Australian Journal of Crop Science AJCS

AJCS 16(03):394-401 (2022) ISSN:1835-2707
doi: 10.21475/ajcs.22.16.03.p3448

Assessment of spatial variability of soil fertility index in a Dystrofic Yellow Latosol

Antônio Lopes do Bonfim Neto[1], Francisco de Assis Oliveira[2], Eduardo Cezar Medeiros Saldanha[3], Pedro Silvestre Campos[4] and Mario Luiz Ribeiro Mesquita[1*]

[1]**Universidade Estadual do Maranhão, Campus Bacabal – MA, Brasil**
[2]**Universidade Federal Rural da Amazonia, Instituto de Ciências Agrárias, Belém –PA Brasil**
[3]**Eduardo Cezar Medeiros Saldanha, Desenvolvimento Técnico Yara, Brasil**
[4]**Universidade Estadual do Maranhão, Campus Bacabal e Programa de Pós-Graduação em Agricultura e Ambiente, Brasil**

***Corresponding author: mario-mesquita51@hotmail.com**

Abstract

The lack of knowledge about the variability of the soil fertility index often leads to the use of inadequate amounts of fertilizers for certain areas within the crop. Aiming to analyze a set of variables from the soil chemical analysis, the multivariate analysis was applied in order to condense these variables into a smaller group of factors without important information loss to create a Soil Fertility Index. Soil sampling was carried out in a total area of approximately 740 ha, with a sampling grid of 5 ha, making a total of 148 sampling points. Four factors: (1) exchangeable aluminum (Al), aluminum saturation (m%), base saturation (V%), calcium, pH, Mg and P. Al and m% (2) the CTC and OM variables (3) Fe and Mn variables and (4) B, S and Zn were extracted from the factorial analysis explaining 74.79% of the total variance of the data, which is satisfactory by the criterion of the variance percentage. The analysis showed that in the sample dataset, an intersection was occurred in the mean dataset forming three centroids which coincides with the number of properties where the set of samples was collected. The use of multivariate analysis proved to be efficient for the proposed study, since the analysis of variance could not show efficiency due to the interrelations between the variables causing bias in the results. Based on the universe analyzes studied here, approximately 97% of the sampled area presented satisfactory or high soil fertility levels, which leads to the use of reduced amounts of fertilizers in most of the growing area.

Keywords: Soil fertility index; factor analysis; canonical discriminant analysis.
Abbreviations: SFI_soil fertility index; FA_factor analysis; CDA_canonical discriminant analysis; KMO_Kaiser-Meyer-Olkin test.

Introduction

The activity of producing grains of rice, corn and soybeans in the municipality of Santarém, state of Pará in the Eastern Amazon, is carried out in production units that use agricultural mechanization, chemical fertilizers, and lime to soil acidity and pesticides. The use of no-tillage as well as precision agriculture is still in its infancy stage (Teixeira et al., 2019)

The depletion of soil nutrients for several decades of cultivation, without replacement, as well as the exploration of new areas with low fertility soils, makes Brazilian agriculture increasingly dependent on massive application of fertilizers (Castro et al., 2020).

In most cases, the variability of arable soils is not considered by farmers, who choose to apply homogeneous soil management practices, especially chemical fertilization for the total growing area, compelling farmer to use inadequate amounts of fertilizers for certain areas within the cropping fields.

The excessive use of fertilizers, generally above the real needs, in cropping fields increases production costs and is a potential source of pollution of surface and groundwater over the years (Ahmed et al., 2017; Faulcon et al., 2019).

Saldanha et al. (2013) identified high variability of chemical attributes in Western Amazon soils, due to variations in the level of agricultural practices applied by farmers, particularly, the amount of fertilizer used and the periods of application of lime to soil acidity.

According to Silva et al. (2015), the soil is the result of the interaction of geological, topographical and climatic factors, among others, which together confer their own chemical and physical characteristics and properties.

Soil formation factors act naturally and vary, according to management, ranging from a few square meters to thousands of hectares, making the chemical attributes of the soil not randomly organized, but rather with some spatial structure (Faulcon et al., 2019)

Therefore, if homogeneous management practices induce spatial variability of soil chemical variables then controlling these variables can be decisive to improve soil fertility and consequently to increase crop productivity. This could promote a better balance of soil quality indicator levels.

