PARTITIONING OF FERTILIZATION MANAGEMENT ZONES FOR MAIZE BASED ON SOM-FCM

CHEN, H. – YI, S. J.* – WANG, X. – LI, W.

College of Engineering, Heilongjiang Bayi Agricultural University
Daqing 163319, China
(phone: +86-4596-819-216)

**Corresponding author*
e-mail: yishujuan_2005@126.com; phone: +86-199-6550-2030

(Received 25th Feb 2025; accepted 22nd Apr 2025)

Abstract. At present, many clustering algorithms have been applied to the calculation of farmland fertilization management zoning, among which the fuzzy c-means (FCM) algorithm is the most widely used, but the FCM algorithm is sensitive to the initial clustering center, which affects the calculation efficiency. In order to improve the efficiency of clustering calculation, especially for the clustering calculation of large datasets such as normalized difference vegetation index (NDVI), an algorithm combining the self-organizing feature map neural network (SOM) and FCM is proposed. The algorithm first uses SOM for preliminary clustering to determine the cluster center, then applies it as the initial cluster center of the FCM, and finally uses the FCM to further cluster the data. This method effectively combines the advantages of SOM and the FCM algorithm, solves the problem of setting the initial clustering center in the FCM, and overcomes their respective limitations. The NDVI data of maize was obtained through the ground remote sensing detection system. The FCM algorithm and the SOM-FCM algorithm were used to partition the fertilization management zones of maize. The sum of squared errors (SSE) and silhouette coefficient (SC) were used to evaluate the effectiveness of partitioning the fertilization management zones. The experimental results show that the SOM-FCM algorithm is better than the FCM algorithm for different NDVI data volumes. The maximum SSE difference is 639, and the maximum difference of SC value is 0.048. As NDVI data volume continues to grow of NDVI data volume, SOM-FCM algorithm shows better performance than FCM algorithm.

Keywords: *fuzzy c-means, Greenseeker, remote sensing, SSE, Silhouette coefficient*

Introduction

With the rapid development of precision agriculture technology, the scientific and practical division of farmland management has become one of the important research topics in agricultural production (Barman et al., 2021). As one of the important food crops in the world, the fertilization management of maize directly affects yield, quality, and environmental sustainability. However, due to the spatial heterogeneity of soil properties, topographic characteristics and climatic conditions, traditional fertilization methods often find it challenging to meet the specific needs of different regions, quickly leading to fertilizer waste or environmental pollution (Ameer et al., 2022; Ohana-Levi et al., 2021). Therefore, studying the zoning of precise fertilization management based on spatial variability is particularly important (Pawase et al., 2023). Fertilization management partitioning is based on information sources such as soil nutrient content, crop yield, and canopy spectral characteristics. K-means clustering, fuzzy c-means (FCM) clustering, hierarchical clustering, and other algorithms are used to calculate partitions. Among these algorithms, the FCM algorithm, with its unique ability to support fuzzy classification, is the preferred choice for farmland management zoning, underscoring its importance in this field. In order to optimize farmland management,

DOI: http://dx.doi.org/10.15666/aeer/2303_58415857

Moharana collected 122 soil samples, analyzed the nutrient content and its spatial variability, and partitioned the study area into four management areas using geostatistics and the FCM clustering algorithm (Moharana et al., 2020). The results showed significant differences in soil nutrients in each management area. Fertilization suggestions based on the management area could significantly reduce the amount of fertilizer, reduce environmental impact and improve crop yield. Vendruculo uses the FCM algorithm to predict the optimal number of partitions for management tasks in precision agriculture applications based on geo-referenced yield and grain moisture data sets (Vendrusculo and Kaleita, 2011). Through the long-term study of the data from the biological station in Michigan, the optimal number of partitions obtained by the FCM algorithm is 8 to 10, which provides important support for farmland management. Termin built a spatiotemporal dynamic clustering model based on the FCM algorithm to generate nitrogen management areas in the growing season of citrus orchards to optimize crop production and reduce environmental pollution (Termin et al., 2023). Five variables were selected to describe the spatiotemporal variability of canopy nitrogen content in four citrus plots. The dynamic clustering model was used to analyze the relevant variables, and the management areas were defined by comparing the spatial correlation of the clustering map, which provided support for precise nitrogen management in specific places and times.

Although the FCM algorithm performs well in dealing with data uncertainty and fuzziness, its clustering effect is often affected by the selection of initial clustering centers and parameter settings, which may lead to the instability of clustering results and local optimal solutions (Wang et al., 2017; Ming-Chuan and Don-Lin, 2001). In order to solve the above problems, this paper proposes a SOM-FCM algorithm combining self-organizing map (SOM) and FCM clustering, which can realize the fine division of maize fertilization management zoning. The SOM algorithm can reduce the data dimension, maintain the topological structure of data, and provide a more stable and effective initial clustering center for the FCM algorithm to optimize the clustering process and improve the accuracy and stability of clustering results. Combining the advantages of these two algorithms can more accurately partition the maize fertilization management zones and provide a scientific basis for precision agriculture management.

Materials and methods

Experimental site

The experimental site is located at 17 operation station in the fourth management area of Zhaoguang farm, Heilongjiang Province, China, as shown in *Figure 1*. Zhaoguang farm is located east of Daxing'an Mountains and north of Xiaoxing'an Mountains in Heilongjiang Province. The area has undulating mountains with an altitude of 240-330 meters. The climate type of the region is a cold temperate continental monsoon climate, which is characterized by wind and dry in spring, hot and humid in summer, rapid cooling in autumn and lasting and severe cold in winter. The primary soil type for planting crops is black soil, and its natural conditions are conducive to collecting spectral information on crop canopy. The plot area for maize NDVI data collection is 41.2 ha, and the collection time is June 18, 2019. During the collection of NDVI data, the maize plants were at the 6-leaf stage. Based on the BBCH growth scale for maize, this corresponds to the leaf development stage, identified by the BBCH code 16.

The variety of maize is Demeiya3, which takes about 110 days from emergence to maturity. The planting method adopts large ridges and double rows, with a ridge width of 1.1 meters and a planting density of about 67,500 plants per hectare. The planting date is May 7, 2019. The fertilization plan includes two stages: base fertilizer and topdressing. The base fertilizer mainly uses high nitrogen compound fertilizer, while combining organic fertilizer to improve soil structure; Topdressing is divided into two key stages. During the jointing stage (6-9 leaf stage), urea is applied at a rate of about 180 kg/ha to promote stem development, while during the trumpet stage (10-12 leaf stage), urea is applied at a rate of about 240 kg/ha to meet the nutritional requirements for panicle differentiation.

