Revista Brasileira de Ciência do Solo

Division - Soil In Space and Time | Commission – Pedometrics

Delineation of potential management zones for cowpea by factorial kriging and spatial Fuzzy C-Means of yield and soil chemical and texture properties

João Fernandes da Silva Júnior[(1)*], Thiago Thomé da Silva[(1)], Devid Jackson da Silva Sousa[(1)], Rose Luiza Moraes Tavares[(3)], Benedito Dutra Luz de Souza[(4)] and Daniel Pereira Pinheiro[(2)]

[(1)] Universidade Federal Rural da Amazônia, *Campus* de Capanema, Capanema, Pará, Brasil.
[(2)] Universidade Federal Rural da Amazônia, Instituto de Ciências Agrárias, Belém, Pará, Brasil.
[(3)] Universidade de Rio Verde, Fazenda Fontes do Saber, Rio Verde, Goiás, Brasil.
[(4)] Agropecuária Milênio, Tracuateua, Pará, Brasil.

*** Corresponding author:**
E-mail:
joao.fernandes@ufra.edu.br

Received: December 03, 2023

Approved: June 4, 2024

How to cite: Silva Júnior JF, Silva TT, Sousa DJS, Tavares RLM, Souza BDL, Pinheiro DP. Delineation of potential management zones for cowpea by factorial kriging and spatial Fuzzy C-Means of yield and soil chemical and texture properties. Rev Bras Cienc Solo. 2025;49:e0230158.
https://doi.org/10.36783/18069657rbcs20230158

Editors: José Miguel Reichert and Quirijn de Jong van Lier.

ABSTRACT: In precision agriculture, accurate delineation of management zones and understanding spatial variability of soil properties and crop yields are critical for optimizing resource allocation and improving productivity. Spatial variability of different environmental factors (soil and plants) is evident in several studies. Associations between the texture and chemical properties of the soil and cowpea yield have been tested, but a large, unexplained variance of ranges between kriged maps is usually reported. This suggests that a deeper exploration into the soil properties of these spatial interactions may help develop our understanding on how to reduce the number of soil property maps to delineate management zones and simplify interpretation. The main objective of this study was to investigate whether factorial kriging analysis can be used as an auxiliar variable to cokriging of soil properties and cowpea yield, and what is the potential of Spatial Fuzzy c-Means associated with factorial kriging analysis to delineate management zones. This study employed factor maps and spatial clustering to classify the cowpea field in management zones based on a multivariate and geostatistical analysis using soil texture and chemical properties. From Farmer, 66 soil samples were collected at a layer of 0.00-0.20 m, at points with a regular spacing of 12 m, at Agropecuária Milênio in the municipality of Tracuateua, Pará State, to make the technology applicable to the most common data available to farmers. It also used Spatial Fuzzy c-Means to generate estimated maps. Only the kriged maps of soil properties were inefficient in delineating management zones. Factor maps and Spatial Fuzzy c-Means were efficient in delineating the two management zones. Factorial kriging analysis can be used in cokriging to estimate soil properties and the cowpea field. The proposed method is a practical tool to delineate management zones, performing better and more efficiently compared with soil multiple property maps. The optimal number of management zones for cowpea cultivation was determined to be two. This encompasses soil management, yield considerations, and site-specific choices, all aimed at mitigating the impacts of precision agriculture on high productivity.

Keywords: precision agriculture, geostatistics, site-specific management, multivariate analysis, Amazon.

INTRODUCTION

Cowpea (*Vigna unguiculata* (L.) Walp.), is a grain legume native from Africa and is a primary source of protein for millions of people in North and Northeast Brazil and other parts of the developing world. Cowpea is a legume crop with enormous nutritional, agronomic and economic value (Osipitan et al., 2021; Abebe and Alemayehu, 2022), and consequently, researchers are investigating how precision agriculture (PA) can leverage the spatial variability of soil properties resulting from the intricate interplay between soil characteristics, weather patterns, and management practices to optimize crop input management and promote both economic viability and environmental sustainability (Ahmad et al., 2021; Karydas et al., 2023). In addition, the studies by Guedes Filho et al. (2010) and Lipiec and Usowicz (2018) identified spatial correlations between crop yields and certain physical and chemical soil properties, suggesting the potential division of the area into distinct land management zones. Thus, it is important to detect areas with higher or lower productivity for the sustainable management of crops and localized application of fertilizers. While significant research has explored the application of precision agriculture (PA) in major Brazilian crops like coffee (Kazama et al., 2021), corn (Rodrigues and Corá, 2015; Anselmi et al., 2021), soybean (Umbelino et al., 2018), alfalfa (Rossi et al., 2018), sugar cane (Sanches et al., 2019), peach (Oldoni et al., 2019), a recent study by Anago et al. (2023) highlights the potential of PA for cowpea in African farms. Their stdudy demonstrates how geostatistical tools can be integrated with Diagnosis and Recommendation Integrated System (DRIS) to optimize cowpea production.

This technology offers valuable insights for significantly increasing yields of understudied Brazilian crops like cowpea, using an appropriate quantity of these inputs (water, energy, fertilizers and pesticides). Furthermore, the cowpea transcends its role as a simple food source within the Brazilian food chain. It emerges as a crop with remarkable resilience, demonstrating an ability to thrive under diverse climatic and soil conditions. Additionally, its exceptional efficiency in utilizing fertilizers makes it a valuable asset for sustainable agricultural practices. According to Silva et al. (2020), cowpeas are sensitive to extreme environmental conditions, requiring specific edaphic and climatic characteristics for optimal growth and development. Variations in fertility and nitrogen use efficiency are challenges for cowpea; e.g., nitrogen fertilization at sowing can limit biological nitrogen fixation in bean cultivation in Brazil, impacting yield due to low soil nitrogen content and minimal input from resource-poor farmers (Barros et al., 2018).

Another limitation includes high interannual variability, susceptibility to prolonged droughts, and low levels of technification, significantly impacting crop production (Santos et al., 2021). Furthermore, cowpeas are sensitive to damage caused by diseases, including anthracnose, angular leaf spot, common bacterial blight, and rust (Ganascini et al., 2019). These diseases can significantly impact cowpea production in the country. For this reason, management zone delineation in cowpea is important to propose more assertive management, thus avoiding production losses.

In PA, it is essential to delineate management zones (MZs) that can express the combination of homogeneous factors. The MZs are subdivisions of a field, each characterized by the relative homogeneity of crops and/or environmental parameters (Doerge, 1999; Khan et al., 2020), which, therefore, differ in the need for specific input treatment rates. When it comes to determining MZs based on actual crop growth patterns, yield maps are valuable tools in precision agriculture (Georgi et al., 2018). Many approaches have been described in the literature for delineating management zones using yield maps derived from specific inputs (e.g., fertilization, irrigation, pesticides, yield). However, the complex combination of these factors often proves difficult to understand. The most usual approaches to determine MZs using yield maps are: 1) the empirical method, which uses frequency distribution of yield and expert knowledge to divide the field into three or four (Khan et al., 2020); and 2) cluster analysis, such as K-means and Fuzzy C-means

(Javadi et al., 2022); and 3) the iterative self-organizing data analysis technique (Yang, 2020) or 4) determination of management zones from normalized and standardized equivalent productivity maps (Suszek et al., 2011; Schenatto et al., 2017).

Although the methods of empirical classification are simpler, cluster analysis allows for greater differentiation between management zones or production classes. Empirical methods are used mainly when the target variable (usually yield) is used to create MZs. For attributes that are correlated to the target variable for creation of MZs, clustering methods are usually used. Castrignanò et al. (2017) successfully combined sensors with multivariate geostatistics to determine MZs in an agricultural field. Oldoni et al. (2019) were able to delineate management zones in a peach orchard using multivariate and geostatistical analyses.

Recently, Gavioli et al. (2016) and Abdelaal et al. (2021) proposed the determination of management zones using Fuzzy C-means, spatial correlation analysis, principal component analysis (PCA) named as (MULTISPATI-PCA) analysis making them more viable for implementation from the viewpoint of field operations. Multivariate analysis has advantages over other methods in that many factors can be integrated into the analysis (John et al., 2021). In this line, our proposal is to evaluate the potential of the method Spatial Fuzzy c-Means (SFCM) to delineate MZs, considering the optimization of delineation by spatial factorial analysis.

There is a gap in studies on the management of spatial variability in areas with Cowpea that have studied applications of precision agriculture with a multivariate geostatistical approach, based on the strategy that takes into account method Spatial Fuzzy c-Means (SFCM) and factorial kriging analysis (FKA) (Matheron, 1982).

However, tools as SFCM to determine MZs from multivariate analysis can be used to assist applications of PA. We hypothesized that SFCM and factorial kriging analysis (FKA) are effective techniques to delineate potential management zones from soil properties and cowpea yield. When used with the score’s factors that the explained variance major, these techniques may optimize the soil properties maps from ordinary kriging.

This study had two goals: (i) to investigate whether Factors can be applied consistently as an auxiliar variable to cokriging of soil properties and cowpea yield, (ii) to investigate the potential of integrating fuzzy c-means clustering (SFCM) with factorial kriging analysis to delineate management zones based on soil texture and chemical properties using soil samples collected from the layer of 0.00-0.20 m, which represents a typical depth for agricultural practices in the northeastern region of Pará State, Brazil.

MATERIALS AND METHODS

Description of the study site and field procedures

The experiment was carried out from June to September 2018, in an area belonging to the company Agropecuária Milênio, in the community of Fátima, located in Tracuateua, Pará State, Brazil. The area is located at 01° 08’ 67” S and 46° 58’ 33” W, in an *Latossolo Amarelos Álico* with medium texture according to the Brazilian Soil Classification System (SiBCS) (Santos et al., 2018) and classified as Ferralsol (Dystric Arenic Xanthic), according to the World Reference Base for Soil Resources (IUSS Working Group WRB, 2022).

The climate is tropical Awi type, according to the Köppen classification system, with high annual rainfall rates, around 2.543 mm, with a short dry season between August and December (monthly precipitation around 60 mm) and average annual temperature around 27.7 °C.

The fallow area with evergreen equatorial forest has smooth wavy relief and is used for cowpea plantations. It was prepared with mechanical mowing, preferably with a knife roller. The remains of plant mass were incorporated into the soil, at a layer of approximately 0.00-0.20 m, to better protect the soil against surface erosion. The BR3-Tracuateua cultivar of cowpea was chosen for this experiment. On July 6, 2018, the seeds were hand-sown with a spacing of 0.50 m between rows and a density of 15 plants per linear meter. The experiment was conducted in a field (1.2 ha) after one year-fallow, where cowpea had been previously grown.

Before the installation experiment, the area had been preserved with natural vegetation. Agricultural operations in the field were predominantly mechanized. When the experiment started, grain crops were sown directly on natural vegetation, and chemical fertilizers and liming had been used for more than a decade.

The area had been cultivated for one year with beans in a conventional system under spontaneous vegetation. Weeds were controlled with hand weeding every 15 days. Pests and diseases were monitored and controlled weekly, and when necessary, chemical control with insecticides and fungicides was used. Cowpea (*Vigna unguiculata*) was planted under natural precipitation conditions and without irrigation. No lime and/or gypsum or fertilization were used in this experiment.

