affected are the people for whom the extreme events are unforeseen incidents. Thus, analysis of historical data is very important for extracting information related to changing precipitation patterns.

The study is motivated by various phenological changes in precipitation (e.g., sudden, uncertain, increase, or decrease). The uncertainty and inconsistencies associated with the precipitation adversely affect the stakeholders of different sectors, dependent farmers, health and safety (not limited to living creatures but also the environment). The study aspect is of considerable practical importance for water resource management planning, flood and drought risk assessment. Therefore, precipitation is very important and its variability needs to be studied. Understanding the impact of precipitation variability on water resources is a very challenging and profound task. Consequently, the objective of this article is to investigate spatial-
temporal change and regional analysis of precipitation using gridded monthly precipitation in the Mahanadi River Basin.

Aher et al. [1] studied precipitation variability and spatio-temporal analysis in the upper Godavari River basin using daily precipitation data (2000–2014) which was then correlated with satellite-derived TRMM data and found a positive correlation between TRMM and rainfall data over three half-decades with microlevel precipitation variability. Huang et al. [5] studied the upper portions of the Hongshui River basin (UHRB) to determine the spatial and temporal variability of precipitation using three precipitation indices (PCI, PCD, and PCP) by analyzing daily precipitation data (1959–2015) and found high PCI in the northeast of UHRB and low PCI in the southwest of UHRB. PCP reveals early rainfall in the eastern UHRB and PCD indicates more dispersed rainfall in the western UHRB. Yin et al. [25] have studied the Huai River basin for frequency analysis of extreme 5-day and 10-day precipitation and spatio-temporal changes using EOF and hydrological frequency analysis by analyzing daily precipitation data (1960–2014) and found a weak upward trend for the two extreme events and are closely associated with drought/flooding. Yuan et al. [26] investigated the spatial and temporal variation of seasonal and annual precipitation in the Poyang Lake basin using statistical techniques and continuous wavelet transform by analyzing monthly precipitation data (1960–2008). Tayeb Raziei [15] has studied Iran for regionalizing precipitation time variability and precipitation regime using precipitation data of 155 synoptic stations during 1990–2014, with the application of S-mode PCA and Hierarchical cluster analysis preceded by T-mode PCA.

Many researchers have practiced their research on the Mahanadi River Basin by employing different regression and classification methods on various observational or gridded data records. Singh et al. [22] investigated the middle Mahanadi river basin and discovered significant spatiotemporal variation, including seasonal shifts, increasing dry spells, and extreme weather events, to better understand changes in precipitation characteristics and patterns on a regional scale agro-climatic zone. Sahu et al. [19] have studied the Mahanadi river basin to explore the combined and relative effect of “Indian monsoon and Indo-pacific climatic modes” on the flood frequency analysis and high streamflow events and found that ENSO and ENSO Modoki were significant during JJA and SON on river extremities. Sahu et al. [16] have studied the Mahanadi river basin to explore the effectiveness of DBSCAN for precipitation regionalization and found good agreement in clusters, but underestimating the homogeneity characteristics, the findings were assessed using a popular unsupervised hierarchical cluster analysis.

In this study, a data record length is segmented and is then applied to a sequential spatial pattern analysis-based PCA approach, which is also referred to as the Eigen-based technique. A dataset of 230 grid station points with a record length of 117 years (1901–2017) is segmented into three layers, each of which spans 39 years: the first segment layer spans 1901–1939, the second segment layer spans 1940–1978, and the third segment layer spans 1979–2017. A pattern delineation using Spatial(S)-mode of principal component analysis for extracting information about different patterns of time variability. Here in Spatial(S)-mode, the magnitude of monthly precipitation is considered as observations and station points as spatial field variable. The resulting strong spatial pattern indicates high component loadings among interconnected stations and loadings congruent to zero are associated with dispersed stations. The Mahanadi River basin
has very vast climatological characteristics for which the segmental approach of spatial (S)-mode will portray some different precipitation patterns and unveil different aspects of the Mahanadi's precipitation regimes.

**Study Area And Data**

The Mahanadi is a major east-flowing river in India (see Figure-1), which is east-centrally located having a geographical coordinate of 19°20'N to 23°35'N latitude and 80°30’ E to 86°50’E longitude. The thalweg is 858 kilometers long from origin to false point, with a catchment area of 141,600 km². The river was known as Kanaknandani during the ancient era and Mahanadi during the Mahabharat era. The river originates from Pharsiya village, Sihawa town in Dhamtari district of Chhattisgarh and merging into the delta region of the Bay of Bengal, located at Jagatsinghpur, Odisha.

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 [13, 14, 21]. A pre-analysis data check for stationarity, independence, and spatial independence was performed to assess the data quality and accuracy. A blind assumption of the above pre-analysis data checks may mislead the analysis and conclusion of the test results. The stationarity of the dataset was made using a non-parametric Mann-Kendall trend test [7, 9], lag-1 to lag-5 autocorrelation coefficient for independence check and Moran’s I test [10] for spatial independence. The MK test result indicates 200 amongst 230 station points as statistically insignificant and only 13% (30 station points) lying below the 5% confidence interval. Since most units of observation have no significant trends, it is therefore reliable to infer that there is no significance at the regional scale and that the data are stationary (readers can also refer to Sahu et al. [20] for wavelet synopsis-based trend detection). The Independence test (Auto-correlation coefficient of lag-1 to lag-5) suggests only 5 station points (including GP_9, 10, 16, 69, 183) above the threshold limits (horizontal dash line) in the correlogram plot. Thus, the data may be considered to be time-dependent. The Spatial independence (Moran’s I test) results indicate a statistically significant p-value (2.2e-16) with a positive z-score (23.609) and the Moran’s Index (0.867) which is positive. The above statistics suggest that the dataset tends to cluster spatially, in simple words, ‘high-value cluster near other high value’ and visa-versa considering the dataset to be spatially independent.