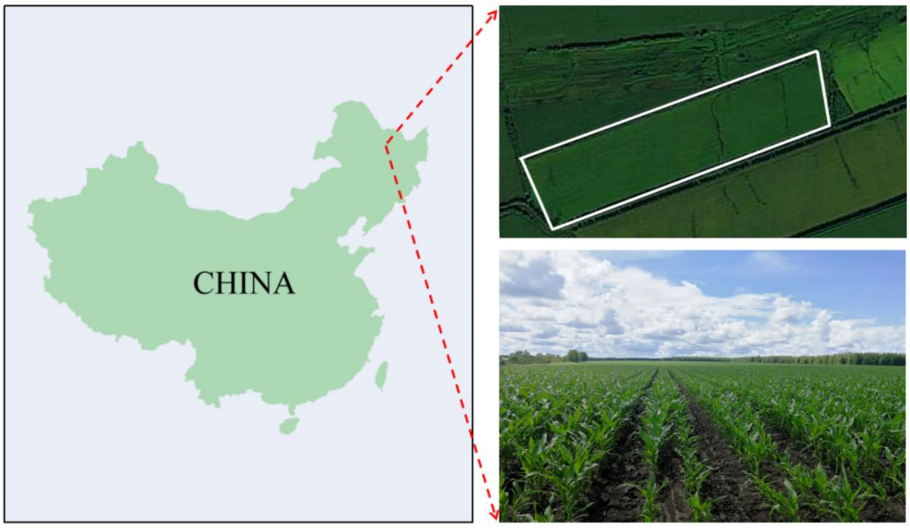


Figure 1. *Experimental site (Zhaoguang Farm, China)*

Data collection

The NDVI data acquisition of maize is based on the multi-component integrated ground remote sensing detection system. The core of the system is composed of six Trimble Greenseeker plant canopy spectral detectors, which are equipped with dual channel narrow-band led active light source modules and can emit 650 nm (FWHM ± 10 nm) red light band and 770 nm (FWHM ± 10 nm) near-infrared band active light source (Karatas and Karademir, 2023; Zsebő et al., 2024). The spatial positioning relies on the AG332 GNSS positioning system (horizontal accuracy ± 2.5 cm), the NDVI recorder stores the collected data, and the real-time visualization processing is realized through the 10-inch industrial touch screen. The full-time operation ability and real-time data output characteristics of Greenseeker provide continuous quantitative indicators for dynamic crop growth monitoring, and then support the precise variable rate fertilization decision based on spatial variability analysis (Karatas and Karademir, 2023). The working principle of Greenseeker sensor is shown in *Figure 2*.

In the experiment, the ground remote sensing detection system was installed on the case 2254 tractor, and the NDVI data of maize was collected during the middle tillage period. The NDVI data was used to partition the fertilization management zones, and the targeted fertilization work was carried out according to the zoning results, as shown in *Figure 3*.

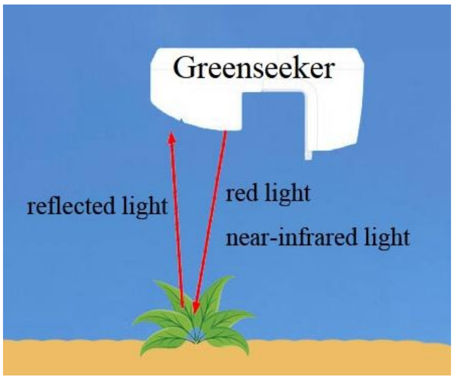


Figure 2. *Schematic diagram of working principle of Greenseeker*

Figure 3. *Ground-based remote sensing detection system for the acquisition of maize canopy NDVI data*

The variable fertilization system installed behind the tractor uses the tractor’s built-in hydraulic system as the driving source, which drives the hydraulic pump to move the fertilization shaft, thereby controlling the amount of fertilizer applied. Adjust the fertilizer application amount according to the divided fertilization zones. The installation position of the hydraulic-driven variable fertilization control system is shown in *Figure 4*.

The management partition program can analyze and process the NDVI data collected by Greenseeker in real-time and calculate the management partition of farmland. To ensure the integrity and traceability of the collected data, NDVI and other data are synchronously recorded in the can data recorder during the collection process. The data collected in the experiment includes longitude, latitude, and NDVI values. The data is collected once per second. The high-frequency data acquisition method can capture the subtle changes in different areas of the crop canopy and provide data support for the subsequent research on the prediction of canopy spectral information. The tractor runs at a speed of about 5 km/h, which is equivalent to about 1.39 m/s. Due to the data collection frequency of once per second, the geographical interval between the data points collected each time is approximately 1.39 meters. This high-frequency data

DOI: http://dx.doi.org/10.15666/aeer/2303_58415857

collection ensures consistency between temporal and spatial resolution, enabling experimental results to reflect the spatial distribution characteristics of NDVI values accurately. The stable speed of the tractor provides uniformity and continuity of data collection, thereby avoiding the problem of uneven distribution of data points caused by speed fluctuations. Some of the data collected in the experiment are shown in *Table 1*.

