

**UNIVERSIDADE FEDERAL DE VIÇOSA**

**CALISTO MANUEL MÁQUINA**

**USO DA GEOESTATÍSTICA UNIVARIADA E MULTIVARIADA PARA  
ANÁLISE DE ATRIBUTOS DO SOLO**

**VIÇOSA – MINAS GERAIS**

**2021**

**CALISTO MANUEL MÁQUINA**

**USO DA GEOESTATÍSTICA UNIVARIADA E MULTIVARIADA PARA  
ANÁLISE DE ATRIBUTOS DO SOLO**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Doctor Scientiae*.

Orientador: Nerilson Terra Santos

Co-orientador: Gérson Rodrigues dos Santos

**VIÇOSA – MINAS GERAIS**

**2021**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade  
Federal de Viçosa - Campus Viçosa**

T

M297u Máquina, Calisto Manuel, 1984-  
2021      Usó da geoestatística univariada e multivariada para análise  
de atributos do solo / Calisto Manuel Máquina. – Viçosa, MG,  
2021.

50 f. : il. (algumas color.) ; 29 cm.

Inclui apêndices.

Orientador: Nerilson Terra Santos.

Tese (doutorado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 42-47.

1. Geologia - Métodos estatísticos. 2. Agricultura de  
precisão. I. Universidade Federal de Viçosa. Departamento de  
Estatística. Programa de Pós-Graduação em Estatística Aplicada  
e Biometria. II. Título.

CDD 22. 551.028

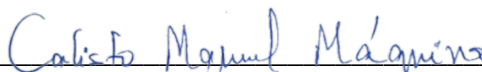
**CALISTO MANUEL MÁQUINA**

**USO DA GEOESTATÍSTICA UNIVARIADA E MULTIVARIADA PARA  
ANÁLISE DE ATRIBUTOS DO SOLO**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Doctor Scientiae*.

APROVADA: 22 de fevereiro de 2021.

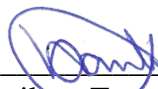
Assentimento:



---

Calisto Manuel Máquina

Autor



---

Nerilson Terra Santos

Orientador

*A Deus, fonte de proteção  
Divina, em nome de Jesus;  
aos meus pais, exemplos de  
maior apoio; e à minha filha.*

## AGRADECIMENTOS

A Deus, pela dádiva da vida e por me permitir realizar tantos sonhos nesta existência, em nome de Jesus Cristo, meu Salvador.

Aos meus pais Manuel do Rosário Máquina e Maria de Lurdes Muavihaca, pelo amor, incentivo e apoio incondicional.

À Universidade Federal de Viçosa, pela oportunidade de realizar o Curso.

Ao Professor Nerilson Terra Santos, pela orientação, pelo apoio e pela confiança.

Aos funcionários do PPESTBIO e, especialmente, ao Júnior Pires, pelo exemplo de profissionalismo.

Aos Professores Marcelo Costa e Daniel Marçal pela disponibilização dos dados utilizados nesta pesquisa.

Aos meus colegas do PPESTBIO, em especial a Zeferino, Lucas e Maurício, pelo companheirismo.

À minha filha Evelyn, pelo amor incondicional.

Este trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

## RESUMO

MÁQUINA, Calisto Manuel, D.Sc., Universidade Federal de Viçosa, fevereiro de 2021. **Uso da geoestatística univariada e multivariada para análise de atributos do solo.** Orientador: Nerilson Terra Santos. Co-orientador: Gérson Rodrigues dos Santos