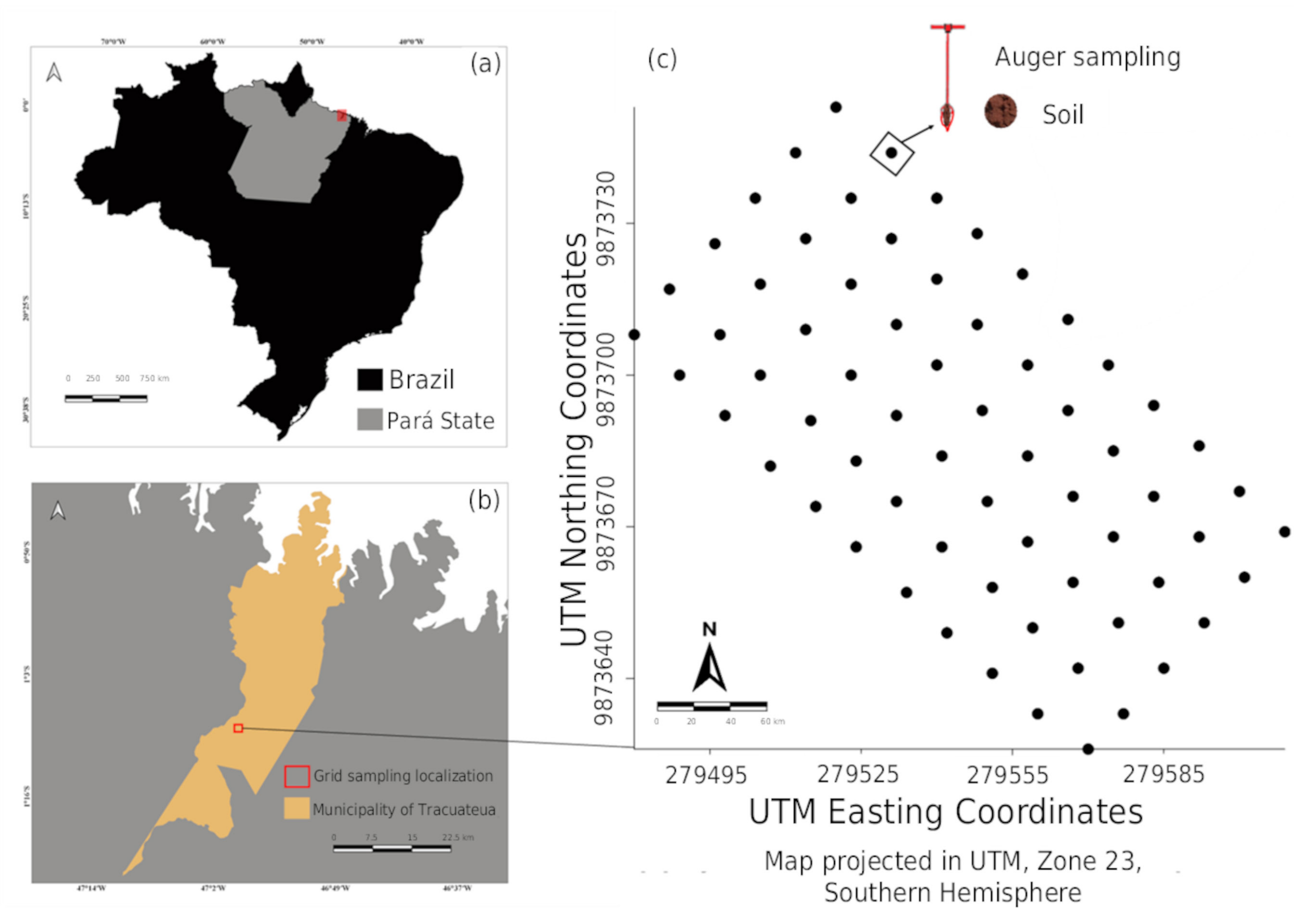

Figure 1. Location of the study area in Brazil (a) and Tracuateua (b), and the sampling experimental grid established on the cowpea farms, with 66 georeferenced samples (c).

Sampling scheme and cowpea yield assessment

Soil samples were collected at a layer of 0.00-0.20 m, with the aid of a Dutch auger and were collected in a 12 × 12 m, had a total of 66 soil samples georeferenced using a Total Station Geodetic Gt2 (Figure 1). At 70 days after emergence, in this geographic position were made of all plants that were inside the 3 × 3 m (9 m^2) sampling cell corresponding to the 66 georeferenced cowpea grain samples (Figure 2a). Cowpea yield was determined at the time of harvest and dry grain processing, with humidity corrected to 13 %. Yield was calculated using the ratio of weight in kg of the grains obtained in each 3 × 3 m sampling cell (Figure 2b) and the area of 9 m^2. After that, these data were transformed into kg ha^{-1}.

Laboratory analyses

Soil samples were sent to the soil laboratory of the Federal Rural University of Amazon, *Campus* of Capanema, where they were air-dried and passed through a 2 mm sieve to obtain air-dried fine soil (ADFS). Chemical analyses (extractable P, K^+, Na^+, Al^{3+}, Ca^{2+} and Mg^{2+}) and texture analyses (grain size analysis to obtain sand, silt and clay fractions) of soil followed the method recommended by Claessen (1997).

Univariate data analysis

Descriptive analysis statistics

Descriptive analysis was applied to the texture and chemical properties of the soil and plant attributes (grain weight/hectare) were evaluated by descriptive statistics (minimum, maximum, mean, standard deviation, coefficient of variation, coefficient of skewness and kurtosis and normality test). The dispersion classification was evaluated by using the Warrick and Nielsen (1980): low variability (CV <12 %), average variability (12≤ CV ≤62 %) and high variability (CV >62 %).

Geostatistical analysis of soil properties and Cowpea yield

Geostatistical analysis is a powerful tool for evaluating the spatial relationships between soil and plant variables and covariables. The geostatistical analysis was based on the regionalized variable theory, using the variables obtained from the soil and cowpea yield.

Mathematical models were adjusted to experimental semivariograms to analyze the spatial dependence structure of the texture and chemical properties of soil and cowpea yield. The selection of the best model was based on (1) minimizing the sum of the squared

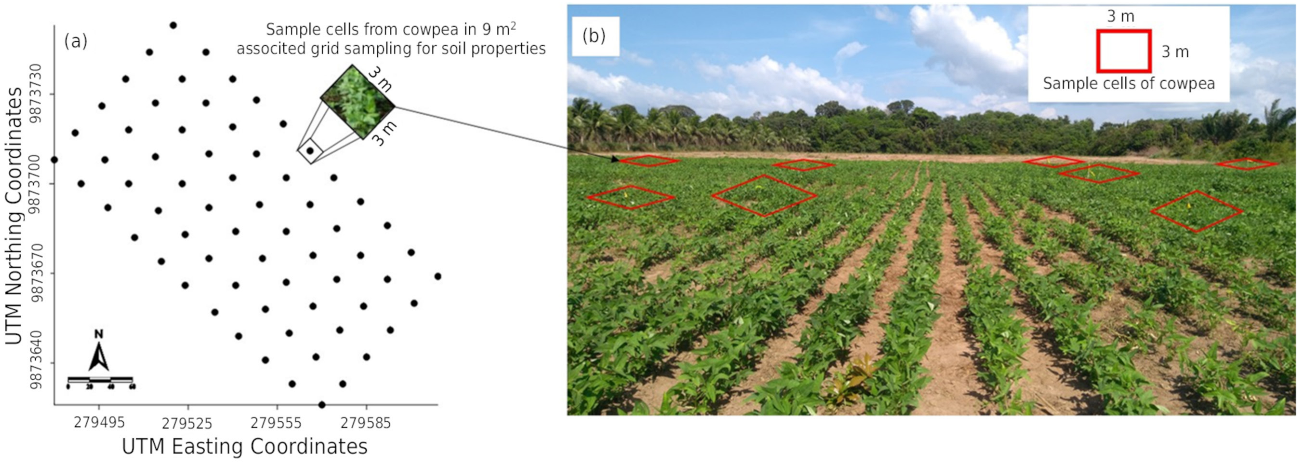

Figure 2. Overview of the grid sampling (a) and cell sampling for measure of cowpea yield (b).

residuals; (2) maximizing the coefficient of determination (R^2), and taking into account the assumption of stationarity of the intrinsic hypothesis (Matheron, 1963), estimated through equation 1.

$$\hat{Y}(h) = \frac{1}{2N(h)} \mathrm{Var}\left[Z(x_i) - Z(x_i + h)\right] \quad \text{Eq. 1}$$

in which: N(h) is the number of pairs of points; Z(xi) and Z (x_i + h) are regionalized variables separated by a distance h.

The semivariogram is represented by the graph of γ(h) versus h. To refine our mathematical model, we adjusted it to the calculated values of γ(h). This process allowed us to estimate the parameters of the theoretical semivariogram model: the Nugget effect (C_0), the Sill ($C_0 + C_1$), and the Range (a). We evaluated four different models: spherical, exponential, Gaussian, and linear. All variographic analyses were conducted using GS+ Version 9.0 (Robertson, 1998).

After obtaining the experimental semivariogram, a permissible theoretical mathematical model was selected to represent it. Then, using the parameters of this theoretical model, the values of soil properties and Cowpea yield at a specific unsampled geographic position in the study area were estimated using ordinary kriging (OK) from soil properties and cowpea yield, as demonstrated in equation 2

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \quad \text{Eq. 2}$$

in which: $Z(x_0)$ is the estimate from kriging at the point x_0; $Z(x_i)$ is the values measured in x_i, i = 1, 2,..., N; λ_i is the kriging weights calculated based on the adjusted semivariogram, assigned to neighboring values $z(x_i)$ to estimate $Z(x_0)$.

Multivariate data analysis

Factorial analysis (FA)

The dataset was analyzed by correlation matrix to evaluate the relationships between all soil properties before to apply the multivariate analysis. Factor analysis is a multivariate statistical method extensively used to reorganize the soil properties into fewer underlying factors (also called common factors) to retain as much information as possible as contained in the original soil properties.

In contrast to the original soil properties, the factors are independent of each other. Substituting these factors for the original soil properties can effectively simplify a large dataset. Factors are derived through an eigenvalue analysis of the correlation matrix, with factor loads and factor scores serving as the primary measures of FA.

The first step involves standardizing the raw data and computing a correlation matrix of the variables based on these standardized variables. The second step is to estimate the factor loads, which indicate the degree of association between the factor and the variables. Factor loads range from -1 to +1, with higher absolute values indicating a stronger relationship between the factor and the variable. Additionally, Liu et al. (2003) suggested categorizing factor loads as strong, moderate, and weak, corresponding to absolute load values in the ranges of >0.75, 0.75–0.50, and 0.50–0.30, respectively.

The last step linearly transforms the factors associated with the initial set of loads per factor rotation to maximize the variable variances and to obtain a better interpretable loading pattern. The four factors can be used to extract other optimized vectors. The latter were obtained using the principal components by means of axis rotation. In this study, varimax rotation was used. The factor scores are calculated for each individual case to represent the contribution of each factor in each case.

The FA was applied to determine the factors that control the regional yield of cowpea and the resulting factors in the main texture and chemical properties of the soil. The factor was extracted by principal components, and only eigenvalues greater than one were considered (Kaiser, 1958). Factorial load matrix was rotated to obtain factors that were not correlated by varimax rotation. In this study, the scores of the common factors of the variables were considered: P, Al^{3+}, K^+, Na^+, Ca^{2+}, Mg^{2+}, Sand, Silt, and Clay.

Multivariate geostatistical modelling

Geostatistical analysis is a powerful tool for evaluating the spatial relationships between soil and plant variables and covariables. Geostatistical analysis was based on the regionalized variable theory, using the variables obtained from the soil and cowpea yield. Then, an analysis was carried out based on FA scores to classify the interrelationship of the variables measured. Multivariate geostatistical methods combine the advantages of geostatistical techniques and multivariate analysis, incorporating spatial or temporal correlations and multivariate relationships to detect and map different sources of spatial variation at different scales.