**Methodology**

A Sequential spatial pattern analysis based on PCA is also referred to as the Eigen-based technique [12]. The Spatial(S)-mode and Temporal(T)-mode were respectively used to analyze the patterns of precipitation time variability and the patterns of precipitation regimes. The study is carried out in two-part (1) Segmental analysis, which is completely based on eigenvalue decomposition of the time series data for different timeslots. (2) Pattern delineation which is based on the maximum loading value approach. The identified patterns of the different timeslot segments were then analyzed for their dispersions of the annual precipitation observed at different station points using similarities and dissimilar characteristics of inter-cluster and between clusters respectively. Validation of the regionalized pattern for distinctness and a pairwise comparison of CDFs using the Kolmogorov-Smirnov ‘D’ statistic test [3].
3.1 Segmental Analysis: patterns of precipitation time variability and regionalization

Data preparation before applying sequential spatial pattern analysis (SSPA), which is a prerequisite for principal component analysis. (1) Data segmentation. The first task of data segmentation is to view the changes in the three-timeslot layers. This perspective view of analyzing precipitation can help us to justify the impact of a warming climate on the patterns of precipitation. (2) Data normalization, the second task of data normalization for which the objective is set to dimensionalize the dataset and to stabilize the tail of the distribution (close to normal distribution) by transforming the dataset to low skewed properties. Transforming dataset to low skewed characteristics by subjecting them to the log, the square root, or the cubic root transformation. The cubic root transformation results in the lowest skewed value. Precipitation time variability patterns across the basin for which a cubic root transformed monthly precipitation time series data of matrix (M x N), where M = 230 represents station points and N = 468 (39 x 12 years) denotes the record length (observations) were implemented to Spatial(S)-mode of PCA. The Kaiser-Mayer-Olkin (KMO) [6] measure for sampling adequacy of cubic root transformed data, the KMO value for all three time-series slots appears 0.99, implying 'marvellous' for principal component analysis implementation. In PCA implementation, an eigenvector is assigned to all the components and eigendecomposition results in as many components as variables in the dataset. An application of an adjacent tool or analysis to retain or decide on the optimum number of components for further assessment, such as the scree plot [2] in conjunction with parallel analysis [4] and accomplishing the estimated eigenvalues with the use of North’s thumb rule criteria [11].

A graphical representation of eigenvalue to that of the eigenvector is known as a scree plot in terms of PCA, whose optimism lies in its elbow. The scree plot's main disadvantage is its interpretation, which is intuitive and arbitrary, implying only a few components (say two or three). A sole criterion for deciding the optimum number of components is not feasible because eigenvalues are subject to bias and errors, and a blind assumption may mislead the test results. A Parallel analysis, which is a supplement to the scree plot for better interpretation and accounting for the feasibility of eigenvalue by adjusting the bias and errors. The rationale of parallel analysis is to figure out that factor accounting for maximum variance, whose probability existence is not expected by chance. North's thumb rule criteria state “if the sampling error of an eigenvalue has a magnitude comparable to or larger than the linear spacing to its closest eigenvalue, then these two eigenvectors are mixed”, In other words, it is a function of the degree of separation between the eigenvalues. The below equation suggests Lawlay's [8] formula for estimating the standard error.

\[
\Delta \lambda_i = \left( \frac{2}{n^*} \right)^{1/2} \lambda_i \quad (1)
\]

Where,

\( n^* \) = effective sample size,

\( \lambda_i \) = eigenvalue for \( i^{th} \) order.
Equation – 1 suggests that 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. Also, the error limit at 95% confidence interval associated with each eigenvalue can be defined as $\text{Error} = \lambda_i \pm 1.96\Delta\lambda_i$

### 3.2 Pattern delineation (maximum loading value approach)

The orthogonal rotation of the retained components, obtained after implementing the cubic root transformed dataset to a sequential spatial pattern analysis with the use of varimax rotation criteria. The objective is to extract more localized information beneficial for pattern interpretation. For analyzing the spatio-temporal variability and the effect of a warming climate, components (accounting maximum variance) from each of the timeslot's segments is considered. The variability or shifting of patterns can be visualized by plotting the rotated loading vectors of delineated patterns on a map using loading value (\(\parallel\)) as a threshold. The Mahanadi river basin is regionalized into a few distinctive sub-regions characterized by different patterns of precipitation time variability for each timeslot segment. The maximum loading value approach results in satisfactory regionalization, where all the stations are classified and the station boundaries do not overlap. Readers can also refer to Sahu et al. [16] for DBSCAN based regionalization and other statistics for spatial and temporal variations for the Mahanadi basin. The identified sub-regions can also be studied for regional frequency analysis, which is a supplementary content excluded in this study, refer Sahu et al. [17] for l-moment based regional frequency analysis.

Temporal(T)-mode of Sequential spatial pattern analysis to identify the patterns with different precipitation regimes. The objective of this approach is to identify those station points whose characteristics of annual precipitation regardless of the magnitudes in each station point are similar. To ascertain the objective of this approach, the monthly precipitation data was changed to its relative form, i.e., monthly relative precipitation. Then a Temporal-mode of SSPA was applied to the \((N \times M)\) matrix of a relative monthly dataset. Here \(M = 230\) refers to a number of station points and \(N = 12\) refers to monthly relative precipitation. The phenomenon used in this approach is only a kind of normalization to ascertain the over-weighting concerning high precipitation values to the station points. The varimax rotation of the principal component scores was only used for extracting the localized information for interpreting the spatial patterns of monthly relative precipitation.

