Pesquisa Agropecuária Brasileira

Soil Science/ Original Article

ISSN 1678-3921

Journal homepage: www.embrapa.br/pab

For manuscript submission and journal contents, access: www.scielo.br/pab

Geostatistics and multivariate analysis to determine experimental blocks for sugarcane

Abstract – The objective of this work was to define experimental blocks for sugarcane experiments using geostatistical techniques, principal component analysis, and clustering techniques applied to soil properties. For this, data of soil chemical properties from a sugarcane experiment were used. Geostatistical techniques were applied to identify the spatial variability of these properties and to estimate the values for non-sampled locations through kriging. The principal components analysis was used for dimensional reduction, and, with the new variables obtained, the cluster analysis was performed using the k-means method to determine the experimental blocks with two to five replicates. Of the 12 analyzed variables, 10 showed spatial dependence. The principal component analysis allowed reducing the dimensionality of the data to two variables, which explained 82.27% of total variance. The obtained blocks presented irregular polygonal shapes, with different formats and sizes, and some of them showed discontinuities. The proposed methodology has the potential to identify more uniform areas in terms of soil chemical properties to allocate experimental blocks for sugarcane.

Index terms: experimental design, field experimentation, kriging, principal component analysis, spatial variations.

Geoestatística e análise multivariada para determinação de blocos experimentais para cana-de-açúcar

Resumo – O objetivo deste trabalho foi definir blocos experimentais para experimentos com cana-de-açúcar, com uso de técnicas de geoestatística, análise de componentes principais e técnicas de agrupamento aplicadas às propriedades do solo. Para isso, foram utilizados dados de propriedades químicas do solo de um experimento com cana-de-açúcar. As técnicas de geoestatística foram aplicadas para identificar a variabilidade espacial dessas propriedades e estimar os valores para locais não amostrados por meio de krigagem. A análise de componentes principais foi aplicada para redução dimensional, e, com as novas variáveis obtidas, realizou-se a análise de agrupamento pelo método *k-means*, para determinar os blocos experimentais com duas a cinco repetições. Das 12 variáveis analisadas, 10 apresentaram dependência espacial. A análise de componentes principais permitiu reduzir a dimensionalidade dos dados para duas variáveis, que explicaram 82,27% da variância total. Os blocos obtidos apresentaram formas poligonais irregulares, com diferentes formatos e tamanhos, e alguns mostraram descontinuidade. A metodologia proposta tem potencial para identificar áreas mais homogêneas em termos de propriedades químicas do solo, para alocar blocos experimentais de cana-de-açúcar.

Termos para indexação: delineamento experimental, experimentação de campo, krigagem, análise de componentes principais, variações espaciais.

Gustavo Henrique da Silva[1] ✉, Kaléo Dias Pereira[2], Antonio Policarpo Souza Carneiro[1], Matheus de Paula Ferreira[1], Gérson Rodrigues dos Santos[1] and Luiz Alexandre Peternelli[1]

[1] Universidade Federal de Viçosa, Departamento de Estatística, Avenida Peter Henry Rolfs, s/nº, Campus Universitário, CEP 36570-900 Viçosa, MG, Brazil. E-mail: gustavo.silva2@ufv.br, policarpo@ufv.br, matheusferreiraufv@gmail.com, gerson.santos@ufv.br, peternelli@ufv.br

[2] Universidade Federal de Viçosa, Departamento de Engenharia Florestal, Avenida Purdue, s/nº, Campus Universitário, CEP 36570-900 Viçosa, MG, Brazil. E-mail: kaleo.pereira@ufv.br

✉ Corresponding author

Received
May 12, 2023

Accepted
April 25, 2024

How to cite
SILVA, G.H. da; PEREIRA, K.D.; CARNEIRO, A.P.S.; FERREIRA, M. de P.; SANTOS, G.R. dos; PETERNELLI, L.A. Geostatistics and multivariate analysis to determine experimental blocks for sugarcane. **Pesquisa Agropecuária Brasileira**, v.59, e03373, 2024. DOI: https://doi.org/10.1590/S1678-3921.pab2024.v59.03373.

Introduction

Brazil stands out in sugarcane (*Saccharum officinarum* L.) production, which is expected to reach 652.9 million of tons in the 2023/2024 crop season, representing an increment of 6.9% in relation to that of 2022/2023 (Acompanhamento..., 2023). However, the increased production area also increases environmental impacts, which, added to climate changes, presents a great challenge to producers (Pittelkow et al., 2015).

In this context, agricultural experimentation emerges as an important tool to improve crop productivity. Among the basic principles of experimentation, local control is key to enhance experiment efficiency by dividing the known heterogeneous environment into more homogeneous sections (Costa et al., 2007). This procedure aims to reduce experimental error in order to raise experimental precision through the systematic control of sources of variation.