Figure 4. *Layout of hydraulic drive variable fertilization control system*

Table 1. *Details of the acquired data (partial)*

Date	Time	Longitude	Latitude	Mean NDVI
2019-6-18	14:38:39	126.6272425	48.0360503	0.462
2019-6-18	14:38:40	126.6272369	48.0360717	0.452
2019-6-18	14:38:41	126.6272405	48.0360869	0.453
2019-6-18	14:38:42	126.6272425	48.0360503	0.418
		…		
2019-6-18	18:18:58	126.6267755	48.0368535	0.417
2019-6-18	18:18:59	126.6267587	48.0368497	0.414
2019-6-18	18:19:00	126.6267470	48.0368446	0.401
2019-6-18	18:19:01	126.6267361	48.0368393	0.389

Delimitation of fertilization management zones

FCM algorithm

Ruspini first proposed the concept of fuzzy partition based on fuzzy set theory in 1972, which laid an important theoretical foundation for the subsequent research and application of the fuzzy clustering algorithm (Ruspini, 1972). On this basis, Dunn extended the traditional hard c-means (HCM) clustering algorithm to the field of fuzzy partition in 1973 and proposed a preliminary FCM clustering framework, which introduced the concept of membership to describe the degree of belonging of data

DOI: http://dx.doi.org/10.15666/aeer/2303_58415857

points to different categories (Dunn, 1973). Subsequently, Bezdek Further developed Dunn's work and systematically proposed the FCM algorithm by extending the objective function to a more general form, and described the algorithm in detail in their research in 1984 (Bezdek et al., 1984). This improvement enhances the algorithm's flexibility and significantly improves its adaptability in practical applications. As a classical soft partition method, the FCM algorithm is important in fuzzy clustering. By optimizing an objective function based on membership, the algorithm can calculate the membership value of all clustering centers for each sample in the data set to realize the flexible data division. Its core advantage is that it can effectively deal with the uncertainty and overlap problems in the data set, so it has been widely used in pattern recognition, data mining and image segmentation (Nayak et al., 2015; Kalyani et al., 2018). For the data set $X = [x_1,...,x_n]$, the objective function of the FCM algorithm is given as *Equation 1*.

$$\mathrm{J}_{FCM}(U,V) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} \left\| x_j - v_i \right\|^2 \tag{Eq.1}$$

where U is the membership matrix, V is the cluster center matrix, c the number of cluster centers, n is the number of data, m is the fuzzy weight index, u_{ij} is the membership ($0 \le u_{ij} \le 1$; i = 1,2,...,c; j = 1,2,...,n), $\left\| x_j - v_i \right\|^2$ is the Euclidean distance from the j-th data to the i-th cluster center v_i.

The constraint conditions for the objective function of the FCM algorithm to reach the minimum value are *Equations 2* and *3*.

$$v_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} x_j}{\sum_{j=1}^{n} u_{ij}^{m}} \tag{Eq.2}$$

$$u_{ij} = \left(\sum_{k=1}^{c} \left(\frac{\left\| x_j - v_i \right\|}{\left\| x_j - v_k \right\|} \right)^{2/(m-1)} \right)^{-1} \tag{Eq.3}$$

FCM algorithm achieves data clustering by iteratively optimizing the minimum value of the objective function (*Eq. 1*). The implementation process can be summarized as follows: firstly, initialize the number of cluster centers C, the stop threshold value Ɛ and the fuzzy weight coefficient m, and randomly generate the initial cluster centers; Then, the fuzzy membership matrix and cluster center matrix are dynamically updated based on Euclidean distance, and the change of cluster center is continuously detected during the iteration process. If the change is lower than the preset threshold, the calculation is terminated, and the membership matrix and cluster center matrix representing the fuzzy division of the sample are output. Otherwise, the matrix update is repeated until the convergence conditions are met. The algorithm's output reflects the soft classification attribution of samples through fuzzy membership and provides a precise spatial location of cluster centers.

Determination of the number of clusters c using the elbow method

Mean distortions is an index used to evaluate the quality of clustering. It refers to the average value of the sum of the distances between all data points and the cluster centers in its group. The average distortion reflects the tightness of the data in a cluster. The smaller the average distortion, the tighter the data structure in the cluster, the greater the average distortion, and the looser the data structure in the cluster. When clustering a data set, the average distortion degree will decrease as clusters increase. When the number of clusters reaches a critical point, the average distortion degree will significantly improve and decline slowly. At this point, the critical point is the point with better clustering performance. The elbow method is based on the above principle to calculate the average distortion degree and determine the optimal number of clusters in the partition. The specific calculation formula of the average distortion degree is given as *Equation 4*.

$$MD = \sum_{i}^{c} \frac{\sum_{p \in Ci} \left\| \mathrm{p} - \mathrm{m_i} \right\|}{n} \tag{Eq.4}$$

where *MD* is mean distortion, *c* is the number of clusters, *Ci* is the dataset of the i-th cluster, P is a specific data point within the i-th cluster, m_i is the cluster center of *Ci*, *n* is the data quantity in the i-th cluster, $\left\| p - m_i \right\|$ is the Euclidean distance from a data of the i-th cluster to the cluster center.

SOM-FCM algorithm

Self Organizing Map (SOM) is a neural network model based on unsupervised learning proposed by Kohonen. Its core idea is to simulate the function of self-organizing feature mapping in the brain's nervous system, thereby achieving automatic classification of input data patterns and topological structure preservation (Kohonen, 1990, 2013; Kohonen et al., 1996). Through in-depth research on SOM learning rules, it can be found that the model exhibits excellent automatic classification ability under unsupervised conditions and can gradually reveal the potential structure and distribution characteristics within the data through the self-organizing learning process of input data (Miljković, 2017). This feature gives SOM broad application prospects in pattern recognition, visualization, and high-dimensional data analysis. SOM continuously adjusts the connection weight coefficients between neurons through a competitive learning mechanism so that these weights can gradually reflect the similar relationship between input samples during the training process and map complex patterns in high-dimensional input space to low-dimensional output layers while maintaining consistency in topological proximity relationships. This mapping achieves dimensionality reduction of data and visually displays the spatial distribution characteristics of classification results (Cottrell et al., 2018). The typical topology structure of the SOM neural network is shown in *Figure 5,* where the dynamic updating process of connection weights between the input layer and the output layer constitutes the core computing mechanism of the model, providing important theoretical support and practical tools for understanding the inherent laws of complex datasets. This study utilizes longitude, latitude, and NDVI data as inputs to the SOM neural network, with the outputs serving as the initial clustering centers. The collaborative effect of three input data enables the joint representation of spatial coordinates and ecological features. Although NDVI is the main feature variable, longitude and latitude as spatial constraints are crucial for maintaining the topological structure of geographic distribution.