### Results

#### 4.1 Statistics of Spatial (S)-mode PCA implementation

Implementation of the cubic root transformed data to Spatial(S)-mode of principal component analysis, for which the first 5-component whose eigenvalue > 1 are shown in Table-1. The summary statistics of PCA implementation suggest that the first un-rotated component accounts for 90.20% of segment-1, 89.0% of segment-2, and 87.80% of segment-3 (see Figure-3). The Scree plot in association with parallel analysis, which is subjectively used for assessing the standard error (sampling error) of the resulting eigenvalue from implementation and suggests the optimal number of components to be retained for further analysis (see Figure-2). Figures-4, and 5 show the spatial distribution of rotated loadings of retained principal components and the principal component scores. The summary statistics of varimax rotated loadings of PCs as tabulated
in Table-2, suggest 42.49% for the first varimax-rotated loading patterns of a segment-1 (1901–1939) explaining the southeast and some parts of northwest Mahanadi, whereas 18.16% for the second patterns explaining southeast Mahanadi with high positive loading.

The third pattern accounts for 34.24%, featuring the northeast Mahanadi, explaining most of the negative loadings. Segment-2 (1940–1978) for which the first pattern accounts for 39.07% comparably equals to the first pattern of Segment-1 but completely different loading characteristics i.e., negative loadings explaining northeast and northwest Mahanadi, whereas the second and third patterns account for 20.75% and 33.78% variance explaining southeast and southwest with some part of northwest Mahanadi respectively featuring high positive loadings. For Segment-3 (1979–2017), the second pattern explains both types of loadings (positive and negative) featuring 28.52% of variance explaining south and southeast Mahanadi. The first pattern accounts for 35.75% of the variance with high positive loading representing western Mahanadi, whereas the third pattern featuring negative loadings of north and northeast Mahanadi explains only 28.12% of the total variance.

4.2 Characterization of core origin

A fundamental concept of the EOF coefficient was assessed to lighten the information lying behind the emerging of the core points and to clearly understand the negative and positive characteristics of core points and rotated loadings. Results of the EOF coefficient (see Figure-4), where each station is assigned to a component upon which it loads highly. The patterns as depicted by retained varimax rotated loadings highlight the cores of high positive and negative loadings of different timeslot segments. Two high negative core loading emerged from the Raipur district of Chhattisgarh, with spatial coordinates of 82°E x 21°25’N during (1901–1939) and 82°E x 21°N during (1979–2107), accounting for 34.25% and 28.13% of the total variance, respectively. The three-adjoining district of Raipur including Durg, Dhamtari, and Mahasamund encountered points of high positive core loading spatially located at 81°30’E x 21°N during (1940–1978), 81°45’E x 20°30’N during (1901–1939), and 82°30’E x 21°15’N (1901–1939) respectively accounting 33.78%, 18.16%, and 42.49% of the total variance, whereas one positive core emerging point is identified at Raigarh and Korba district spatially located at 83°E x 22°15’N during (1979–2017) accounting 35.75% of the total variance.

Cuttack district in Odisha state has been identified as the core of negative loading, which is spatially located at 85°30’E x 20°30’N and accounts for 39.07% of total variance during (1940–1978), whereas Nayagarh and Khordha districts spatially located at 85°15’E x 20°30’N were found to be a core point of positive loadings during (1940–1978), accounting for 20.75% of the total variance. Finally, the second pattern of segment-3 (1979–2017) accounted for 28.52% of the variance with (undefined) dispersed core point (refer to Table-3). Sahu et al. [18] studied the Mahanadi basin using the same dataset, intending to regionalize the basin into homogeneous sub-regions. They identified four patterns, where the third pattern was similar to what was expected in this study by the 1st and 2nd components of segment-2 (1940–1978).

4.3 Statistics of Temporal (T)-mode PCA implementation

For studying the patterns of precipitation regime, the monthly relative precipitation is then implemented to a Temporal(T)-mode of sequential spatial pattern analysis. The first three un-rotated components whose
eigenvalue > 1 as an output of the implementation is shown in Table-4, for which segment-1 (1901–1939) accounts for 87.40%, segment-2 (1940–1978) accounts for 77.57%, and segment-3 (1979–2017) accounts for 80.71% of the variance. Further, parallel analysis for component retention found three-component from segment-2, two-components from segment-1 and segment-3. Figure-6 shows the spatial distribution of varimax rotated loadings of retained principal components. The spatial distribution of monthly relative precipitation can be explained by the first two components of segment-1 and segment-3, whereas the first three components for segment-2.

The test results indicate that for segment-1, April, May, September, October, November, and December (July and August) have negative (positive) loadings on component-1, implying spatial pattern variability of spring (April and May), autumn (September, October, and November), summer (July and August), and early winter (December) precipitation, accounting for 42.93% of the variance. The upper and the middle portion of the Mahanadi are featured with negative loadings suggesting reasonable precipitation during the spring (Apr & May) and autumn (Sep, Oct, & Nov), along with some noticeable precipitation during early winter (Dec). The lower portion of the Mahanadi is featured with high positive loadings suggesting reasonable precipitation during summer (Jul & Aug). January, February, and March (June) have positive (negative) loadings for component-2 accounting for 36.07% of the variance, indicating spatial pattern variability of winter (Jan & Feb) and representative precipitation variability during early spring (Mar) and summer (Jun), adding a noticeable amount of precipitation to the annual total. The north Mahanadi is featured with positive loadings, receives reasonable precipitation during winter (Jan & Feb) and the rest (with negative loading) receives noticeable precipitation during early spring (Mar) and early summer (Jun) (see Tables – 5 and 6). Sahu et al. [18] have used sequential spatial pattern analysis in connection with DBSCAN for regionalizing the same basin and dataset and found three homogeneous precipitation regimes. The patterns formed were completely different from what was expected in this study.