Thus, it is necessary to investigate if homogeneous soil fertility management practices optimize agricultural production, recurrent practices in conventional agriculture.
Simultaneous analysis of soil chemical variables measured on one experiment or sampling unit can be made by multivariate statistical methods. These methods allow for a reduction in the dimension of analyzes with soil variable multiple responses in order to simplify their understanding, visualization and interpretation, aside from obtaining sufficient details for an adequate representation of the results (Yeater et al., 2015).
Assessment of soil variability using multivariate statistical analyses was carried out by several authors with satisfactory results, including (Freitas et al., 2014; Vasu et al., 2017; Bhunia et al., 2018; Carvalho et al., 2018; Ousmenku et al., 2018; Hou et al., 2021). However, there are no reports on application of multivariate statistical methods to assess soil variability in the eastern Amazon.
Therefore, if homogeneous management practices induce the spatial variability of chemical attributes in the soil then controlling these variables can be decisive for the improvement of soil fertility and consequent increase in the crop productivity due to a better balance of soil quality indicator levels.
In this context, the objective of this study was to assess variability of soil chemical attributes by means of multivariate analysis to condense variables into a smaller group of factors without significant loss of information, aiming to create a Soil Fertility Index that will be used as a decision tool for a more rational soil fertilization of the soils. The resulting groups of variables will be used for discriminant analysis to confirm the existence of statistical difference between the variables in each group.

Results and discussion

Factor analyses

The adequacy of the factor analysis was determined by the KMO and Bartlett tests. The KMO test (0.658), indicated that the variables are correlated and the factorial model presented a good level of adequacy to the data (Table 1). Values of this test below 0.50 are unacceptable (Watson, 2017). In the Bartlett test, the overall significance of the correlation matrix was evaluated, which presented a statistic of 2392.692, indicating that the correlations are significant at the 1% level of probability, that is, the correlation matrix is not identity (Table 1).
In Table 2, the first three columns are the results for the five extracted factors, which are the factor loadings for each variable in each factor. The fourth column provides the statistic, detailing the degree to which each variable is "explained" by the four components, called commonality. Of the last two columns, the first is the sum of the column of squared factor loadings (eigenvalues) and indicates the relative importance of each factor in explaining the variance associated with the set of analyzed variables. The sums of the five factors are 5.021, 2.391, 1.657, 1.470 and 1.428, respectively. As expected, the factor solution extracted the factors in the order of their importance, with factor 1 explaining the largest portion of the variance (31.38%), factor 2 explaining 14.94%, factor 3 explaining 10.35%, factor 4 explaining 9.185% and factor 5 explaining 8.92%. The four factors explain 74.79% of the total variance of the data, which is satisfactory by the criterion of percentage of variance.
Based on the work of Santana (2007), the total portion of the variance explained in this study, by the factorial solution (11,967) can be compared with the total variation of the set of variables, which is represented by the trace of the factorial matrix.
The trace is the total variance to be explained, obtained by the sum of the eigenvalues of the set of variables (sum of the total eigenvalues, first column of Table 2, given that each variable has a possible eigenvalue equal to 1.0.
The total sum of the eigenvalues trace percentages extracted for the factorial solution serves as an index to determine the degree of adequacy of the factorial solution in relation to what all the variables represent. The index for this solution shows that 74.796% of the total variance is represented by the information contained in the factorial matrix of the solution in terms of the four factors. The index is considered high, and the variables are closely related to each other as expected (Table 2)
The sum of the factor loadings of the squared factors generates the commonality (Table 3). The commonality size is a useful index to assess how much of the variance in a given variable is explained by the factor solution. Large commonalities indicate that a large portion of the variance in a variable was extracted by the factor solution. A small commonality, lower than 0.50, shows that a good part of the variance contained in a variable is not explained by the factors (Santana, 2007).
The selection of significant variables that must be part of a factor is chosen based on the magnitude of the factor loading (Table 3). Thus, they can be chosen by looking from left to right along each line and selecting the highest valued loads. Adopting this process, factor 1 has seven significant loads; factor 2, two significant loads, factor 3, two significant loads, factor 4, three significant loads, and factor 5, two.
The first factor includes the variables exchangeable aluminum (Al), aluminum saturation (m%), base saturation (V%), calcium, pH, Mg and P. Al and m%, with negative values, are antagonistic to other variables of the factor, thus being coherent with the knowledge of soil studies. Thus, to maintain a satisfactory level of nutrition for the plants, it is necessary to reduce the levels of Al and m%. This factor can be called the "Base saturation" dimension, since the factors influenced by these variables are related to an increase in electrical charges in the soil solution.
The second factor included the CTC and OM variables. These variables are related to the cation exchange capacity, where organic matter plays a fundamental role in the adsorption of nutritive elements by electrical charges and cation exchanges. The combination of these variables improves soil quality, which is sought after in soybean crops using no-till. This dimension of conduct is linked to the provision of services, which names the factor as "Cationic exchanges".
The third factor was only composed of Fe and Mn variables. The factors showed antagonistic signs, with Fe showing a negative sign. Although these elements have opposite electrical charges, there are no reports that prove competition for coupling sites, although Fe presents toxicity to plants at high levels in the soil. This factor can be represented with the "Iron Toxicity" dimension.
The fourth factor, represented by B, S and Zn, showed a negative sign for the first two variables, although there are