Several researchers (Aggelopooulou et al., 2013; Buttafuoco et al., 2015) have described multivariate geostatistical methods in detail for the delineation of MZs. Geostatistics provides a semivariogram of data within a spatial dependency structure, including spatial and temporal covariance functions. As expected, these semivariogram models are called spatial or temporal structures and are defined in terms of the correlation between any two separate points spatially or temporally. Semivariograms provide a means of quantifying the commonly observed relationship between sample values and sample proximity (Lin et al., 2008). The second-order stationarity semivariogram C(h) of the regionalized variable, Z(x), is defined as:

$$\hat{Y}(h) = \frac{1}{2N(h)} \mathrm{Var}\left[Z(x_i) - Z(x_i + h)\right] \qquad \text{Eq. 3}$$

in which: *h* is the distance of separation between pairs of points; Var represents the variance between tail and head between pairs of points; $Z(x_i)$, and $Z(x_i + h)$ are regionalized variables separated by a distance h.

An experimental semivariogram for the lag interval distance in class h, is given by equation 4.

$$\hat{Y}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[Z(x_i) - Z(x_i + h)\right]^2 \qquad \text{Eq. 4}$$

in which: $\hat{Y}(h)$ are the semivariance values for the distance h; $N(h)$ represents the number of pairs of points separated by the lag distance h; $Z(xi)$. Similarly, the spatial correlation or cross semivariograms ($\hat{y}_{\alpha\beta}$) between two variables can be defined according to equation 5.

$$\hat{y}_{\alpha\beta} = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[Z_\alpha(x_i) - Z_\alpha(x_i + h)\right] \cdot \left[Z_\beta(x_i) - Z_\beta(x_i + h)\right] \qquad \text{Eq. 5}$$

in which: α and β represent the different regionalized variables, specifically the factor scores from the Factorial Kriging Analysis (FKA) and cowpea yield. The experimental cross semivariogram ($\hat{y}_{\alpha\beta}(h)$) can be calculated using equatoin 6.

$$\hat{y}_{\alpha\beta(h)} = \sum_{u=1}^{S} y_{\alpha\beta}^{u}(h) = \sum_{u=1}^{S} b_{\alpha\beta}^{u} g^{h}(h) \qquad \text{Eq. 6}$$

in which: S is the number of spatial scales, $b^u{}_{\alpha\beta}$ are the coefficients associated with each spatial scale u, and $b^u(h)$ represents the elementary semivariogram functions for each scale. This multivariate linear spatial model enables efficient regionalization of the set of random functions, allowing for the manipulation of complex, multivariate spatial data (Wackernagel, 2003).

A set of second-order stationarity regionalized variables, $\{Z_i(x);\ i=1,...,N\}$, can be decomposed into a set of spatial components, $\{Z_i^u(x);\ i=1,...,N;\ u=1,...,S\}$, according to equation 7.

$$Z_i(x) = \sum_{u=1}^{S} Z_i^u(x) + m_i \qquad \text{Eq. 7}$$

in which: i represent the different regionalized variables; $N(h)$ is the number of regionalized variables; i represents the different spatial scales; and S is the dominion area; m_i is $E[Z_i(x)]$.

Then, the set of spatial components Z_i^u can be decomposed into sets of spatially unrelated factors (Goovaerts, 1992; Wackernagel, 2003). During the modeling of the experimental semivariogram, the coefficients of the nugget effect (C_0), the sill (C_0+C_1), and the range (a) were menores measured. The spherical, exponential, and Gaussian models were also tested.

The best model was selected based on several criteria: 1) These criteria included linear and angular regression coefficients; 2) Additionally, the regression coefficient between observed and estimated values was considered; 3) The residual sum of squares (RSS) was used to assess model fit and finally; and 4) we included linear and angular regression coefficients from cross validate.

Ordinary kriging of factors scores

After modeling the direct and cross semivariograms, interpolation was performed using ordinary kriging (OK) using the factors scores from FA. This methodology consists of performing a weighted average of the neighboring samples (Equation 6), with the weights of each neighboring sample determined by semivariance as a function of h (Equation 2), resulting in an estimate without trend and with minimum variance (Matheron, 1963), according to equation 8.

$$\hat{Z}(x_0) = \sum_{i=1}^{N} \lambda_i(x) Z(x_i) \qquad \text{Eq. 8}$$

in which: $\hat{Z}(x_0)$ is the estimative from OK in the point not sampled at position (x_0) ; N is the number of values used for estimation; $\hat{Z}(x_1)$ is the value observed in point i_i, in which i=1,2,...n; and λ_i is the weight associated with each value and is the value observed at point i.

Co-kriging

Co-kriging (CK) is a method to estimate that extends ordinary kriging by incorporating additional observed variables, known as covariates. These covariates are often correlated with the variable of interest and are used to enhance the accuracy of the interpolation. Unlike regression and universal kriging, co-kriging does not require the secondary information be available at all prediction locations. Covariate can be measured at the same points as the target variable (co-located samples), at different points, or both. The CK is commonly used when the covariate is less expensive to measure than the target variable.

The value Z_0^*, which is unknown, can be expressed as a linear combination of N values from two or more regionalized variables. In the case of two-variable co-kriging, where the input data are only available at wells, equation 9 is the general equation used.

$$Z_0^* = \sum_{i=1}^{n} \lambda_i Z_i + \sum_{i=1}^{n} \beta_j t_j \quad \text{Eq. 9}$$

in which: z_0* represents the predicted value of Z_i at location 0. The values $Z_{i,...,}Z_n$ represent the actual measurements collected at *n* nearby locations. Additionally $t_1,...,t_m$ are secondary data at n nearby locations. Finally, $\lambda_1,...,\lambda_n$ and $\beta_1,...,\beta_n$ are the weights that need to be calculated to perform CK effectively.

The objective of CK in this study is to preliminarily assess the existing spatial correlation between the variables (factors) and soil properties. This correlation should be substantial to ensure the consistency of interpretations for spatial multivariate data in delineating management zones.

Assessment factors as auxiliar to cokriging

To assess the use of factors as an auxiliary covariate, the existence of a spatial dependence structure of the factors was evaluated by modeling of direct semivariograms and cross semivariogram using factors to check the potential use of the factors for cokriging estimates with the factors for estimation of soil properties and cowpea yield with higher loads. We leveraged kriging to link the insights from factorial analysis with the spatial patterns revealed by geostatistics. This involved using only factor scores with eigenvalues greater than 1. The potential of the regionalized factors for estimating cowpea yield was assessed by calculating the correlation between the factor scores maps and the cowpea yield maps. These correlations were then visualized using a scatter plot generated with SAGA GIS 2.3.

Yield-factors-based management zones using Spatial Fuzzy c-Means (SFCM)

The SFCM approach to delineation of potential management zones-based soil properties uses a kriged map (composite R, G, B) from the fusion of three kriged maps based on the scores of the principal components, and this fusion returns only one map, reducing the dimensionality. However, the RGB stack maps were classified by Fuzzy C-Means (FCM) clustering algorithm, which was first introduced by Dunn (1973) and later improved by Bezdek (1981). Let X = x_1, $x_{2,}$..., x_n are the n pixels of the image to be partitioned into C clusters, where x_j represents a feature of a vector. The algorithm is an iterative optimization process that minimizes the cost function defined by the equations 10 and 11.

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left(d_{ij} / d_{ik} \right)^{(2/m-1)}} \quad \text{Eq. 10}$$

$$C_j = \sum_{i=1}^{n} \left(\frac{\left(\left(\mu_{ij} \right)^m \right) x_i}{\left(\mu_{ij} \right)^m} \right) \quad \text{Eq. 11}$$

in which: d_{ij} is the distance between *ith* data and *jth* cluster center; *C* represents the number of clusters; *m* is the fuzziness index; μ_{ij} is the fuzziness index; *n* is the number of data points; and C_j represents the jth cluster center.

Considering the tendency of neighboring pixels to exhibit similar values, there is a higher probability of them being assigned to the same cluster. To harness this spatial information, the following steps are: STEP 1 - Selection of data for grouping; STEP 2 - Standardize the columns; STEP 3 - Use a non-spatial method to select appropriate parameters (k, m and β); STEP 4 - Use the selected parameters to determine the alpha value for the spatial method; STEP 5 - See Groups, if all obtained groups are stable; STEP 6 - Assess classification quality and spatial consistency; STEP 7 - Investigate the results of spatial and non-spatial classification.

Yield-based management zones

We located the MZs derived from the cowpea yield map using the methodology adopted by Milani et al. (2006). We assumed the cowpea yield map could cluster in three zones according to your yield obtained from the pixel values. However, the pixels were reclassified according to criteria: Low, Medium and High, as shown in table 1.

Criteria for assessing Management Zones based on SFCM

Following the spatial fuzzy c-means classification, two raster stacks, each containing three factor layers as RGB stack images, were used as input. The first stack utilized F3, F2, and F1 factors (R = F3, G = F2, B = F1), while the second employed F4, F2, and F1 factors (R = F4, G = F2, B = F1). These factor selections were based on their highly explained total variance. Subsequently, we employed five different criteria to evaluate potential management zones.

1. Variance analysis (ANOVA) between management zones was obtained according to Milani et al. (2006) from Cowpea Yield.

2. The uncertainty of SFCM using the Local Indicators of Spatial Association-LISA called local Moran's I statistic (Anselin, 1995) using R package "Geocmeans" (Jérémy and Apparicio, 2021) in R 4.1.2 environment (R Development Core Team, 2021). Moran index from RGB stack imageries as described above. When the results are negative, values represent the pixels with errors that decrease local accuracy, and high positive values indicate good local accuracy.

3. Explained inertia (ei): If the values are close to 1, it indicates a good classification by SFCM.

4. The silhouette index, proposed by Rousseeuw (1987), is a measure used to assess the quality of pixel classification. Values close to 1 indicate that pixels are well classified, meaning they are closer to objects of the same class than to objects of different classes.

5. Spatial inconsistency: values near to zero indicate a good classification.

The overall workflow is illustrated in the flowchart presented in figure 3.

RESULTS AND DISCUSSION

Descriptive statistics

Table 2 shows that K^+, Na^+, Al^{3+}, clay, and Cy have normally distributed after a Shapiro–Wilk test of normality with non-significant skewness. Although the occurrence of non-normality to P, Ca^{2+}, Mg^{2+}, sand, and silt, it is common for data obtained from nature (field) (Webster, 1985).

Soil texture class was defined as a sandy soil and according to the Pará State's handbook for fertilization (Brasil et al., 2020), the means of soil chemical properties content were low in the study area. All soil properties had medium coefficients of variation between 20.73 to 62.00, except sand that had 5.82, indicating low variability according to Warrick and Nielsen's criteria (1980).

Table 1. Criteria for defining management zones based on a cowpea yield map

Management zones	Management requirements	Zones
Low yield zone	<95 % in relation to medium	Low
Medium yield zone	95 % Medium yield 105 %	Medium
High yield zone	<105 % in relation for medium	High

Figure 4 shows the negative values of correlations between Al^{3+} with Ca^{2+} and Mg^{2+}, of -0.72 and -0.62, respectively, suggesting divergent trends of ions in the soil. These correlations are common in tropical acidic soils (Cahyono et al., 2020). The value of r = 0.67 between Ca^{2+} and Mg^{2+} suggests a medium and positive relationship between them. Exchangeable potassium had a similar relationship with Mg^{2+} and Na^{+}, with r = 0.54 and r = 0.58, respectively (Figure 4). Others properties have slow correlations, and the places in white color in figure 4 did not show a significant correlation at ($p<0.01$). The negative values of correlations between sand with clay and silt were -0.45 and -0.34, respectively. Gubiani et al. (2021) found correlation values similar to our results.