The first three components explain segment-2, such that April, May, and October (July and August) have negative (positive) loadings on component-1, implying spatial pattern variability of spring (Apr & May), summer (Jul & Aug), and representative precipitation variability during mid-autumn (Oct) accounting for 30.35% of the total variance. The tail or the lower portion of Mahanadi is featured with positive loading, receives reasonable precipitation during summer (Jul & Aug), while the rest portion (with negative loading) receives precipitation during the spring (Apr & May) and autumn (Oct). For Component-2, January, February, and March (September) have positive (negative) loadings suggesting spatial pattern variability of winter (Jan & Feb), and representative precipitation variability during early spring (Mar) and early autumn (Sep) accounting for 31.36% of the total variance. The upper portion of the Mahanadi streamlines featured with positive loading receives reasonable precipitation during winter (Jan & Feb) and early spring (Mar), whereas the lower portion of Mahanadi streamlines with negative loadings receives noticeable precipitation to its annual total during early autumn (Sep). Finally, for component-3, in which November, and December (June) with negative (positive) loadings suggesting spatial pattern variability of late autumn (Nov), early winter (Dec), and early summer (Jun) precipitation accounting for 15.86% of the total variance. The upper and middle portions of the Mahanadi have negative loadings, indicating reasonable precipitation during late autumn (Nov) and early winter (Dec), whereas the lower portion has positive loadings, indicating noticeable precipitation during early summer (Jun) (see Tables 5 and 6).
Table 1
Decomposition analysis of the Cubic-root transformed data matrix of (M x N) using Spatial(S)-modes of Sequential spatial pattern analysis for the first five components whose Eigenvalue is > 1.

| Segment (1) | Component (2) | Eigenvalue (3) | Eigen-Low (4) | Eigen-High (5) | Cumulative Variance (6) |
|-------------|---------------|----------------|---------------|----------------|-------------------------|
| 1901–1939   | 1             | 207.53         | 193.96        | 221.10         | 90.20                   |
|             | 2             | 7.58           | 7.09          | 8.08           | 93.50                   |
|             | 3             | 3.12           | 2.92          | 3.32           | 94.90                   |
|             | 4             | 2.30           | 2.15          | 2.45           | 95.90                   |
|             | 5             | 1.22           | 1.14          | 1.30           | 96.40                   |
| 1940–1978   | 1             | 204.62         | 191.24        | 218.00         | 89.00                   |
|             | 2             | 7.70           | 7.20          | 8.21           | 92.30                   |
|             | 3             | 2.99           | 2.80          | 3.19           | 93.60                   |
|             | 4             | 2.15           | 2.01          | 2.29           | 94.60                   |
|             | 5             | 1.21           | 1.13          | 1.29           | 95.10                   |
| 1979–2017   | 1             | 201.88         | 188.68        | 215.07         | 87.80                   |
|             | 2             | 7.81           | 7.30          | 8.32           | 91.20                   |
|             | 3             | 2.90           | 2.71          | 3.09           | 92.40                   |
|             | 4             | 2.17           | 2.03          | 2.31           | 93.40                   |
|             | 5             | 1.20           | 1.12          | 1.28           | 93.90                   |

Note: - Column no. (4) & (5) represent Sampling error range.
Table 2
-Selected components from each segment for studying the spatio-temporal and the effect of warming climate along with Un-rotated, Varimax-rotated, and parallel analysis characteristics.

| Segment (1) | Component (2) | Adjusted Eigenvalue (3) | Estimated Bias (4) | Variance (5) | Cumulative Variance (6) | Variance (7) | Cumulative Variance (8) |
|-------------|---------------|--------------------------|--------------------|--------------|--------------------------|--------------|--------------------------|
| 1901–1939   | 1             | 205.694                  | 1.836              | 90.20        | 90.20                    | 42.49        | 42.49                    |
|             | 2             | 5.827                    | 1.757              | 3.30         | 93.50                    | 18.16        | 60.65                    |
|             | 3             | 1.422                    | 1.697              | 1.40         | 94.90                    | 34.25        | 94.90                    |
| 1940–1978   | 1             | 202.794                  | 1.828              | 89.00        | 89.00                    | 39.07        | 39.07                    |
|             | 2             | 5.953                    | 1.752              | 3.30         | 92.30                    | 20.75        | 59.82                    |
|             | 3             | 1.296                    | 1.696              | 1.30         | 93.60                    | 33.78        | 93.60                    |
| 1979–2017   | 1             | 200.046                  | 1.831              | 87.80        | 87.80                    | 35.75        | 35.75                    |
|             | 2             | 6.049                    | 1.757              | 3.40         | 91.20                    | 28.52        | 64.27                    |
|             | 3             | 1.198                    | 1.700              | 1.20         | 92.40                    | 28.13        | 92.40                    |

Adjusted eigenvalues > 1 indicate the dimensions to retain (3 components retained)

Table 3
Core point description

| Segment          | Components | Variance | Spatial Core location | Location Description | Nature of Core                      |
|------------------|------------|----------|-----------------------|----------------------|-------------------------------------|
| 1901–1939        | 1          | 42.49    | 82°30'E x 21°15'N     | Mahasamund           | Positive loading                     |
|                  | 2          | 18.16    | 81°45'E x 20°30'N     | Dhamtari             | Positive loading                     |
|                  | 3          | 34.25    | 82°00'E x 21°15'N     | Raipur               | Negative loading                     |
| 1940–1978        | 1          | 39.07    | 85°30'E x 20°30'N     | Cuttack              | Negative loading                     |
|                  | 2          | 20.75    | 85°15'E x 20°30'N     | Nayagarh/Khordha     | Positive loading                     |
|                  | 3          | 33.78    | 81°30'E x 21°00'N     | Durg                 | Positive loading                     |
| 1979–2017        | 1          | 35.75    | 83°00'E x 22°15'N     | Raigarh/Korba        | Positive loading                     |
|                  | 2          | 28.52    | 84°00'E x 20°00'N     | Khandhamal           | Dual (+/-) loading                   |
|                  | 3          | 28.13    | 82°00'E x 21°00'N     | Raipur               | Negative loading                     |