Regarding the control of environment variability, the choice between a randomized complete block design and a completely randomized design depends on whether the plot-to-plot variation is smaller than that of the block-to-block (Clewer & Scarisbrick, 2013), considering that the efficiency of an experiment depends on defining blocks as uniform as possible. Any unwanted variation within the blocks may maximize confounding factors in relation to the treatments.

To support experiment planning, geostatistics is an alternative that can be used to identify the spatial structure of soil properties through kriging interpolation (Oliver & Webster, 2014; Carneiro et al., 2016a, 2016b; Silva et al., 2017; Bhunia et al., 2018; Amaral & Justina, 2019).

The objective of this work was to define experimental blocks for sugarcane experiments using geostatistical techniques, principal component analysis, and clustering techniques applied to soil properties.

Materials and Methods

For the study, the used data were those of soil fertility collected in the research by Ferreira (2020), with the support of Centro de Pesquisa e Melhoramento da Cana-de-Açúcar, an institution for sugarcane research and improvement of Universidade Federal de Viçosa. The sugarcane experimental area, a 42x80 m plot, covering 3,360 m^2, was located in the municipality of Oratórios, in the state of Minas Gerais, Brazil.

The area was subjected to a systematic sampling, in a 4x9regular grid, with 36 sampling points (Figure 1). Point density was approximately 0.01 point per square meter, a value considered intermediate when compared with those found in the literature (Pasini et al., 2021; Adão et al., 2022).

Soil samples were collected in October 2019, at a depth between 0–20 cm, properly stored, and, then, sent to the municipality of Viçosa, also in the state of Minas Gerais, for analyses. The following 12 soil chemical properties were evaluated: hydrogen potential (pH), phosphorus, potassium, magnesium, calcium, aluminum, potential acidity (H+Al), total exchangeable bases, effective cation exchange capacity (CTC_t), cation exchange capacity at pH_7 (CTC_T), aluminum saturation index, and base saturation index. The used extractors were: Mehlich-1 for K and P; KCl 1.0 mol L^{-1} for Ca, Mg, and Al; and calcium acetate 0.5 mol L^{-1} at pH 7 for H+Al (Donagema et al., 2011).

Shapiro-Wilk's test, at a 5% significance level, was applied to check whether the distribution of the variables met normality assumption. Additionally, histograms and boxplot graphs for each analyzed variable were used to complement the analysis of data distribution. The boxplot was specifically used to detect and remove outliers as recommended by Smiti (2020). According to Santos et al. (2017), because they are considered inconsistent values, outliers can impair the quality of the variogram and geostatistical interpolation.

The base package of the R software (R Core Team, 2020), version 4.0.2, was used, together with the geoR package, version 1.8.1, to identify the spatial dependence of the variables and to fit a model.

When spatial dependence was observed, the variograms were subjected to the variofit function of the geoR package. The coefficients of the models were estimated using the methods of ordinary least squares or weighted least squares (Cressie, 1985).

In order to evaluate the quality of the fit, the Jackknife cross-validation technique was carried out using the xvalid function of the geoR package. For this, the following aspects of cross-validation were used: angular coefficient of the regression between estimated and observed values equal or near 1, mean of the estimation error near zero, mean of the standardized error near zero, and variance of the standardized estimation error near 1 (Mendoza Hernández, 2021).

After model fitting, the spatial dependence index (SDI) suggested by Biondi et al. (1994) was calculated in order to determine the degree of intensity of spatial dependence, using the following equation:

$$SDI = C_1 / C_1 + C_0 \times 100$$

where C_1 is the contribution, and C_0 is the nugget effect.

In the absence of spatial dependence, interpolation can be performed using other non-stochastic methods, among which the inverse distance weighted estimation stands out (Salekin et al., 2018; Chen et al., 2019; Shukla et al., 2020).

For the interpolation of the data of soil chemical properties (36 sampled values for each attribute), the ordinary kriging technique was carried out using the obtained adjusted variogram. This type of kriging was chosen because it is a popular method that provides the best unbiased linear estimative according to Bai & Tahmasebi (2021).

For the principal component analysis, the collected and estimated data were used, resulting in 729 coordinate points. The first k components that explained 80% or more of the total accumulated variance were chosen (Jolliffe & Cadima, 2016). Afterwards, clustering was performed by parameterizing the algorithm in order to find clusters in the same number of suggested experimental blocks (two, three, four, and five). Using the results of the clustering analysis, maps of the experimental area were generated.

