iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IX Month of publication: September 2025

DOI: https://doi.org/10.22214/ijraset.2025.74183

Cluster Analysis for Sustainable Soybean Production in India

Ram Manohar Patel[1], Kamal Bunkar[2], Chhaya Arya[3]
[1, 2]Institute of Computer Science, Vikram University, Ujjain, Madhya Pradesh, India
[3]International Institute of Professional Studies, Vikram University, Ujjain, Madhya Pradesh, India

Abstract: Clustering spatial growth and dynamics is crucial for understanding output and concentration patterns for optimizing resources and promoting sustainable development. India, with Madhya Pradesh (MP) as the leading soybean state, exhibits significant growth and instability patterns. Thus, we explored the spatiotemporal patterns of soybean growth and instability since 1964. Nationally, the pooled growth of soybean is positive, accompanied by an unstable area and production, along with medium yield stability. However, despite stable and positive decadal growth, yield growth has recently declined to negative. In MP as well, these variables grew positively and outpaced national growth; however, they have declined to negative in the last decade. Stability has increased to a stable state but is currently at a medium level. The clustering of MP districts identified 5, 5 and 4 optimum clusters using the 'Elbow' method, with high precision in the growth of these variables. The clustered maps highlighted the Malwa Plateau, Jhabua Hills, Nimar Valley, Vindhiyan Plateau, Gird Region, Northern Hills, and Harda and Betul districts as potential high-growth regions, whereas the Kymore Plateau stands out as a low-growth region. Supervised clustering revealed a more stable soybean area in western MP over the last decade, whereas production stability generally remained low throughout the state. Therefore, developing climate-resilient technologies, improving extension services and formulating clusterwise policies should be prioritized to boost sustainable soybean production in India.

Keywords: Soybean, Agroclimatic Zones, K-Means clustering, Compound Annualized Growth Rates (CAGR), Cuddy Della Instability Index (CDVI)

I. INTRODUCTION

Sustainable agriculture retains stability and requires not only stimulation of production growth but also reduction of instability. Agriculture is susceptible to many biotic and abiotic factors that create production instability, ultimately leading to food insecurity [1]. Agriculture plays a vital role in India, ensuring food security for 65% of the population. Food security is a global challenge and critical for developing and populous countries such as India [2].

The soybean cultivation is suitable in varied agroecological zones, but the expansion in central India over the last five decades is unparalleled due to the low cost of cultivation [3,4]. Initially, cultivation began in rainfed fallow lands; later, it replaced fallow lands and less profitable competing crops (cotton, maize, sorghum, pearl millet, and other minor crops) in Madhya Pradesh (MP) [5]. The MP, known as the 'Soya State', has an area of 5.97 million hectares and a production of 6.33 million tons from 2022-23, contributing 42.35% of the country's total production. Despite the growth in area and production, breaking the jinx of low yield remains a significant challenge. Soybean yield oscillates between 1.0 and 1.3 tons per hectare, which is not impressive compared to world yield and its climatic potential yield (about one-third) [6]. The total soybean area and production of the country have increased to 408.26 and 945.38 times from 1970-71 to 2023-24, with compound annual growth rates (CAGRs) of 11.15% and 12.2%, respectively. Over the last one and a half decades (2011-12 to 2023-24), it has increased at a rates of 1.84 and 1.11, respectively, which implies that the production increase is due to the increase in area. Nearly 99% of the country's soybean was cultivated under rain-fed conditions from 2021–22. It has an area share of 62.81% in rain-fed oilseeds and 42.06% of total oilseeds, along with a 32% share in total oilseed production, 26% of total vegetable oil and a two-thirds share of total oil cakes in the country from 2021-22 [7]. Globally, soybean area and production are growing at annual rates of 2.48% and 3.6%, respectively [8]. Soymeal leads the protein meal market, while soybean oil is the second-largest contributor to the vegetable oil market, followed by palm oil. Soybean contributes nearly two-thirds of the total protein and one-fourth of the total edible oil produced worldwide [9]. Worldwide, soybean consumption for food uses has been increasing at a rate of 3.35% annually, with India ranking third after China and Japan [8]. The growth of a variable signifies its past performance and growth analysis is generally used for its trend analysis, indicating its performance [10]. Thus, crop growth and instability are crucial indicators of crop production performance. Studying crop production growth and instability is essential for enabling the government to formulate suitable production and trade policies to minimize imports.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

Many studies on the growth and instability of soybean in India are available, but most of them focus either on the country or state level. District-level studies analyzing the growth and instability of soybean in India are very rare. The available district-level studies are limited to small areas, short time periods or similar constraints. No comprehensive study on district level analysis of a state has been conducted since the introduction of soybean in India. This study performed spatiotemporal cluster analysis of Madhya Pradesh districts with different CAGR values and instability indices to understand the concentration patterns of soybean growth and dynamics, helping decision-makers formulate clusterwise policies for optimizing resources and promoting sustainable soybean development.

II. DATA AND METHODOLOGY

A. Data Sources

The soybean was introduced in India during the Green Revolution period. The data is available since 1964 with missing values that were interpolated using graduated average of preceding and succeeding values. Many districts were divided over the time and their areas and productions were adjusted based on area share. During the initial periods, the year in which a district had at least 20 hectares of soybean area was considered the base year for the analysis of the district. The secondary data obtained from the Directorate of Agriculture and Cooperation (GOI), ICRISAT database and Department of Agriculture, Government of Madhya Pradesh, India. The district-wise data of area (A), production (P) and yield (Y) of soybean in MP were collected from 1964 to 2023. MP is divided into eleven agroclimatic zones (ACZs), with districts under them [11], as depicted in Fig. 1.