Factorial analysis

Factor analysis (FA) for soil properties showed that the first four factors are those whose eigenvalues are greater than 1, extracting cumulative explained variance of 79.71 % of the studied soil properties' variability as follows: 31.61, 19.05, 16.17 and 12.88 % for factors: F1, F2, F3 and F4, respectively (Table 3). This shows these four factors have potential to explain the 9 original soil properties. The results obtained here are in harmony with those from John et al. (2021), who found F1 to F5 could describe and explain approximately 78 % of the total variability of soil properties.

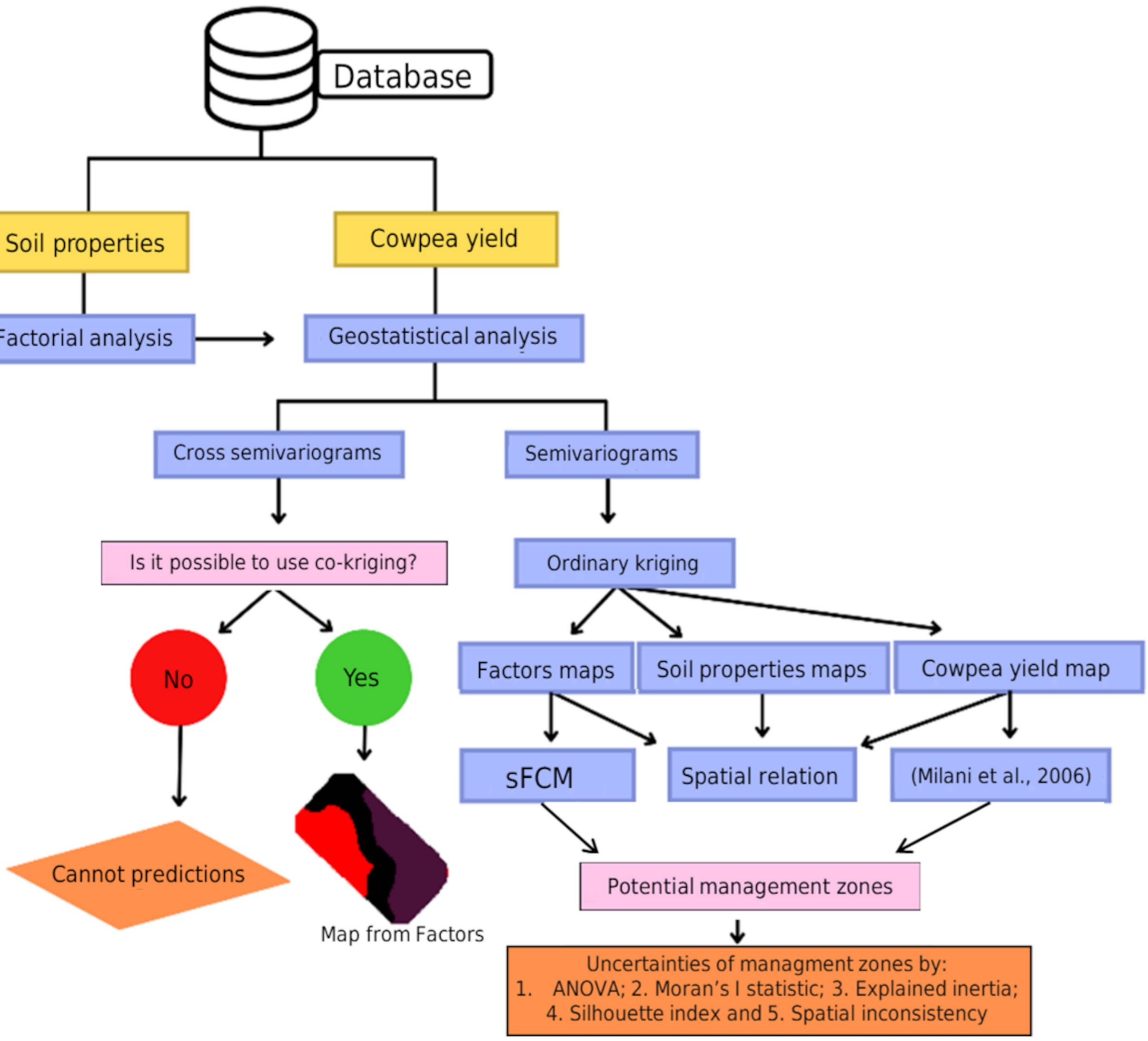

Figure 3. Flowchart showing the sequence of methodologies applied for delineation of potential management zones for cowpea by factorial kriging and Spatial Fuzzy c-Means (SFCM) of yield and soil chemical and texture properties.

Table 2. Descriptive statistics of soil properties and cowpea yield (n = 66)

Properties	Min	Max	Mean	SD	CV	Sk	Kurtosis	p-value
Cy (kg ha^{-1})	10.76	110.19	64.97	20.36	31.34	-0.20	0.42	>0.100*
P (mg dm^{-3})	0.01	0.09	0.04	0.02	38.59	0.56	-0.50	0.040
K^+ (cmol$_c$ dm^{-3})	0.00	0.17	0.07	0.04	55.33	0.29	-0.01	>0.100*
Na^+ (cmol$_c$ dm^{-3})	0.00	0.02	0.01	0.01	45.57	-0.56	0.04	>0.100*
Al^{3+} (cmol$_c$ dm^{-3})	0.38	1.19	0.87	0.18	21.22	-0.62	0.11	0.098*
Ca^{2+} (cmol$_c$ dm^{-3})	0.08	1.15	0.35	0.22	62.50	1.68	2.92	<0.010
Mg^{2+} (cmol$_c$ dm^{-3})	0.07	0.60	0.19	0.10	51.96	1.51	3.58	<0.010
Sand (g kg^{-1})	505.00	892.00	835.29	48.62	5.82	-5.00	33.11	<0.010
Silt (g kg^{-1})	8.00	171.00	49.56	21.39	43.16	2.90	15.50	<0.010
Clay (g kg^{-1})	60.00	160.00	110.00	22.80	20.73	0.19	-0.21	>0.100*

Min: Minimun; Max: Maximum; Cy: Cowpea yield; SD: standard deviation; CV: coefficient of variation; Sk: Skewness; p-value: * normal distribution by the Shapiro-Wilk test when detected p-value>0.05.

Table 3 shows that Ca^{2+}, Mg^{2+} and Al^{3+} were strong correlated with F1 with (r = -0.78; r = -0.81, and r = 0.85), respectively. On the other hand, K^+ and Na^+ strongly correlated with F2, with r = 0.90 and r = 0.72, respectively. While F3 was correlated with silt (r = -0.71), and F4 was correlated with clay (r = -0.75). Sánchez-Navarro et al. (2021) reported a positive correlation between Ca^{2+}, Mg^{2+}, P and K^+ and F1 in their study of nutritional diagnosis for cultivation of *Vigna unguiculata*.

Factor F1 can explain the effect these nutrients as factors limiting the crop yield in relationship with liming practices that provide exchangeable ions Ca^{2+} and Mg^{2+}, in addition to the antagonistic and harmful effect of Al^{3+} (Table 3). These cations contributions to F1 can be explained due the intensive application of synthetic fertilizers, which provides additional spatial variability of soil chemical properties. Cai et al. (2019) reported the increased addition of synthetic fertilizers can explain soil acidity.

The F2 also had chemical properties (cations: K^+ and Na^+) similar to those observed for F1. Moreover, our findings showed that soil chemical properties can explain 50.66 % of data variance (F1 + F2). Meanwhile, soil texture (Silte and Clay) explains 29.05 % of

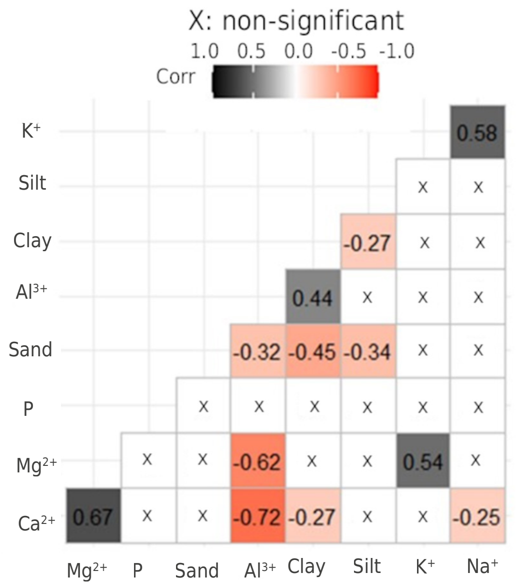

Figure 4. Correlogram based on Pearson correlation (r) between soil properties. Black and red colors symbolize positive and negative correlations, respectively.

Table 3. Correlation coefficients between soil properties and each of the factors scores

	Factor 1	Factor 2	Factor 3	Factor 4
Eigenvalues	2.84	1.71	1.45	1.15
Cumulative variance (%)	31.61	50.66	66.83	79.71
Explained variance (%)	31.61	19.05	16.17	12.88
Correlations				
P	-0.19	-0.07	0.58	-0.12
K^+	-0.50	0.72	-0.27	-0.08
Na^+	-0.12	0.90	-0.13	0.10
Al^{3+}	0.85	0.28	-0.01	0.03
Ca^{2+}	-0.78	-0.42	-0.15	-0.23
Mg^{2+}	-0.81	0.08	-0.20	-0.42
Sand	-0.44	0.16	0.66	0.31
Silt	-0.07	-0.29	-0.71	0.49
Clay	0.57	0.03	-0.12	-0.75
	$Al^{3+} + Ca^{2+} + Mg^{2+}$	$K^+ + Na^+$	Silt	Clay
	Chemical properties		Soil texture	

data variance (F3+F4). Factorial analysis reduces the dimensionality of multiple maps from soil chemical properties. Oldoni et al. (2019) also found a reduction in the variability of soil using geostatistical and multivariate analysis. Our findings support our initial hypothesis, demonstrating the potential utility of interpreting spatial variability in the nutrient index for cowpea production. This aligns with the study of Anago et al. (2023), who highlighted the promise of similar approaches in African contexts.

Geostatistical analysis

The modeling of the cross semivariograms indicated a significant and positive spatial correlation between Ca^{2+} and Mg^{2+} with a negative correlation between Al^{3+} and F1 (Figures 5b, 5d, and 5h). The cross semivariograms indicated a positive spatial correlation between K^+ and F2 (Figure 5j). The cross semivariograms indicated negative spatial correlation between silt and F3 (Figure 5o), and this result indicated that F1, F2, and F3 have the potential to be a useful auxiliary variable for the co-kriging method to improve the forecast accuracy of these soil properties and cowpea yield. Arreño et al. (2017) reported the cokriging technique is superior to kriging technique in spatial prediction, given the use of multivariate spatial data.

Soil Ca^{2+} semivariogram provides a description of its spatial dependence and provides some insights into possible processes that affect its spatial distribution. A spherical model fits well the experimental semivariogram, with a high coefficient of determination ($R^2 = 0.90$), a low $C_0 / (C_0 + C_1)$ ratio of 10.33 %, and an effective range of 70 m. The spatial modeling shown in figures 5a and 5b, in which the Ca^{2+} cross semivariogram with F1, was also well adjusted by an exponential model, with a high coefficient of determination ($R^2 = 0.94$), low $C_0 / (C_0 + C_1)$ ratio of 10.32 % and effective range of 60 m (Figure 5b).