Figure 1. Sugarcane (*Saccharum officinarum*) experimental area in the municipality of Oratórios, in the state of Minas Gerais, Brazil. The yellow points indicate the locations where the chemical properties of the soil were analyzed in a 4x9 regular grid.

Source: adapted from Ferreira (2020).

Results and Discussion

In terms of spatial distribution, most of the variables showed a better fit to the spherical model (Figure 2). In addition, all variables presented a nugget effect, except the base saturation index, which showed a null value until the third decimal place (Table 1). A pure nugget effect was only found for H+Al and CTC_t, which were properly addressed using the inverse distance weighted estimation. In general, the range estimated for spatial dependence was below 100 m, with an average of 57 m, which is equivalent to 71.6% of the largest dimension of 80 m of the experimental area. Souza et al. (2014) concluded that increasing the number of samples changes the results of the geostatistical analysis and widens their range.

According to the SDI (Table 1), 80% of the soil attributes presented a moderate spatial dependence. However, CTC_t and the base saturation index showed a strong dependence, which is related to their smaller nugget effect when compared with the C_1 contribution value obtained for each of these variables. The values found for the inverse distance weighted estimation in the present study were higher for P, Ca, and Mg and lower for K in comparison with those reported by Carvalho et al. (2002). Almeida & Guimarães (2016), studying the soil of a coffee (*Coffea arabica* L.) crop, verified a high spatial dependence only for pH.

For H+Al and CTC_t, it was not possible to identify spatial dependence, being necessary to use the interpolator weighted by inverse distance. The map for H+Al showed some similarity to the one obtained