Table 1. Kaiser-Meyer-Olkin (KMO) and Bartlett tests for the Soil Fertility Index (SFI).

Kaiser-Meyer-Olkin measure of sampling adequacy.		0.658
Bartlett's sphericity test	Chi-square aprox.	2392.692
	Df	120
	Sig.	0.000

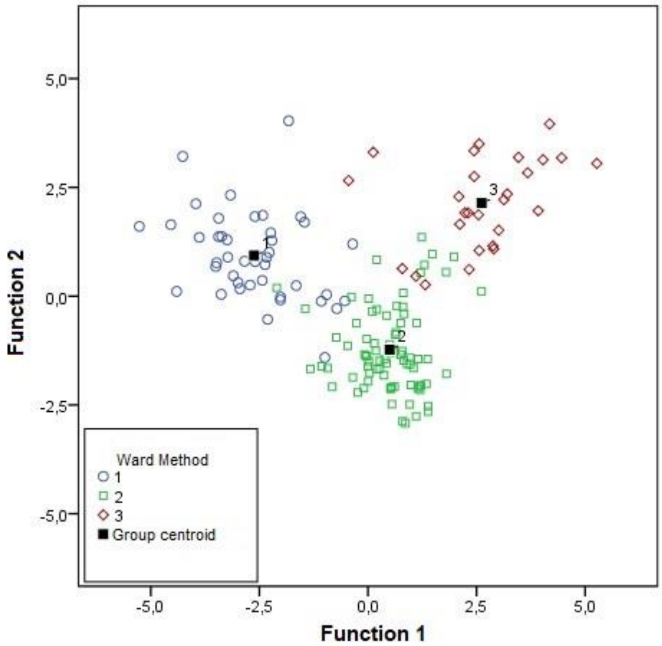

Fig 1. Canonical discriminant function of the sample points of the three groups with their respective centroids showing the intersection in the cloud of fertility data of three rural properties.

Table 2. Results of the eigenvalues for the extraction of component factors and total variance explained by the factors for the Soil Fertility Index (SFI)

Component	Eigenvalues (λ) and initial variances			Variances after rotation		
	Total	% of variation	% cumulative	Total	% of variation	% cumulative
1	5.735	35.845	35.845	5.021	31.383	31.383
2	1.972	12.328	48.173	2.391	14.944	46.327
3	1.645	10.279	58.452	1.657	10.355	56.682
4	1.379	8.617	67.069	1.470	9.185	65.868
5	1.236	7.726	74.796	1.428	8.928	74.796
6	0.945	5.909	80.704			
7	0.798	4.987	85.691			
8	0.584	3.648	89.339			
9	0.512	3.203	92.542			
10	0.467	2.919	95.461			
11	0.367	2.297	97.757			
12	0.212	1.324	99.081			
13	0.095	0.596	99.677			
14	0.032	0.200	99.877			
15	0.017	0.103	99.980			
16	0.003	0.020	100,000			

Table 3. Factor loading matrix (α) after orthogonal rotation by the Varimax method for the Soil Fertility Index (SFI).