For Mg^{2+}, there was a high coefficient of determination ($R^2 = 0.90$), an average $C_0 / (C_0 + C_1)$ ratio of 48.81 %, and an effective range of 55 m. The cross Mg^{2+} semivariogram with F1 was also well adjusted by a spherical model, with a high coefficient of determination ($R^2 = 0.81$), average $C_0 / (C_0 + C_1)$ ratio of 44.53 %, and effective range of 60 m (Figures 5c and 5d). For the sample spacing of this study, no spatial dependence structure could be found for Na^+, and no spatial dependence was found in the modeling of the cross semivariogram between sand and clay with factor 4. However, with factors 1, 2, and 3, kriged maps can be made with these variables that have higher factor loads using these first three factors as auxiliary variables due to the cross semivariograms

(Figures 5b, 5d, 5h, 5j, 5m and 5o). This indicates the potential to create kriged maps for factors 1, 2, and 3 by incorporating these correlated variables as auxiliary information. By leveraging the spatial relationships between these factors and other variables, we can create robust kriged maps for factors 1, 2, and 3, potentially leading to effective delineation of management zones.

Our results reveal a crucial finding: a spatial dependence structure exists between the most influential factors and the strongly correlated soil properties. This allows us to leverage just three maps of the first factor scores (instead of ten individual soil property maps) to effectively explain these properties spatial continuity. This not only simplifies the process of delineating management zones but also identifies the most coherent and practical approach for this task. Consequently, we can gain a deeper understanding of the spatial patterns in soil properties that are simultaneously linked to cowpea yield. The conclusion provided by Anago et al. (2023) strongly suggests that this ease interpretation of spatial variability is crucial in nutrient management in cowpea production.

Exchangeable Ca^{2+} and Mg^{2+} were the nutrients that had a direct negative and significant effect on yield (Figures 4a and 4c), indicating these elements would also be high in the soil in the area of low yield (in red), although the average content is within the range of 0.10 to 0.72 $cmol_c$ dm^{-3} and 0.11 to 0.29 $cmol_c$ dm^{-3} - considered as suitable for the crop (Malavolta, 1987).

The Ca^{2+} and Mg^{2+} show an antagonistic behavior for Al^{3+}. Although these ions are the cause of toxicity. In this study, cowpea was resistant to Al^{3+}, because in the zone with the great content, this ion had higher Cy values. According to Yang et al. (2013), the observed behavior can be explained by *Vigna unguiculata* tolerance to high concentrations of this metal, owing to its complexation with organic acids exuded by the root system and because of the plant genotype, which can promote the ability to adapt to adverse physical and chemical conditions, minimizing problems caused by low yield in acidic soils.

Therefore, the microrelief (that is, small variations in the relief) governs the flow of water, modifying the properties of the soil, e.g., texture and structure, which, in turn, govern pore distribution. Therefore, they dominate factors such as leaching. O'Geen (2013) clearly described, in detail, the influence of texture and structure on the relationship between soil moisture, soil water flow, and soil properties. In fact, although there was no evidence of similarity in the spatial continuity pattern between soil properties maps (Figures 6a, 6b, 6c, 6d, 6e, 6f, 6g, and 6h). This results in contradictory evidence to our hypothesis while proposing to delineate management zones. However, when these maps are integrated using factorial analysis and SFCM, this hypothesis is reasonable since it is possible to see the clustering of these maps in primaries MZs (Figures 6n and 6o). Figure 6n and 6o show the probable distribution of MZs using RGB stack maps from F1, F2 and F3, because showed in blue the zones that could be explained by F1, and in green the zones that could be explained by F2, and in red the zones that F3 could explain.

However, these results demonstrate the potential of FKA that may be directly applied to obtain MZs integrating three scopres maps from FKA representative of soil properties because there is possible to reduce the multiple maps with different rangers for one simple RGB stack maps multivariate. It not escaped our notice that is the first insight of potential tool to extract information's not obtained by geostatistics modelling using univariate data. Morari et al. (2009) working with multivariate geostatistics in delineating management zones, reported that multivariate geostatistics provides local information which may be useful for site-specific management.

There was a strong spatial pattern of F1 with Cy (Figures 6i and 6p); although it explains 28 %, it can be used instead of three maps of soil properties (e.g., Al^{3+}, Ca^{2+}, Mg^{2+}). This shows the potential of combining geostatistics with factor analysis, as this can help to interpret the spatial continuity of soil and plant attributes. This indicates that the spatial

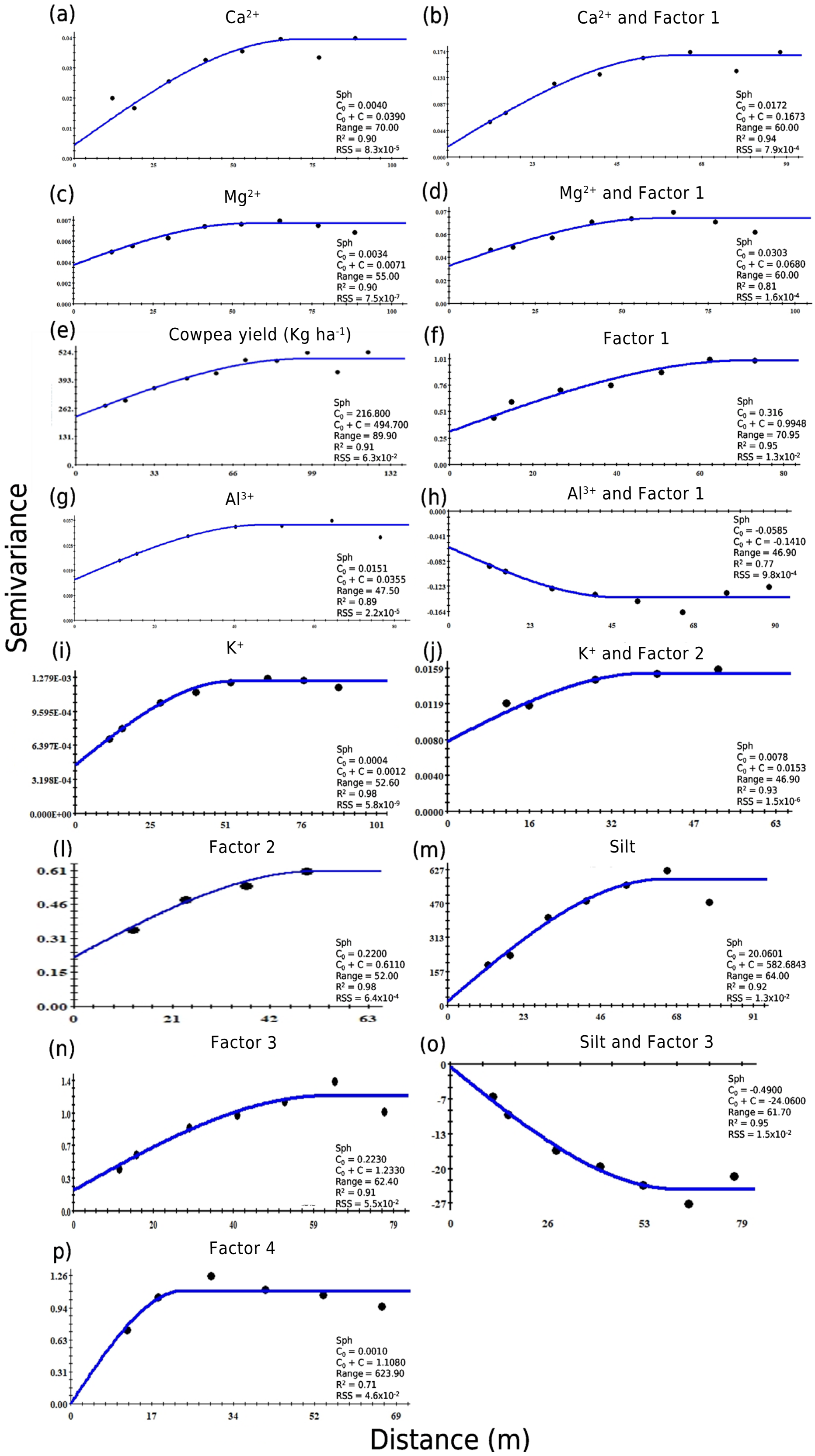

Figure 5. Direct and cross experimental semivariograms (black points) and semivariogram models (blue solid line) for soil and yield variables and first four principal components. Model: Sph – Spherical; C_0 - nugget effect; C - Contribution; C0 + C - Sill; RSS - residual sum of squares; R^2 - coefficient of determination of the cross validation. Direct semivariogram of Ca^{2+} (a); cross semivariograma between Ca^{2+} and Factor 1 (b); direct semivariogram of Mg^{2+} (c); cross semivariograma between Mg^{2+} and Factor 1 (d); direct semivariogram of Cowpea yield (e); direct semivariogram of Factor 1 (f); direct semivariogram of Al^{3+} (g); cross semivariograma between Al^+ and Factor 1 (h); direct semivariogram of K^+ (i); cross semivariograma between K^+ and Factor 2 (j); direct semivariogram of Factor 2 (l); direct semivariogram of silt (m); direct semivariogram of Factor 3 (n); cross semivariograma between Silt and Factor 3 (o); and direct semivariogram of Factor 4 (p).

variability of Factor 1 can be used to provide further insights into the causes and effects of high and low yield more accurately and into adequate site-specific management of cowpea fields. These results are consistent with the findings reported by Oliveira et al. (2019), i.e., the association of PCA and cluster analysis with the basic texture and chemical properties of the soil is perfectly feasible for the definition of management zones. However, the delineation of management zones based in factorial analysis confirm our hypothesis that MZs can be delineated by reduction the number of soil properties indicators from multivariate analysis.