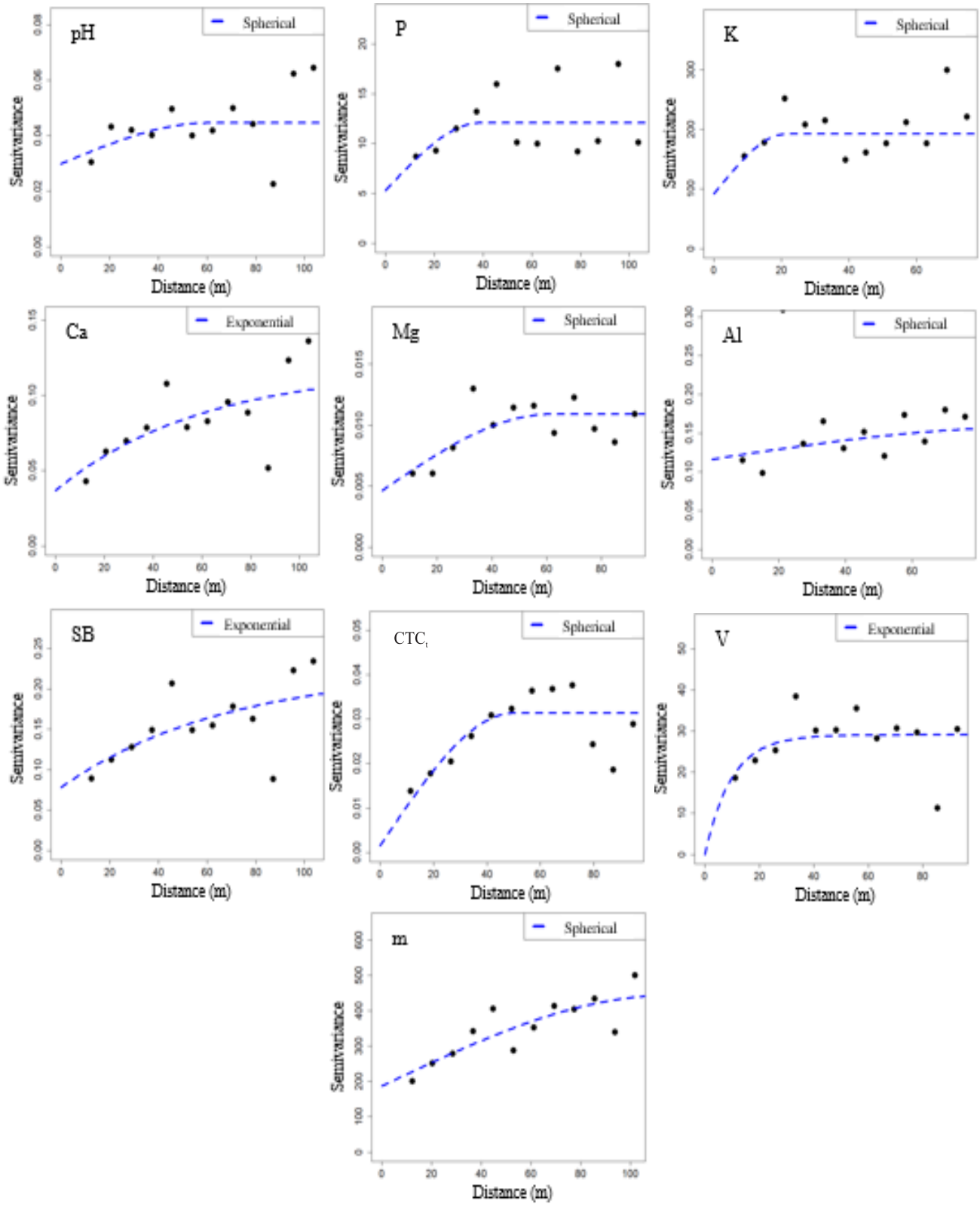

Figure 2. Semivariograms of the soil properties from a sugarcane (*Saccharum officinarum*) experimental area in the municipality of Oratórios, in the state of Minas Gerais, Brazil. SB, total exchangeable bases; CTC_t, effective cation exchange capacity; V, base saturation index; and m, aluminum saturation index.

via ordinary kriging, but not that of CTC_t. During the estimation process, the root mean square error calculated for CTC_t and H+Al was 0.665 and 0.619, respectively.

According to the results of the principal component analysis, the first four latent variables explained more than 90% of the total variance in the data. In most academic works, principal components (PCs) are had as the set of latent variables that explain at least 70% of total variance (Ferreira, 2018), which is why, in the present study, the first two principal components (PC1 and PC2) that, together, explain more than 80% of data variability were selected.

To determine the experimental blocks, PC1 and PC2 were used in the k-means clustering algorithm, taking into account the information of all chemical variables. The formed groups consisted of the blocks with a greater uniformity considering both PCs (Figure 3), evidencing the number of blocks as a function of their color and shapes.

It should be noted that dividing the experimental area into a larger number of blocks will increase the uniformity within the blocks, but reduce the area of each block, limiting the number of treatments to be tested, as well as the size of the experimental units. Therefore, the choice of the number of blocks should consider the number of treatments and the studied crop.

Regarding the number of suggested experimental blocks, when the experimental area was divided into only two blocks (n=2), the spatial continuity of the red and blue blocks became evident (Figure 3). The blocks presented distinct areas, with the red block being larger than the blue one. Moreover, the shape of these blocks were not polygons with straight sides, but curves that followed the spatial variability of the terrain.

In the case of three blocks (n=3), blocks with non-regular shapes and distinct areas were observed. Although the k-means algorithm classified some points belonging to the green block within the pink block, the practical situation in the field may ignore these few points and treat the three obtained blocks as continuous.

Considering four blocks in the experimental area (n=4), distinct shapes and areas were also verified. The blue block stood out due to its obvious discontinuity, with two relatively large parts that should receive a replicate of each treatment in experiment planning.

When the experimental area was divided into five blocks (n=5), blocks with a similar size and shape to those of n=4 were observed. As the number of blocks increased, the discontinuity of the small area also increased. Therefore, for practical purposes, it might be better for the researcher to mark and not use these small areas if they are not large enough to implement an experimental plot. Normally, to define the size of an experimental unit, the researcher can carry out a uniformity test using methodologies such as the method of the maximum curvature of the coefficient of variation (Cargnelutti Filho et al., 2016).

The values obtained in the classification of the 729 points referring to the selected PC1 and PC2 are shown

Table 1. Geostatistical parameters estimated for the semivariograms describing the spatial variability of the evaluated soil chemical variables and the methods used to obtain them[(1)].

Variable	C_0	C_1	Level	Reached	Model	Method[(2)]	IDE (%)[(3)]
pH	0.02	0.012	0.045	56.00	Spherical	WLS	34.83
P	5.270	6.823	12.094	38.85	Spherical	WLS	56.41
K	91.340	100.690	192.032	21.36	Spherical	WLS	52.44
Ca	0.037	0.081	0.121	60.01	Exponential	WLS	68.64
Mg	0.005	0.006	0.011	61.02	Spherical	OLS	57.41
Al	0.120	0.042	0.160	93.00	Spherical	WLS	25.00
SB	0.079	0.150	0.231	67.93	Exponential	OLS	65.50
CTC(t)	0.002	0.029	0.031	50.52	Spherical	WLS	95.08
V	<0.001	29.062	29.062	9.70	Exponential	OLS	100.00
m	186.831	255.980	442.816	114.87	Spherical	OLS	57.81

[(1)]SB, total exchangeable bases; CTC_t, effective cation exchange capacity; V, base saturation index; m, aluminum saturation index; C_0, nugget effect; and C_1, contribution. [(2)]WLS, weighted least squares; and OLS, ordinary least squares. [(3)]IDE, inverse distance weighted estimation, whose values were classified as: IDE ≥ 0.75, high; 0.25 ≤ IDE < 0.75, moderate; and IDE < 0.25, low.

in Table 2. The total sum of squares was the same for all blocks, a result that was already expected since this value is calculated based on the variance of the scores of PC1 and PC2.