To better understand the dynamics, the analysis period is divided into six decadal periods [PC (Pre-Commercialization): 1964-1970, I: 1971-1980, II: 1981-1990, III: 1991-2000, IV: 2001-2010, V: 2011-2022, and VI (Pooled): 1964-2022] for the MP state and periods [I: 1970-1980, II: 1981-1990, III: 1991-2000, IV: 2001-2010, V: 2011-2022, and VI (Pooled): 1964-2022] for the country-level analyses. However, district-level analysis is conducted for the last decade (2011-2022) (Period-A) and the pooled period (Period-B).

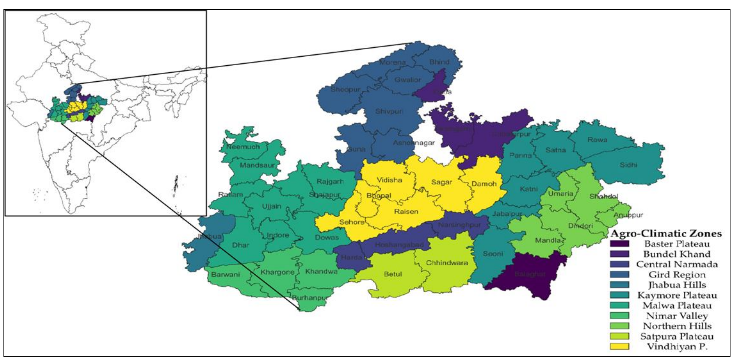

Fig. 1. District-wise agro-climatic zones (ACZs) of Madhya Pradesh

B. Analytical Framework

1) Estimation of Growth Rates: Growth rates are estimated using two approaches, namely, the linear growth rate (LGR) and the compound annual growth rate (CAGR). To avoid seasonal and cyclical fluctuations, CAGR is more appropriate for assessing growth rates in APY. The CAGR is used to assess trends in growth rates of APY of soybean in the districts of MP. The growth rates were estimated using the following three ways,

i) Exponential trend model [12,13].

$$\text{Exponential trend equation: } Y = ab^t \tag{1}$$

ii) Semi-logarithmic form of CAGR is expressed as equation (Eq. 2) below. where Y is the area (000 ha)/production (000 tons)/productivity (kg/ha); t is the time period in year; a is a constant; b = (1 + r) is a regression coefficient; and r is the compound annual growth rate.

$$\text{Log } Y = \log a + t \log b \tag{2}$$

iii) Percent compound growth rate (*r*): The growth rate computed by regression coefficient 'b'. Hence, the CAGR is estimated by equation (Eq. 3). The t test at the 95% probability level was used to check the significance of 'b' in the APY of soybean, with the null hypothesis (H0) stating no trend in the growth rate and the alternative hypothesis (H1) stating a trend in the growth rate.

$$r = [(\text{Antilog of } b) - 1] \times 100 \tag{3}$$

2) Instability Analysis: Sustainable agriculture is driven by high growth and low risk or instability in agricultural produce. Instability is a crucial decision parameter, dependent on many factors (such as weather, production technologies, improved and hybrid varieties, etc.), which affects price stability and production dynamics in cases of production instability [14]. Agricultural instability is measured mainly by the coefficient of variation (CV) [15], the Cuddy–Della–Valle Index (CDVI), and Coppock's Instability Index (CII) [16]. CDVI is common in several studies and a better measure for assessing agricultural instability than CV [17]. CV has the limitation of overestimating instability that is resolved by CDVI by detrending the time series data, thereby representing the correct direction of instability [18]. Thus, CDVI is utilized here to compute the instability. The lower the CDVI value is, the lower the instability or the higher the stability, and vice versa. The CDVI is computed as follows [19, 20]:

$$CV = \frac{Standard\ Deviation}{Mean} * 100 \quad (4)$$

$$CDVI = CV\sqrt{1 - R'^2} \quad (5)$$

where R'^2 is the adjusted coefficient of determination of the time trend regression adjusted for its degree of freedom. The CDVI values can be categorized into three instability groups as [21]

Low Instability	< 15
Medium Instability	15 - 30
High Instability	> 30

3) Cluster Analysis: Compared with spatial dynamics, cluster analysis make proper adjustments and utilizing soybean production technologies and resources by identifying priority areas for specific crop production [22]. This study analyzes the spatiotemporal expansion, growth and instability dynamics of soybean. Thus, K-means clustering technique is employed due to its wider adaptability, high processing efficiency, and suitability for large and continuous datasets [23]. It is an unsupervised partitioning classification technique uses euclidean distance as a linkage method and modifies the centroid constantly through multiple iterations for better clustering results [24, 25]. The improved clustering results require a confirmed k value a priori. The basic principle of clustering is to divide the n samples into k clusters and minimize the intra-cluster sum of squares to its centroid, which satisfies the clustering equation (Eq. 6), where x is the sample value, Y is the cluster set, yj is an individual cluster in the cluster set, and Cj is the cluster center in the cluster yj [26, 27].

$$arg_Y\, min \sum_{j=1}^{k} \sum_{x \in y_j} ||x - z_j||^2 \quad (6)$$

K-means clustering analysis was performed on the soybean growth and instability of APY data via Geoda 1.22 software. The default technique is KMeans++, with the initial re-run set to 150 and a maximum of 1,000 iterations. The optimum number of clusters was obtained by the 'Elbow' method of the "ClusterMap's K-Means Clustering" module of the QGIS 3.34.10 software. No data transformation (or normalization) was performed (i.e., 'Raw' data) due to univariate clustering analysis and unitless data (ratios and percentages) [28]. The clustering results were evaluated using the ratio of the total between-cluster sum of squares (variance) (BSS) to the total sum of squares (variance) (TSS) (Eq. 7). A higher value of the ratio signifies a higher degree of inter-cluster separation and intra-cluster compactness, suggesting better clustering results.

$$\frac{BSS}{TSS} = \frac{\sum_{j=1}^{k} n_j (\bar{c}_j - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad (7)$$

III. RESULTS AND DISCUSSIONS

At the outset, an attempt has been made to evaluate the trend, CAGRs, CDVI instability of the APY of soybean in India, MP state and districts of MP. Additionally, spatiotemporal unsupervised clustering of CAGR and supervised clustering of CDVI at the district level were performed to comprehend the dynamics from the country to the district level. This includes classifying the soybean districts to identify potential areas and explore future prospects for soybean expansion. The triennium-ending (TE) mean is considered for the study period to mitigate annual shocks.