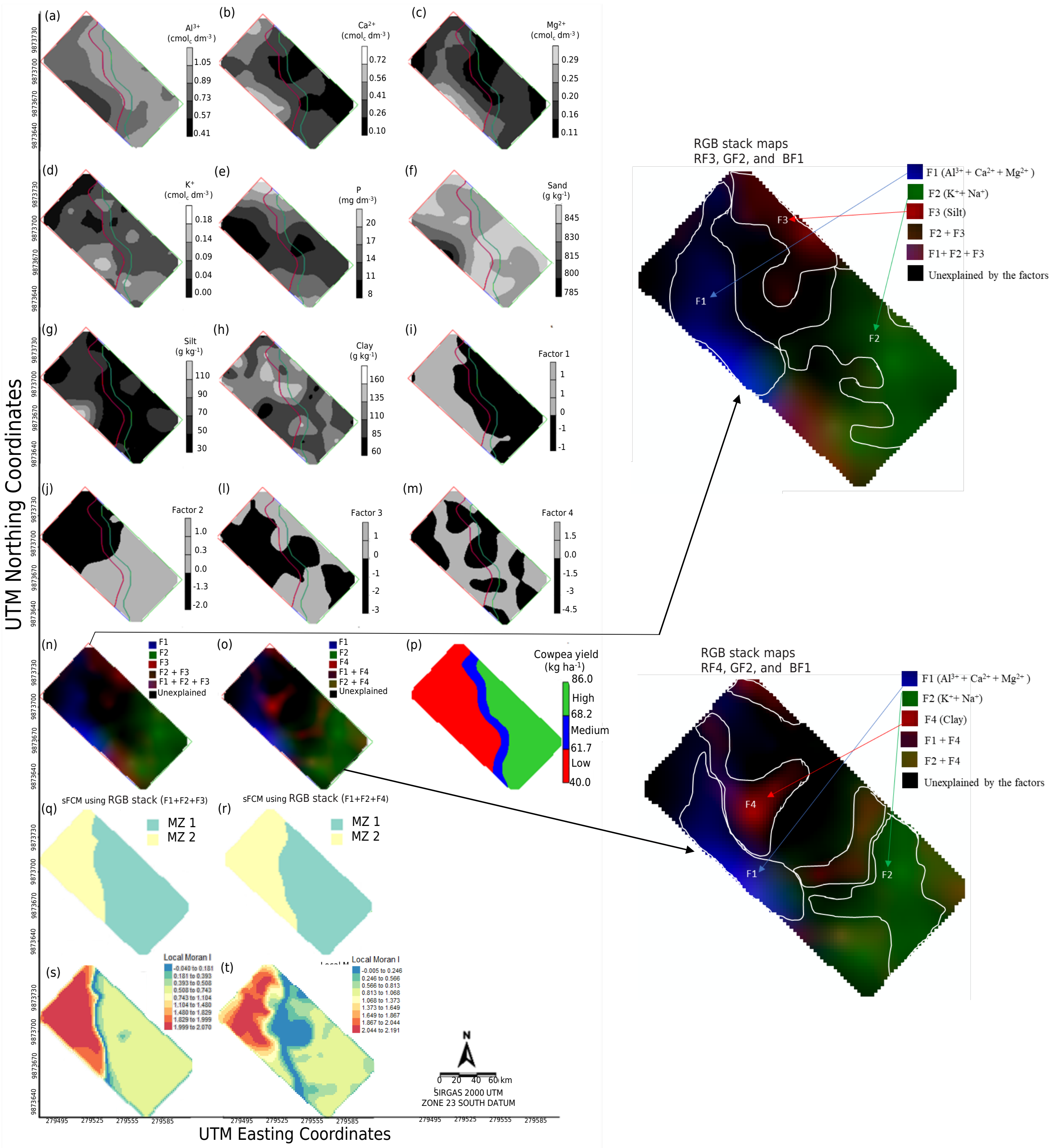

Figure 6. Predicted maps: Al^{3+}: aluminium (a); Ca^{2+}: calcium (b); Mg^{2+}: magnesium (c); K^{+}: potassium (d); P: phosphorus (e); sand (f); silt (g); clay (h); factor 1 (i); factor 2 (j); factor 3 (l); factor 4 (m); RGB stack from Factor 1,2 and 3 (n); RGB stack from Factor 1, 2 and 4 (o); management zones based on cowpea yield (p); management zones based on spatial fuzzy c-means from Factor 1, 2 and 3 (q); management zones based on spatial fuzzy c-means from Factor 1, 2 and 4 (r); uncertainties of management zones by results of local Moran index from Factor 1, 2 and 3 s); uncertainties of management zones by results of local Moran index from factor 1, 2 and 4 (t).

Potential of kriged maps of factor scores for estimation of cowpea yield

The potential of the factors to estimate cowpea yield was assessed on the basis of the correlation between the raster image, graphically represented by using a scatter plot with SAGA 2.3. There is a linear negative relationship between cowpea yield and F1 with $R^2 = 0.61$ (Figure 7a). It means the higher and the more negative the values of the scores of F1, the greater the cowpea yield. There is a linear positive relationship between cowpea yield and F2 with $R^2 = 0.48$ (Figure 7b). Our study found a weak linear correlation between cowpea yield and factor 3 with $R^2 = 0.17$ (Figure 7c). There is also no linear relationship between cowpea yield and factor 4 with $R^2 = 0.00$ (Figure 7d).

However, the most important factors for the prediction models were F1, F2 and F3. These results support our hypothesis that the performance factors for the prediction Cy is crucial for many practical applications and supports the use of factors to delineate MZs. Thus, this proves that the use of these factors is useful to understand the spatial continuity of yield as related to soil properties. Bevington et al. (2019) concluded that the principal component and FKA are efficient in regression model to predict dependent variables from independent variables.

In general, all the first three factors have a degree of spatial correlation with yield, indicating that the latter can be estimated, since these factors provide information on the soil properties responsible for the spatial variability of yield. However, these results confirm our hypothesis that the FKA associated with SFCM has a high potential to assist in the delineation of MZs. This is consistent with the results reported by Aliyu et al. (2020), who showed that the combination of geostatistics with principal component analyses and K-means cluster analyses was successful in the delineation of soil nutrient management zones.

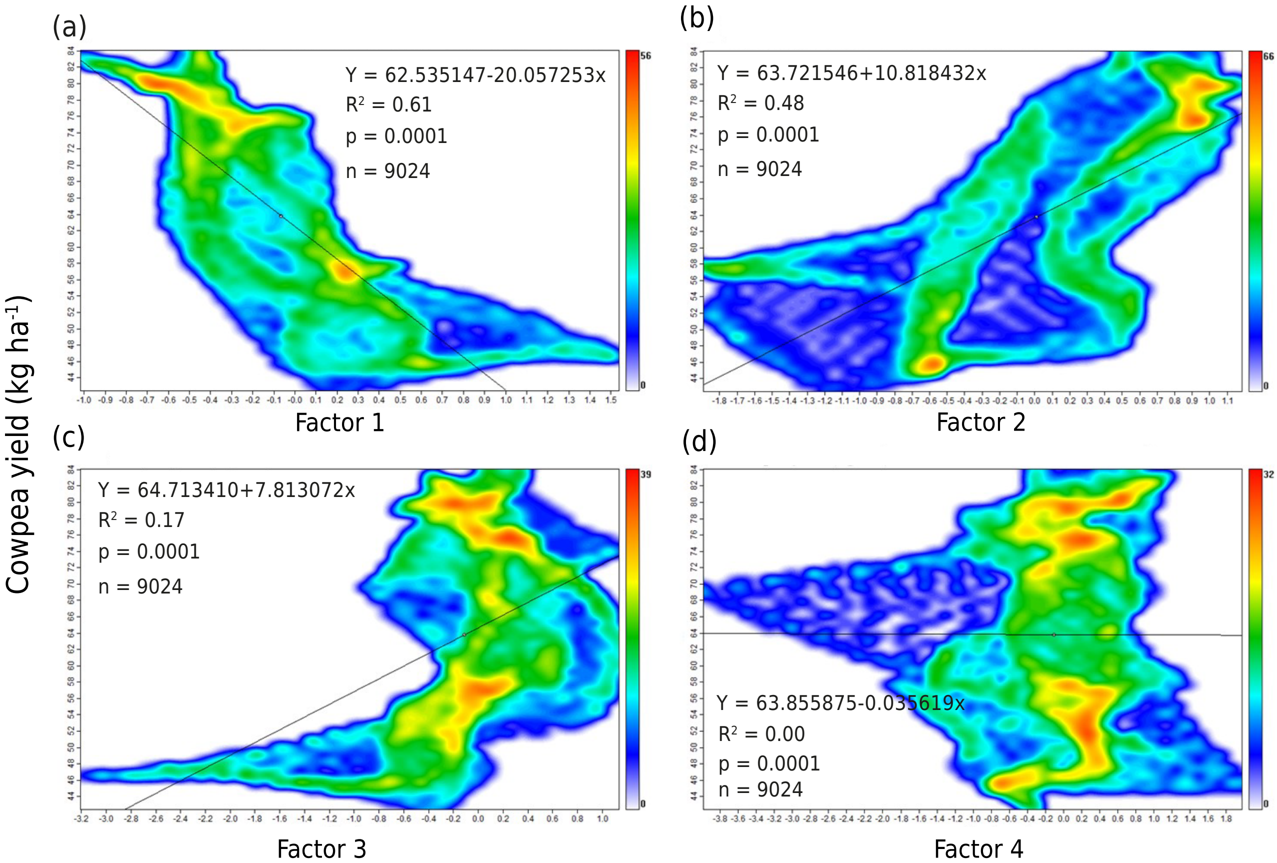

Figure 7. Scatter plot correlation between interpolated maps of cowpea yield with factors: 1, 2, 3 and 4. Correlation between interpolated cowpea yield and factor 1 (a); correlation between interpolated Cowpea yield and factor 2 (b); correlation between interpolated Cowpea yield and factor 3 (c); correlation between interpolated Cowpea yield and factor 4 (d).

Evaluation of management zones by cowpea yield map

The potential management zones (MZs) identified from the cowpea yield map were significantly different, as determined by ANOVA and Tukey's tests. These MZs were categorized as low, medium, and high potential zones (Figure 8). Our findings are similar to those of Schenatto et al. (2017), who delineated MZs for soybean yield using ANOVA. These MZs are valuable for assessing their relationship with soil-based MZs.

Evaluation of management zones by Spatial Fuzzy c-Means (SFCM)

Figures 6a to 6h demonstrate that soil property maps alone are insufficient for delineating management zones (MZs). This limitation is primarily due to the high variability among soil properties and the absence of consistent spatial patterns. These challenges become even more pronounced when multiple maps are integrated, especially at varying spatial scales, underscoring the complexity of accurately defining MZs. See in figures 6a to 6h that the eight soil property maps were reduced to four factors maps (figures 6i to 6m), and these were reduced to just two RGB stack maps, and finally, for the management zone map. The major drawbacks of these multiple kriged maps are different ranges result in distinctly different management zones (Peralta and Costa, 2013). When considering the four factor maps (Figures 6i to 6m), a reduction in dimensionality is evident. While some zones can be identified, precise interpretation for effective application in precision agriculture may be challenging. This suggests the need for an integrated approach, combining multiple factors into a single raster map, which we call the RGB stack. Evidence supporting this interpretation is improved by integrating factors maps, as can be seen in figures 6n and 6o, and such the datasets with nine soil properties maps were reduced for two maps. These results provide the possible explanation for when each factor had a major influence. In figure 6n and 6o, the places of blue color represent the potential of F1 to explain the effect and causes of Al^{3+}, Ca^{2+} and Mg^{2+}. The places of green color represent the potential of F2 to explain the effect and causes of K^+ and Na^+, which repeat for the other factors (Figures 6n and 6o enlarged).

Figures 6q and 6r show that RGB stack maps were classified into two zones by SFCM. This result is consistent because it has been the practical solution to seek a balance between the total variance and the number of zones instead of the best number of zones. Xiaohu et al. (2016) proposed a similar strategy for delineating the optimal number.

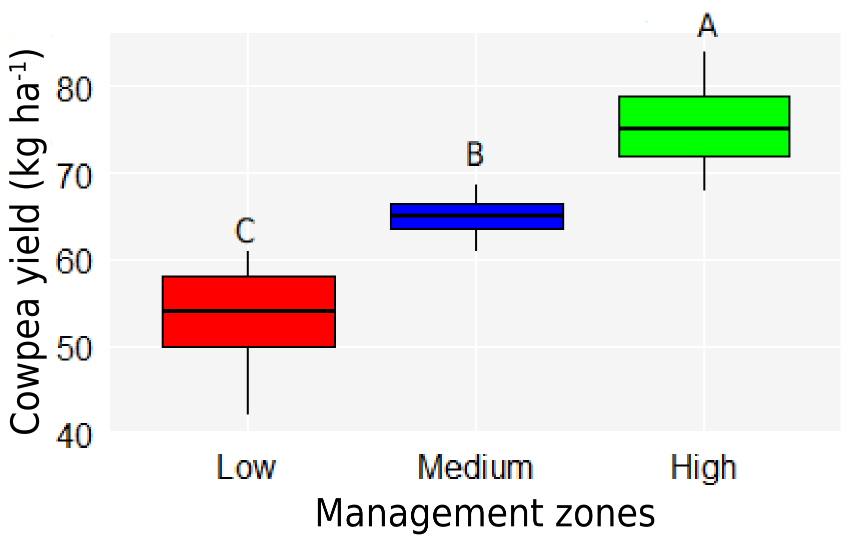

Figure 8. Box-plot and the results of one-way ANOVA analysis between the zones of cowpea yield. Management zones (MZs) with high productivity (a), management zones (MZs) with medium productivity (b) and management zones (MZs) with low productivity (c). Different letters represent a significant difference at $p<0.05$.