As previously discussed, the experimental blocks obtained by clustering presented different-sized areas (Table 3), which should be considered when planning an experiment since they imply certain restrictions.

Ferreira (2020) used 14 m^2 plots for a selection experiment of sugarcane, in alignment with Leite et al. (2009) and Igue et al. (1991). Considering the blocks with smaller areas, the division of the experimental area into two, three, four, or five blocks would allow testing 90, 46, 43, and 29 treatments in experiments with sugarcane, respectively.

In an experiment with corn (*Zea mays* L.), Assis & Silva (1999) concluded that the ideal experimental plot should vary from 0.75 to 6.77 m^2. Considering a 5.0 m^2 plot, it would be possible to establish 252, 130, 121, and 81 experimental units within the smallest block when dividing the area into two, three, four, or five blocks, respectively.

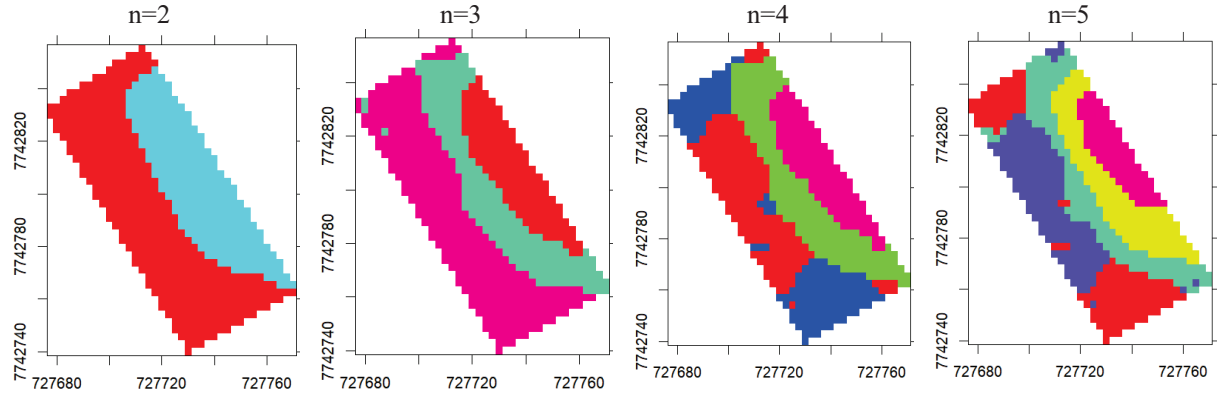

Figure 3. Definition of clustered blocks via the k-means algorithm as a function of the number of replicates (n), considering the spatial variability of soil chemical attributes in a sugarcane (*Saccharum officinarum*) experimental area.

Table 2. Lack of uniformity between and within blocks as a function of the sums of squares associated with principal components 1 and 2 for dividing a sugarcane (*Saccharum officinarum*) experimental area into two, three, four, and five blocks.

Parameter[(1)]	Number of blocks			
	n=2	n=3	n=4	n=5
SQ_T				7,185.97
SQ_E	4,254.07	5,030.79	5,699.24	6,013.00
SQ_{Yellow}	N	N	N	242.77
SQ_{Blue}	1,215.84	N	329.64	263.37
SQ_{Pink}	N	1,201.08	387.07	206.66
SQ_{Green}	N	496.27	445.87	207.97
SQ_{Red}	1,716.06	457.82	324.15	252.20
N_{Yellow}	N	N	N	149.00
N_{Blue}	276.00	N	178.00	181.00
N_{Pink}	N	364.00	134.00	86.00
N_{Green}	N	216.00	213.00	160.00
N_{Red}	453.00	149.00	204.00	153.00

[(1)]SQ_T, total sum of squares; SQ_E, total sum of squares between the blocks; SQ_{Yellow}, SQ_{Blue}, SQ_{Pink}, SQ_{Green}, and SQ_{Red}, sum of squares of the yellow, blue, pink, green, and red blocks, respectively; N_{Yellow}, N_{Blue}, N_{Pink}, N_{Green}, and N_{Red}, number of points classified in the yellow, blue, pink, green, and red blocks, respectively; and N, not applicable.

