MAUSAM, **77**, 1 (January 2026), 109-118

MAUSAM

DOI: https://doi.org/10.54302/pw6a8b54
Homepage: https://mausamjournal.imd.gov.in/index.php/MAUSAM

UDC No. 551.577.3(540.49)

Statistical assessment of precipitation patterns using time series clustering: Insights from the Bundelkhand region of Central India

SAMIR BARMAN[1], RAMASUBRAMANIAN V.[2*], SANTOSHA RATHOD[3], BISHWA BHASKAR CHOUDHARY[1], AVIJIT GHOSH[1], SADHNA PANDEY[1], GAURENDRA GUPTA[1] and MRINMOY RAY[4]

[1]*ICAR - Indian Grassland and Fodder Research Institute, Jhansi, India – 284003*

[2]*ICAR - National Academy of Agricultural Research Management, Hyderabad, India – 500030*

[3]*ICAR - Indian Institute of Rice Research, Hyderabad, India – 500030*

[4]*ICAR - Indian Agricultural Statistics Research Institute, New Delhi, India – 110012*

(*Received 25 October 2024, Accepted 15 April 2025*)

***Corresponding author's email: ram.vaidhyanathan@gmail.com**

सार – वर्षा एक अत्यंत महत्वपूर्णजलवायु कारक है जो पारितंत्र, खेती और पानी के संसाधन पर असर डालता है, खासकर भारत में, जहाँ खेती अत्याधिकवर्षा पर निर्भर करती है। इस जलवायुतत्व में होने वाले बदलाव अनुसंधान करने वालों और पॉलिसी बनाने वालों के लिए बहुत आवश्यक हैं क्योंकि यह फैसले लेने की प्रक्रिया पर असर डालता है। इस अध्ययनका उदेश्य बुंदेलखंड इलाके में, जिसमें भारत में उत्तर प्रदेश और मध्य प्रदेश के कुछ हिस्से शामिल हैं, 1951 से 2023 तक वर्षा के प्रतिरूप का विश्लेषण करना और टाइम सीरीज़ (TS) समूहीकरणविधियों का उपयोग करके भविष्य के प्रवृत्ति का अनुमान लगाना है। TS समूहीकरणजो सिंगुलर स्पेक्ट्रम विश्लेषण (SSA) जैसी टेक्निक का उपयोग करके जलवायु डेटा में प्रतिरूप की पहचान करती है, वर्षा के डेटा के प्रवृत्ति, मौसम और स्वसहसंबंद के बारे में आवश्यक जानकारी देती है। इन प्रतिरूप की समूहीकरण करके, यह अध्ययनवर्षा में होने वाले बदलावों की समझ को बेहतर बनाने की कोशिश करती है, जिससे बेहतर खेती की योजना और फैसले लेने के लिए एक आधार मिलता है। अगले 24 महीनों के लिए वर्षा का अनुमान भी लगाया गया है। नतीजे बुंदेलखंड इलाके में दो अलग-अलग ज़ोन (क्लस्टर) की पहचान करते हैं, और दोनों ज़ोन (दक्षिणी और उत्तरी) में, समय के साथ वर्षा के प्रतिरूप में कमी देखी गई है। दक्षिणी हिस्से में कुल वर्षा ज़्यादा होती है, और वर्षा में उतार-चढ़ाव उत्तरी हिस्से की तुलना में कम हैं। बुंदेलखंड इलाके में सालाना वर्षा पूरे समय बहुत अनियमित रही है, महीनों में असमान रूप से बंटी हुई है, और वर्ष के एक-तिहाई हिस्से में अत्याधिक होती है।

ABSTRACT. Rainfall is a critical climatic factor influencing ecosystems, agriculture, and water resources, particularly in India, where agriculture relies heavily on rain-fed conditions. Fluctuations in this climatic element hold significant importance for researchers and policymakers as these influences decision-making processes. The study aims to analyze precipitation patterns in the Bundelkhand region, which includes parts of Uttar Pradesh and Madhya Pradesh in India, from 1951 to 2023, and to project future trends using time series (TS) clustering methods. TS clustering, which identifies patterns in climatic data using techniques like Singular Spectrum Analysis (SSA), offers valuable insights into trends, seasonality, and autocorrelations of rainfall data. By clustering these patterns, the study seeks to enhance understanding of rainfall variability, providing a basis for improved agricultural planning and decision-making. The projections of the rainfall have also been made for the next 24 months. The result identifies two different zones (clusters) in the Bundelkhand region, and in both zones (Southern and Northern), a decrease in rainfall patterns has been noted over time. Total rainfall is higher in the southern part, and fluctuations in rainfall are less compared to the northern part. The yearly rainfall in the Bundelkhand region is highly irregular throughout the period, unevenly distributed across the months, and highly concentrated in one-third of the year.

Key words – Climate change, Precipitation patterns, Rainfall variability, Singular spectrum analysis (SSA).

1. Introduction

Climate change has posed a significant challenge to agricultural production systems by altering the environmental conditions in which crops grow and livestock thrive (de Luis *et al*., 2011; Choudhary and Sirohi, 2020). This phenomenon triggers various effects, including shifts in temperature patterns, altered precipitation cycles and an increased occurrence of extreme weather events. According to the Intergovernmental Panel on Climate Change's (IPCC) Sixth Assessment Report, the average global surface temperature has risen by 1.1 °C from the period 1850-1900 to 2001-2021, with an accelerated rate of increase observed after the 1970s. With each additional 0.5 °C rise in temperature, the frequency of hot extremes, extreme precipitation and droughts is anticipated to increase. Furthermore, continued warming is projected to deplete the Earth's water reserves in the near future (IPCC, 2011). Precipitation helps mitigate the adverse effects of temperature by maintaining regional water balance, thus playing a crucial role in agricultural production (Panda and Sahu, 2019). However, erratic rainfall, such as unpredictable timing, intensity, or distribution, can lead to significant disruptions in crop growth, water resource management, and food security. For regions dependent on rain-fed agriculture, erratic rainfall can cause either droughts or floods, or both of which have severe economic and social consequences. Understanding the variation in rainfall is essential for developing accurate climate models, informing policy decisions, and implementing adaptive strategies to mitigate the impacts of climate change. Research on these variations can help identify trends, predict future patterns, and guide the development of resilient agricultural practices and water management systems, ultimately contributing to enhanced food security and sustainable development.

Variations in rainfall across any region can be captured by analyzing the time series (TS) clusters of this component and characterizing them at specific points or areas distributed over a geographical region. TS clustering identifies distinct patterns in climatic data, providing valuable insights for various purposes (Aghabozorgi *et al*., 2015; Alqahtani *et al*., 2021). It helps detect anomalies, identifies discrepancies, integrates tools for recognizing dynamic changes in TS, and supports prediction, recommendations, and pattern discovery over various domains.