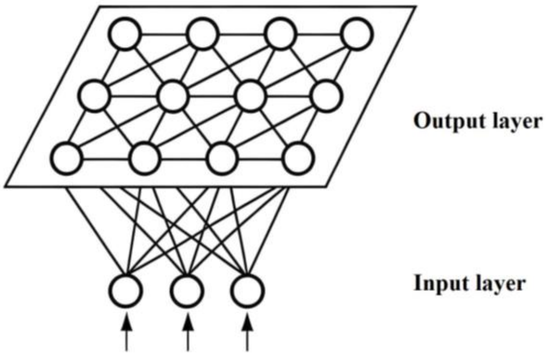


Figure 5. *Schematic diagram of typical SOM neural network topology structure*

The process of providing initial clustering centers for FCM algorithm using SOM algorithm includes the following steps: First, data preprocessing is performed to standardize the raw data to eliminate dimensional differences between different features, thereby ensuring that the input data is suitable for SOM algorithm training; Next, perform SOM network initialization to construct a two-dimensional SOM grid, where each node corresponds to a weight vector, and randomly initialize or initialize the weight vector of each node through heuristic methods; Then enter the SOM training process, input the input data one by one into the SOM network. For each input data point, calculate its distance from the weight vectors of all nodes (usually using Euclidean distance), find the node closest to it, that is, the best matching unit (BMU), and adjust the weight vectors of BMU and its neighboring nodes to make them closer to the input data points. The adjustment formula is given as *Equation 5*.

$$w_i(t+1) = w_i + \eta(t) \cdot h_{BMU,i}(t) \cdot (x - w_i(t)) \tag{Eq.5}$$

where w_i is the weight vector of the i-th node in time, $\eta(t)$is the learning rate, $h_{BMU,i}(t)$ is the neighborhood function used to control the influence range of BMU and its neighboring nodes, x is the current input data point.

After SOM training is completed, the weight vector of each node can be regarded as a representative description of the input data. Next, the initial cluster centers are extracted based on the topological structure of the SOM grid. The specific method includes uniformly selecting c nodes (c is the target cluster number of the FCM algorithm) directly from the SOM grid and using their weight vectors as the initial cluster centers. Alternatively, density-based methods can be used to select weight vectors based on the activation frequency of a node (i.e. the number of input data points mapped to that node), with nodes with higher activation frequencies being more suitable as initial clustering centers; After obtaining c cluster centers, use them as the initial cluster centers for FCM. Finally, the extracted initial cluster centers will be passed to the FCM algorithm as its initial conditions, and the FCM algorithm will iteratively optimize based on this to obtain more accurate and stable clustering results. The results output by the SOM-FCM algorithm are assigned partition numbers for each collection point, ultimately dividing the entire plot into four different growth areas. This allows the variable fertilization control system installed at the back of the tractor to adjust the fertilization amount according to the growth of the different regions. At present, this study focuses on the division of management zones, with specific fertilizer application rates fluctuating up and down (5% -10%) based on the benchmark fertilizer application rate. While ensuring the total fertilizer application rate, increasing the fertilizer application rate in areas with poor growth and reducing the fertilizer application rate in areas with good growth can improve fertilizer utilization efficiency. The flowchart of using the SOM-FCM algorithm for NDVI data clustering is shown in *Figure 6*.

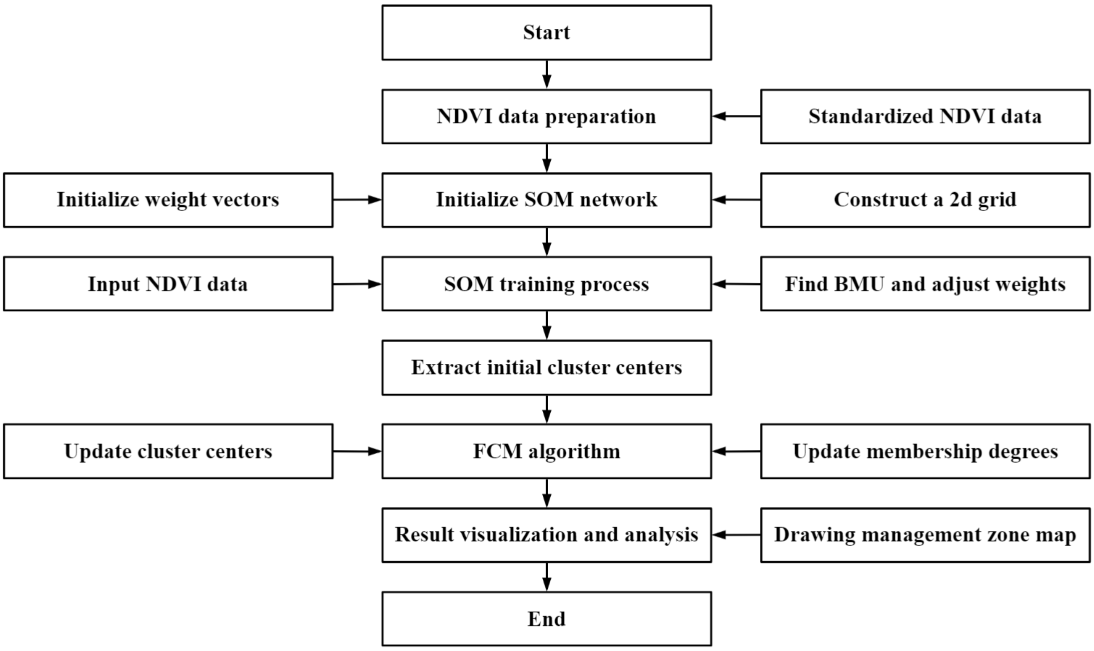


Figure 6. *Flowchart of SOM-FCM algorithm*

Evaluation of clustering performance

The sum of squares of errors (SSE) is a commonly used clustering quality evaluation metric that measures the degree of closeness between data points and their cluster centers (Hassan et al., 2021). Its core idea is to reflect the compactness of clustering results by calculating the sum of squared distances from each data point to its cluster center (Gul and Rehman, 2023). The smaller the SSE value, the closer the data point is to the center of its cluster and the better the clustering effect. Due to the lack of actual labels as references, SSE has become an important internal evaluation criterion that can directly evaluate the quality of clustering results based on data distribution characteristics in unsupervised learning. The FCM algorithm aims to optimize data partitioning to minimize SSE, thereby achieving tight clustering of data points within clusters and clear differentiation between clusters. However, SSE also has limitations, such as being sensitive to the selection of cluster numbers and more suitable for evaluating spherical clusters, which may not be accurate enough for clusters with complex shapes. Therefore, in practical applications, other evaluation methods, such as the contour coefficient, are usually combined to comprehensively judge the clustering effect and select the optimal number of clusters. The calculation formula for SSE is given as *Equation 6*.

$$\mathrm{SSE} = \sum_{i=1}^{c} \sum_{p \in C_i} \left| p - m_i \right|^2 \tag{Eq.6}$$

where the SSE is the sum of squared errors, c is the number of partitions, C_i is the dataset of the i-th partition, p represents certain data in the partition, and m_i is the center of partition C_i.