Variables	Factor					Commonality
	F1	F2	F3	F4	F5	
Al	-0.939	-.034	-.029	-.022	.057	.887
m%	-0.937	-.080	-.046	.049	.065	.893
V%	0.934	.103	.102	.188	.032	.929
Ca	0.839	.389	.123	.228	.011	.922
Ph	0.799	.056	.008	.478	.098	.880
Mg	0.744	.335	-.030	-.258	-.102	.744
P	0.551	-.126	.291	.182	.421	.615
CTC	.197	0.947	.047	-.023	.005	.938
MO	.182	0.947	.023	.004	.036	.931
Fe	-.125	.092	-0.792	.104	.102	.673
Mn	.012	.243	0.772	.178	.140	.707
B	.006	-.239	.486	-0.633	-.245	.754
S	-.051	.268	-.027	-0.512	.083	.344
Zn	.222	.229	.273	0.510	-.243	.495
K	.122	.183	-.062	.090	,818	.729
Cu	.189	.124	-.009	.278	-.630	.525
Sum of squared eigenvalue	5.0213	2.3910	1.6568	1.4696	1.4284	11.967
Trace percent (%)	31.383	14.943	10.355	9.1854	8.9280	74.796

Table 4. Values of original and standardized factor scores and the Soil Fertility Index (SFI). (reduced).

Obs	Original fator score					Standardized fator score					SFI
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	FP1	FP2	FP3	FP4	FP5	
127	1.3129	0.4533	1.8204	0.8981	2.1036	0.90756	0.60227	0.76744	0.69591	1.00000	0.81221
124	1.3593	0.3895	1.0530	0.3415	1.3823	0.91819	0.58980	0.61193	0.61399	0.88428	0.76877
129	1.7160	0.2260	0.1916	0.4103	1.0151	1.00000	0.55783	0.43737	0.62412	0.82537	0.76676
102	0.8633	0.8767	2.1830	0.7310	0.2003	0.80446	0.68510	0.84091	0.67131	0.69464	0.75620
125	0.9199	0.2525	1.5888	0.9012	1.7583	0.81745	0.56301	0.72050	0.69636	0.94459	0.75350
137	1.3673	1.8764	-1.0395	0.5756	-0.2541	0.92004	0.88064	0.18790	0.64844	0.62175	0.74185
114	0.3571	2.0606	1.7329	0.2837	0.6489	0.68840	0.91667	0.74970	0.60549	0.76661	0.74164
123	0.4841	0.7050	2.9681	1.6307	-0.5371	0.71752	0.65151	1.00000	0.80372	0.57634	0.73717
128	1.4009	-0.5751	0.6819	0.3519	1.6602	0.92775	0,40114	0.53672	0.61551	0.92887	0.73019
105	0.6384	1.9651	2.0439	0,9737	-2.3652	0.75290	0.89799	0.81273	0.70703	0.28304	0.72845
2	1.5092	0.8499	-0.0787	-1.1346	0.5643	0.95256	0.67986	0.38261	0.39674	0.75304	0.72710
126	1.0955	0.0308	0,9916	0.3040	1.1464	0.85771	0,51964	0.59949	0.60847	0.84643	0,72246
146	0.3608	2.3834	0.5647	0.0531	0.4322	0.68925	0.97980	0.51298	0.57154	0.73185	0.71353
135	1.1115	1.4596	-1.3777	0.5500	0.9040	0.86138	0.79910	0.11938	0.64468	0.80754	0.71317
111	0.8241	0.3692	1.2367	1.0079	-0.0486	0.79547	0.58583	0.64916	0.71207	0.65472	0.70629
106	0.9644	-0.4815	1.6539	0.0616	1.0314	0.82766	0,41945	0.73369	0.57279	0.82799	0.70183
118	1.0965	0.7143	-0.9181	0.3213	0.8702	0.85795	0.65332	0.21251	0.61101	0.80212	0.69072
143	0.6837	2.4867	-0.9500	-0.6519	-0.0537	0.76328	1.00000	0.20605	0.46778	0.65390	0.68409
109	0.9354	0.4337	-0.0662	1.1351	-0.1165	0.82099	0.59844	0.38514	0.73079	0.64383	0.68396
. . .											. . .
86	-1.8409	-1.1902	0.7599	0.8386	-0.1016	0.18441	0.28083	0.55253	0.68715	0.64620	0.37150
61	-2.5662	0.6717	-0.1819	1.0906	-0.4506	0.01808	0.64500	0.36169	0.72424	0.59021	0.34592
22	-1.9156	-0.7993	0.4263	-0.2482	-0.9189	0.16726	0.35729	0.48494	0.52719	0.51508	0.33493
66	-2.1877	-0.5204	-1.7210	1.0336	0.0612	0.10487	0.41183	0.04980	0.71585	0.67233	0.30134
53	-2.6451	-0.8944	-0.6385	1.3219	0.2424	0.00000	0.33869	0.26916	0.75828	0.70140	0.28178
Máximo	1.7160	2.4867	2.9681	2.9642	2.1036						0.81221
Mínimo	-2.6451	-2.6259	-1.9668	-3.8302	-4.1294						0.28178