Note: - Spatial Core locations are approximate locations.
### Table 4
Decomposition analysis of the monthly relative precipitation data matrix of \((N \times M)\) using Temporal\((T)\)-modes of Sequential spatial pattern analysis for the first five components whose Eigenvalue is > 1.

| Segment  | Component | Eigenvalue | Eigen-Low | Eigen-High | Cumulative Variance |
|----------|-----------|------------|-----------|------------|---------------------|
| 1901–1939| 1         | 6.22       | 5.64      | 6.80       | 51.83               |
|          | 2         | 3.26       | 2.95      | 3.56       | 79.00               |
|          | 3         | 1.01       | 0.92      | 1.11       | 87.40               |
| 1940–1978| 1         | 5.54       | 5.02      | 6.06       | 46.19               |
|          | 2         | 2.53       | 2.30      | 2.77       | 67.30               |
|          | 3         | 1.23       | 1.12      | 1.35       | 77.57               |
| 1979–2017| 1         | 6.42       | 5.82      | 7.02       | 53.50               |
|          | 2         | 2.29       | 2.08      | 2.51       | 72.60               |
|          | 3         | 0.97       | 0.88      | 1.06       | 80.71               |

Note: - Column no. (4) & (5) represent Sampling error range.

### Table 5
Selected components from each segment for studying the spatio-temporal and the effect of warming climate along with Un-rotated, Varimax-rotated, and parallel analysis characteristics.

| Segment  | Component | Adjusted Eigenvalue | Estimated Bias | Variance | Cumulative Variance | Variance | Cumulative Variance |
|----------|-----------|---------------------|----------------|----------|---------------------|----------|---------------------|
| 1901–1939| 1         | 5.83                | 6.22           | 51.83    | 51.83               | 42.93    | 42.93               |
|          | 2         | 2.98                | 3.26           | 27.17    | 79.00               | 36.07    | 79.00               |
| 1940–1978| 1         | 5.16                | 5.54           | 46.19    | 46.19               | 30.35    | 30.35               |
|          | 2         | 2.25                | 2.53           | 21.11    | 67.30               | 31.36    | 61.71               |
|          | 3         | 1.02                | 1.23           | 10.27    | 77.57               | 15.86    | 77.57               |
| 1979–2017| 1         | 6.03                | 6.42           | 53.50    | 53.50               | 44.51    | 44.51               |
|          | 2         | 2.01                | 2.29           | 19.10    | 72.60               | 28.09    | 72.60               |

Adjusted eigenvalues > 1 indicate the dimensions to retain (3 components retained)
Table 6

-Results of varimax rotated loadings obtained from Temporal (T)-mode of the sequential spatial pattern analysis

| Segment – 1 (1901–1939) | Segment – 2 (1940–1978) | Segment – 3 (1979–2017) |
|-------------------------|-------------------------|-------------------------|
| **Month**               | **Comp.1**              | **Comp.2**              | **Comp.1** | **Comp.2** | **Comp.3** | **Comp.1** | **Comp.2** |
| January                 | 0.11                    | 0.488                   | 0.543      | -0.118     | 0.154      | 0.56       |
| February                | 0.506                   | -0.139                  | 0.567      | 0.123      | -0.173     | 0.495      |
| March                   | -0.102                  | 0.423                   | -0.245     | 0.28       | 0.218      | -0.296     | 0.182      |
| April                   | -0.283                  | -0.156                  | -0.411     | 0.334      | -0.339     |            |
| May                     | -0.358                  | -0.429                  |            |            | -0.378     |            |
| June                    |                        | -0.42                   | 0.645      | 0.235      | -0.18      |            |
| July                    | 0.379                   | 0.106                   | 0.393      | 0.11       | 0.383      | 0.104      |
| August                  | 0.389                   | 0.406                   |            |            | 0.311      | -0.114     |
| September               | -0.307                  | -0.203                  | -0.149     | -0.487     | 0.117      | -0.161     | -0.2       |
| October                 | -0.386                  | -0.338                  |            | -0.3       | -0.381     |            |
| November                | -0.381                  | -0.307                  | -0.378     |            | -0.361     |            |
| December                | -0.28                   | 0.253                   | 0.199      | -0.357     |            | 0.541      |

Segment-3 for which the first two components account for 44.51% and 28.09% of the variance is sufficient for explaining the temporal variability. For component-1, March, April, May, October, and November (June, July, and August) have high negative (positive) loadings, suggesting spatial pattern variability of spring (Mar, Apr, & May), autumn (Oct & Nov), and summer (Jun, Jul, & Aug) precipitation. The upper and the middle portion of Mahanadi are featured with high negative loadings suggesting reasonable precipitation during spring (Mar, Apr, & May) and autumn (Oct & Nov), while the lower portion with high positive loading hits the summer monsoon (Jun, Jul, & Aug) with reasonable high precipitation. The consecutive three-month December, January, and February (September) feature positive (negative) loadings on component-2, suggesting spatial pattern variability of winter (Dec, Jan, & Feb) and early autumn (Sep) precipitation. The lower portion of upper and middle Mahanadi along with lower Mahanadi is featured with negative loading suggesting noticeable precipitation to its annual total during early autumn (Sep), while the upper portion of upper and middle Mahanadi with positive loadings suggesting reasonable precipitation during winter (Dec, Jan, & Feb) (see Tables – 5 and 6).