Table 4. Spatial diagnostic for the optimal number of clusters (MZs) with the Moran I index, explained inertia, silhouette index and Spatial inconsistency applied in RGB stack maps from Factors: Spatial Fuzzy c-Means (SFCM), MZs: Management Zones

Cluster Performance Index	SFCM	Intepretation	Source	Factors	Numbers of MZs
Moran I index (I)	0.97	indicated close to perfect clustering	Anselin (1995)	F123	2
Explained inertia (EI)	0.39	indicated a medium clustering	Rousseeuw (1987		
Silhouette index (SI)	0.65	indicated clusters well defined	Rousseeuw (1987)		
Spatial inconsistency (SC)	0.08	indicated strong consistency	Zhang and Wu (2004)		
Moran I index	0.97	indicated close to perfect clustering	Anselin (1995)	F124	2
Explained inertia	0.28	indicated a medium clustering	Rousseeuw (1987		
Silhouette index	0.59	indicated clusters well defined	Rousseeuw (1987)		
Spatial inconsistency	0.06	indicated strong consistency	Zhang and Wu (2004)		

Figures 6s and 6t show the uncertainty between zones by Moran's I statistic values. Places with negative values in color blue represent the pixels with errors that decrease the local accuracy of SFCM. This observation is consistent with the MZs defined by Cy; there is a greater variability due to these places being transition zones. Anselmi et al. (2021) described that the variance between MZs can be explained by variance in soil properties.

Table 4 presents the performance of the Spatial Fuzzy c-Means (SFCM) method in delineating management zones (MZs). The results indicate that the optimal clustering occurs when k = 2, as evaluated using multiple statistical and spatial assessment tools. For the RGB stack derived from factors 1, 2, and 3, the following metrics were obtained: Moran's I index (I = 0.97), explained inertia (EI = 0.39), silhouette index (SI = 0.65), and spatial consistency (SC = 0.08). These results confirm the ability of the sFCM method to define two distinct MZs based on this specific RGB configuration.

Similarly, when using the RGB stack derived from factors 1, 2, and 4, two MZs were also identified. The performance metrics for this configuration were comparable, with Moran's I index (I = 0.97), explained inertia (EI = 0.28), silhouette index (SI = 0.59), and spatial consistency (SC = 0.06). These values highlight slight differences in spatial variability and clustering quality between the two RGB stacks, reflecting the influence of the selected input factors on the clustering process.

Overall, the RGB stack derived from factors 1, 2, and 3 demonstrated slightly higher clustering quality (based on the silhouette index and spatial consistency) compared to the stack from factors 1, 2, and 4. However, the Moran's I index remained consistent (I = 0.97) across both configurations, indicating similar spatial autocorrelation in the clustering results. These findings underscore the robustness of the SFCM method in delineating MZs and emphasize the importance of selecting appropriate input factors to enhance clustering performance. However, the better cluster numbers (MZs) were two, as indicated by high Silhouette index values, slow Explained inertia and Spatial inconsistency values. However, we obtained better classification performance in two zones (Table 4, and Figures 6q and 6r). Therefore, the farmer needs to be more assertive about the cluster number and spatial uncertainty to obtain satisfactory MZs for adopting technologies for sustainable farming systems.

CONCLUSIONS

This study demonstrated the effectiveness of integrating Spatial Fuzzy c-Means (SFCM) with factorial kriging analysis to delineate management zones based on soil texture and chemical properties in the northeastern region of Pará State, Brazil using soil samples collected from a layer of 0.00-0.20 m, representing a typical layer for agricultural practices in the region. This approach effectively captures the intricate relationships between these properties and cowpea yield, enabling the identification of zones with distinct characteristics suitable for tailored management practices. By translating complex soil information into an interpretable format, this methodology can facilitate decision-making for optimizing cowpea yield, particularly for resource-limited smallholder farmers in Brazil.

Using factor scores in cokriging offers a valuable tool for precision agriculture, enabling more accurate prediction of soil properties and cowpea yield. This approach can guide informed decision-making for optimizing fertilizer applications, irrigation strategies, and other management practices.

Our study highlights the importance of management zone delineation for precision agriculture applications in Brazilian cowpea farms. The high precision achieved through our approach demonstrates the value of this strategy. For cowpea production areas with historically high soil property variability due to fertilizer application, our findings suggest methodologies utilizing multivariate spatial variability indices (regionalized factors) combined with factor kriging analysis are particularly beneficial.

Our multivariate approach offers distinct advantages for precision agriculture in cowpea farms. It overcomes the challenges associated with managing numerous soil property maps and simplifies interpretation through the combined application of multivariate factor analysis, kriging, and the SFCM grouping technique. This comprehensive approach facilitates the delineation of management zones, promoting environmental sustainability and economic competitiveness in cowpea agriculture.

Future research is necessary for a more robust evaluation of the effectiveness of this approach across diverse cowpea farming systems. This evaluation should include measurements of soil bulk density, macro-, and microporosity, particularly in the subsurface, and data collected using electrical conductivity and magnetic susceptibility sensors – tools commonly employed in precision agriculture. However, this study represents a significant first step towards implementing innovative precision agriculture practices for cowpea farms in Brazil.

DATA AVAILABILITY

The study data are not publicly available. Readers seeking further clarification about the research may contact the corresponding author.

FUNDING

Brazilian National Council for Scientific and Technological Development (CNPq) granted the scholarship to the second and third authors.

ACKNOWLEDGMENTS

The authors would like to thank Agronomist Engineer Benedito Dutra Luz de Souza from Agropecuária Milênio for seeds and field support and to Geotechnologies and Pedometrics Research Group (GEOP, UFRA; < https://geopufra.com/en/ >) for team support.

AUTHOR CONTRIBUTIONS

Conceptualization: Devid Jackson da Silva Sousa (equal), João Fernandes da Silva Júnior (equal), and Thiago Thomé da Silva (equal).

Formal analysis: João Fernandes da Silva Júnior (lead).

Investigation: Devid Jackson da Silva Sousa (lead).

Methodology: Devid Jackson da Silva Sousa (equal), João Fernandes da Silva Júnior (equal) and Thiago Thomé da Silva (equal).

Software: Thiago Thomé da Silva (lead).

Supervision: João Fernandes da Silva Júnior (lead).

Writing – original draft: Devid Jackson da Silva Sousa (equal), João Fernandes da Silva Júnior (equal), and Thiago Thomé da Silva (equal).

Writing – review & editing: Benedito Dutra Luz de Souza (equal), Daniel Pereira Pinheiro (equal), João Fernandes da Silva Júnior (equal) and Rose Luiza Moraes Tavares (equal).

REFERENCES

Abdelaal SMS, Moussa KF, Ibrahim AH, Mohamed ES, Kucher DE, Savin I, Abdel-Fattah MK. Mapping spatial management zones of salt-affected soils in arid region: A case study in the east of the Nile Delta, Egypt. Agronomy. 2021;11:2510. https://doi.org/10.3390/agronomy11122510

Abebe BK, Alemayehu MT. A review of the nutritional use of cowpea (*Vigna unguiculata* L. Walp) for human and animal diets. J Agric Food Res. 2022;10:100383. https://doi.org/10.1016/j.jafr.2022.100383

Aggelopooulou K, Castrignanò A, Gemtos T, De Benedetto D. Delineation of management zones in an apple orchard in Greece using a multivariate approach. Comput Electron Agr. 2013;90:119-30. https://doi.org/10.1016/j.compag.2012.09.009

Ahmad A, Ordoñez J, Cartujo P, Martos V. Remotely Piloted Aircraft (RPA) in Agriculture: A pursuit of sustainability. Agronomy. 2021;11:7. https://doi.org/10.3390/agronomy11010007

Aliyu KT, Kamara AY, Jibrin MJ, Huising JE, Shehu BM, Adewopo JB, Mohammed IB, Solomon R, Adam MA, Samndi AM. Delineation of soil fertility management zones for site-specific nutrient management in the maize belt region of Nigeria. Sustainability. 2020;12:9010. https://doi.org/10.3390/su12219010

Anago FN, Agbangba EC, Dagbenonbakin GD, Amadji LG. Continuous assessment of cowpea [*Vigna unguiculata* L. Walp.] nutritional status using diagnosis and recommendation integrated system approach. Sci Rep. 2023;13:14446. https://doi.org/10.1038/s41598-023-40146-0

Anselin L. Local indicators of spatial association—LISA. Geogr Anal. 1995;27:93-115. http://dx.doi.org/10.1111/j.1538-4632.1995.tb00338.x

Anselmi AA, Molin JP, Bazame HC, Corrêdo LP. Definition of optimal maize seeding rates based on the potential yield of management zones. Agriculture. 2021;11:911. https://doi.org/10.3390/agriculture11100911

Arreño G, Kim KK, Kang C, Choi S. A study on prediction comparison of kriging and cokriging using PCA. J Korean Stat Soc. 2017;19:1721-32. https://doi.org/10.37727/jkdas.2017.19.4.1721

Barros RLN, Oliveira LB, Magalhães WB, Médici LO, Pimentel C. Interaction of biological nitrogen fixation with sowing nitrogen fertilization on common bean in the two seasons of cultivation in Brazil. J Plant Nutr. 2018;6:774-81. https://doi.org/10.1080/01904167.2018.1426016

Bevington J, Scudiero E, Teatini P, Vellidis G, Morari F. Factorial kriging analysis leverages soil physical properties and exhaustive data to predict distinguished zones of hydraulic properties. Comput Electron Agr. 2019;156:426-38. https://doi.org/10.1016/j.compag.2018.11.034

Bezdek JC. Pattern recognition with fuzzy objective function algoritms. New York: Plenum Press; 1981. https://doi.org/10.1007/978-1-4757-0450-1

Brasil EC, Cravo MS, Viégas IJM. Amostragem de solo. In: Brasil EC, Cravo MS, Viégas IJM, editors. Recomendações de adubação e calagem para o Estado Pará. Belém: Embrapa Amazônia Oriental; 2020.

Buttafuoco G, Castrignanò A, Cucci G, Rinaldi M, Ruggieri S. An approach to delineate management zones in a durum wheat field: Validation using remote sensing and yield mapping. Precis Agric. 2015;15:241-8. https://doi.org/10.3920/978-90-8686-814-8_29

Cahyono P, Loekito S, Wiharso D, Afandi, Rahmat A, Komariah, Nishimura N, Senge M. Patterns of nutrient availability and exchangeable aluminum affected by compost and dolomite in red acid soils in Lampung, Indonesia. Int J Geomate. 2020;19:173-9. https://doi.org/10.21660/2020.76.87631

Cai A, Xu M, Wang B, Zhang W, Liang G, Hou E, Luo Y. Manure acts as a better fertilizer for increasing crop yields than synthetic fertilizer does by improving soil fertility. Soil Till Res. 2019;189:168-75. https://doi.org/10.1016/j.still.2018.12.022

Castrignanò A, Buttafuoco G, Quarto R, Vitti C, Langella G, Terribile F, Venezia A. A combined approach of sensor data fusion and multivariate geostatistics for delineation of homogeneous zones in an agricultural field. Sensors. 2017;17:2794. https://doi.org/10.3390/s17122794

Claessen MEC. Manual de métodos de análise de solo. 2. ed. Rio de Janeiro: Embrapa Solos; 1997.