A. *Soybean Scenarios in India*

Since 1970, the area and production of soybean in India have been continuously increasing. The area has expanded from 0.033 to 12.83 million hectares; production has grown from 0.018 to 13.68 million tons; and productivity has risen from 556 to 1066.33 kg per hectare from 1972TE to 2023TE, respectively. The increases in area, production, and productivity have been approximately 390.10, 744.49, and 0.92 times greater since 1970, respectively (Table 1). Productivity, however, has oscillated at approximately 1–1.3 tons per hectare over the last three decades.

Despite this low average yield compared with the global average yield (2.75 tons/ha; 2023-24 TE) [8], the country has the potential to achieve yield of over 3 tons per hectare by adopting advanced and improved technologies, as evidenced by frontline demonstrations (FLDs) conducted nationwide [29]. Numerous factors contribute to this yield gap, including the limited adoption of advanced technologies; inadequate management practices; abiotic, biotic and edaphic stresses; the untimely availability of inputs; and more [30, 11]. The maximum area (13.26 million hectares), production (14.98 million tons) and productivity (1352.9 kg/ha) were observed from 2023-24, 2022-23 and 2012-13, respectively (Fig. 2).

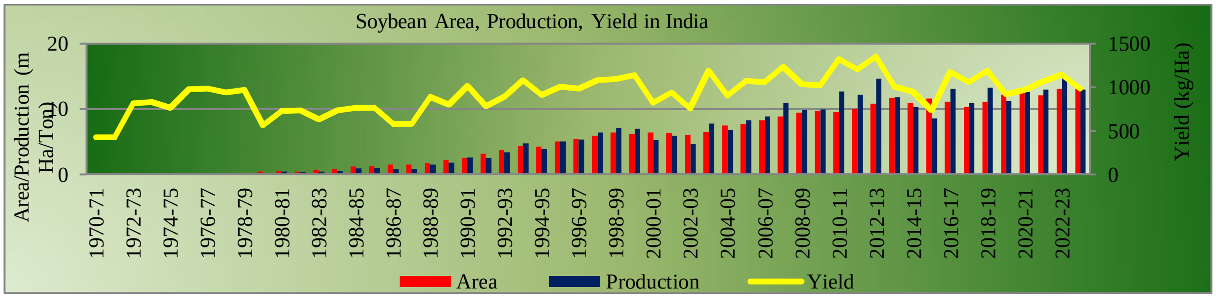

Fig. 2. Trends in soybean area, production and yield in India

The CAGR for area and production is positive and significant at the 1% level, except in the last decade, when production was positively insignificant. Initially, area expansion and production growth were relatively high and have slowly decreased, with the least growth occurring in the last decade. Additionally, yield growth follows a similar trend but remains insignificant, except in the second-to-last decade, where it increased significantly at the 10% level of significance and became negative in the last decade. The pooled growth rates of area (11.15%), production (12.27%), and yield (1.01%) are positively significant at the 1% level. This signifies that the time trend has significantly affected the per unit change in APY (Table 1).

The CDVI measure of instability is common and superior to other methods. Thus, we measured the stability of soybean in India over time via the CDVI (Fig. 2). The area was highly stable in all decades except in the first decade and pooled period, when it was medium and highly unstable, respectively. Production exhibits low instability in the first, fourth and fifth decades; medium instability in the second and third decades; and high instability in the pooled period. However, yield demonstrates medium instability in all decades except in the third and fourth decades, where it has low instability. The pooled instability revealed that the soybean area, production and productivity have not stabilized after six decades of cultivation (Table 1).

Table 1. Decadal growth and instability analysis of India

	Triennium Ending Decadal Mean						Times Rise since 1970
	Mean72TE	Mean80TE	Mean90TE	Mean2000TE	Mean10TE	Mean23TE	
Area(mha)	0.033	0.47	2.184	6.377	9.616	12.83	390
Production (mTons)	0.018	0.341	1.985	6.5	10.869	13.68	745
Yield(kg/ha)	556	756.914	902.711	1020.117	1130.555	1066.33	0.92
CAGR							
	Period-I (1970-80)	Period-II (1981-90)	Period-III (1991-00)	Period-IV (2001-10)	Period-V (2011-23)	Period-VI (1970-23)	
Area	38.29*	17.87*	8.08*	5.87*	1.84*	11.15*	
Production	44.56*	20.98*	9.85*	9.39*	1.11	12.27*	
Yield	4.52	2.64	1.64	3.32***	-0.72	1.01*	
CDVI Instability							
Area	20.99	12.45	6.66	4.33	5.42	34.09	
Production	13.19	19.60	16.95	12.72	14.18	35.79	
Yield	25.75	16.92	12.36	13.13	15.20	17.63	

B. Soybean scenarios in Madhya Pradesh

Soybean in MP has been grown since 1964–65, during the Green Revolution era (Fig. 3). However, APY began gaining momentum since 1980-81.

The area and production have been continuously increasing, with the highest area of 6.69 mha, in 2020-21 and a production of 8.41 mtons recorded in 2012-13. Despite this, production experienced mainly two significant setbacks in 2002-03 and 2020-21, the latter occurred during the COVID-19 outbreak. After 2012-13, a declining trend was observed until 2020-21, followed by a rebound in production. However, no major setbacks were observed in the area trend (Fig. 3). Uneven climatic conditions (especially monsoons and drought) during the pod-filling stage affect soybean in the central India more than they do in other regions. Furthermore, the peasant community requires short-duration varieties due to the early withdrawal pattern of the monsoon and takes advantage of other short-duration crops, such as potatoes, between soybean and wheat cropping. Therefore, they need short-duration varieties such as JS 95-60, which cover more than 95% of the soybean area. However, this variety is susceptible to the adverse effects of biotic and abiotic factors, which lead to lower yields [31].