Research on TS clustering has been done using three primary approaches. Primary approaches differ based on whether they involve direct analysis of raw data, indirect uses of features extracted from the raw data, or the indirect approach of model construction from the raw data

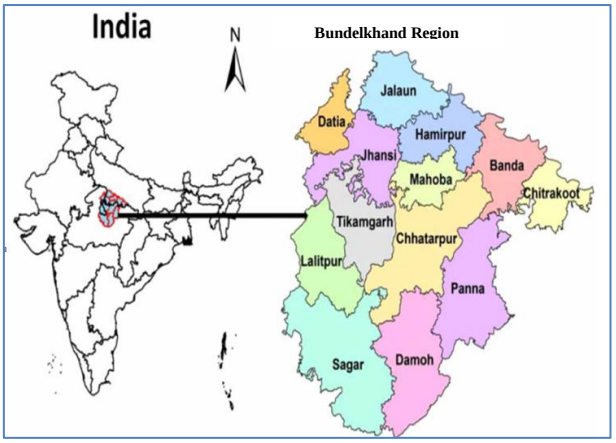

Fig. 1. Bundelkhand Region

Recently, Palacios Gutiérrez *et al*. (2023) suggested a new approach for clustering of TS data by utilizing decomposed TS components (trend, seasonal, and autoregressive components of residuals) through Singular Spectrum Analysis (SSA), focusing on feature extraction from the data. The SSA is a robust method in TS analysis and for TS forecasting, offering diverse applications such as nonparametric decomposition, filtering, parameter estimation, and forecasting within a broad spectrum of fields (Golyandina and Korobeynikov, 2014). In this TS clustering method, the TS is first decomposed, and then its three components (trend, seasonality, and residuals) are reconstructed to enable the extraction of descriptive parameters. Feature vectors are then constructed from these calculated parameters to represent each TS. Finally, the TS are clustered using unsupervised learning algorithms for further analysis.

The rainfall plays a vital role in the agriculture of the Bundelkhand region in India, mainly because it is a semi-arid area characterized by low and erratic precipitation (Jana *et al*., 2017; Jain *et al*., 2020; Som and Dey, 2022). Agriculture in Bundelkhand relies heavily on rain-fed conditions, with minimal irrigation facilities available. Adequate rainfall is essential for crop growth, soil fertility, and the replenishment of groundwater resources, which are often scarce in this region (Pandey *et al*., 2021). Monthly rainfall data on the Bundelkhand region (see Fig. 1) were used to validate the study. In Bundelkhand, the average number of rainy days is 52 days, according to the India Meteorological Department (IMD); however, in recent years, this number has decreased to just 24 days. Average rainfall in the area ranges from 800 to 900 mm, but its occurrence and the distribution of precipitation lack a consistent pattern. This unpredictability has left farmers unprepared for sowing, resulting in recurring drought issues nearly every year, even in years of relatively good rainfall (Sandhu *et al*., 2016). This study aims to identify evolving precipitation patterns and interpret them across

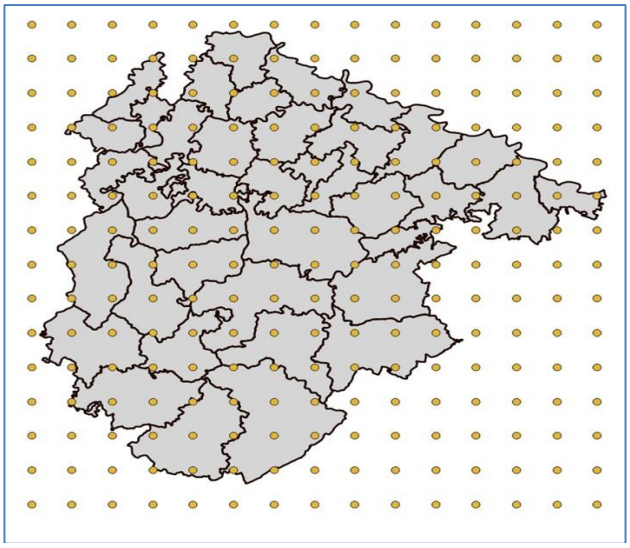

Fig. 2. Spatial Distribution of the points

different zones in Bundelkhand from 1951 to 2023, ultimately projecting precipitation for the next 24 months using TS clustering methods.

2. Data and methodology

2.1. *Data description*

Monthly gridded rainfall data (0.25° x 0.25°) in millimeters from the IMD, covering 225 grid locations in the Bundelkhand region, were used to validate the study (see Fig. 2). The grids include points distributed across 13 districts within the Bundelkhand region and nearby areas, covering monthly data from January 1951 to December 2023, with 876 observations at every grid point.

2.2. *Methodology*

Research on TS clustering generally revolves around three main approaches: direct analysis of raw data, indirect use of features derived from raw data, or indirect clustering through model construction from raw data. The TS clustering of precipitation data in the Bundelkhand region was done using the principals involved in extracting features from the data, introduced by Palacios Gutiérrez *et al*. (2023), with necessary reorientations to the data considered.

2.2.1. *Singular spectrum analysis (SSA) method*

The fundamental SSA algorithm breaks down an original TS into the sum of its components, including trend, seasonal component, and noise, without requiring prior knowledge about the TS structure. Let us consider a real-valued TS, $X_N = \{x_t\}_{t=1}^{N}$ of length $N(N > 2)$. The SSA algorithm consists of four steps: embedding, singular value decomposition (SVD), eigentriple grouping, and diagonal averaging. In addition, the first two steps are combined in the decomposition phase, while the last two are integrated into the reconstruction phase (Golyandina and Korobeynikov, 2014).

Step 1 - Embedding Step: The main aspect of the SSA algorithm is characterized by the window of length L, where $1 < L < N$. This step involves mapping the original TS into a sequence of K lagged vectors, where $K = N - L + 1$, each of size L, as follows:

$$X_j = \left(x_j, x_{j+1}, \dots, x_{j+L-1}\right)^T \quad (1)$$

for $j = 1,2, \dots, K$ to form the trajectory matrix X of the series X_N as

$$X = [X_1, X_2, \dots, X_K] = \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{pmatrix} \quad (2)$$

Step 2 - SVD: In this step, the SVD of the trajectory matrix X is computed using the eigenvalues and eigenvectors of the matrix $S = XX^T$:

$$X = [X_1 + X_2 + \cdots + X_L] = \sum_{i=1}^{L} \sqrt{\lambda_i} u_i v_i^T \quad (3)$$

where $u_1, u_2, \dots, u_L$ are eigenvectors (left singular vectors) of S, $v_i = X^T u_i / \sqrt{\lambda_i}$ are the factor vectors, λ_i are non-zero eigenvalues of S and $X_i = \sqrt{\lambda_i} u_i v_i^T (i = 1,2, \dots, L)$ are the elementary matrix of rank 1. Here, the collection (λ_i, u_i, v_i) is referred to as i^{th} eigentriple (ET) of the SVD, where $i = 1,2,..,L$, that encapsulates all information about each of the corresponding components X_i.