A produção agrícola é uma das vias utilizadas para a melhoria da economia não só de cidadãos comuns, mas também de um país, devido à crescente procura de alimentos associada ao crescimento populacional. Assim, a aposta de grandes agricultores está na melhoria das técnicas de produção agrícola, tendo em conta não só o ganho financeiro, mas também a diminuição do impacto ambiental. A análise da variabilidade dos atributos do solo tem sido um dos focos para melhor controlar o uso dos insumos agrícolas. Contudo, poucos trabalhos focaram o efeito que a variabilidade espacial conjunta desses atributos tem no delineamento de zonas de manejo. Pelo exposto, o objetivo deste trabalho foi utilizar a geoestatística multivariada para o delineamento de zonas de manejo, considerando a variabilidade espacial conjunta de atributos que sabiamente têm efeito na produtividade de uma cultura. Foram identificados pares de atributos químicos que, por meio do algoritmo *Fuzzy c-means*, permitiram o mapeamento de zonas de manejo. O número ótimo de zonas de manejo de cada par de atributos, considerando três cenários diferentes, nomeadamente 100%, 75% e 50% de pontos amostrados dos atributos primários, foi determinado após a avaliação conjunta das funções Índice de Performance *Fuzzy* (FPI) e Entropia de Partição Modificada (MPE). O índice Kappa indicou concordância significativamente entre zonas de manejo dentro do mesmo par de atributos, o que atestou o uso da cokrigagem ordinária para o delineamento de zonas de manejo.

**Palavras-chave:** Agricultura de precisão. Delineamento de zonas de manejo. Estatística Espacial.

## ABSTRACT

MÁQUINA, Calisto Manuel, D.Sc., Universidade Federal de Viçosa, February, 2021. **Use of univariate and multivariate geostatistics to analyze soil attributes.** Advisor: Nerilson Terra Santos. Co-adviser: Gérson Rodrigues dos Santos

Farming production is one of the paths used to improve the economy not only of ordinary citizens but also of a country due to the growing demand for food associated with the rise in the population. Thus, the bet of large farmers is on improving farming production techniques, considering not only the financial gain but also the reduction of the environmental impact. The analysis of the variability of soil attributes has been one of the focuses to better control the use of agricultural inputs. However, few studies have focused on the effect that the combined spatial variability of these attributes has on the design of management zones. From the above, the objective of this work was to use multivariate geostatistics for the design of management zones, considering the joint spatial variability of attributes that wisely affect the productivity of a crop. Pairs of chemical attributes were identified and through the *Fuzzy c-means* algorithm, allowed the mapping of management zones. The optimal number of management zones for each pair of attributes, considering three different scenarios, namely 100%, 75%, and 50% of sampled points of the primary attributes, was determined after the joint evaluation of the Fuzzy Performance Index (FPI) and Modified Partition Entropy (MPE). The Kappa index indicated substantial and significant agreement between management zones within the same pair of attributes, which attested to the use of ordinary “cokriging” for the design of management zones.

**Keywords:** Precision agriculture. Management zones. Spatial Statistics.

## SUMÁRIO

<b>1. Introdução Geral.....</b>	<b>9</b>
1.1. Contextualização .....	9
1.2. Objetivo.....	10
1.3. Justificativa do trabalho .....	10
1.4. Hipótese do trabalho .....	11
1.5. Estrutura do trabalho .....	11
<b>2. Revisão Bibliográfica .....</b>	<b>12</b>
2.1. Abordagem da geoestatística univariada .....	12
2.1.1. Krigagem ordinária .....	14
2.2. Abordagem da geoestatística multivariada.....	18
2.2.1. Modelo Linear de Correogionalização .....	19
2.2.2. Cokrigagem ordinária .....	20
2.3. Validação cruzada.....	22
2.4. Delineamento de zonas de manejo .....	23
<b>3. Geoestatística multivariada como ferramenta alternativa no delineamento de zonas de manejo.....</b>	<b>26</b>
3.1. Introdução.....	26
3.2. Material e métodos.....	27
3.2.1. Descrição da área de estudo e coleta de dados.....	27
3.2.2. Análise estatística .....	27
3.2.2.1. Análise univariada dos atributos .....	27
3.2.2.2. Abordagem da Geoestatística Multivariada .....	28
3.2.2.3. Delineamento de zonas de manejo .....	30
3.3. Resultados e discussão .....	31
3.3.1. Estatística descritiva dos atributos .....	31
3.3.2. Análise de variabilidade espacial multivariada.....	33
3.3.2.1 Delineamento de zonas de manejo .....	36
<b>4. CONCLUSÕES .....</b>	<b>41</b>
<b>REFERÊNCIAS .....</b>	<b>42</b>
<b>APÊNDICES.....</b>	<b>48</b>
<b>APÊNDICE A – RESUMO DO ARTIGO ANÁLISE DE DADOS DE SOBREVIVÊNCIA ESPACIALMENTE CORRELACIONADOS .....</b>	<b>49</b>