Initially, in 1975, productivity increased to 1121 kg/ha but suddenly decreased from 1977-78 to 407 kg/ha. Since then, productivity has been increasing and reached its peak yield of 1362 kg/ha in 2012-13. From 1964-65 to 2022--23, the APY increased by approximately 1687, 4128 and 145 times, respectively. However, productivity over the last three decades has fluctuated around 1 ton per hectare. Recently, productivity has fallen to its lowest level since 1980-81, with 502 kg/ha in 2020-21 (Fig. 3). Poor management practices, specifically nutrient management for this oil- and protein-rich crop, lower adoption rates for advanced agricultural technologies by farmers in MP than in those in Maharashtra, and increasingly unfavorable climatic conditions experienced in MP over the past few years may lead to lower yields in the state [32].

The decadal CAGRs (Table 2) of area and production are positive and significant at the 1% level of significance for all periods except for periods PC (Pre-Commercialization) and V (last decade), where growth were negative and insignificant for both area and production. The highest growth rates were observed in Period I, followed by Period II. The CAGRs for the pooled period (Period VI) for area and production are 14% and 16%, respectively. Growth in the first period (Period PC) was negative for both area and production due to the introductory crop initially grew in current fallow lands during the rainy season [5], which took time to be adopted as a commercial crop. In the last decade (Period V), the crop peaked in 2012-13, followed by uneven weather conditions and the COVID-19 outbreak, leading to negative growth. The growth rates of yield are positive in all periods except for periods I and V, but the growth rates for only periods IV and VI are positively significant at the 1% level. The negative growth in periods I and V implies that production growth was lower than area growth was, as is evident from Table 2. The pooled period demonstrated significant positive growth rates for area, production, and yield.

The decadal instability analyses (Table 2) revealed high instability for area and production during periods PC, I and VI. However, the area is highly stable and production has moderate instability in all other periods except IV. The yield shows high instability only in period I, low instability in periods III and IV, and moderate instability in the remaining periods. Period IV is highly stable for APY. The pooled period revealed highly unstable areas and production and medium yield stability.

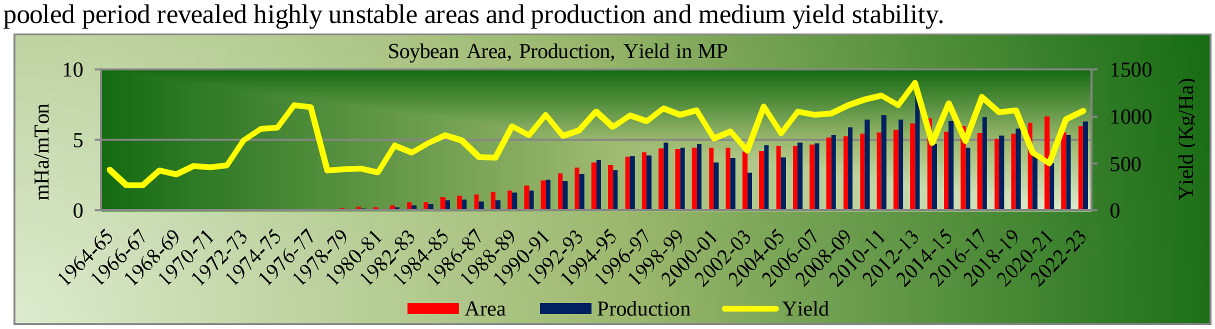

Fig. 3. Trends in the soybean area, production and yield in Madhya Pradesh

Table 2: Decadal growth and instability analysis of Madhya Pradesh

	Period-PC	Period I	Period II	Period III	Period IV	Period V	Period VI (Pooled)
CAGR							
Area	-8.233	47.080*	19.621*	6.038*	3.327*	-0.069	14.29*
Production	-2.495	37.798*	23.416*	7.126*	8.737*	-2.931	16.00*
Yield	6.253	-6.304	3.172	1.027	5.236*	-2.864	1.50*
Instability							
Area	32.19	35.04	12.46	6.24	3.83	8.53	34.09
Production	34.37	30.59	20.92	17.27	12.69	23.71	35.79
Yield	21.02	40.39	18.12	12.79	12.42	26.54	17.63

Note: 1) * - 1% level of significance. 2) Period PC (**P**re-**C**ommercialization): 1964-70, Period I: 1971-80, Period II: 1981-90, Period III: 1991-2000, Period IV: 2001-10, Period V: 2011-22, Period VI (Pooled CAGR): 1964-22

C. Soybean scenarios in districts of Madhya Pradesh

Growth Clustering. The variation in growth of APY is evident from Fig. 4 for both: Period A (2011-22) and Period B (Pooled Period: 1964-2022). k-means clustering of APY growth assess regional growth and development better.

Area growth clustering. The clustering trend of area growth for Period A shows concentrated and positive growth in entire Malwa Plateau and Jhabua Hills; partially in Nimar Valley (except Barwani and Burhanpur districts), Vindhiyan Plateau (only Bhopal, Sehore & Vidisha), and Gird Region (only Guna, Ashok Nagar & Shivpuri). These regions have CAGRs of 0.7% to 7.73%. However, other regions have negative growth (Fig. 4a). The Kymore P. had the lowest area growth (ranging from -44.2% to -36.48%). Seoni, Jabalpur, Mandla, Katni and Rewa are low-growth districts; among them, Rewa experienced the least growth (-44.2%) over the last decade. In Period B, area growth was positive and spread across the state. However, negative growth is found only in Balaghat, Katni, Bhind, Datia and Mandla. A high CAGR (14–18%) is observed in parts of the Malwa, Vindhiyan, Gird Region and Nimar Valley (Fig. 4d). During the last decade, only 18 (37.5%) districts had positive area growth, whereas 33 (62.5%) districts had negative area growth. Moreover, only 15 (~31%) districts have a CAGR greater than 1, and 33 (~69%) districts have a CAGR less than 1. Thus, about one-third districts experienced increasing area growth. The clustering in 5 optimal clusters using K-Means gives the best results and evaluations give 97.24% and 95.36% precision (with respect to total sum-of-squares percentage) for Periods A and B, respectively.