In experiments with several treatments and/or large experimental units, however, the researcher may choose the incomplete block design, which, although more difficult to analyze than the complete block design, is compensated by gain in experimental precision (Pimentel-Gomes, 2022).

The coefficient of variation (CV) within each block was calculated for the five following soil chemical properties that presented the highest CV in the total area (Table 4): Ca, Mg, Al, total exchangeable base, and base saturation index. It should be noted that a high CV reflects a great variability in the experimental area.

For n=2, the red block was homogeneous for almost all of the five variables evaluated. Four variables presented a CV classified as low, below 10%, and only the base saturation index showed a value classified as medium. However, for the blue block, this homogeneity was lower since almost all CV values were classified as medium.

As the number of experimental blocks increased, the CV of the variables decreased. For n=5, the highest

Table 3. Area of the blocks and their respective proportions in relation to the total area for the different numbers of blocks (n) formed in a sugarcane (*Saccharum officinarum*) experimental area[(1)].

Blocks	n=2		n=3		n=4		n=5	
	A (m^2)	P (%)	A (m^2)	P (%)	A (m^2)	P (%)	A (m^2)	P (%)
Yellow	N	N	N	N	N	N	637.0	18.9
Blue	1,260	37.5	N	N	758.0	22.6	743.0	22.1
Pink	N	N	1,723	51.3	609.0	18.1	408.0	12.1
Green	N	N	985.0	29.3	976.0	29.0	708.0	21.0
Red	2,100	62.5	652.0	19.4	1,017.0	30.3	864.0	25.7

[(1)]A, area of each block; P, proportion of the area occupied by the block in relation to the total area; and N, not applicable.

Table 4. Coefficients of variation in the total area and within each block for the following soil chemical variables: calcium (CV_{Ca}), magnesium (CV_{Mg}), aluminum (CV_{Al}), sum of bases (CV_{SB}), and base saturation index (CV_V), considering the different number of blocks (n).

Blocks	CV_{Ca} (%)	CV_{Mg} (%)	CV_{Al} (%)	CV_{SB} (%)	CV_V (%)
Total area	95.78[(1)]	49.55[(1)]	37.14[(1)]	48.94[(1)]	49.07[(1)]
		n=2 blocks			
Red	7.69	7.83	7.02	6.48	12.63
Blue	19.00	18.15	4.88	14.93	18.56
		n=3 blocks			
Red	13.78	13.87	3.11	10.88	16.54
Green	13.48	9.78	4.95	8.63	14.45
Pink	5.56	5.59	6.23	4.84	10.97
		n=4 blocks			
Pink	13.07	13.41	2.99	10.43	16.24
Green	13.92	10.02	4.69	8.82	14.83
Blue	6.55	5.46	7.08	4.68	13.37
Red	4.43	5.78	5.11	3.97	9.42
		n=5 blocks			
Pink	11.64	11.21	1.66	8.67	13.60
Yellow	11.45	9.35	3.54	7.97	13.01
Green	10.77	6.22	4.69	5.65	11.44
Blue	3.57	5.07	4.72	3.46	9.04
Red	11.45	9.35	3.54	7.97	13.01

[(1)]Values calculated from the initial 36 sample points.

CV of 13.60% was observed for the base saturation index, a value classified as medium. Considering the other variables and blocks obtained, most of the CVs were lower than 10%.

The initial experimental area showed a great variation in soil chemical attributes, reaching a CV of 95.78% (Table 4). However, when the proposed methodology was used, it was possible to obtain more uniform experimental blocks with a CV classified as medium in the most extreme case.

Conclusions

1. The proposed methodology, using geostatistical techniques, principal component analysis, and clustering techniques, can be used to divide the sugarcane (*Saccharum officinarum*) experimental area into uniform blocks based on soil chemical properties.

2. Using the k-means algorithm, the experimental area can be divided into two, three, four, or five blocks with a high uniformity.

3. The use of regular-shaped blocks is not adequate to standardize the sugarcane experimental area.

Acknowledgments

To Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), for research support through a master's degree scholarship (code 001).

References

ACOMPANHAMENTO DA SAFRA BRASILEIRA [DE] CANA-DE-AÇÚCAR: safra 2023/24: segundo levantamento, v.11, n.2, agosto 2023.

ADÃO, A. da S.; FERNANDES, H.C.; SANTOS, N.T.; MARTINS, F.C.M.; PEREIRA, P.S.; SOUZA, L.M.R. de; OLIVEIRA, Z.R.C.R. de. Análise da correlação dos atributos físicos do solo com os componentes de rendimento de grãos de milho em diferentes sistemas de cultivo. **Research, Society and Development**, v.11, e48611226059, 2022. DOI: https://doi.org/10.33448/rsd-v11i2.26059.