Silhouette Coefficient (SC) is an unsupervised intrinsic clustering evaluation index proposed by Rousseeuw in 1986, which comprehensively considers cohesion and separation (Rousseeuw, 1987). It can evaluate the performance of different clustering algorithms based on a given dataset without actual labels or compare the impact of the same algorithm on clustering results under different operating modes. Two key values can be calculated for each data point i in the cluster: a(i) and b(i). Among them, a(i) represents the average distance between data point i and other data points in its cluster, reflecting the degree of closeness between the point and members of the same cluster, which is called intra-cluster dissimilarity. The smaller a(i), the more suitable data point i will be assigned to the current cluster. Furthermore, b(i) represents the minimum average distance between data point i and all data points in another cluster, reflecting the degree of distance between that point and other clusters, which is called inter-cluster dissimilarity. The larger b(i), the less belonging data point i is to other clusters. Based on these two values, the SC evaluates clustering quality by comprehensively measuring intra-cluster compactness and inter-cluster separability. The value range of SC is usually between [-1, 1], and the closer the value is to 1, the better the clustering effect; The closer the value is to -1, the poorer the clustering effect; A value close to 0 indicates that the data point may be located at the boundary of two clusters (Dudek, 2020). The definitions of a(i), b(i) and SC can be found in *Equations 7* and *8*.

$$SC(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \tag{Eq.7}$$

$$SC(i) = \begin{cases} 1 - \dfrac{a(i)}{b(i)}, a(i) < b(i) \\ \quad 0 \quad , a(i) = b(i) \\ \dfrac{b(i)}{a(i)} \quad , a(i) > b(i) \end{cases} \tag{Eq.8}$$

As an evaluation index that comprehensively measures intra-cluster compactness and inter-cluster separability, SC can not only help us judge the quality of clustering results but also be used to select the optimal number of clusters or compare the performance of different clustering algorithms (Ünlü and Xanthopoulos, 2019; Dinh et al., 2019; Shahapure and Nicholas, 2020; Azimi et al., 2017). Due to its computational dependence on point-to-point distance, the computational cost is higher when the data scale is large. In practical applications, it is necessary to combine other evaluation indicators (such as SSE) for comprehensive analysis in order to obtain reliable clustering evaluation results.

Results and discussion

Calculation of the number of management zones

Calculate the average distortion of the collected maize canopy NDVI data and plot its changes in *Figure 7*. Using the elbow rule, it can be seen that for maize NDVI data, when the number of partitions is greater than 3, the average distortion does not significantly decrease, and the curve is relatively smooth, indicating that the number of partitions above 3 is reasonable for maize NDVI data. In agricultural production, as the number of management zones increases, the requirements for working machinery during variable operations will also correspondingly increase, which will inevitably lead to an increase in related costs. Therefore, choosing an appropriate number of management zones is necessary to achieve the best comprehensive effect of zone accuracy and economic benefits. Based on the conditions and experience of the fertilization site, this paper determines that the number of management zones is 4.

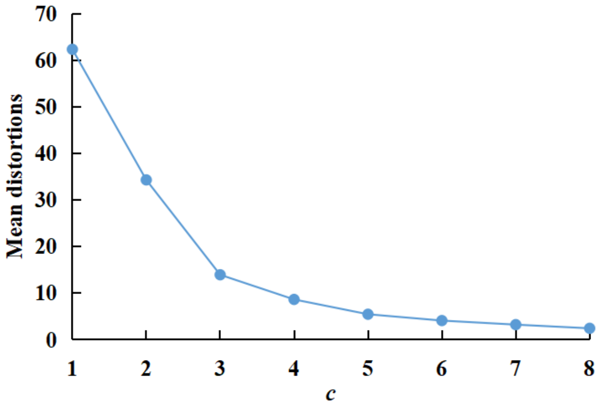


Figure 7. *Mean distortions*

Comparison of the FCM algorithm and SOM-FCM algorithm

Use the scikit learn library in Python to write clustering algorithm programs. Scikit learn library is a Python-based machine learning tool library that has become a commonly used tool for data mining and analysis due to its simplicity and efficiency. We implemented two partitioning algorithms, FCM and SOM-FCM, using the Scikit learn library and calculates their results in managing to partition. We have developed a clustering performance evaluation program to assess the clustering performance of these two algorithms. In the evaluation program, the clustering performance of the two algorithms was quantitatively analyzed by calculating the internal evaluation indicators SSE and SC. Subsequently, a comparison chart of the trends of SSE and SC under different data volumes was drawn based on the calculation results, and the performance differences of the two algorithms in different scales of data were shown in *Figures 8* and *9*.

From the experimental results, it can be seen that both the FCM algorithm and SOM-FCM algorithm show a gradually increasing trend in SSE values with the increase of data volume, which is in line with the general rule of SSE increasing with the expansion of data size in clustering analysis because more data points will lead to a more significant cumulative effect of error square sum. However, under the same amount of data, the SSE value of the SOM-FCM algorithm is consistently lower than that of the FCM algorithm, indicating that SOM-FCM outperforms the traditional FCM algorithm in clustering performance. Specifically, when the data volume is 4,000, the SSE value of SOM-FCM is about 2% lower than FCM, and as the data volume increases, this gap gradually widens. For example, when the data volume reaches 12,000, the SSE value of SOM-FCM is about 10% lower than FCM. This phenomenon indicates that SOM-FCM can better capture the spatial distribution characteristics of data and improve clustering accuracy by combining self-organizing maps (SOM) for data preprocessing, especially on large-scale datasets. There is also a specific difference in the SSE value growth rate between the two algorithms. The FCM algorithm has a relatively higher growth rate, reflecting its limitations in handling complex data distributions. At the same time, SOM-FCM exhibits stronger robustness and adaptability, further verifying the superiority of this method in NDVI data clustering. SOM-FCM achieves better clustering performance under different data volumes and demonstrates better stability and performance improvement potential when dealing with large-scale data.