Production Growth Clustering. The production growth for Period A is negative in all districts, except Khargone (4.16), Ashoknagar (3.06), Indore (1.76), Rajgarh (1.04), Khandwa (0.75) and Vidisha (0.51). The higher production growth region is concentrated in the Malwa Plateau (except Neemuch and Mandsaur districts), Nimar Valley, Jhabua Hills, Vindhyan Plateau (except Raisen and Damoh), Northern Hills (only Dindori and Anuppur) and Harda (Central Narmada Valley), as well as Ashok Nagar district (Gird Region) (Fig. 4b). However, the lowest production zone is Kaymore Plateau, with Mandla district. The pooled period has positive production growth in all the zones (Fig. 4e); however, only five districts, namely, Balaghat (-10.58), Katni (-5.27), Bhind (-3.78), Mandla (-0.99) and Datia (-0.27), experienced negative growth. The high-production growth clusters belong to the Malwa Plateau (Mandsaur, Neemuch, Ratlam, Ujjain and Dhar), Vindhyan Plateau (Vidisha, Sagar and Damoh), Gird Region (Guna, Ashok Nagar and Sheopur) and Khandwa district of the Nimar Valley. Production growth clustering is similar to area growth. In the last decade, about 15% (7) of districts have shown positive growth, whereas 85% (41) districts have shown negative growth. Also, the production growth of only 8% (4) districts was more than 1, while 92% (44) districts have less than 1. Hence, only about 10–15% districts had increasing growth. Clustering gives five optimum clusters for both periods, with precisions 96.67% and 95.64% for Periods A and B, respectively.

Yield growth clustering. The yield growth variability for period A has an opposite trend compared to area or production. Yield growth is generally high and positive in regions with low and negative areas and production growth, such as Kaymore P. and Northern Hills. It is significantly positive only in Panna and Katni at 10% significance level, while it is significantly negative in Betul, Khandwa, Jhabua, Mandsaur, Guna, and Sidhi at 10% significance level; however, other districts show insignificant growth (Fig. 4c). Growth in period B is significantly positive in all regions but generally higher in western MP (Fig. 4f); however, it is negative in 4th cluster. The growth is positive in only 1/3rd (16) districts, while 2/3rd (32) districts have negative growth. However, 11 (~23%) districts have more than one yield growth and 37 (~77%) districts have less than one yield growth. The optimum four yield clusters (for both periods) were obtained with precisions of 92.27% and 92.08% for periods A and B, respectively.

Instability clustering. The supervised clustering of soybean (APY) instability provides insight into the dynamics and volatility in the regions. The higher the CDVI value is, the greater the instability and vice-versa (Fig. 5).

Area Instability. The clustering of area instability for Period A shows high stability in the Malwa Plateau, Sehore and Bhopal (Vindhiyan P.), and Chhindwara (Satpura P.). Medium stability is found in Shajapur (Malwa), Jhabua Hills, Khargone & Burhanpur (Nimar Valley), Harda (Central Narmada Valley), Vidisha and Sagar (Vindhiyan P.), Dindori & Anuppur (Northern Hills); and Guna, Ashok Nagar, Sheopur & Shivpuri (Gird Region). The rest districts have unstable area dynamics (Fig. 5a). The pooled period (Fig. 5d) shows unstable area across the state. This implies that the area in Period B was more volatile and uncertain than in Period A. In the last decade, only 11 (23%) districts had highly stable areas, 13 (27%) had medium stability and half (24%) of the districts had unstable areas.

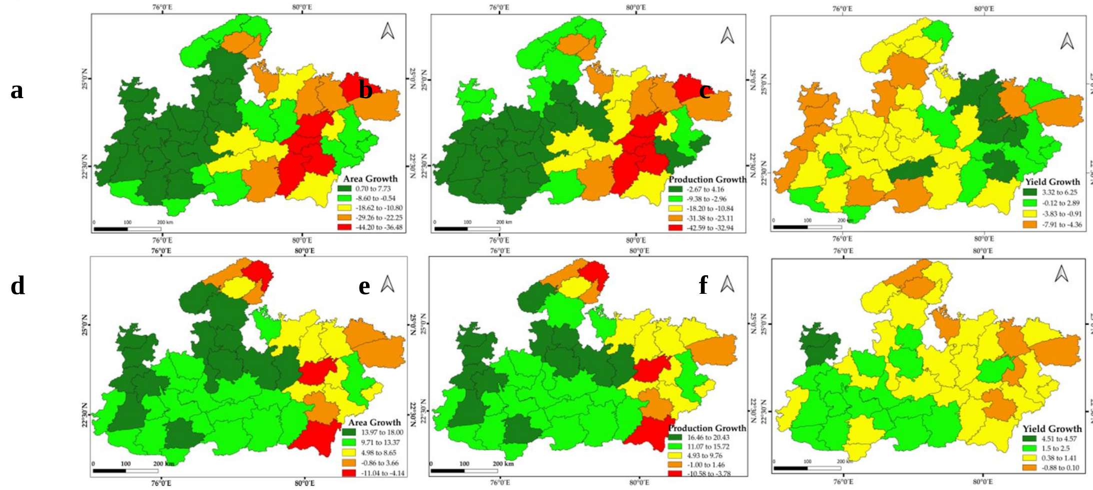

Fig. 4. Spatiotemporal clustering of soybean area, production, and yield growth rates in MP. Period A (2011--22): a-area, b-production, c-yield growth clusters; Period B (1964--2022) – pooled period: d-area, e-production, f-yield growth rate clusters.