Step 3 - Eigentriple grouping: The set of indices $\{1,2,..,L\}$ is partitioned into m disjoint subsets $I_1, I_2, \dots, I_m$ where each subset $I = \{i_1, i_2, \dots, i_p\}$. It results in the matrix X_I corresponding to each subset I defined as

$$X_I = X_{i_1} + X_{i_2} + \cdots + X_{i_p}.$$

Following the computation of the resultant matrices for the groups $I = I_1, I_2, \dots, I_m$, the decomposition is described by equation (2):

$$X = X_{I1} + X_{I2} + \cdots + X_{\mathfrak{Z}} \quad (4)$$

Step 4 - Diagonal averaging: At this step, each matrix X_{I_j} of the grouped decomposition (4), where $j = 1,2,..,m$, is transformed into a new series of length N.

Furthermore, applying diagonal averaging to a resultant matrix X_{I_j} results in a *reconstructed series* $\tilde{X}^j = \left(\tilde{x}_1^{(j)}, \tilde{x}_2^{(j)}, \ldots, \tilde{x}_N^{(j)}\right)$. As a result, the original series $(x_1, \ldots, x_N)$ is decomposed into the sum of m reconstructed series as follows:

$$x_n = \sum_{j=1}^{m} \tilde{x}_n^{(j)}, n = 1,2, \ldots, N \quad (5)$$

The complete decomposition of the original TS is analogous to identifying the eigentriples in the SVD of the trajectory matrix X. These eigentriples correspond to the trend, periodic or seasonal components, and noise, and are categorized based on their component types.

Recommendations for selecting the window length can be derived from the (approximate) separability conditions. For instance, if the goal is to identify a known periodic (or seasonal) component with one or more periods from a TS, it is recommended to select a window length L that will be proportionate to the highest period and a sufficiently large L ($L \approx N/2$). Golyandina & Korobeynikov (2014) discussed the selection of SSA parameters in detail. The selection of length L also depends on the steps involved in extracting the features from the data. The first stage of *Sequential SSA* involves extraction of the trend from TS, with a minimum length of L, divisible by the period. The second stage involves the detection and extraction of the periodic components from the residue of the first stage, using a window length $L \approx N / 2$.

2.2.2. *Pattern identification and construction of feature vector for each TS*

Features are derived from the entire series to identify complete TS exhibiting similar patterns, such as comparable trends, stationarity, and autoregressive patterns. This involves the extraction of three components from the initial TS through sequential SSA, followed by modeling these components in a manner that allows the extraction of their associated characteristics as follows:

(*i*) Trend component: The extracted trend series, T_{trend}, is fitted using a linear regression as $T_{trend} = \mu + \beta t$, and the parameters $\mu$and β are estimated, where t is time.

(*ii*) Seasonality component: The extracted seasonal series, $T_{seasonal}$, is fitted as $T_{seasonal} = c_1 sin(2\pi t/\omega) + c_2 cos(2\pi t/\omega)$, and the parameters $c_1$and $c_2$are estimated using a linear regression, where ω is the period.

(*iii*) Residual component: From the extracted residual component $T_{residual}$, an $AR(T)$ is obtained and autocorrelations $\varphi_1, \varphi_2, \varphi_3$ and φ_ω are calculated, where ω represents the period.

Therefore, employing sequential SSA on each initial TS involves a two-stage process. Firstly, the trend-related component is isolated by selecting the smallest possible window length Lthat can be divided by the identified period ω to ensure the smoothness of the TS containing a periodic component. Secondly, by utilizing the maximum window length L ($L \approx N/2$ and divisible by ω) for enhanced separability, the seasonality of the residual of the first stage is extracted to create a new TS for the residual portion. After decomposing each initial series using sequential SSA and modeling the extracted TS, a feature vector is constructed. This vector consists of the estimated parameters for the i^{th} series, denoted as $C_i = \{\mu_i, \beta_i, c_{1i}, c_{2i}, \varphi_{1i}, \varphi_{2i}, \varphi_{3i}, \varphi_{\omega i}\}$.

2.2.3. *Clustering of TS based on feature vectors*

Various methods are available to evaluate the similarity or grouping of any two TS. In this investigation, the Euclidean distance is used as a similarity measure, and *hkmeans* (hierarchical k-means) clustering algorithm, an unsupervised learning algorithm, is employed to cluster feature vectors. During this, the normalized reconstructed C_i vector, scaled to a standard deviation of 1 and centered to a mean of 0, has been used for clustering.

The *hkmeans* technique is a hybrid approach, which is a combination of *hierarchical* and *k-means* clustering methods. It comprises three steps: first, *hierarchical* clustering is performed, and the tree is segmented into *k*-clusters; next, the center (*i.e*., the mean) of each cluster is determined; finally, *k-means* is computed by using the set of cluster centers as the initial cluster centers.

2.2.4. *SSA forecasting*

A TS via SSA can be successfully forecasted if the following conditions are met: the series has a well-defined structure that can be identified using an algorithm, a continuation method based on this structure is accessible, and the structure remains consistent over the forecast period, see Marques *et al*., 2006; Hassani *et al*., 2009; Kalantari, 2021; Movahedifar *et al*., 2023. With these conditions in place, a forecasting model can be developed. The complete algorithm for SSA forecasting has been described by Golyandina *et al*. (2001).

The two methods of SSA forecasting are recurrent and vector forecasting. When the TS can be represented by a linear recurrent relation (LRR), and if it is minimal, the recurrent SSA forecasting method is optimal (Golyandina and Zhigljavsky, 2020). In this approach, the model is defined by a Linear Recurrent Formulae (LRF), which is applied to the last L-1 terms of the initial TS to extend it.