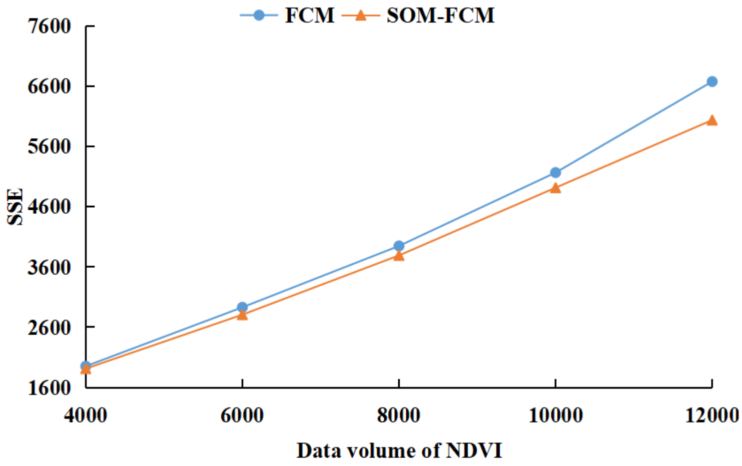


Figure 8. *Comparison of the SSE values of the two algorithms*

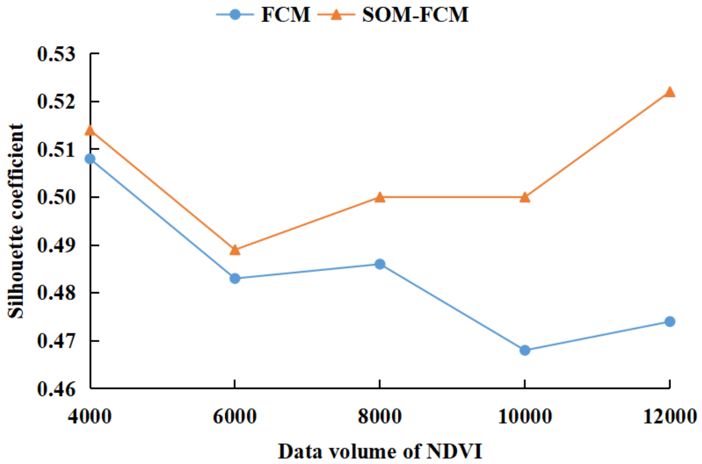


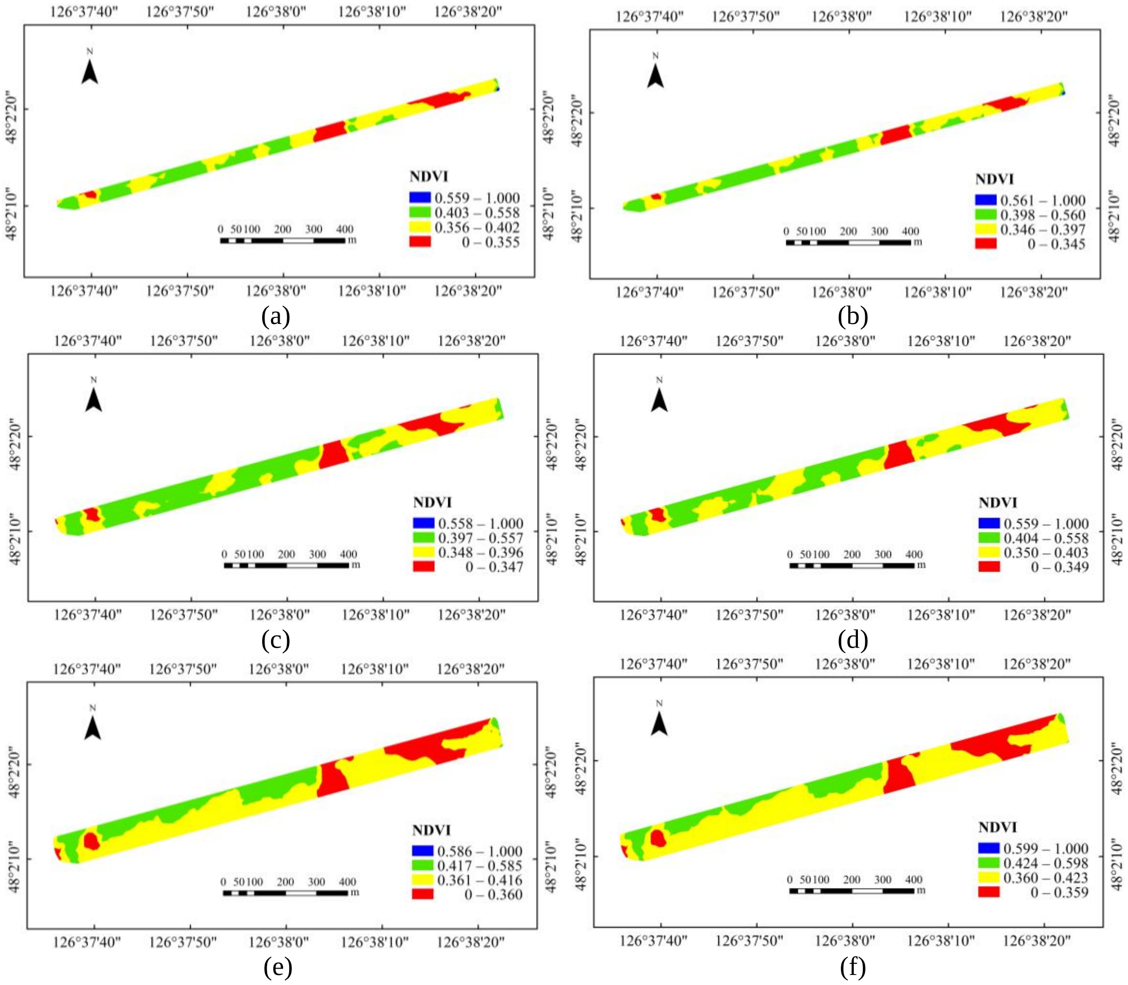
Figure 9. *Comparison of the SC values of the two algorithms*

Based on the SC calculation results, the superiority of the SOM-FCM algorithm in NDVI data clustering can be further verified. From the trend of SC value changes, the