Production Instability. The medium instability is found only in Indore, Dhar, Ratlam and Vidisha; however, other districts have unstable production during period A (Fig. 5b). Period B has unstable production across the state (similar to area) (Fig. 5e). In both periods, production was more volatile and uncertain than it was in the surrounding area. Period A have no districts with high stability for production; only 4 (~8%) districts have medium stability, whereas other 44 (~92%) districts have highly unstable production.

Yield Instability. Yield is not highly stable in any period. Medium instability is observed for period A in Sheopur, Morena, Gwalior, Datia, Vidisha, Damoh, Katni, Panna, Rewa, Balaghat, and Anuppur. Compared to western MP, the stability (medium) in the last decade has shifted more towards the eastern-north MP (Fig. 5c). However, for the pooled period, medium instability is more prominent in the east and west MP than in the northern MP, as reported by Dhar, Indore, Dewas, Shajapur, Panna, Rewa, Dindori, Balaghat, Datia and Sheopur districts. The remaining districts from both periods exhibited high instability for yield (Fig. 5f). In the last decade, no districts had high yield stability; only 13 (~27%) districts had medium stability, while 35 (~73%) districts have had highly unstable yields.

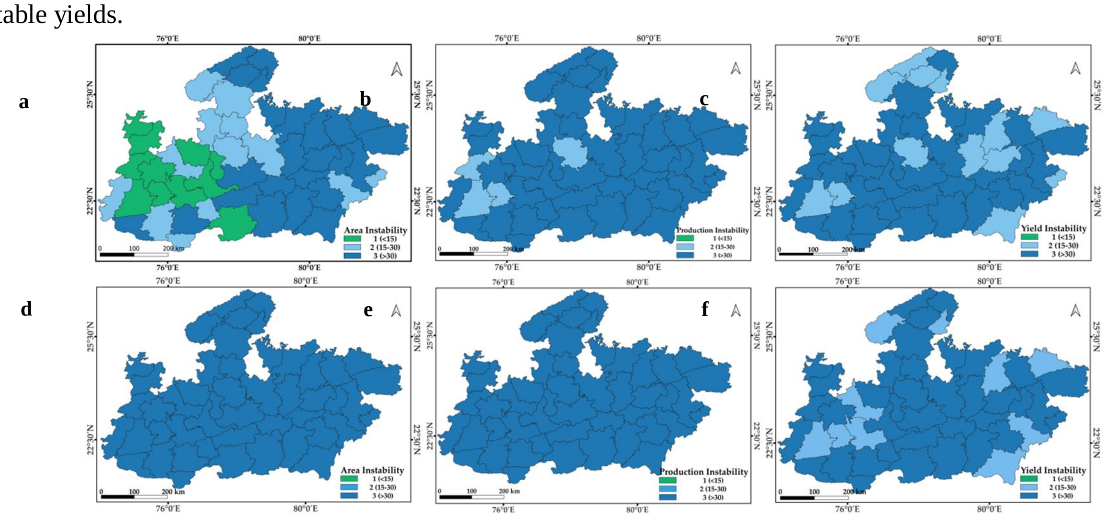

Fig. 5: Spatiotemporal clustering of the soybean area, production, and yield instability in MP. Period-A (2011-22): a-area, b-production, c-yield instability clusters; Period-B (~1964-2022) – Pooled Period: d-area, e-production, f-yield instability clusters.

D. Cluster Sensitivity Analysis

Unsupervised clustering of growth data categorizes the districts with similar growth patterns. The “Elbow” criteria of the “Scree Plot” [33] present 5, 5, and 4 clusters for A, P, Y growth, respectively, for both periods A and B (Fig. 6). The K-means algorithm uses the 'Euclidean' distance function.

The default parameter set (Table 3) of the K-Means algorithm provides the best results with the 'K-Means++' method compared with the 'Random' initialization method. The cluster evaluations of APY growth are 97.24%, 96.67% and 92.27%, respectively, and 95.36%, 95.64% and 92.08%, respectively, precision (with respect to total sum-of-squares percentage) for Periods A and B (Table 3).

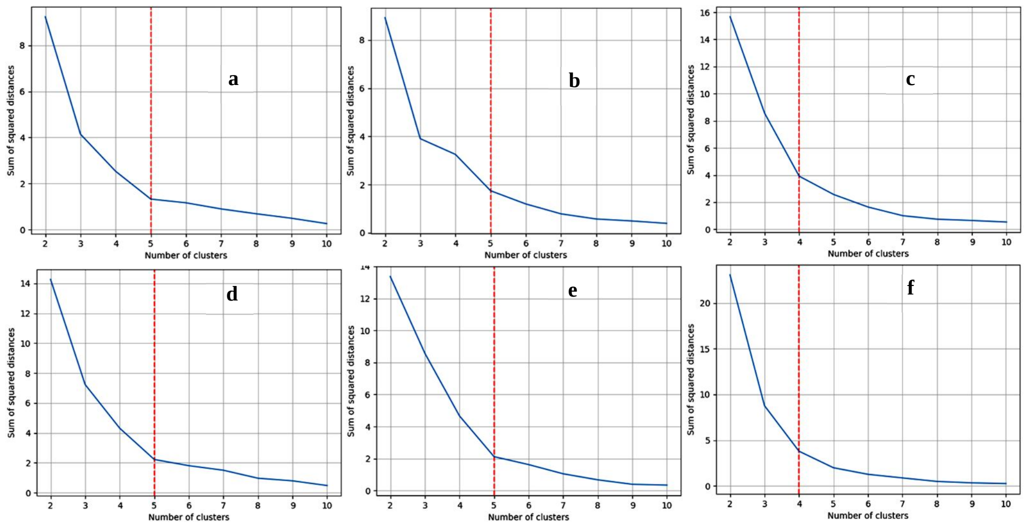

Fig. 6: Number of clusters identified by the 'Elbow' method for soybean APY growth (CAGR) in MP Period A (2011-22): a-area, b-production, and c-yield growth; Period B (~1964-2022) – pooled period: d-area, e-production, and f-yield growth.