4.4 Characterization of identified patterns

The patterns of the different timeslot segments were then analyzed for their dispersions of the annual precipitation observed at different station points using similarities and dissimilar characteristics of inter-cluster and between clusters respectively. Figures-8 and 10 depict a boxplot in which the median’s relative distance is close to the center, indicating symmetric distribution for the clusters of all the timeslot segments except the third cluster of segment-1 (1901–1939) with negatively skewed properties. The defined clusters of
the segments also showed good compactness, suggesting that inter-cluster precipitation variability is homogeneous in different sub-clusters. The third cluster has the largest interquartile range (iqr), indicating a larger surface area and a lower degree of homogeneity. It is the most distinct sub-cluster of the Mahanadi basin and contains the major thalwegs of the river. The wider interquartile range of the third cluster is characterized by phenological changes in precipitation magnitudes from the high-altitude range (500m – 700m) to the streamlines.

Figures-7 and 9 show the application of empirical cumulative distribution functions (eCDFs) to the identified clusters to validate their distinctness and a pairwise comparison of CDFs using the Kolmogorov-Smirnov 'D' statistic test [3] as illustrated in Figures-8 and 10. On a visual interpretation of the CDF plot, the CDFs of all three sub-clusters appear similar and close to each other, as shown in Figures-7 and 9. So the D-statistics of the KS test is used for pairwise comparison of all the CDFs, indicating that the clusters identified are distinct from one another. The results of D-statistics indicate that the p-value of all the test pairs (except 2pairs) is congruent to zero, suggesting that the clusters are different from one another. Excepted pairs are the cluster 2–3 for segment-1 (1901–1939) and segment-2 (1940–1978) having homogeneity p-value of 0.595 (59.5%) and 0.655 (65.5%) respectively (see Figures-8a and 8b). Figure-10 shows a cluster comparison for different timeslot segments. Figure 10a represents the cluster-1 for different segments, suggesting high homogeneity between segment-1 and 2, and segment-2 and 3 with p-value 0.549 (54.9%) and 0.643 (64.3%) respectively. Similarly, cluster-2 has high homogeneity among all segments with a p-value of 1 (100%) see Fig. 10b. While cluster-3 has medium-range homogeneity between segment-2 and 3 with p-value 0.305 (30.5%) see Fig. 10c.

The boxplot analysis of the identified patterns of precipitation for their dispersion among the clusters is shown in Figure-11. According to the boxplot's interpretation of compactness, the inter-cluster variations for the identified patterns of precipitation regime are medium to low, implying possible homogeneity to acceptable homogeneity in their identified regimes. The boxplot of the three timeslot segments illustrated in Figure-11 suggests well-defined, compact, and distinct precipitation regimes covering a plain terrain, high mountains, and delta region of 230 station points. Some station points have deviated from the median line, e.g., June and September (cluster-1), August and September (cluster-2), June and October (cluster-3) in Figure-11a. June, September, and October months for clusters - 1, 2, and 3 respectively, see Figure-11b. Whereas October month for cluster-1 and for cluster-3 (September and October) see Figure-11c, indicating precipitation variations in magnitude at different locations. Overall, the boxplot indicates that the patterns of the precipitation regimes are well-defined and homogeneous.

Clusters-1 and 3 (upper and middle Mahanadi) have similar precipitation marching months (June to October) for all segments. The boxplot in Figure-11 suggests an almost symmetric distribution of precipitation towards the march of months, followed by a dry season starting in November and ending in May. The spatially scattered station points in clusters-1, and 3, are featured with intense precipitation in the summer monsoon (JJA) and early autumn (SO), suggesting the summer monsoon be the main rainy season. Cluster-2 (lower Mahanadi) features wider boxplots for all segments, indicating a lower degree of homogeneity and rapid changes in precipitation values. Larger march of the month (May to November) compared to cluster-1 and 3 with a symmetric distribution. May, June, July, August, September, October, and November are the rainiest months in the lower or the southernmost portion of the Mahanadi basin. In this precipitation regime, the
summer monsoon and autumn are the two main rainy seasons, as well as the late spring (May) contributing noticeable precipitation to its annual total. Further, followed by a dry season that starts in December and ends in April, see Figure-11c.

**Conclusion**

Eigen-based sequential spatial pattern analysis, for studying the spatial-temporal variation of the Mahanadi basin and the maximum loading value approach for regionalizing the identified pattern. The objective of regionalizing the identified patterns is to study the aging effect. The aging effect of 117 years of gridded rainfall data of resolution 0.25° x 0.25° was studied in three segment layers. Investigation of precipitation characteristics to assess the spatio-temporal variation at the regional scale along with regionalization is presented here.

Observations from this investigation include:

1. Segment-1 (timeslot 1901–1939), PCA implementation results in a 90.20% variance from the 1st component and similarly the 1st component variance of 89.0% and 87.80% of segment-2 and segment-3 respectively. The PCA implementation results suggest a decrease in proportion by 1.33% chronologically.

2. Eigenvalue decomposition (Parallel analysis) results suggest the first three components be retained for further analysis for all segments with maximum eigenvalue for the 1st component are 205.69, 202.79, and 200.04 respectively for segments-1, 2, and 3. The analysis predicts an eigenvalue decrease of 1.41–1.35%.

3. Varimax rotation result statistics are as follows: 42.49%, 18.16%, and 34.35% of segment-1, 39.07%, 20.75%, and 33.78% of segment-2, and 35.75%, 28.52%, and 28.13% of segment-3. Concerning components of segment-1, the variance of component-1 and 3 was reduced by more than 6% and more than 10% for component-2.