SC value of the FCM algorithm shows a fluctuating downward trend with the increase in data volume. Although there is a slight rebound at 8,000 and 12,000 data volumes, it shows a certain degree of instability and a downward trend overall, reflecting the weakened ability of the FCM algorithm to capture inter-cluster separation when processing large-scale data. The SC value of the SOM-FCM algorithm is higher than the FCM algorithm in all data volumes and shows a more stable upward trend overall. When the data volume reaches 12,000, the SC value reaches 0.522, significantly better than the FCM algorithm’s 0.474. This indicates that SOM-FCM, by combining the preprocessing mechanism of SOM, can more effectively optimize the distribution characteristics of data, thereby improving intra-cluster tightness and inter-cluster separation, making clustering results more reasonable and reliable. Based on the analysis results of SSE, it can be seen that SOM-FCM reduces the sum of squared errors and improves the contour coefficient, indicating that this method has a stronger comprehensive performance in balancing clustering compactness and separability. Especially in large-scale data scenarios, the advantages of SOM-FCM are more pronounced, as the increasing trend of its SC value complements the relatively low growth of its SSE value, further demonstrating the robustness and adaptability of the algorithm in processing complex NDVI data.

Based on the partition results obtained from the clustering program, use ArcGIS software to draw fertilizer management partition maps for different NDVI data volumes. Firstly, load the preprocessed NDVI data and the partition results generated by cluster analysis, then interpolate spatially based on the geographic information (latitude and longitude) of the NDVI data, and finally assign different colours to NDVI values in different ranges according to the partition results. The specific fertilizer management partition diagram is shown in *Figure 10*.

Using ArcGIS to calculate the proportion of each partition, the SOM-FCM algorithm and FCM algorithm show specific differences in partition partitioning based on the proportion of the four partition areas of the two methods under different data volumes, and this difference becomes more pronounced as the data volume increases. Combining the numerical changes of SSE and SC, it can be found that the SOM-FCM algorithm outperforms FCM in overall clustering performance. The SSE value of the SOM-FCM algorithm is always lower than that of FCM, indicating that it performs better in reducing intra-cluster errors. The SC value of the SOM-FCM algorithm is significantly higher than FCM at 8000 and 12000 data volumes, especially reaching 0.522 at 12000 data volume, indicating that its clustering results have better compactness and separation. From *Figure 10*, it can be seen that the two partitioning algorithms exhibit differentiated characteristics at different data scales. At the 4000 data level, there is a significant difference in the block-level distribution of the second partition (marked in green) between the two algorithms, and there is a difference in the number of blocks in the second partition. When the data scale expands to the 8000 and 12000 levels, although the difference in block numbers tends to converge, the difference in partition coverage area persists. This phenomenon conforms to the typical characteristics of partitioning strategies that vary with data size, where differences are easily amplified in small-scale data volumes and tend to approach large-scale data volumes. Reflected in the proportion of partition area, the first partition proportion of the SOM-FCM algorithm gradually increases with the data volume increase (from 19.95% to 29.76%). In contrast, the proportions of the second and third partitions are relatively stable or slightly fluctuating, showing stronger robustness and consistency. In contrast, the partition ratio of FCM changes more dramatically, especially at a data volume of 12000. The first partition ratio (30.42%) is close to the SOM-FCM algorithm. However, the third partition ratio (28.07%) is significantly higher than the SOM-FCM algorithm (23.46%), indicating that FCM's partition partitioning is not precise enough at high data volumes, which can easily lead to blurred boundaries of specific clusters. SOM-FCM performs better in quantitative indicators and demonstrates higher rationality and stability in the distribution of partition area proportions, making it more suitable for managing partition partitioning of NDVI data.



(a) (b) (c) (d) (e) (f)

Figure 10. *Comparison of the management zones obtained using the FCM algorithm and the SOM-FCM algorithm. (a) FCM algorithm when the data volume reaches 4,000; (b) SOM-FCM algorithm when the data volume reaches 4,000; (c) FCM algorithm when the data volume reaches 8,000; (d) SOM-FCM algorithm when the data volume reaches 8,000; (c) FCM algorithm when the data volume reaches 12,000; (d) SOM-FCM algorithm when the data volume reaches 12,000*

Conclusions

In this study, we proposed a hybrid clustering algorithm that combines SOM and FCM to improve the efficiency and accuracy of management zone delineation for maize fertilization based on NDVI data. The SOM-FCM algorithm addresses the sensitivity of

FCM to initial cluster centers by utilizing SOM to provide stable and effective initial centroids, thereby enhancing the overall clustering performance. Experimental results demonstrate that the proposed method outperforms traditional FCM across various NDVI dataset sizes, as evidenced by lower SSE and higher SC values. Specifically, the maximum differences in SSE and SC between SOM-FCM and FCM were 639 and 0.048, respectively, with the performance gap widening as data volume increased. This indicates that SOM-FCM is advantageous for handling large-scale and complex agricultural datasets. Furthermore, the spatial distribution of management zones generated by SOM-FCM exhibited greater consistency and robustness than FCM, making it a more reliable tool for precision agriculture applications. These findings highlight the potential of integrating unsupervised learning techniques like SOM with traditional clustering algorithms to achieve more accurate and efficient data-driven decision-making in agricultural management. Future research could further explore the applicability of SOM-FCM in other crop types or environmental conditions to validate its versatility and scalability. This study partitioned the management zones based on the SOM-FCM and used a variable fertilization control system installed behind the tractor for fertilization. Only floating fertilization was carried out based on the benchmark fertilization amount. Subsequent research will combine yield to study the specific application of the required fertilization amount.

Acknowledgements. This research was funded by the National Key Research and Development Program (2023YFD1501005-06), Heilongjiang Bayi Agricultural University Support Program for San Heng San Zong (ZRCPY202014), Heilongjiang Bayi Agricultural University Scientific Research Startup Project (XDB202405).

REFERENCES

[1] Ameer, S., Cheema, M. J. M., Khan, M. A., Amjad, M., Noor, M., Wei, L. (2022): Delineation of nutrient management zones for precise fertilizer management in wheat crop using geo-statistical techniques. – Soil Use and Management 38: 1430-1445.

[2] Azimi, R., Ghayekhloo, M., Ghofrani, M., Sajedi, H. (2017): A novel clustering algorithm based on data transformation approaches. – Expert Systems with Applications 76: 59-70.

