PAB
Pesquisa Agropecuária Brasileira

ISSN 1678-3921

Journal homepage: www.embrapa.br/pab

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

Soil Science/ Original Article

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).

DOI: 10.1590/S1678-3921.pab2024.v59.03373

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

DOI: 10.1590/S1678-3921.pab2024.v59.03373

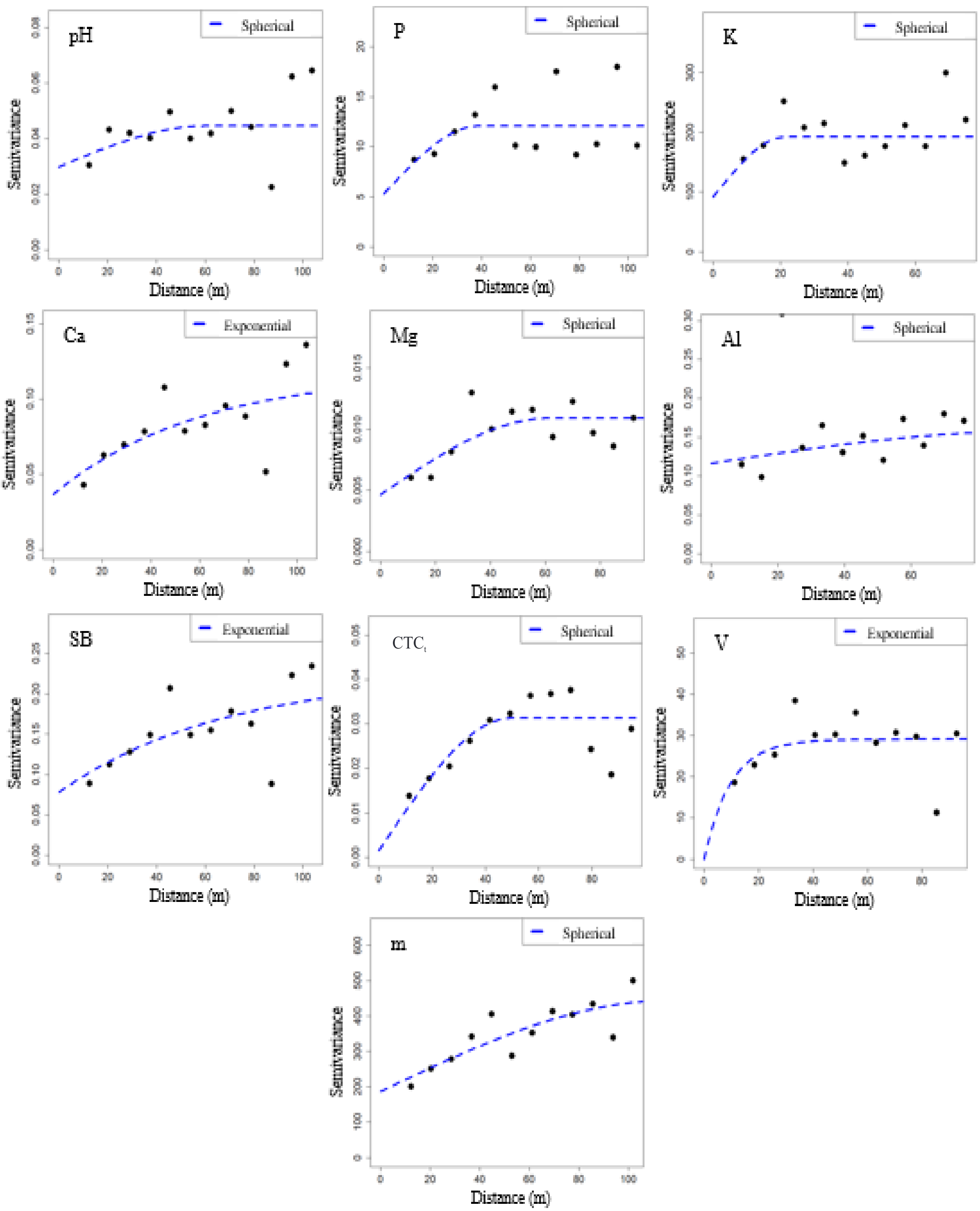

Figure 2. Semivariograms of the soil properties from a sugarcane (*Saccharum officinarum*) experimental area in the municipality of Oratórios, in the state of Minas Gerais, Brazil. SB, total exchangeable bases; CTC$_t$, effective cation exchange capacity; V, base saturation index; and m, aluminum saturation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