4. The pattern size for segment-1 is 83, 44, and 103 grids, and the pattern size for segment-2 is 96, 51, and 83 grids. The pattern-1 plays well in between 83–96 grids, whereas pattern-2 size increases by 61.36% and pattern-3 size decrease by 32.03%.

5. For all segments, patterns-1 and 3 have a similar march of months for the rainy season (June to September) and summer is the main rainy season. Similarly, pattern-2 has an extended march of the month (May to November) with summer and autumn as the main rainy season.

6. Interpretation of ecdf suggests, as the cluster size decreases, the proportion above mean annual precipitation increases for cluster-3. Here, cluster size decreases by 32.03%, and the proportion above mean annual precipitation increases by 34.03%. No such dominating changes were observed in clusters-1 and 2.

7. Boxplot suggests that cluster-2 was uniform throughout the segments regardless of size. Segment-3 exhibits identical clusters (heterogeneous to each other). For cluster-1, the segment pair 1–2 and 2–3 have inter-cluster similarities with a similarity index of 54.9% and 64.3% respectively. Whereas the segment pairs 1–3 have no such connection.

Concluding remark of investigation: The above observation is the result of the irregular oscillation of sea surface temperature i.e., the Indian Ocean dipole (IOD). The Indian monsoon rainfall is influenced by a system of oscillating sea surface temperatures in which the western Indian Ocean becomes alternately warmer and then colder than the eastern part of the ocean. The clusters formed by the patterns of precipitation variability
for the different segments are the local-dependent phenological changes that emerge because of the substantial effect of hydrology, climatology, and local topography. The spatial patterns of precipitation regimes were influenced by the summer monsoon and the moist current prevailing from the Indian Ocean. The influence was evident from the results of the boxplot (figures-8, 10, and 11), which accurately reflects the precipitation variability in the amount of magnitude at different locations. The regionalized patterns of the precipitation obtained using the maximum loading value approach have different applications and uses. For water resource management in regions with known precipitation variability, the future prediction of precipitation can be simplified. This can be furnished by defining regional time-series for each of the identified clusters which are then associated with predictors (climatological indices) and accomplished by a stochastic and statistical method. Similarly, clusters with known precipitation regimes are useful in activities that include agricultural planning, farming calendar, planting time of different crops, and rain-fed and dry farming practices.

The limitations of this study are based on the results of the precipitation regionalization, which may be substantially different for different methods and the different sizes of clusters characterized by the different time periods and spatial resolution. Therefore, sensitivity analysis of the results is characterized by different spatial resolutions and a different number of station points over the entire basin.

Declarations

**Ethical statement:** The submitted manuscript with the title “Density-Based Spatial Clustering of Application with Noise approach for regionalization and its effect on Hierarchical CA“ is an original work for peer review and publication. This manuscript has not been sent to any other publication or journal for any review and will not be sent unless a clear consent is received from your side.

**Data and Code availability statement:** we acknowledge the data/code share policy and the same will be available for the readers and the researchers that support the results and analysis presented in this paper.

**Consent to participate:** I (Corresponding author) gave my consent to participate in a related research study.

**Consent to publication:** The Authors grant consent to the publisher and declare that any person named as co-author of the contribution is aware of the fact and has agreed to be so named.

**Author contributions:** Conceptualization, R.T.S, M.K.V, and I.A; Supervision and Review, M.K.V, and I.A; Methodology, R.T.S and M.K.V; Writing-original draft, R.T.S; Writing-editing, R.T.S, and M.K.V.

**Name of the supporting fund organization:** No funding.

**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.
References

1. Aher, S., Shinde, S., Gawali, P., Deshmukh, P. and Venkata, L.B., 2019. Spatio-temporal analysis and estimation of rainfall variability in and around upper Godavari River basin, India. Arabian Journal of Geosciences, 12(22), pp.1-16. https://doi.org/10.1007/s12517-019-4869-z

2. Cattell, R.B., 1966. The scree test for the number of factors. Multivariate behavioral research, 1(2), pp.245-276. https://doi.org/10.1207/s15327906mbr0102_10

3. Conover, W.J., 1998. Practical nonparametric statistics. Vol-350. John Wiley & Sons.

4. Horn, J.L., 1965. A rationale and test for the number of factors in factor analysis, Psychometrika, 30, pp.179–185. https://doi.org/10.1007/BF02289447

5. Huang, Y., Wang, H., Xiao, W., Chen, L.H., Yan, D.H., Zhou, Y.Y., Jiang, D.C. and Yang, M.Z., 2018. Spatial and temporal variability in the precipitation concentration in the upper reaches of the Hongshui River basin, southwestern China. Advances in Meteorology, 2018. https://doi.org/10.1155/2018/4329757

6. Kaiser, H.F., 1960. The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, pp.141–151. http://dx.doi.org/10.1177/001316446002000116

7. Kendall, M., 1975. Rank Correlation Methods. Griffin, London, 202pp.

8. Lawley, D.N., 1956. Tests for significance for the latent roots of covariance and correlation matrices. Biometrika 43(1-2), pp.128–136. https://doi.org/10.1093/biomet/43.1-2.128

9. Mann, H.B., 1945. Nonparametric Tests Against Trend, Econometrica, 13(3), pp.245-259. http://dx.doi.org/10.2307/1907187

10. Moran, P.A., 1950. Notes on continuous stochastic phenomena. Biometrika, 37(1/2), pp.17-23. https://doi.org/10.1037/2332142

11. North, G.R., Bell, T.L., Cahalan, R.F. and Moeng, F.J., 1982. Sampling errors in the estimation of empirical orthogonal functions. Monthly weather review, 110(7), pp.699-706. https://doi.org/10.1175/15200493(1982)110%3C0699:SEITEO%3E2.0.CO;2

12. Pearson, K., 1901. On Lines and Planes of Closest Fit to Systems of Points in Space. Philosophical Magazine, 2(11), pp.559–572. https://doi.org/10.1080/14786440109462720

13. Rajeevan, M., Bhave, J., Kale, J.D., and Lal, B., 2005. Development of a high resolution daily gridded rainfall data for the Indian Region (version 2), Meteorol. Monogr. Climatol. 22/2005, India Meteorol. Dep., New Delhi.