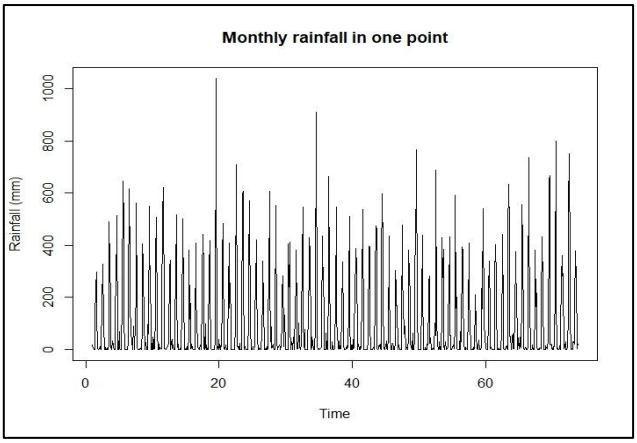

Fig. 3. Average Monthly rainfall (mm) at a point from 1951 to 2023

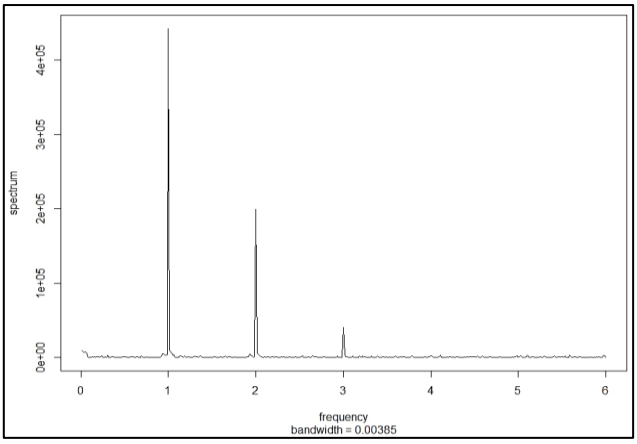

Fig. 4. Periodogram of initial TS

Let, I be the chosen set of eigentriple associated with the grouping set of SSA, $U_i \in R^L, i \in I$ is the corresponding eigenvectors of the chosen eigentriple and $\underline{U_i} \in R^{L-1}$ is their first $l-1$ coordinates corresponding to U_i. Also, π_i be the last coordinate of the U_i, $\nu^2 = \sum_{i \in I} \pi_i^2$ and $\tilde{X} = \{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_N\}$ be the reconstructed TS by I.

The recurrent forecasting algorithm can be represented as

$$z_i = \begin{cases} \tilde{x}_i & for \quad i = 1, \dots, N \\ \sum_{l=1}^{L-1} a_i y_{i-l} & for \quad i = N+1, \dots, N+h \end{cases}$$

where h is the forecast horizon and the vector of linear coefficients $R = (a_1, \dots a_{L-1})$ is defined as:

$$R = \frac{1}{1-\nu^2} \sum_{i \in I} \pi_i \underline{U_i}.$$

The numbers $z_{N+1}, \dots, z_{N+h}$ are the h step ahead recurrent forecast.

3. Results and discussion

The study's validation utilized monthly gridded rainfall data (0.25 x 0.25) in millimeters (mm) collected from IMD, covering 225 grid locations in the Bundelkhand region of India. These grids encompass points distributed across 13 districts in the Bundelkhand region and its nearest areas, see Fig. 2. The dataset spans from January 1951 to December 2023 for each grid point, totaling 876 observations. In Bundelkhand, the average rainy days, according to IMD, are 52 days, but in recent years, it has been limited to 24 days. Agriculture in this region heavily relies on rainfall as most cultivated lands here are rain-fed, and the irrigation facilities are limited. Although the average precipitation is 800-900 mm, its occurrence lacks a definite pattern, leading to unpreparedness among farmers for sowing. Consequently, the region faces drought-related challenges almost yearly, even in years with good rainfall (Sandhu *et al*., 2016).

As depicted in Fig. 3, the monthly average rainfall data within the grid points in the Bundelkhand region exhibits varying seasonality. Peak rainfall is observed during the monsoon months, with a decrease in rainfall during the rest of the time. A fluctuation pattern (in trend) is observed over time, as highlighted later in Fig. 7.

It has been found that the trend of the TS is intricate, suggesting a need for a sequential decomposition. Initially, the trend is isolated; to address such a dynamically changing trend, we smooth it out. The window of length is set to 12, as suggested by Golyandina and Korobeynikov (2014). Since we are dealing with monthly data, the horizontal axis of the periodogram goes from 0 to 6, as shown in Fig. 4. The periodogram reflects a long-term trend, with peaks at 1, 2, and 3 corresponding to annual, semi-annual, and quarterly patterns, respectively.

After performing the first decomposition, if we look at Figs. 5 and 6, we can see that the first set of information mainly represents the trend. The other sets contain details that change quickly, suggesting they are not part of the trend. Interestingly, when we use this main set to put the data back together, it gives us the extracted trend, as seen in Fig. 7. Seasonality is then extracted from the residual produced during the first stage of decomposition, following the trend extraction. Since the TS has a length of N=876 the window length L is set to half its value, *i.e*., 438, to ensure good separability. It ensures that $L \leq N/2$ and is divisible by 12. Fig. 9 illustrates multiple steps where approximately equal eigenvalues result in pairs of eigenvectors corresponding to sine waves. It shows three almost regular polygons, ET1 - 2, ET3 - 4, and ET5-6