[3] Barman, A., Sheoran, P., Yadav, R. K., Abhishek, R., Sharma, R., Prajapat, K., Singh, R. K., Kumar, S. (2021): Soil spatial variability characterization: delineating index-based management zones in salt-affected agroecosystem of India. – Journal of Environmental Management 296: 113243.

[4] Bezdek, J. C., Ehrlich, R., Full, W. (1984): FCM: fuzzy c-means algorithm. – Computers & Geoscience 10: 191-203.

[5] Cottrell, M., Olteanu, M., Rossi, F., Villa-Vialaneix, N., N. (2018): Self-organizing maps, theory and applications. – Revista de Investigacion Operacional 39: 1-22.

[6] Dinh, D.-T., Fujinami, T., Huynh, V.-N. (2019): Estimating the Optimal Number of Clusters in Categorical Data Clustering by Silhouette Coefficient. – In: Chen, J. et al. (eds.) Knowledge and Systems Sciences: 20th International Symposium, KSS 2019, Da Nang, Vietnam, November 29–December 1, 2019, Proceedings. Springer, Singapore, pp. 1-17.

[7] Dudek, A. (2020): Silhouette Index as Clustering Evaluation Tool. – In: Jajuga, K. et al. (eds.) Classification and Data Analysis: Theory and Applications. Springer, Cham, pp.19-33.

[8] Dunn, J. C. (1973): A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. – Journal of Cybernetics 3: 32-57.

[9] Gul, M., Rehman, M. A. (2023): Big data: an optimized approach for cluster initialization. – Journal of Big Data 10: 120.

[10] Hassan, B. A., Rashid, T. A., Mirjalili, S. (2021): Performance evaluation results of evolutionary clustering algorithm star for clustering heterogeneous datasets. – Data in Brief 36: 107044.

[11] Kalyani, C., Ramudu, K., Reddy, G. R. (2018): A review on optimized K-means and FCM clustering techniques for biomedical image segmentation using level set formulation. – Biomedical Research 29: 3660-3668.

[12] Karatas, M., Karademir, E. (2023): Management of nitrogen stress in cotton (Gossypium hirsutum L.) using Greenseeker technology. – Journal of Applied Life Sciences and Environment 55(4): 441-456.

[13] Kohonen, T. (1990): The self-organizing map. – Proceedings of the IEEE 78: 1464-1480.

[14] Kohonen, T. (2013): Essentials of the self-organizing map. – Neural Networks 37: 52-65.

[15] Kohonen, T., Oja, E., Simula, O., Visa, A., Kangas, J. (1996): Engineering applications of the self-organizing map. – Proceedings of the IEEE 84: 1358-1384.

[16] Miljković, D. (2017): Brief review of self-organizing maps. – 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 22-26 May 2017, pp. 1061-1066.

[17] Ming-Chuan, H., Don-Lin, Y. (2001): An efficient fuzzy c-means clustering algorithm. – Proceedings 2001 IEEE International Conference on Data Mining, 29 Nov.-2 Dec. 2001, pp. 225-232.

[18] Moharana, P. C., Jena, R. K., Pradhan, U. K., Nogiya, M., Tailor, B. L., Singh, R. S., Singh, S. K. (2020): Geostatistical and fuzzy clustering approach for delineation of site-specific management zones and yield-limiting factors in irrigated hot arid environment of India. – Precision Agriculture 21: 426-448.

[19] Nayak, J., Naik, B., Behera, H., Fuzzy (2015): C-Means (FCM) Clustering Algorithm: A Decade Review from 2000 to 2014. – In: Jain, L. C. et al. (eds.) Computational Intelligence in Data Mining-Volume 2: Proceedings of the International Conference on CIDM, 20-21 December 2014. Springer, New Delhi, pp. 133-149.

[20] Ohana-Levi, N., Ben-Gal, A., Peeters, A., Termin, D., Linker, R., Baram, S., Raveh, E., Paz-Kagan, T. (2021): A comparison between spatial clustering models for determining N-fertilization management zones in orchards. – Precision Agriculture 22: 99-123.

[21] Pawase, P. P., Nalawade, S. M., Bhanage, G. B., Walunj, A. A., Kadam, P. B., Durgude, A. G., Patil, M. R. (2023): Variable rate fertilizer application technology for nutrient management: a review. – International Journal of Agricultural and Biological Engineering 16: 11-19.

[22] Rousseeuw, P. J. (1987): Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. – Journal of Computational and Applied Mathematics 20: 53-65.

[23] Ruspini, E. (1972): Optimization in sample descriptions: data reduction and pattern recognition using fuzzy clustering. – IEEE Transactions on Systems, Man and Cybernetics 2: 541.

[24] Shahapure, K. R., Nicholas, C. (2020): Cluster quality analysis using Silhouette score. – IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), 6-9 Oct. 2020, pp. 747-748.

[25] Termin, D., Linker, R., Baram, S., Raveh, E., Ohana-Levi, N., Paz-Kagan, T. (2023): Dynamic delineation of management zones for site-specific nitrogen fertilization in a citrus orchard. – Precision Agriculture 24: 1570-1592.

[26] Ünlü, R., Xanthopoulos, P. (2019): Estimating the number of clusters in a dataset via consensus clustering. – Expert Systems with Applications 125: 33-39.

[27] Vendrusculo, G. L., Kaleita, F. A. (2011): Modeling zone management in precision agriculture through fuzzy C-means technique at spatial database. – Louisville, Kentucky, August 7-10, 2011. ASABE, St. Joseph, MI.
[28] Wang, Y., Cheng, J., Zhang, L., Tang, Q.-H. (2017): Sensitivity Analysis and Optimization of FCM Initial Clustering Centers. – In: Lin, H. Z. et al. (eds.) 3rd Annual International Conference on Social Science and Contemporary Humanity Development (SSCHD 2017). Atlantis Press, Paris, pp. 506-513.
[29] Zsebő, S., Bede, L., Kukorelli, G., Kulmány, I. M., Milics, G., Stencinger, D., Teschner, G., Varga, Z., Vona, V., Kovács, A. J. (2024): Yield prediction using NDVI values from Greenseeker and MicaSense cameras at different stages of winter wheat phenology. – Drones 8: 88.