ALMEIDA, L. da S.; GUIMARÃES, E.C. Geoestatística e análise fatorial exploratória para representação espacial de atributos químicos do solo, na cafeicultura. **Coffee Science**, v.11, p.195-203, 2016.

AMARAL, L.R. do; JUSTINA, D.D.D. Spatial dependence degree and sampling neighborhood influence on interpolation process for fertilizer prescription maps. **Engenharia Agrícola**, v.39, p.85-95, 2019. Special issue. DOI: https://doi.org/10.1590/1809-4430-Eng.Agric.v39nep85-95/2019.

ASSIS, J.P. de; SILVA, P.S.L. e. Tamanho e forma ideais da unidade experimental em ensaio com milho. **Agropecuária Técnica**, v.20, p.42-50, 1999.

BAI, T.; TAHMASEBI, P. Accelerating geostatistical modeling using geostatistics-informed machine learning. **Computers & Geosciences**, v.146, art.104663, 2021. DOI: https://doi.org/10.1016/j.cageo.2020.104663.

BHUNIA, G.S.; SHIT, P.K.; CHATTOPADHYAY, R. Assessment of spatial variability of soil properties using geostatistical approach of lateritic soil (West Bengal, India). **Annals of Agrarian Science**, v.16, p.436-443, 2018. DOI: https://doi.org/10.1016/j.aasci.2018.06.003.

BIONDI, F.; MYERS, D.E.; AVERY, C.C. Geostatistically modeling stem size and increment in an old-growth forest. **Canadian Journal of Forest Research**, v.24, p.1354-1368, 1994. DOI: https://doi.org/10.1139/x94-176.

CARGNELUTTI FILHO, A.; ALVES, B.M.; TOEBE, M.; FACCO, G. Tamanhos de unidades experimentais básicas e de parcelas em tremoço branco. **Ciência Rural**, v.46, p.610-618, 2016. DOI: https://doi.org/10.1590/0103-8478cr20150756.

CARNEIRO, J.S. da S.; FARIA, Á.J.G. da; FIDELIS, R.R.; SILVA NETO, S.P. da; SANTOS, A.C. dos; SILVA, R.R. da. Diagnóstico da variabilidade espacial e manejo da fertilidade do solo no Cerrado. **Scientia Agraria**, v.17, p.38-49, 2016a. DOI: https://doi.org/10.5380/rsa.v17i3.50096.

CARNEIRO, J.S. da S.; SANTOS, A.C.M. dos; FIDELIS, R.R.; SILVA NETO, S.P. da; SANTOS, A.C. dos; SILVA, R.R. da. Diagnóstico e manejo da variabilidade espacial da fertilidade do solo no cerrado do Piauí. **Revista de Ciências Agroambientais**, v.14, 2016b.

CARVALHO, J.R.P. de; SILVEIRA, P.M. da; VIEIRA, S.R. Geoestatística na determinação da variabilidade espacial de características químicas do solo sob diferentes preparos. **Pesquisa Agropecuária Brasileira**, v.37, p.1151-1159, 2002. DOI: https://doi.org/10.1590/S0100-204X2002000800013.

CHEN, Z.-Y.; ZHANG, T.-H.; ZHANG, R.; ZHU, Z.-M.; YANG, J.; CHEN, P.-Y.; OU, C.-Q.; GUO, Y. Extreme gradient boosting model to estimate $PM_{2.5}$ concentrations with missing-filled satellite data in China. **Atmospheric Environment**, v.202, p.180-189, 2019. DOI: https://doi.org/10.1016/j.atmosenv.2019.01.027.

CLEWER, A.G.; SCARISBRICK, D.H. **Practical statistics and experimental design for plant and crop science**. Chichester: John Wiley & Sons, 2013. 346p.

COSTA, R.B. da; RESENDE, M.D.V. de; SILVA, V.S. de M. e. Experimentação e seleção no melhoramento genético de TECA (*Tectona grandis* Lf). **Floresta e Ambiente**, v.14, p.76-92, 2007.

CRESSIE, N. Fitting variogram models by weighted least squares. **Journal of the International Association for Mathematical Geology**, v.17, p.563-586, 1985. DOI: https://doi.org/10.1007/BF01032109.

DONAGEMA, G.K.; CAMPOS, D.V.B. de; CALDERANO, S.B.; TEIXEIRA, W.G.; VIANA, J.H.M. (Org.). **Manual de métodos de análise de solo**. 2.ed. rev. Rio de Janeiro: Embrapa Solos, 2011. 230p. (Embrapa Solos. Documentos, 132).