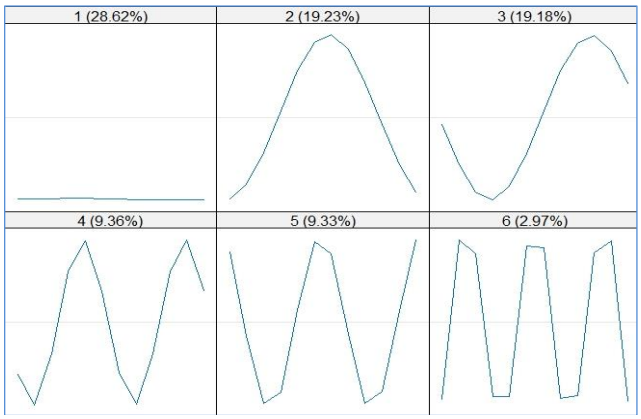

Fig. 5. Eigenvectors in the first stage (L=12) of the initial TS

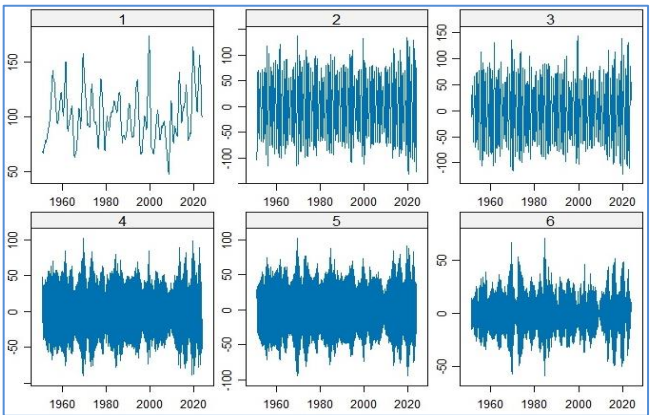

Fig. 6. Elemental reconstructed series in the first stage

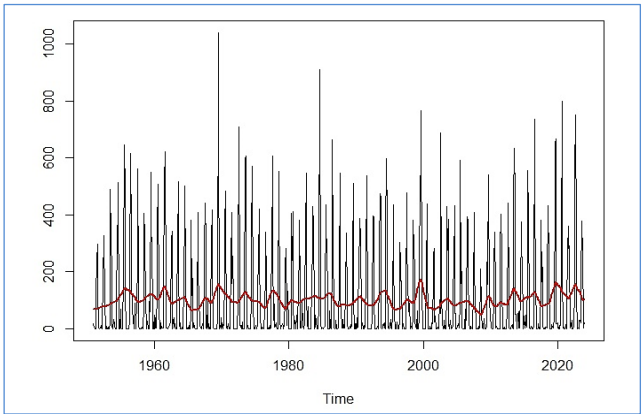

Fig. 7. Initial series and estimated trend at the first stage

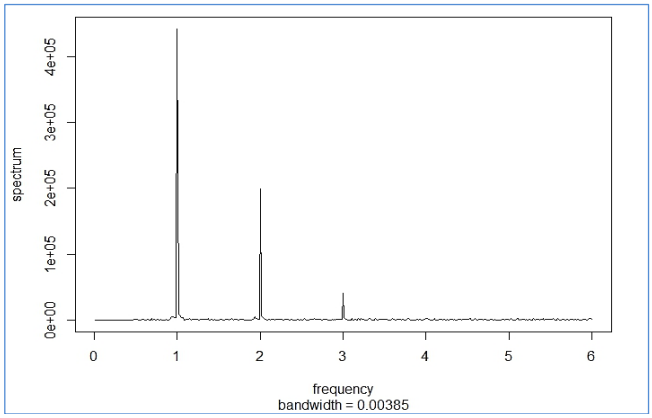

Fig. 8. Second-stage periodogram of the first-stage residuals

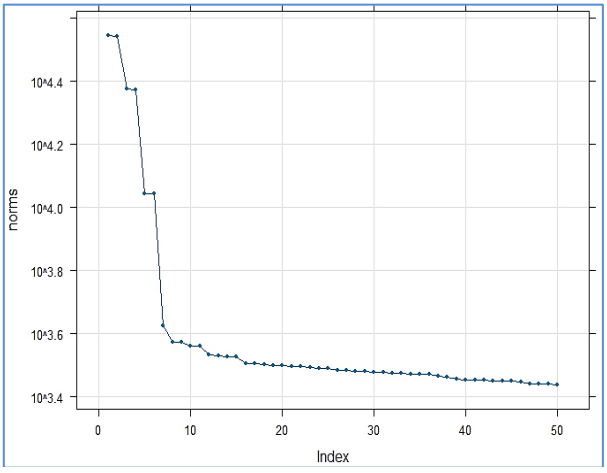

Fig. 9. Second stage Eigenvalues (L=432)

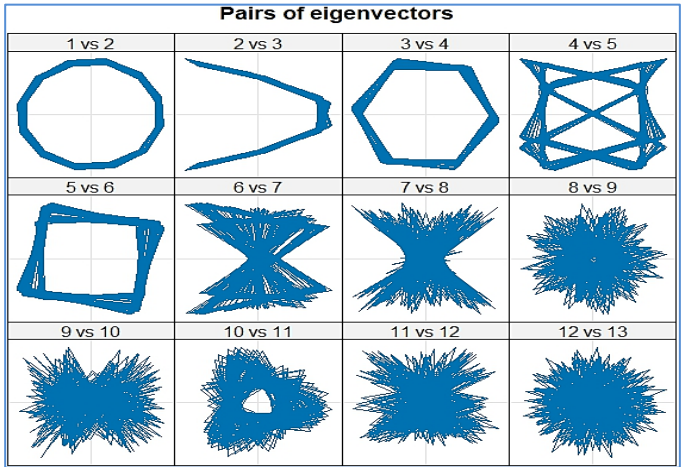

Fig. 10. Second stage scatter plot of eigenvectors (L=432)

corresponding to periods 12, 6, and 4, which occur due to seasonality, see Fig. 10, and are clearly explained by the periodogram (Fig. 8). The final sequential SSA decomposition of the initial TS is displayed in Fig. 11.

After figuring out the TS components, *viz*., trend, seasonality component, and noise in the data employing sequential SSA, we then figure out the details for each of these parts. The values we figure out for each part are