<b>APÊNDICE B – USO DE ANÁLISE MULTIVARIADA NO DELINEAMENTO DE ZONAS DE MANEJO .....</b>	<b>50</b>
--	-----------

## 1. Introdução Geral

### 1.1. Contextualização

Nos últimos tempos, o setor agrônômico vem conhecendo maior crescimento devido ao constante avanço tecnológico que tem, desse modo, originado uma nova forma de prática de agricultura, denominada *Agricultura de Precisão*. A prática da agricultura de precisão permite ao agricultor o uso dos insumos agrícolas com maior precisão, aumento da lucratividade e redução dos impactos ambientais.

Diferentes autores, a exemplo de Molin *et al.* (2015), definem agricultura de precisão como um conjunto de ferramentas e tecnologias aplicadas ao gerenciamento agrícola baseado nas variabilidades espacial e temporal da unidade produtiva, para o aumento do retorno econômico e a redução dos impactos ambientais. Em geral, a definição da agricultura de precisão está associada ao uso das tecnologias de informação para o acompanhamento mais preciso dos fatores de produção.

A pressuposição da agricultura de precisão é que tanto a variabilidade espacial dos fatores que afetam a produtividade e a qualidade das culturas agrícolas quanto o manejo diferenciado em cada ponto de produção agrícola podem otimizar a lucratividade. Existem diferentes tipos de análises que podem ser consideradas para a verificação dessa pressuposição. Um desses tipos de análise se baseia na variabilidade espacial dos atributos químicos e físicos do solo. Muitos trabalhos descrevem os benefícios encontrados dentro da agricultura de precisão ao considerarem a análise da variabilidade espacial como fase importante para a percepção das práticas de manejo do solo. Por exemplo, Sabino *et al.* (2011) afirmam que o conhecimento da variabilidade espacial das propriedades de fertilidade do solo pode contribuir para a definição de melhores estratégias de manejo. Beckett (1971) e Cambardella e Karlen (1999) evidenciam a importância de levar em conta a avaliação preliminar da variabilidade espacial dos atributos químicos e físicos do solo para a prática da agricultura de precisão.

Uma estratégia efetiva para o manejo de atributos de solo consiste na definição de áreas dentro do campo de produção, onde os fatores de produção apresentam padrões similares. Essas áreas são também designadas por zonas de manejo (MOLIN; CASTRO, 2008; MOLIN *et al.*, 2015; KHAN *et al.*, 2020). Essa estratégia de manejo permite que o

custo dos insumos agrícolas e o impacto ambiental sejam minimizados, através do manejo diferenciado em diferentes áreas dentro do mesmo campo de produção agrícola.

A produtividade agrícola, a qual está diretamente relacionada aos atributos químicos e físicos do solo, requer que sejam observadas formas adequadas para o seu manejo, visando otimizar a produtividade. A avaliação da variabilidade espacial simultânea do conjunto de atributos químicos e físicos pode ser a melhor alternativa para caracterizar a fertilidade do solo e permitir a definição de subáreas dentro do campo de produção, onde os fatores de produção agrícola apresentam características homogêneas.

## **1.2. Objetivo**

Neste trabalho, procurou-se identificar grupos de atributos químicos, físicos e físico-químicos correlacionados entre si e mapeou-se a possível estrutura de dependência espacial presente em cada grupo de atributos, visando satisfazer o seguinte objetivo:

- Avaliar se a geoestatística multivariada pode auxiliar na definição de zonas de manejo.

## **1.3. Justificativa do trabalho**

O conhecimento *a priori* da variabilidade espacial de determinado atributo do solo de interesse, em um campo de produção agrícola, é de suma importância para o seu manejo adequado. No entanto, é comum dentro do mesmo campo de produção agrícola a existência de outros atributos do solo que afetam diretamente o atributo de interesse, o que torna imprescindível a sua inclusão na análise, para melhor percepção e manejo de fertilidade do solo.