FP: Standardized factor, SFI: Soil Fertility Index.

Table 5. Results of the M of Box test.

M of box		1243.973
F	Aprox.	3.649
	df1	272
	df2	19883.629
	Sig.	.000
Tests the null hypothesis of equal population covariate matrices.		

Table 6. Wilks' Lambda Hypothesis Test of sample groups.

Function test	Lambda de Wilks	Chi-square	Df	Sig.
1 to 2	.078	351.558	32	.000
2	.341	147.819	15	.000

Table 7. Classification by Ward's method grouping soil samples into sample groups for each property.

Ward Method			Association to predicted group			Total
			1	2	3	
Original	Countinng	1	37	5	0	42
		2	2	73	4	79
		3	1	3	23	27
	%	1	88.1	11.9	.0	100.0
		2	2.5	92.4	5.1	100.0
		3	3.7	11.1	85.2	100.0
Note: 89.9% of the original cases were grouped correctly.						

no reports of interdependence of these nutrients in their availability to plants. This factor was named "metabolic and sanitary regulators", since such elements improve the resistance and metabolism of plants.

The fifth and last factor is represented by the variable K and Cu. The name of this factor is "potassium nutrition", since this element has a very high charge for the factor.

In Table 4, the five original factor scores can be positive or negative. A positive sign indicates that the soil in the sampled region is in a satisfactory nutritional balance for the plants and that a negative sign means a soil nutritional imbalance, even though, the effects of the positive forces outweigh the effects of the negative forces in the first 16 samples observed (first highlighted area with SFI from 0.70183 to 0.81221). In the last 4 positions of Table 4 (second highlight area with SFI from 0.28178 to 0.34592), the negative sign means that the effects of the positive forces are outweighed by the effects of the negative forces.

Sorting the SFI values, sixteen sample points were obtained that presented SFI > 0.70 (the SFI mean was 0.56). One hundred and twenty-eight samples showed an intermediate degree of fertility, with SFI between 0.35 and 0.70. The other four sampling points showed a low degree of fertility.

It is observed that most of the sampled points (86%) were classified in the category of "satisfactory" fertility levels, which do not present much nutritional deficiency, guaranteeing reasonable productivity of the cultivated species.

The SFI aims to assess soil fertility in the three properties under study and was used to test the hypothesis.

Canonical discriminant analysis

This statistical technique classifies individuals or objects into mutually exclusive groups based on a set of independent variables. Separation is the first step of this analysis, and the exploratory part of the analysis consists of looking for characteristics that allocate objects in different previously defined groups. In the analysis, Box's M test was significant at 1%, rejecting the null hypothesis, which characterizes that the covariance matrices between sample groups are not equal (Table 5).

The null hypothesis test is rejected (significant at 1%) since the two functions in the three groups are not equal (Table 6), agreeing with the Factor Analysis result (Table 4).

Figure 1 shows that in the sample data set there is an intersection in the data cloud, where 5 samples originally belong to rural property 1 are in group 2. Of the 79 sampling points of rural property 2, only six samples were decharacterized from group 2 and of the 27 sampling points of rural property 3, four do not belong to group 3.

Table 7 shows that 89.9% of the sampled cases were correctly classified, with group 1 reaching 88.1%, group 2 with 92.4% and group 3 with 85.2%. The analyzed functions serve to classify the soil samples that might not be contained in the sample group of each property.

In general, the multivariate analysis of this mass of data suggests that only 20 ha is considered to have a low level of fertility, while 640 ha have satisfactory levels of fertility and 80 ha indicated to have a high level of soil fertility.

This result serves as a tool for technicians and farmers to adopt decision-making measures, acting on the set of variables that define soil fertility levels.