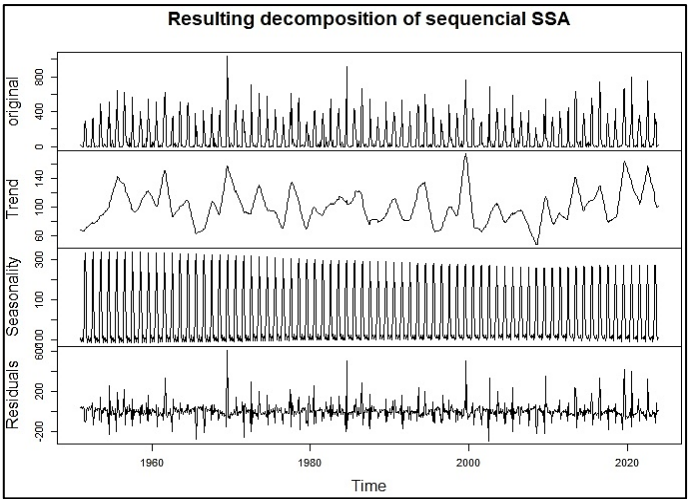

Fig. 11. Decomposed series

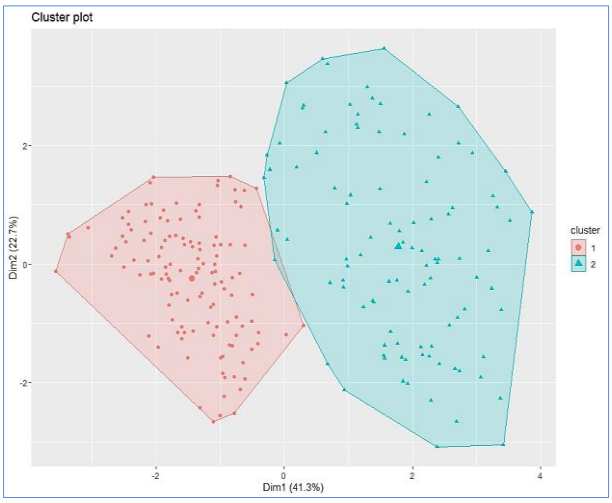

Fig. 12. Clusters of the TS

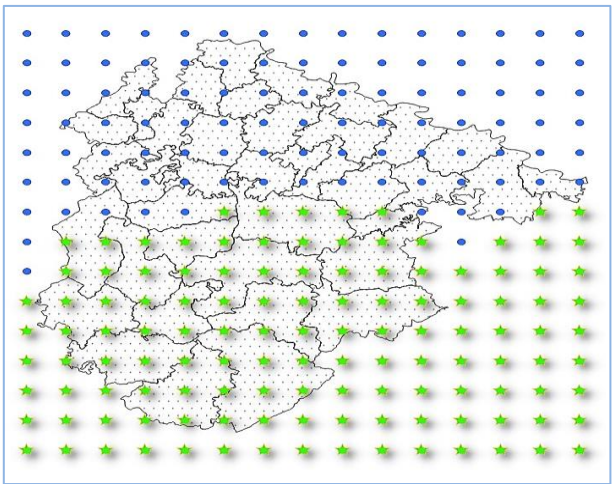

Fig. 13. Spread of grid points in Bundelkhand region. Green Star: Cluster 1 and Blue Dot: Cluster 2.

listed in Table 1, which shows how each component of the initial data is represented. We repeat this process for the sets of TS data for every point, creating a feature list (a set consisting of all feature vectors of the TS) for each one. Finally, clustering was done using the *hkmeans* approach. The final clusters are shown in Fig. 12. The results show that the Bundelkhand region has been partitioned into two clusters.

Fig. 13 shows that Cluster 1 encompasses the southern part of the Bundelkhand region, covering the areas of Sagar, Damoh, Panna, the southern part of Chhatarpur, a small part of Tikamgarh, and the southern part of Lalitpur. Cluster 2 primarily includes the northern part of the Bundelkhand region, covering the areas of Jhansi, Jalaun, Hamirpur, Banda, Chitrakoot, Mahoba, Datia, and the northern part of Tikamgarh, Lalitpur, and Chhatarpur.

TABLE 1

Parameters associated with SSA extracted components, used to obtain feature vector for a particular spatial location

Trend			μ	β
			99.3987	0.0035
Seasonality			c_1	c_2
			-122.40	-103.90
Residuals	φ_1	φ_2	φ_3	φ_{12}
	-0.1390	-0.0316	-0.0939	-0.0264

To gather details about the groups obtained, we examine the centroids of each group. These centroids represent the average value of each characteristic used in defining the feature vectors for each cluster. As a result, Cluster 1 is associated with areas in the Bundelkhand region that receive higher rainfall than the other clusters. The seasonal component of these TS usually correlates/reflects cyclical variations characterized by c_1 = -111.97 and c_2 = -92.84, with an amplitude of 145.456. On the contrary, Cluster 2 includes points in the Bundelkhand region where the monthly average rainfall is lower than that in Cluster 1, and these values exhibit a slight decrease over time. The seasonal component, in this case, has an amplitude of cyclic variation of 103.735, defined by c_1 = -82.46 and c_2 = -62.94. Also, it has been found that the primary autocorrelation of the residuals is negative for both clusters.

By analyzing the patterns derived from the centroid series of each cluster, it becomes evident that the Bundelkhand region is divided into two distinct zones. The innovative trend analysis (see, Chowdari *et al*., 2023) over annual rainfall indicates that the Northern part (Cluster 2) has a decreasing trend in rainfall over time, while a similar decreasing trend is observed in the Southern part (Cluster 1) but slightly less than the North,

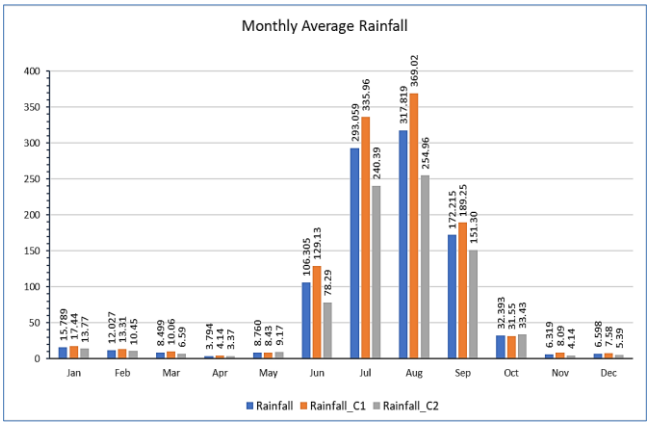

Fig. 14. Monthly average rainfall (mm) in Cluster 1 and Cluster 2

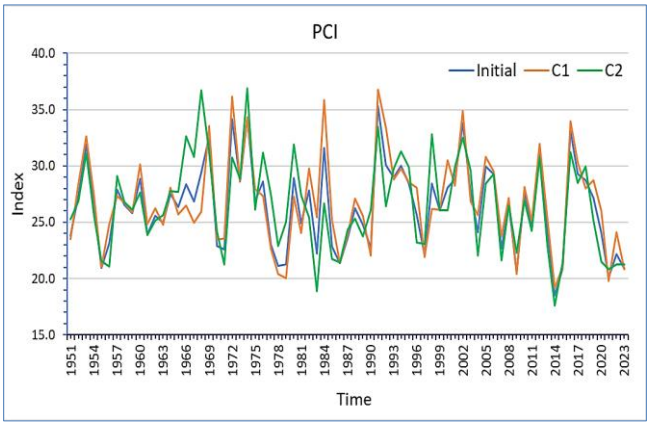

Fig. 15. Precipitation Concentration Index (PCI)

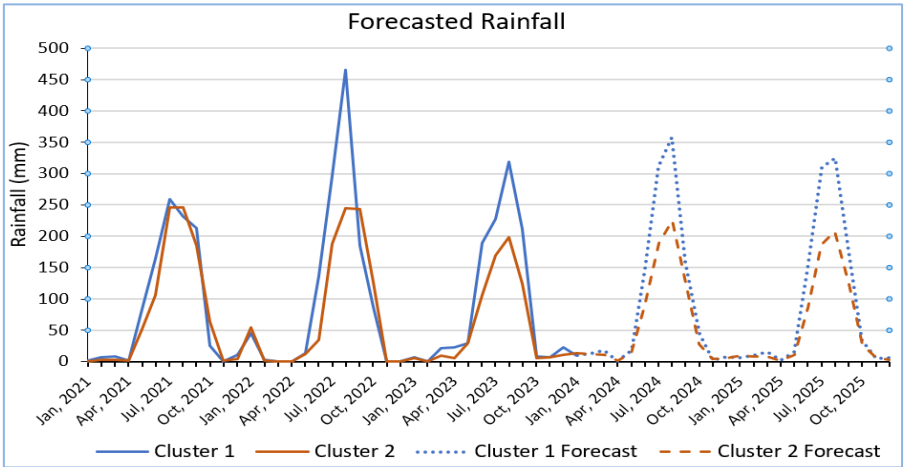

Fig. 16. 24-month ahead forecasted rainfall (mm) for both clusters.