Table 3. Cluster sensitivity analysis of soybean growth

Clustering Parameters	Area		Production		Yield	
Period – A: 2011-22						
Method:	KMeans	KMeans	KMeans	KMeans	KMeans	KMeans
Number of clusters:	5	5	5	5	4	4
Initialization method:	Random	KMeans++	Random	KMeans++	Random	KMeans++
Initialization re-runs:	150	150	150	150	150	150
Maximum iterations:	1000	1000	1000	1000	1000	1000
Transformation:	Raw	Raw	Raw	Raw	Raw	Raw
Precision Percentage (%)	97.24%	97.24%	96.65%	96.67%	92.17%	92.27%
Period – B (Pooled Period): 1964-22						
Method:	KMeans	KMeans	KMeans	KMeans	KMeans	KMeans
Number of clusters:	5	5	5	5	4	4
Initialization method:	Random	KMeans++	Random	KMeans++	Random	KMeans++
Initialization re-runs:	150	150	150	150	150	150
Maximum iterations:	1000	1000	1000	1000	1000	1000
Transformation:	Raw	Raw	Raw	Raw	Raw	Raw
Precision Percentage (%)	93.35%	95.36%	93.81%	95.64%	92.08%	92.08%

IV. CONCLUSION AND POLICY OPTIONS

This study explored soybean growth and dynamics in India from 1964-2023. Nationally, soybean area and production have increased at CAGRs of 11.15% and 12.27%, respectively, since 1970. However, yields stagnated around 1 - 1.3 tons per hectare. Decadal growth is positive in all periods except for yield in Period V; with initially greater growth, it has gradually declined to its lowest value in the last decade. The APY became highly stable over the last two decades, with the area becoming stable faster than production and productivity. However, in the pooled period, area and production are highly unstable, whereas yield has medium stability due to greater fluctuations.

In Madhya Pradesh, soybean area, production and productivity are increasing at rates of 14.29%, 16%, and 1.5% CAGR, respectively, outperforming national growth; however, productivity has plateaued at approximately 1 ton per hectare. Except for the PC and last decade periods, growth in area and production was significantly positive but steadily decreasing, turning negative in the last decade, whereas yield growth was negative only during the first and last decade periods. The stability of all three variables gradually increased and became highly stable in period IV but rose again. The area is becoming stable more rapidly than production and yield. The pooled (VI) period shows a highly unstable area and production but medium stability for yield.

Unsupervised clustering of Madhya Pradesh districts identified optimal clusters for area, production, and yield growth using the K-Means algorithm with high precision. The spatiotemporal clustered map highlighted the Malwa Plateau, Nimar Valley, Jhabua Hills, Vindhiyan Plateau, Northern Hills, Gird Region, and Harda and Betul districts as potential high-growth regions and the Kymore Plateau as a low-growth region. The yield growth trend reversed, with higher yields in western MP during the pooled period than during the last decade. Supervised clustering of instability indicated that the soybean area is more stable in western MP during period A, whereas production generally has low stability throughout the state. However, during period B, both area and production were unstable and volatile, with no clear pattern of yield stability in either period.

These findings, along with an understanding of the spatiotemporal growth and dynamics of soybean, lead to the formulation of cluster-specific policies for rapidly adopting new technologies and extension activities in low-contributing regions with the aim of improving productivity, ensuring food security, and providing sustainable solutions for reducing yield gaps, such as developing climate-resilient technologies, enhancing extension services, and expanding soybean cultivation in other seasons to boost productivity.

REFERENCES

[1] Ayele A, Worku M, Bekele Y. Trend, instability and decomposition analysis of coffee production in Ethiopia (1993-2019). Heliyon. 2021; 20:7(9):e08022. doi:10.1016/j.heliyon.2021.e08022. PMID:34589632; PMCID: PMC8461354.

[2] Chattopadhyay M, Mitra SK. Comparative decision models for anticipating shortage of food grain production in India. Theoretical and Applied Climatology. 2018; 131: 523–530. https://doi.org/10.1007/s00704-016-1961-0

[3] Patel RM, Sharma AN, Sharma P. GIS based pest-weather model to predict the incidence of Girdle beetle (Oberiopsis brevis) in Soybean crop. Journal of Agrometeorology. 2021; 23(2): 183-188.

[4] Sharma P, Patel RM, Dupare BU. Profitability of Soybean vis-a-vis Major Kharif Crops in Madhya Pradesh. Soybean Research. 2015; 13(2): 65-77.

[5] Patel RM, Sharma P, Bunkar K, Ramesh A, Khandekar N. Decadal cropping pattern and expansion dynamics of Soybean in Madhya Pradesh, India. Soybean Research. 2023; 21(1): 110-123 (2023).

[6] Patel RM, Sharma P, Sharma AN. Weather Based Prediction Models for Forewarning Tobacco Caterpillar, Spodoptera Litura (Fabricius) Larval Population in soybean. Journal of the Indian Society of Agril. Statistics. 2020; 74(2): 129–1

[7] Sharma P, Dupare BU, Patel RM. Soybean Supply Chain Management in India: A Review. Soy. Res.. 2023; 21(2):1-16

[8] FAO-AMIS. https://app.amis-outlook.org/#/market-database/custom-query

[9] Nataraj V, Kumar S, Kumawat G, Shivakumar M, Rajput LS, Ratnaparkhe MB, Ramteke R, Gupta S, Satpute GK, Rajesh V, Kamble V, Chandra S. Charcoal Rot Resistance in Soybean: Current Understanding and Future Perspectives. In: Disease Resistance in Crop Plants. 2019; [Wani S. (eds)] Springer, Cham, pp: 241-259.