A geoestatística multivariada, uma técnica pouco explorada para análise da variabilidade espacial de atributos do solo, permite a incorporação na modelagem de atributos secundários que apresentam correlação significativa com o atributo de interesse. Sendo assim, o estudo procura demonstrar a importância do uso da técnica de geoestatística multivariada no delineamento de zonas de manejo, visando contribuir nas metodologias aplicadas em agricultura de precisão.

#### **1.4. Hipótese do trabalho**

O uso da geoestatística multivariada pode melhorar a utilização da variabilidade espacial dos atributos do solo e auxiliar na definição das zonas de manejo do solo.

#### **1.5. Estrutura do trabalho**

Esta tese está organizada em três capítulos: Capítulo 1, onde é apresentada a introdução geral do trabalho. Neste capítulo são apresentados os seguintes elementos: *Contextualização do tema*, *Objetivo do trabalho*, *Justificativa do trabalho* e *Hipótese do trabalho*. Esses quatro elementos foram previamente apresentados.

O Capítulo 2 apresenta uma revisão sobre *Metodologias Geoestatísticas Univariada e Multivariada*, *Validação Cruzada* e *Delineamento de Zonas de Manejo*.

No Capítulo 3 é feita a aplicação da metodologia de geoestatística multivariada para o delineamento de zonas de manejo, utilizando dados reais de atributos de solo. São apresentadas as conclusões do trabalho, elaboradas de acordo com a hipótese estabelecida.

O trabalho também apresenta resumos dos artigos publicados durante a formação, subdivididos em dois apêndices: No Apêndice A é apresentado o resumo do artigo intitulado *Análise de Dados de Sobrevivência Espacialmente Correlacionados*, publicado na “Revista Brasileira de Biometria”; e o Apêndice B traz o resumo do artigo intitulado *Uso de Análise Multivariada no Delineamento de Zonas de Manejo*, publicado na “Revista Científica”.

## 2. Revisão Bibliográfica

### 2.1. Abordagem da geoestatística univariada

Nos estudos envolvendo dados de fertilidade do solo, procuram-se analisar os padrões espaciais ali presentes. Uma das técnicas que se ocupa nesse propósito é a geoestatística, devido à sua capacidade de detectar e mapear os padrões espaciais das variáveis regionalizadas (JOURNEL, A. G.; HUIJBREGTS, C. J., 1978; ISAACS; SRIVASTAVA, 1989; GOOVAERTS, 1997; WACKERNAGEL, 2003).

A técnica geoestatística tem sido muito utilizada por incorporar o efeito de dependência espacial, além de permitir a análise de múltiplas propriedades do solo simultaneamente, como explicam Mcbratney e Pringle (1999). Esta técnica se baseia no estabelecimento de um modelo para a dependência espacial para uma variável  $Z$  sendo estudada. Esta variável pode representar, por exemplo, a produção de uma cultura agrícola, um atributo (físico ou químico) do solo da região em estudo. Contudo, o valor da variável deve conter informações relativas à posição geográfica de onde foi obtida. Por essa razão, Matheron (1965) criou a teoria das variáveis regionalizadas.

A teoria de Matheron (1965) supõe que existe uma variável, digamos  $Z(x)$ , para cada uma das  $x$  possíveis posições dentro de uma área  $D$  em estudo. Além disso, esta teoria supõe que há um mecanismo de probabilidade que rege a dependência espacial entre os valores de  $Z$  de quaisquer duas posições  $x_i$  e  $x_i + h$ .

A modelagem da dependência espacial pode ser realizada por meio de estatísticas, calculadas a partir dos pareamentos dos valores  $Z(x_i)$  e  $Z(x_i + h)$ . Uma dessas estatísticas é a semivariância, a qual é definida, segundo Webster e Oliver (1990), Isaacs e Srivastava (1989) e Wackernagel (2003), como:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (1)$$

em que  $Z(x_i)$  e  $Z(x_i + h)$  representam os valores dos atributos observados nas posições  $x_i$  e  $x_i + h$ , e  $N(h)$  representa o número de pares de pontos amostrais, separados pela distância  $h$ . Pares de pontos mais próximos tendem a apresentar valores mais similares entre si do que aqueles pontos que se encontram mais distantes entre si.