Doerge TA. Management zone concepts. Canada: Potash & Phosphate Institute; 1999. Avaiable from: http://www.ipni.net/publication/ssmg.nsf/0/C0D052F04A53E0BF852579E500761AE3/$FILE/SSMG-02.pdf.

Dunn JC. A fuzzy relative of the ISODATA Process and its use in detecting compact well-separated clusters. J Cybern. 1973;3:32-57. https://doi.org/10.1080/01969727308546046

Ganascini D, Laureth JCU, Mendes IS, Tokura LK, Sutil EL, Villa B, Alovisi AMT, Caon IL, Mercante E, Coelho SRM. Analysis of the production chain of bean culture in Brazil. J Agric Sci. 2019;11:256-67. https://doi.org/10.5539/jas.v11n7p256

Gavioli A, Souza EG, Bazzi CL, Guedes LPC, Schenatto K. Optimization of management zone delineation by using spatial principal components. Comput Electron Agr. 2016;127:302-10. https://doi.org/10.1016/j.compag.2016.06.029

Georgi C, Spengler D, Itzerott S, Kleinschmit B. Automatic delineation algorithm for site-specific management zones based on satellite remote sensing data. Precis Agric. 2018;19:684-707. https://doi.org/10.1007/s11119-017-9549-y

Goovaerts P. Factorial kriging analysis: a useful tool for exploring the structure of multivariate spatial soil information. J Soil Sci. 1992;43:597-619. https://doi.org/10.1111/j.1365-2389.1992.tb00163.x

Gubiani PI, Almeida TA, Mulazzani RP, Pedron FA, Suzuki LEAS, Pereira CA. Shaking settings to reduce the breakdown of Entisol fragile particles in texture analysis. Rev Bras Cienc Solo. 2021;45:e0210066. https://doi.org/10.36783/18069657rbcs20210066

Guedes Filho O, Vieira SR, Chiba MK, Nagumo CH, Dechen SCF. Spatial and temporal variability of crop yield and some rhodic hapludox properties under no-tillage. Rev Bras Cienc Solo. 2010;34:1-14. https://doi.org/10.1590/S0100-06832010000100001

IUSS Working Group WRB. World Reference Base for Soil Resources. International soil classification system for naming soils and creating legends for soil maps. 4th edition. Vienna, Austria: International Union of Soil Sciences; 2022.

Javadi SH, Guerrero A, Mouazen AM. Clustering and smoothing pipeline for management zone delineation using proximal and remote sensing. Sensors. 2022;22:645. https://doi.org/10.3390/s22020645

Jérémy G, Apparicio P. Apport de la classification floue c-means spatiale en géographie: Essai de taxinomie socio-résidentielle et environnementale à Lyon. Cybergeo: Eur J Geogr. 2021;972:1-26. https://doi.org/10.4000/cybergeo.36414

John K, Afu SM, Isong IA, Aki EE, Kebonye NM, Ayito EO, Chapman PA, Eyong MO, Penížek V. Mapping soil properties with soil-environmental covariates using geostatistics and multivariate statistics. Int J Environ Sci Technol. 2021;18:3327-42. https://doi.org/10.1007/s13762-020-03089-x

Kaiser HF. The varimax criterion for analytic rotation in factor analysis. Psychometrika. 1958;23:187-200. https://doi.org/10.1007/BF02289233

Karydas C, Chatziantoniou M, Stamkopoulos K, Iatrou M, Vassiliadis V, Mourelatos S. Embedding a precision agriculture service into a farm management information system-ifarma/PreFer. Smart Agric Technol. 2023;4:100175. https://doi.org/10.1016/j.atech.2023.100175

Kazama EH, Silva RP, Tavares TO, Correa LN, Estevam FNL, Nicolau FEA, Maldonado Júnior W. Methodology for selective coffee harvesting in management zones of yield and maturation. Precis Agric. 2021;22:711-33. https://doi.org/10.1007/s11119-020-09751-1

Khan H, Farooque AA, Acharya B, Abbas F, Esau TJ, Zaman QU. Delineation of management zones for site-specific information about soil fertility characteristics through proximal sensing of potato fields. Agronomy. 2020;10:1854. https://doi.org/10.3390/agronomy10121854

Lin YP, Yeh MS, Deng DP, Wang YC. Geostatistical approaches and optimal additional sampling schemes for spatial patterns and future sampling of bird diversity. Glob Ecol Biogeogr. 2008;17:175-88. https://doi.org/10.1111/j.1466-8238.2007.00352.x

Lipiec J, Usowicz B. Spatial relationships among cereal yields and selected soil physical and chemical properties. Sci Total Environ. 2018;633:1579–90. https://doi.org/10.1016/j.scitotenv.2018.03.277

Liu CW, Lin KH, Kuo YM. Application of factor analysis in the assessment of roundwater quality in a blackfoot disease area in Taiwan. Sci Total Environ. 2003;313:77-89. https://doi.org/10.1016/S0048-9697(02)00683-6

Malavolta E. Nutrição mineral das plantas. Campinas: Fundação Cargill; 1987.

Matheron G. Pour une analyse krigeante des données regionalisées. Fontainebleau: Centre de Geostatistique; 1982. Avaiable from: https://cg.ensmp.fr/bibliotheque/public/MATHERON_Rapport_00233.pdf.

Matheron G. Principles of geostatistics. Econ Geol. 1963;58:1246-66. https://doi.org/10.2113/gsecongeo.58.8.1246

Milani L, Souza EG, Uribe-Opazo MA, Filho AG, Johann JA, Pereira JOA. Unidades de manejo a partir de dados de produtividade. Acta Sci Agron. 2006;28:591-8. https://doi.org/10.4025/actasciagron.v28i4.937

Morari F, Castrignanò A, Pagliarin C. Application of multivariate geostatistics in delineating management zones within a gravelly vineyard using geo-electrical sensors. Comput Electron Agr. 2009;68:97-107. https://doi.org/10.1016/j.compag.2009.05.003

O'Geen AT, Dahlgren RA, Swarowsky A, Tate KW, Lewis DJ, Singer MJ. Research connects soil hydrology and stream water chemistry in California oak woodlands. Calif Agric. 2010;64:78-84. https://doi.org/10.3733/ca.v064n02p78

Oldoni H, Terra VSS, Timm LC, Reisser Júnior C, Monteiro AB. Delineation of management zones in a peach orchard using multivariate and geostatistical analyses. Soil Till Res. 2019;191:1-10. https://doi.org/10.1016/j.still.2019.03.008

Oliveira JF, Mayi S, Marchão RL, Corazza EJ, Hurtado SC, Malaquias JV, Tavares Filho J, Brossard M, Guimarães MF. Spatial variability of the physical quality of soil from management zones. Precis Agric. 2019;20:1251-73. https://doi.org/10.1007/s11119-019-09639-9

Osipitan AO, Fields JS, Lo S, Cuvaca I. Production systems and prospects of cowpea (*Vigna unguiculata* (L.) Walp.) in the United States. Agronomy. 2021;11:2312. https://doi.org/10.3390/agronomy11112312

Peralta NR, Costa JL. Delineation of management zones with soil apparent electrical conductivity to improve nutrient management. Comput Electron Agr. 2013;99:218-26. https://doi.org/10.1016/j.compag.2013.09.014

R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: http://www.R-project.org/.

Robertson GP. GS+ GeoStatistics for the environmental sciences: GS+ user's guide. Plainwell: Gamma Design Software; 1998.

Rodrigues MS, Corá JE. Management zones using fuzzy clustering based on spatial-temporal variability of soil and corn yield. Eng Agric. 2015;35:470-83. https://doi.org/10.1590/1809-4430-Eng.Agric.v35n3p470-483/2015

Rossi R, Pollice A, Bitella G, Labella R, Bochicchio R, Amato M. Modelling the non-linear relationship between soil resistivity and alfalfa NDVI: A basis for management zone delineation. J Appl Geophy. 2018;159:146-56. https://doi.org/10.1016/j.jappgeo.2018.08.008

Rousseeuw P. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53-65. https://doi.org/10.1016/0377-0427(87)90125-7

Sanches GM, Paula MTN, Magalhães PSG, Duft DG, Vitti AC, Kolln OT, Borges BMMN, Franco HCJ. Precision production environments for sugarcane fields. Sci Agric. 2019;76:10-7. https://doi.org/10.1590/1678-992X-2017-0128

Sánchez-Navarro V, Zornoza R, Faz A, Fernández JAT. Cowpea crop response to mineral and organic fertilization in SE Spain. Processes. 2021;9:822. https://doi.org/10.3390/pr9050822

Santos HG, Jacomine PKT, Anjos LHC, Oliveira VA, Lumbreras JF, Coelho MR, Almeida JA, Araújo Filho JC, Oliveira JB, Cunha TJF. Sistema brasileiro de classificação de solos. 5. ed. rev. ampl. Brasília, DF: Embrapa; 2018.

Santos JPO, Bulhões LEL, Cartaxo PHA, Gonzaga KS, Freitas ABTM, Ribeiro JKN, Pereira MCS, Dias MS, Xavier MA, Dantas EA. Interannual variability of productive aspects of bean culture in a municipality in the Semi-arid region of Alagoas, Brazil. Sci Electron Arch. 2021;14:26-32. https://doi.org/10.36560/14120211204

Schenatto K, Souza EG, Bazzi CL, Gavioli A, Betzek NM, Beneduzzi HM. Normalization of data for delineating management zones. Comput Electron Agr. 2017;143:238-48. https://doi.org/10.1016/j.compag.2017.10.017

Silva AV, Silva Filho JF, Silva MCT, Vaz NCA, Silva MLG. Edaphoclimatic aptitude and agricultural production environments of the bean culture. Sci Electron Arch. 2020;13:102-12. https://doi.org/10.36560/131020201114

Suszek G, Souza EG, Uribe-Opazo MA, Nobrega LHP. Determination of management zones from normalized and standardized equivalent productivity maps in the soybean culture. Eng Agric. 2011;31:895-905. https://doi.org/10.1590/S0100-69162011000500007

Umbelino AS, Oliveira DG, Martins MPO, Reis EF. Definições de zona de manejo para soja de alta produtividade. Rev Cienc Agrar. 2018;41:674-82. https://doi.org/10.19084/RCA18092

Wackernagel H. Multivariate geostatistics: An introduction with applications. 3rd ed. Berlin, Heidelberg: Springer-Verlag; 2003.

Warrick AW, Nielsen DR. Spatial variability of soil physical properties in the field. In: Hillel D, editor. Applications of soil physics. New York: Academic; 1980. p. 319-44.

Webster R. Quantitative spatial analysis of soil in the field. Adv Soil Sci. 1985;3:1-70. https://doi.org/10.1007/978-1-4612-5090-6_1

Xiaohu Z, Li J, Xiaolei Q, Jianxiu Q, Juan W, Yan Z. An improved method of delineating rectangular management zones using a semivariogram-based technique. Comput Electron Agr. 2016;121:74-83. https://doi.org/10.1016/j.compag.2015.11.016

Yang C. Remote sensing and precision agriculture technologies for crop disease detection and management with a practical application example. Engineering. 2020;6:528-32. https://doi.org/10.1016/j.eng.2019.10.015

Yang LT, Qi YP, Jiang HX, Chen LS. Roles of organic acid anion secretion in aluminium tolerance of higher plants. Biomed Res Int. 2013;2013:173682. https://doi.org/10.1155/2013/173682