TABLE 2

Timeseries parameters for monthly rainfall of Bundelkhand region

	Cluster 1			Cluster 2	
Trend	μ = 98.313*	β= -0.011*	Trend	μ = 77.954*	β= -0.021*
Seasonality	c_1 = -111.971*	c_2 = -92.843*	Seasonality	c_1 = -82.461*	c_2 = -62.943*

see Table 4. Similar results have also been reported for monthly patterns in Table 2.

Moreover, from Table 3 and Fig. 14, the CV value is higher in the Northern part (Cluster 2), indicating that the fluctuations in average rainfall in different months are more than in the Southern part. 90% of the rainfall has happened mainly during the rainy season.

Furthermore, when analyzing the annual rainfall data across the region, it is evident that Cluster 1 experiences an average annual rainfall of 1124.74 mm with a coefficient of variation (CV) of 20.23%. In contrast, Cluster 2 receives an average of 813.19 mm annually with a CV of 22.42%. These findings highlight a notable disparity in rainfall between the southern and northern regions, with the southern region experiencing higher average rainfall.

where P_i represents the monthly rainfall for the month i, i ranges from 1 for January to 12 for December. Fig. 15 depicts the pattern of temporal change. According to Oliver's (1980) categorization, a PCI value between 16 and 20 implies irregular rainfall concentration within one-third of the year, while a PCI value exceeding 20 suggests highly irregular rainfall distribution throughout the period. In the Bundelkhand region, the annual PCI indicates a strong uneven rainfall distribution, with PCI values surpassing 20 in all the years except 2013-14. In simpler terms, the annual rainfall in Bundelkhand is not evenly distributed throughout the year but is heavily concentrated within a third of it.

TABLE 3

Monthly Average Rainfall (mm) in the Bundelkhand region

	Overall	Overall_CV %	Cluster_1	Cluster_1_CV %	Cluster_2	Cluster_2_CV%
Jan	15.92	106.55	17.58	116.43	13.88	110.94
Feb	12.19	130.19	13.50	132.82	10.59	146.20
Mar	8.39	138.40	9.89	140.88	6.54	159.77
Apr	3.64	124.96	3.89	124.08	3.34	166.80
May	8.47	122.23	8.13	145.48	8.88	109.56
Jun	105.67	70.19	128.29	70.03	77.90	79.20
Jul	294.34	29.48	337.47	32.87	241.38	30.37
Aug	318.56	30.27	369.72	31.92	255.75	36.00
Sep	172.21	55.80	188.93	58.10	151.68	57.48
Oct	32.75	117.57	31.88	107.69	33.82	141.23
Nov	6.31	213.24	8.10	215.42	4.10	245.19
Dec	6.45	195.81	7.37	207.34	5.32	199.45

TABLE 4

Innovative Trend Analysis (ITA) over Annual Rainfall

Cluster 1		Cluster 2	
Slope (s)	-1.8146*	Slope (s)	-3.6411*
Slope SD	0.18229	Slope SD	0.12650
Change Point (year)	1985	Change Point (year)	1971

Additionally, the Precipitation Concentration Index (PCI) was utilized to evaluate the annual and seasonal pattern of rainfall distribution using the following formula:

$$PIC_{Annual} = \frac{\sum_{i=1}^{12} P_i^2}{\left(\sum_{i=1}^{12} P_i\right)^2} \times 100$$

Finally, the future projection, *i.e.*, the 24-month rainfall forecast for both clusters, was computed using the recurrent SSA forecasting method, as previously described. The forecasted values (see Fig. 16) suggest that the rainfall pattern is anticipated to be slightly higher than in 2023 but lower than in 2021 & 2022. It also indicates that rainfall in the Cluster 1 region will remain higher than in the Cluster 2 region, in line with past trends observed in 2022 and 2023.

4. Conclusions

This study highlights the utility of TS clustering in identifying distinct rainfall patterns across the Bundelkhand region of Central India. By analyzing rainfall data trends, seasonality, and noise components from 1951 to 2023, the study has identified two distinct climatic zones: South and North Bundelkhand. The analysis reveals a general decline in rainfall over time, with the Southern zone receiving higher rainfall and exhibiting less fluctuations compared to the Northern zone. These findings underscore the irregular and uneven distribution of rainfall in Bundelkhand, with the majority of precipitation concentrated in just one-third of the year. To address these challenges, a comprehensive approach is needed, including the development of water conservation infrastructure like rainwater harvesting systems and check dams, and promoting drought-resistant crops along with water-efficient irrigation practices tailored to the Southern and Northern zones. Establishing an early warning system for localized weather forecasting, promoting sustainable land use practices such as agroforestry and soil moisture management, and fostering climate-smart infrastructure and community-based adaptation strategies will further enhance resilience. Coordinating policies across government levels will ensure a unified regional response, while public-private partnerships can drive innovation and provide funding for climate resilience projects, ensuring Bundelkhand region to adapt effectively to its changing climate.

Availability of data

The data can be accessed through the IMD website.

Code availability

The basic codes are available on the R platform.

Acknowledgments

The authors express gratitude to the ICAR-National Academy of Agricultural Research Management (ICAR-NAARM), Hyderabad, for their support and provision of necessary facilities that facilitated the successful completion of this study under Professional Attachment Training (PAT) to the first author. The authors also duly acknowledge the ICAR-Indian Grassland and Fodder Research Institute (ICAR-IGFRI), Jhansi, for kind support

in allowing the completion of PAT at ICAR-NAARM, Hyderabad.