Para estabelecer o modelo de dependência espacial, faz-se necessário, inicialmente, construir o semivariograma empírico, em que são plotados os valores da semivariância média para diferentes intervalos de distância  $h$  entre pares de pontos. O ajuste do modelo teórico é, então, realizado nos pontos do semivariograma empírico. Existem vários modelos que podem ser ajustados ao semivariograma empírico. Contudo, os modelos teóricos mais comuns são os modelos *linear*, *esférico*, *exponencial* e *gaussiano* (ISAACS; SRIVASTAVA, 1989; BURROUGH, 1993; GOOVAERTS, 1997; WACKERNAGEL, 2003; WEBSTER; OLIVER, 2007; YAMAMOTO; LAMDIM, 2013).

O modelo linear é utilizado no caso em que não se obtiveram dados distantes o suficiente para detectar independência entre os pontos amostrados e é representado da seguinte forma:

$$\gamma(h) = c_{0lin} + mh \quad (2)$$

em que  $c_{0lin}$  é efeito pepita do modelo linear, e  $h$  é a distância dos pares de pontos amostrais, que é afetada multiplicativamente pelo modelo linear ( $m$ ). No contexto geoestatístico, o termo  $m$  representa o grau de intransição do modelo.

O modelo esférico é utilizado para descrever os semivariogramas e é definido pela seguinte equação:

$$\gamma(h) = \begin{cases} c_{0sph} + c_{sph} \left\{ \left( \frac{3h}{2a} \right) - \frac{1}{2} \left( \frac{h}{a} \right)^3 \right\}, & h \leq a \\ c_{0sph} + c_{sph}, & h > a \end{cases} \quad (3)$$

em que  $c_{0sph} + c_{sph}$  representa o patamar do modelo e geralmente corresponde à variância dos dados da propriedade  $Z$ ,  $c_{0sph}$  é o efeito pepita do modelo esférico,  $c_{sph}$  é a contribuição da semivariância no patamar e  $a$  é o valor da amplitude, que corresponde à distância a partir da qual a dependência espacial não é mais observada.

O modelo exponencial é também utilizado para descrever a estrutura de dependência espacial de um fenômeno e é definido da seguinte forma:

$$\gamma(h) = c_{0exp} + c_{exp} \left\{ 1 - \exp \left( -\frac{h}{a} \right) \right\} \quad (4)$$

em que  $c_{0exp}$  é o efeito pepita do modelo exponencial e  $c_{exp}$  é a contribuição da semivariância no patamar do modelo esférico.

O modelo gaussiano é outro modelo bastante utilizado, principalmente em situações em que ocorrem fenômenos contínuos, e é definido pela seguinte equação:

$$\gamma(h) = c_{0gau} + c_{gau} \left\{ 1 - \exp\left(-\frac{h}{a}\right)^2 \right\} \quad (5)$$

em que  $c_{0gau}$  é efeito pepita do modelo gaussiano e  $c_{gau}$  é a contribuição da semivariância no patamar.

### 2.1.1. Krigagem ordinária

Após o ajuste do melhor modelo teórico sobre os valores de semivariância experimental, é realizada a interpolação dos valores dos atributos, em pontos não amostrados dentro da área considerada no estudo. Existem várias técnicas de interpolação que se baseiam na modelagem da estrutura de autocorrelação espacial, por meio de semivariogramas. Dentro dessas técnicas, a krigagem ordinária é a que mais se destaca não só por levar em conta a estrutura de dependência espacial, mas a facilidade de seu uso.