14. Rajeevan, M., Bhave, J., Kale, J.D., and Lal, B., 2006. High resolution daily gridded rainfall data for the Indian region: Analysis of break and active monsoon spells, Curr. Sci., 91(3), pp.296–306.

15. Raziei, T., 2018. A precipitation regionalization and regime for Iran based on multivariate analysis. Theoretical and applied climatology, 131(3), pp.1429-1448. https://doi.org/10.1007/s00704-017-2065-1

16. Sahu, R.T., Verma, M.K. and Ahmad, I., 2021. 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.

17. Sahu, R.T., Verma, M.K. and Ahmad, I., 2021. Regional Frequency Analysis Using L-Moment Methodology —A Review. In: Pathak K.K., Bandara J.M.S.J., Agrawal R. (eds) Recent Trends in Civil Engineering.
18. Sahu, R.T., Verma, M.K. and Ahmad, I., 2021. 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.

19. Sahu, N., Panda, A., Nayak, S., Saini, A., Mishra, M., Sayama, T., Sahu, L., Duan, W., Avtar, R. and Behera, S., 2020. Impact of Indo-Pacific Climate Variability on High Streamflow Events in Mahanadi River Basin, India. *Water*, 12(7), p.1952. [https://doi.org/10.3390/w12071952](https://doi.org/10.3390/w12071952)

20. Sahu, R.T., Kumar, K. and Verma, M.K., 2021. Trend analysis of rainfall in India with wavelet synopsis [Manuscript submitted for publication]. Civil Engineering Department, National Institute of Technology, Raipur.

21. Shepard, D., 1968. A two-dimensional interpolation function for irregularly spaced data, paper presented at 23rd ACM National Conference, January 1968, Assoc. for Comput. Mach., New York, pp.517-524. [https://doi.org/10.1145/800186.810616](https://doi.org/10.1145/800186.810616)

22. Singh, G., Panda, R.K. and Nair, A., 2020. 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, pp.5-19. [https://doi.org/10.1016/j.jher.2019.11.001](https://doi.org/10.1016/j.jher.2019.11.001)

23. Waller, L.A., 2009. Detection of clustering in spatial data. In: The SAGE Handbook of Spatial Analysis; SAGE Publications: Thousand Oaks, CA, USA, pp.299–320. [https://dx.doi.org/10.4135/9780857020130.n16](https://dx.doi.org/10.4135/9780857020130.n16)

24. Wang, T., Ren, C., Luo, Y. and Tian, J., 2019. NS-DBSCAN: A density-based clustering algorithm in network space. *ISPRS International Journal of Geo-Information*, 8(5), p.218. [https://doi.org/10.3390/ijgi8050218](https://doi.org/10.3390/ijgi8050218)

25. Yin, Y., Pan, X., Yang, X., Wang, X., Wang, G. and Sun, S., 2019. Spatiotemporal Changes and Frequency Analysis of Multiday Extreme Precipitation in the Huai River Basin during 1960 to 2014. *Advances in Meteorology, 2019*. [https://doi.org/10.1155/2019/6324878](https://doi.org/10.1155/2019/6324878)

26. Yuan, L., Yang, G., Li, H. and Zhang, Z., 2016. Spatio-temporal variation analysis of precipitation during 1960-2008 in the Poyang Lake basin, China. *Open Journal of Modern Hydrology*, 6(2), pp.115-127. [http://dx.doi.org/10.4236/ojmh.2016.62010](http://dx.doi.org/10.4236/ojmh.2016.62010)

**Figures**
Figure 1

The Mahanadi River basin
Figure 2

Eigenvalue decomposition of the correlation matrix for the three segments obtained using the spatial-mode of the sequential spatial pattern analysis with the application of Horn's parallel analysis for component retention.
Figure 3

Spatial distribution of loadings for the retained principal components, represented graphically in the ArcGIS platform using kriging interpolation.
Figure 4

Spatial distribution of the core origin for high loadings (positive or negative), a graphical representation of EOF coefficient upon the component it loads highly
Figure 5

PC scores of the varimax-rotated loadings and Regional monthly precipitation time series. X-axis presents the years and the first month of each season
Figure 6

Spatial distribution of rotated PC Scores obtained from implementing Temporal(T)-mode of PCA
Figure 7

Empirical cumulative distribution function of the regionalized patterns for all the three segments, the black dash vertical line represents the mean annual precipitation of the Mahanadi basin (1572mm)

Figure 8

Boxplot and KS ‘D – test’ (pairwise comparison) of regionalized patterns for all three segments
Figure 9

Comparison of empirical cumulative distribution function for each regionalized pattern to its different segment. The black dash vertical line represents the mean annual precipitation of the Mahanadi basin (1572mm)

Figure 10

Boxplot and KS ’D – test’ (pairwise comparison) for each regionalized pattern to its different segment
Figure 11

Boxplots of the monthly march of relative monthly precipitation for the identified regionalized patterns. The boxes illustrate the median, the first, and the third quartiles. The whiskers denote the data exceeding 1.5 times the inter-quartile range, and the dots show the outliers.