Materials and methods

Study area: Location and characteristics

The study was carried out on three farms in the municipality of Santarém (2°26'22'' S and 54°41'55'' W), state of Pará, in the eastern Amazon. The region's climate type is Am, which corresponds to a humid tropical climate, with a short dry season according to Köppen's classification (Alvares et al., 2013; Beck et al., 2018). The region has an average annual rainfall of 2,100 mm, with a season of lower rainfall ranging from one to five months. The average annual temperature is

25°C. The relief of the area varies from flat to slightly undulating. The predominant soil is the Dystrophic Yellow Latosol and the vegetation is of the Lowland Dense Ombrophilous Forest type (Veloso et al., 1991).

We used a database of soil analyses results from samples collected in 2009 and 2010 in three agricultural properties with soybean and rice conventional cultivation. Soil sampling was carried out in an area of approximately 740 ha, with a sampling grid of 5 ha, making a total of 148 sampling points (composite samples). Each composite sample was formed from 10 single samples collected in a 1 m diameter circle within each sampling grid, thus, resulting in a total of 1,480 single collected samples. Soil collections were carried out using an automated stainless steel screw auger with an electric motor, installed in a quadricycle equipped with GPS, at a depth of 0 – 20 cm depth.

The soil samples were sent to the soil laboratory to determine: pH in water (1:2.5), potential acidity, exchangeable calcium, magnesium and aluminum, potassium and phosphorus, according to EMBRAPA (1997). Based on the results we computed, base saturation and aluminum saturation values.

Factor analysis

We used factor analyses to test the hypothesis that the management practices induce the spatial variability of the attributes and are decisive for the increase in soil fertility when properly conducted, and consequently, for the increase in productivity. Therefore, we analyzed the structures of correlations between a large numbers of independent variables, grouping them into a set of factors, facilitating the understanding of the structure of the data cloud. The use of this technique can initially identify the isolated dimensions of the data structure and then determine the degree to which each variable is explained by each factor, enabling the reduction of the data mass (Gama et al., 2007).

A factor analysis model can be presented in matrix form, as in Dillon and Goldstein (1984):

$$X = \alpha F + \varepsilon \qquad (1)$$

Where X = p-dimensional transposed vector of observable variables, denoted by X = (x1, x2, ..., xp)'.

F=q-dimensional transposed vector of unobservable variables or latent variables called common factors, denoted by F = (f1, f2, ..., fq)', where q < p; e= p-dimensional transposed vector of random variables or single factors,

e = (e1, e2, .., ep)'; and

α= matrix (p, q) of unknown constants, called factor loadings.

In the factor analysis model, it is assumed that the specific factors are orthogonal to each other and with all common factors. Normally, E(ε) = E(F) = 0 and Cov (ε, F) = 0.

To confirm the initial structure, the method of orthogonal rotation of the factors called varimax was used, which is a process in which the reference axes of the factors are rotated around the origin until some other position is reached. This method aims to redistribute the variance of the first factors to the others, thus reaching a simpler and theoretically more significant factorial pattern (Santana, 2005; Hair et al., 2009).

According to Santana (2007) the Soil Fertility Index (SFI) was defined as a linear combination of these factor scores and the proportion of variance, explained by factor in relation to the common variance. The mathematical expression is given by:

$$SFI = \sum_{j=1}^{q} \cdot \left[\frac{\lambda_j}{\sum_j \lambda_j}\right] FP_{ij} \qquad (2)$$

Where λ = is the variance explained by the factor and Σλ is the sum total of the variance explained by the set of common factors.

The factorial score was standardized (FP) to obtain positive values from the original scores and to allow for the hierarchization of the samples, since the SFI values are situated between zero and one. The proposed mathematical formula was as follows:

$$FP_i = \left(\frac{F_i - F_{min}}{F_{max} - F_{min}}\right) \qquad (3)$$

Where Fmin and Fmax = are the maximum and minimum values observed for the factor scores associated with the observations of soil samples according to Santana (2007).

To facilitate the interpretation of the results, the following ranges of SFI values were established, grouping the samples according to their degree of importance: SFI values equal to or greater than 0.70 are considered high; values between 0.35 and 0.69 are intermediate and values less than 0.35 are low.