FERREIRA, D.F. **Análise multivariada**. 3.ed. Lavras: UFLA, 2018. p.394.

FERREIRA, M. de P. **Geoestatística e aerofotogrametria aplicadas à seleção de famílias de cana-de-açúcar**. 2020. 72p. Tese (Doutorado) – Universidade Federal de Viçosa, Viçosa.

IGUE, T.; ESPIRONELO, A.; CANTARELLA, H.; NELLI, E.J. Tamanho e forma de parcela experimental para cana-de-açúcar. **Bragantia**, v.50, p.163-180, 1991. DOI: https://doi.org/10.1590/S0006-87051991000100016.

JOLLIFFE, I.T.; CADIMA, J. Principal component analysis: a review and recent developments. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v.374, art.20150202, 2016. DOI: https://doi.org/10.1098/rsta.2015.0202.

LEITE, M.S. de O.; PETERNELLI, L.A.; BARBOSA, M.H.P.; CECON, P.R.; CRUZ, C.D. Sample size for full-sib family evaluation in sugarcane. **Pesquisa Agropecuária Brasileira**, v.44, p.1562-1574, 2009. DOI: https://doi.org/10.1590/S0100-204X2009001200002.

MENDOZA HERNÁNDEZ, M. **Análise geoestatística de multivariada para definição de zonas de manejo em cana-de-açúcar (*Saccharum officinarum*) na Guatemala**. 2021. 60p. Dissertação (Mestrado) – Universidade Federal de Viçosa, Viçosa.

OLIVER, M.A., WEBSTER, R. A tutorial guide to geoestatistics: computing and modelling variograms and kriging. **Catena**, v.113, p.56-69, 2014. DOI: https://doi.org/10.1016/j.catena.2013.09.006.

PASINI, M.P.B.; ENGEL, E.; LÚCIO, A.D.C.; NORA, S.L.D. Selection of interpolators to predict populations of *Tibraca limbativentris* in irrigated rice. **Brazilian Archives of Biology and Technology**, v.64, e21180601, 2021. DOI: https://doi.org/10.1590/1678-4324-2021180601.

PIMENTEL-GOMES, F. **Curso de estatística experimental**. Piracicaba: Fealq, 2022. v.15, 451p.

PITTELKOW, C.M.; LIANG, X.; LINQUIST, B.A.; van GROENIGEN, K.J.; LEE, J.; LUNDY, M.E.; van GESTEL, N.; SIX, J.; VENTEREA, R.T.; van KESSEL, C. Productivity limits and potentials of the principles of conservation agriculture. **Nature**, v.517, p.365-368, 2015. DOI: https://doi.org/10.1038/nature13809.

R CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2020.

SALEKIN, S.; BURGESS, J.H.; MORGENROTH, J.; MASON, E.G.; MEASON, D.F. A comparative study of three non-geostatistical methods for optimising digital elevation model interpolation. **ISPRS International Journal of Geo-Information**, v.7, art.300, 2018. DOI: https://doi.org/10.3390/ijgi7080300.

SANTOS, A.M.R.T.; SANTOS, G.R. dos; EMILIANO, P.C.; MEDEIROS, N. das G.; KALEITA, A.L.; PRUSKI, L. de O.S. Detection of inconsistencies in geospatial data with geostatistics. **Boletim de Ciências Geodésicas**, v.23, p.296-308, 2017. DOI: https://doi.org/10.1590/S1982-21702017000200019.

SHUKLA, K.; KUMAR, P.; MANN, G.S.; KHARE, M. Mapping spatial distribution of particulate matter using kriging and inverse distance weighting at supersites of megacity Delhi. **Sustainable Cities and Society**, v.54, art.101997, 2020. DOI: https://doi.org/10.1016/j.scs.2019.101997.

SILVA, K.A. da; RODRIGUES, M.S.; CUNHA, J.C.; ALVES, D.C.; FREITAS, H.R.; LIMA, A.M.N. Levantamento de solos utilizando geoestatística em uma área de experimentação agrícola em Petrolina-PE. **Comunicata Scientiae**, v.8, p.175-180, 2017. DOI: https://doi.org/10.14295/cs.v8i1.2646.

SMITI, A. A critical overview of outlier detection methods. **Computer Science Review**, v.38, art.100306, 2020. DOI: https://doi.org/10.1016/j.cosrev.2020.100306.

SOUZA, Z.M. de; SOUZA, G.S. de; MARQUES JÚNIOR, J.; PEREIRA, G.T. Número de amostras na análise geoestatística e na krigagem de mapas de atributos do solo. **Ciência Rural**, v.44, p.261-268, 2014. DOI: https://doi.org/10.1590/S0103-84782014000200011.