[10] Tiwari A, Choudhary VK, Shrey R. Analysis on growth rate and decomposition analysis of major cereals grown in kharif season in northern hills of Chhattisgarh. International Journal of Research in Agronomy. 2024; 7(5S):187-193. DOI:10.33545/2618060X.2024.v7.i5Sc.767

[11] Mall RK, Sharma NK, Singh N. Impacts of Climate Change on Agriculture Sector in Madhya Pradesh. 2016. https://doi.org/10.13140/RG.2.1.3010.0247

[12] Dam A, Chatterjee S, Sarkar D, Kumar P. Trends in Production of Pulses and Oilseeds Over Last Two Decades in Eastern India. Economic Affairs. 2023; **68**(04): 2017-2023.

[13] Mishra S, Balai HK, Malviya PK. Rabi Pulse Crops Cultivation in Madhya Pradesh (India): Growth and Decomposition Analysis. Economic Affairs. 2023; **68**(01): 391-395.

[14] Rana MJ, Islam S, Kamruzzaman M. Growth and instability in area, production and productivity of major spices in Bangladesh. Journal of Agriculture and Food Research. 2021; 6: 100216

[15] Wang Z, Zhang E, Chen G. Spatiotemporal Variation and Influencing Factors of Grain Yield in Major Grain-Producing Counties: A Comparative Study of Two Provinces from China. Land. 2023; 12:1810. https://doi.org/10.3390/land1209181

[16] Coppock JD. International Economic Instability. McGraw-Hill, New York. 1962; 523–525

[17] Joshi P, Gautam P, Wagle P. Growth and instability analysis of major crops in Nepal. Journal of Agriculture and Food Research. 2021; 6: 10023

[18] Cuddy JDA, Della-Valle PA. Measuring the Instability of Time Series Data. Oxford Bulletin of Economics and Statistics. 1978; 40(1): 79–85

[19] Rani R, Sah U, Kumar H, Bishnoi R, Udgata AR. Multilevel growth assessment of mungbean production in Madhya Pradesh. Journal of Food Legumes. 2024; 37(2): 190-195. DOI: 10.59797/jfl.v37.i2.194

[20] Ikuemonisan ES, Mafimisebi TE, Ajibefun I, Adenegan K. Cassava production in Nigeria: trends, instability and decomposition analysis (1970–2018). Heliyon. 2020; 6: e05089

[21] Dhungana SM, Regmi PP, Dhakal SC, Devkota N R. Growth instability index, and decomposition effect on production of large cardamom in Nepal. Agriculture & Food Security. 2024; 13:5. https://doi.org/10.1186/s40066-023-00456-y

[22] Zhang Z, Lu C. Clustering Analysis of Soybean Production to Understand its Spatiotemporal Dynamics in the North China Plain. Sustainability. 2020; 12: 6178. doi:10.3390/su1215617

[23] Sinaga KP, Yang MS. Unsupervised K-Means Clustering Algorithm. IEEE Access. 2020; 8: 80716-80727. doi:10.1109/ACCESS.2020.2988796

[24] Li D, Liang J, Wang X, Wu S, Xie X, Lu J. Rice Yield Simulation and Planting Suitability Environment Pattern Recognition at a Fine Scale. ISPRS International Journal of Geo-Information. 2021; 10: 612. https://doi.org/10.3390/ijgi10090612

[25] Mamun MAAl, Nihad SAI, Sarkar MAR, Aziz MA, Qayum MA, Ahmed R, Rahman NMF, Hossain MI, Kabir MS. Growth and trend analysis of area, production and yield of rice: A scenario of rice security in Bangladesh. PLoS ONE. 2021; 16(12): e0261128. https://doi.org/10.1371/journal.pone.026112

[26] Ye S, Song C, Shen S, Gao P, Cheng C, Cheng F, Wan C, Zhu D. Spatial pattern of arable land-use intensity in China. Land Use Policy. 2020; 99: 104845. doi:10.1016/j.landusepol.2020.10484

[27] Abirami B, Selvi RG, Kannan B, Vanitha G, Thavaprakaash N. Application of K-means Clustering Algorithm in Rice Production of Tamil Nadu, India. International Journal of Environment and Climate Change. 2022; 12(11): 1348-1355.

[28] Ji D, Tian J, Zhang J, Zeng J, Namaiti A. Identification and Spatiotemporal Evolution Analysis of the Urban–Rural Fringe in Polycentric Cities Based on K-Means Clustering and Multi-Source Data: A Case Study of Chengdu City. Land. 2024; 13: 1727. https://doi.org/10.3390/land13111727

[29] Dupare BU, Sharma P, Billore SD, Verma SK. Impact of Climate Change on Soybean Cultivation in Malwa and Nimar Region of Madhya Pradesh: Farmers' Perspective. Soybean Research. 2020; 18: 87–97

[30] Prashnani M, Dupare B, Vadrevu KP, Justice C. Toward food security: Exploring the spatiotemporal dynamics of soybean in India. PLoS ONE. 2024; 19(5):e0292005. https://doi.org/10.1371/journal.pone.0292005

[31] Shrivastava A, Pandey SK, Kobayashi S. Influence of Excessive Moisture Stress on Genetic Parameters in Soybean. Soybean Research. 2014; (Special Issue - II): 33–43

[32] Ellanki RE, Reddy SB. Centre for Economic and Social Studies H. Economics and technology of soybean cultivation in Central India. 2016; Volume 43 of CESS monograph

[33] Mamun MAAl, Nihad SAI, Sarker MR, Sarkar MAR, Hossain MI, Kabir MS. Spatiotemporal mapping of rice acreage and productivity growth in Bangladesh. PLoS ONE. 2024; 19(3): e0300648. https://doi.org/10.1371/journal.pone.0300648