The Bartlett's sphericity and the Kaiser-Meyer-Olkin (KMO) tests were performed to assess the suitability of the method to the data sample.

In one hand, the Bartlett's sphericity test assesses the overall significance of the correlation matrix, which tests the null hypothesis that the correlation matrix is an identity matrix.

On the other hand, the KMO test is based on the principle that the inverse of the correlation matrix approximates the diagonal matrix, so it compares the correlations between the observable variables. The mathematical formulas of these tests are as follows (Dillon and Goldstein, 1984).

$$KMO = \frac{\sum_i \sum_j r_{ij}^2}{\sum_i \sum_j r_{ij}^2 + \sum_i \sum_j a_{ij}^2} \qquad (4)$$

Where rij is the sample correlation coefficient between variables xi and xj and aij is the partial correlation coefficient between the same variables, which simultaneously is an estimate of the correlations between the factors, eliminating the effect of the other variables. The aij should assume values close to zero, since the factors are assumed to be orthogonal to each other. Values of this test below 0.50 are unacceptable (Hair et al., 2009). The Bartlett's test of sphericity tests the null hypothesis that the variables are independent, against the alternative hypothesis that the variables are correlated with each other (Santana, 2007).

Canonical discriminant analysis

We used the canonical discriminant analysis in the IBM SPSS Statistics software, version 20 to confirm the data grouping, a technique consists of analyzing a model in which the dependent variable is categorical, which in this case consists of three classification groups (properties), and the independent variables are metric or of an interval nature (chemical attributes of the soil). When three classifications are involved, the technique is referred to the three groups of discriminant analysis, being appropriate to test the hypotheses that the means of the groups of independent variables found for the three groups are equal. These means of the values of the groups' discriminant scores refer to their centroids, with as many centroids as there are groups.

A comparison between these shows how much the groups are separated during the discriminant function test (Gonçalves et al., 2008).

The general equation of the discriminant model is:

$$D = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + \ldots + b_kX_k \quad (5)$$

Where D = discriminant score;

b = is the discriminant coefficient or weight;

D = is a categorical variable;

X = independent variable;

X1, X2, X3, X4,..., Xk are interval and/or ratio variables.

The quantitative variables used for the study were: Hydrogen potential (pH), base saturation (V%), Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Exchangeable aluminum (Al), organic matter (OM), cation exchange capacity (CTC), aluminum saturation (m%), Sulfur (S), Copper (Cu), Iron (Fe), Manganese (Mn), Zinc (Zn) and Boron (B). The dependent variable, were the rural properties.

Conclusions

The use of discriminant analysis, especially factor analysis, with the creation of the SFI, proved efficient for the proposed study, since the analysis of variance could not show efficiency due to the interrelationships between the variables causing bias in the results.

Approximately 97% of the sampled area presented satisfactory to high soil fertility, which induces the use of reduced amounts of fertilizers in most of the cultivated area. Although this type of sampling could be analyzed using geostatistics, which would give more robustness to the results, the absence of geographic coordinates in the database prevents this analysis from being done efficiently.

Acknowledgements

We would like to thank the companies “Missioneira Agrícola” and “Bunge Fertilizantes” for providing the database of soil samples used in this work, and to the agronomists Victor do Amarale Luiz Fernando Freiberger for their support and sharing of information.

References

Ahmed M, Rauf M, Mukhtar, Z, Saeed NA (2017) Excessive use of nitrogenous fertilizers: an unawareness causing serious threats to environment and human health. Environ Sci Pollut R. 24: 26983–26987.

Alvares CA, Stape JL, Sentelhas PC, Gonçalves JLM, Sparovek G (2013) Köppen’s climate classification map for Brazil. Meteorol Z. 22(6): 711–728.

Beck H, Zimmermann N, McVicar TR, Vergopolen N, Berg A, Wood EF (2018) Present and future Köppen-Geiger climate classification maps at 1-km resolution. Scientific Data. 5: .180214.

Bhunia GS, Shitb PK, Chattopadhyay R (2018) Assessment of spatial variability of soil properties using geostatistical approach of lateritic soil (West Bengal, India) Annals of Agrarian Science. 16: 436–443.

Carvalho MAC, Panosso AR, Teixeira EER, Araújo EG, Brancaglioni VA Dallacorta R (2018) Multivariate approach of soil attributes on the characterization of land use in the southern Brazilian Amazon. Soil Till Res. 184: 207–215.