Authors' contributions

Samir Barman: Conceptualization, Methodology, Formal analysis, Writing - review and editing and investigation and Writing - original draft preparation. (*email-samir.barman@icar.org.in*).
Ramasubramanian V: Conceptualization, Methodology, Formal analysis and investigation, Writing - review and editing and Writing - original draft preparation. (*email-rsubramanianv.iasri@icar.org.in*).
Santosha Rathod: Methodology. (*email-santoshagriculture@gmail.com*).
Bishwa Bhaskar Choudhary: Methodology, Formal analysis and investigation. (*email-bishwa606@gmail.com*).
Mrinmoy Ray: Formal analysis and investigation (*email-mrinmoy4848@gmail.com*).
Avijit Ghosh: Writing - original draft preparation (*email-avijitghosh19892@gmail.com*).
Sadhna Pandey: Writing - original draft preparation (*email- sadhnap69@yahoo.co.in*).
Gaurendra Gupta: Writing - review and editing (*email-guptagaurendra@gmail.com*).

Disclaimer: The contents and views presented in this research article/paper are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

Reference

Aghabozorgi, S., Seyed Shirkhorshidi, A. and Ying Wah, T., 2015, "Time-series clustering – A decade review", *Inf. Syst.*, **53**, 16-38. doi : https://doi.org/10.1016/j.is.2015.04.007.

Alqahtani, A., Ali, M., Xie, X. and Jones, M.W., 2021, "Deep Time-Series Clustering: A Review", *Electronics*, **10**, 3001. doi : https://doi.org/ 10.3390/electronics10233001.

Choudhary, B.B. and Sirohi, S., 2020, "Modelling climate sensitivity of agriculture in Transand Upper Gangetic Plains of India", *Theor. Appl. Climatol.*, **142**, 381-391.

Chowdari, K.K., Deb Barma, S., Bhat, N., Girisha, R., Gouda, K.C. and Mahesha, A., 2023, "Trends of seasonal and annual rainfall of semi-arid districts of Karnataka, India: application of innovative trend analysis approach", *Theor. Appl. Climatol.*, **152**, 241-264. doi : https://doi.org/10.1007/s00704-023-04400-9.

de Luis, M., González-Hidalgo, J.C., Brunetti, M. and Longares, L.A., 2011, "Precipitation concentration changes in Spain 1946–2005", *Nat. Hazards Earth Syst. Sci.*, **11**, 1259-1265. doi : https://doi.org/ 10.5194/nhess-11-1259-2011.

Golyandina, N., Korobeynikov, A., 2014, "Basic Singular Spectrum Analysis and forecasting with R", *Comput. Stat. Data Anal.*, **71**, 934-954. doi : https://doi.org/10.1016/j.csda.2013.04.009.

Golyandina, N., Nekrutkin, V.V, Žigljavskij, A.A., Nekrutkin, V.V., Zhigliavskiĭ, A.A., 2001, "*Analysis of time series structure: SSA and related techniques, Monographs on statistics and applied probability*", Chapman & Hall / CRC, Boca Raton, Fla.

Golyandina, N., Zhigljavsky, A., 2020, "*Singular Spectrum Analysis for Time Series*", SpringerBriefs in Statistics, Springer Berlin Heidelberg, Berlin, Heidelberg. doi : https://doi.org/ 10.1007/978-3-662-62436-4.

Hassani, H., Heravi, S. and Zhigljavsky, A., 2009, "Forecasting European industrial production with singular spectrum analysis", *Int. J. Forecast.*, **25**, 103-118. doi : https://doi.org/10.1016/j. ijforecast.2008.09.007.

IPCC: Masson-Delmotte V, Zhai P, Chen Y, Goldfarb L, Gomis MI, Matthews JBR, Berger S, Huang M, Yelekçi O, Yu R, Zhou, B., Lonnoy E, Maycock TK, Waterfield T, Leitzell K, Caud N, 2011, In: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (ed) Climate Change 2021: the physical science basis.

Jain, R., Chand, P., Rao, S.C. and Agarwal, P., 2020, "Crop and soil suitability analysis using multi-criteria decision making in drought-prone semi-arid tropics in India", *Front. Sustain. Food Syst.*, **19**, 271. doi : https://doi.org/10.5958/2455-7145.2020.00036.3

Jana, C., Alam, N.M., Mandal, D., Shamim, M., Kaushal, R., 2017, "Spatio-temporal rainfall trends in the twentieth century for Bundelkhand region, India", *J. Water Clim. Chang.*, **8**, 441-455. doi : https://doi.org/10.2166/wcc.2017.120.

Kalantari, M., 2021, "Forecasting COVID-19 pandemic using optimal singular spectrum analysis", *Chaos, Solitons & Fractals*, **142**, 110547. doi : https://doi.org/10.1016/j.chaos.2020.110547.

Marques, C.A.F., Ferreira, J.A., Rocha, A., Castanheira, J.M., Melo-Gonçalves, P., Vaz, N., Dias, J.M., 2006, "Singular spectrum analysis and forecasting of hydrological time series", *Phys. Chem. Of Earth, Parts A/B/C*, **31**, 1172-1179. doi : https://doi.org/10.1016/j.pce.2006.02.061.

Movahedifar, M., Hassani, H., Kalantari, M., 2023, "Recurrent Forecasting in Singular Spectrum Decomposition", *Eng. Proc.*, **39**, 68. doi : https://doi.org/10.3390/engproc2023039068.

Oliver, J.E., 1980, "Monthly precipitation distribution: a comparative index", *Prof. Geogr.*, **32**, 300-309. doi : https://doi.org/10.1111/ j.0033-0124.1980.00300.x.

Palacios Gutiérrez, A., Valencia Delfa, J.L. and Villeta López, M., 2023, "Time series clustering using trend, seasonal and autoregressive components to identify maximum temperature patterns in the Iberian Peninsula", *Env. Ecol Stat.*, **30**, 421-442. doi : https://doi.org/10.1007/s10651-023-00572-9.

Panda, A. and Sahu, N., 2019, "Trend analysis of seasonal rainfall and temperature pattern in Kalahandi, Bolangir and Koraput districts of Odisha, India", *Atmos. Sci. Lett.*, **20**, e932. doi : https://doi.org/10.1002/asl.932.

Pandey, V., Srivastava, P. K., Singh, S. K., Petropoulos, G. P. and Mall, R. K., 2021, "Drought identification and trend analysis using long-term CHIRPS satellite precipitation product in Bundelkhand, India", *Sustainability*, **13**, 1042. doi : https://doi .org/10.3390 /su13031042.

Sandhu, L.S., Gautam, U.S., Dubey, S.K., Singh, A., Kumar, R.V. and Singh, H.V, 2016, "Agro-climatic Region Centered Research and Development Planning (Central Plateau and Hill region)".

Som, K.S., Dey, M., 2022, "Analysis and forecasting of rainfall trends in semi-arid Bundelkhand region of Madhya Pradesh, India: using statistical methods", *Environ. Earth Sci.*, **81**, 98. doi : https://doi.org/10.1007/s12665-021-10159-x.