Castro NR, Silva AF, Gilio L. (2021). Desempenho e inter-relações do setor de fertilizantes: uma análise segundo a ótica de insumo-produto. Planejamento e Políticas Públicas. 56: 159-189.

Dillon WR, Goldstein M Multivariate Analysis Methods and Applications, Wiley, New York, 1984. 587 p.

Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA. Serviço Nacional de Levantamento e Conservação de Solos (Rio de Janeiro). Manual de métodos de análise de solo. 2.ed. Rio de Janeiro: Embrapa CNPS, 1997, 212p.

Faucon MP, Houben D, Lambers H (2017) Plant Functional Traits: Soil and Ecosystem Services. Trends Plant Sci. 22(5): 385-394.

Freitas L, Casagrande JC, Oliveira IA, Campos MCC (2014) Análise multivariada na avaliação de atributos de solos com diferentes texturas cultivados com cana-de-açúcar. Revista de Ciências Agrárias. 57: 224-33.

Gama ZJC; Santana AC, Mendes FAT, e Khan AS (2007) Índice de desempenho competitivo das empresas de móveis da região metropolitana de Belém. Revista de Economia e Agronegócio. 5(1):127-160.

Gonçalves CA, Dias AT, E Muniz RM (2008) Análise discriminante das relações entre fatores estratégicos, indústria e desempenho em organizações brasileiras atuantes na indústria manufatureira. Revista de Administração Contemporânea. 12(2): 287-311.

Hair J, Black WC, Babin BJ, Anderson RE, Tatham, R.L Análise Multivariada de Dados. 6. ed. Porto Alegre: Bookman, 2009. 687 p.

Hou L, Liu Z, Zhao J, Ma P, Xu X (2021) Comprehensive assessment of fertilization, spatial variability of soil chemical properties, and relationships among nutrients, apple yield and orchard age: A case study in Luochuan County, China. Ecol Indic. 122: 107285.

Oumenskou H, Baghdadi ME, Barakat A, Aquit M, Ennaji W, Karroum LA, Aadraoui M (2019) Multivariate statistical analysis for spatial evaluation of physicochemical properties of agricultural soils from Beni-Amir irrigated perimeter, Tadla plain, Morocco, Geology, Ecology, and Landscapes. 3(2): 83-94.

Saldanha ECM, Silva Júnior ML, Okomura RS, Bonfim Neto AL, Viégas IJM, Fernandes AR (2013) Spatial variability of soil fertility in areas cultivated with grains in the region of Paragominas, Pará state, Brazil. Revista Ciências Agrárias. 56:120-128.

Santana AC. 2005) Elementos de economia, agronegócio e desenvolvimento local. Belém: GTZ; TUD; UFRA, 2005. p.133-142. (Série Acadêmica, 01).

Santana AC (2007) Análise do desempenho competitivo das agroindústrias de polpa de frutas do Estado do Pará. Teoria e Evidência Econômica. 14: 36-62..

Silva ENS, Montanari R, Panosso AR, Correa AR, Tomaz PK, Ferraudo AS (2015 Variabilidade de atributos físicos e químicos do solo e produção de feijoeiro cultivado em sistema de cultivo mínimo com irrigação. Rev Bras Cienc Solo. 39: 598-607.

Teixeira BES, Santos TS, Terra A (2019). A transformação do território a partir do uso da terra no município de Santarém Pará. Nova Revista Amazônica. 7(3): 99-108.

Vasu D, Singh SK, Sahu N, Tiwary P, Chandran P, Duraisami VP, Ramamurthy V, Lalitha M, Kalaiselvi B (2017) Assessment of spatial variability of soil properties using geospatial techniques for farm level nutrient management, Soil Till Res. 169: 25-34.

Veloso HP, Rangel Filho ALR, Lima JCA. Classificação da vegetação brasileira, adaptada a um sistema universal. Rio de Janeiro: IBGE, 1991. 123 p.

Watson JC (2017) Establishing Evidence for Internal Structure Using Exploratory Factor Analysis. Measurement and Evaluation in Counseling and Development, 50(4): 232-238.

Yeater KM, Duke SE, Riedell WE (2015) Multivariate Analysis: Greater Insights into Complex Systems. Agron J. 15 (2):799-810.