O preditor krigagem ordinária num ponto não amostrado  $x_0$  é obtido pela combinação linear dos valores das posições vizinhas à posição  $x_0$ , e os seus ponderadores (pesos), segundo Journel, A. G. e Huijbregts, C. J. (1978), Isaaks e Srivastava (1989), Oliver *et al.* (1989), Webster e Oliver (1990), Vieira (2000), Wackernagel (2003), Webster e Oliver (2007), Wu *et al.* (2009), Konak (2011), Yamamoto e Lamdim (2013), Silva e Costa (2016) e Pham *et al.* (2019), são obtidos pela equação seguinte:

$$\hat{z}(x_0) = \sum_{\alpha=1}^n \omega_{\alpha} z(x_{\alpha}), \quad \alpha = 1, \dots, n \quad (6)$$

em que  $\omega_{\alpha}$ ,  $\alpha = 1, \dots, n$  representa os pesos associados aos valores observados nos pontos da vizinhança do ponto  $x_0$  e  $z(x_{\alpha})$  é o valor da variável  $Z$ , observado nos pontos amostrados (no caso específico, as nossas variáveis são consideradas atributos físicos e químicos do

solo). Um aspeto de maior interesse é que  $\omega_\alpha$  pode assumir vários valores, no entanto os valores considerados ideais são aqueles que resultam em soma unitária, isto é:

$$\sum_{\alpha=1}^n \omega_\alpha = 1, \text{ para } i = 1, \dots, n. \quad (7)$$

A restrição imposta segundo a equação (7) garante que a esperança média do erro da predição seja zero e o preditor não seja enviesado, sendo este um dos pressupostos exigidos para que o modelo de semivariância seja considerado ótimo para efeitos de predição espacial. Outra pressuposição é que o erro do preditor da krigagem ordinária deve apresentar variância mínima, isto é, a diferença entre o valor predito e o verdadeiro valor em determinado ponto deve ser a menor possível.

Seja  $R(x_0)$  a diferença entre o valor predito e o verdadeiro valor no ponto não amostrado, o que corresponde ao erro do preditor da krigagem ordinária, isto é:

$$R(x_0) = \hat{Z}(x_0) - Z(x_0) \quad (8)$$

em que  $\hat{Z}(x_0)$  denota o valor predito e  $Z(x_0)$ , o verdadeiro valor no ponto não amostrado.

Por definição e sob a hipótese de igualdade de variância para todas as variáveis aleatórias, temos:

$$\begin{aligned} V\{R(x_0)\} &= V\{\hat{Z}(x_0) - Z(x_0)\} \\ &= (def.) = \frac{Cov\{\hat{Z}(x_0)\hat{Z}(x_0)\}}{V\{\hat{Z}(x_0)\}} - 2Cov\{\hat{Z}(x_0)Z(x_0)\} + \frac{Cov\{Z(x_0)Z(x_0)\}}{=\tilde{\sigma}^2} \end{aligned} \quad (9)$$

Observamos que  $Cov\{\hat{Z}(x_0)\hat{Z}(x_0)\} = V\{\hat{Z}(x_0)\} = V\{\sum_{\alpha=1}^n \omega_\alpha Z(x_\alpha)\}$  e, de acordo com Isaaks e Srivastava (1989), a variância de uma variável aleatória resultante da combinação linear ponderada de outras variáveis aleatórias é definida da forma que se segue:

$$V\left\{\sum_{\alpha=1}^n \omega_\alpha Z_\alpha\right\} = \sum_{\alpha=1}^n \sum_{\beta=1}^n \omega_\alpha \omega_\beta Cov\{Z_\alpha Z_\beta\} = \sum_{\alpha=1}^n \sum_{\beta=1}^n \omega_\alpha \omega_\beta \tilde{C}_{\alpha\beta} \quad (10)$$

em que  $\tilde{C}_{\alpha\beta} = Cov\{Z_\alpha Z_\beta\}$ .

Transformações algébricas no segundo termo da equação (9) podem resultar na seguinte equação:

$$2Cov\{\hat{Z}(x_0)Z(x_0)\} = 2 \sum_{\alpha=1}^n \omega_{\alpha} \tilde{C}_{\alpha 0} \quad (11)$$

Combinando as equações (9), (10) e (11), tem-se a seguinte equação, que descreve a função da variância do erro da krigagem ordinária:

$$\tilde{\sigma}_R^2 = \tilde{\sigma}^2 + \sum_{\alpha=1}^n \sum_{\beta=1}^n \omega_{\alpha} \omega_{\beta} \tilde{C}_{\alpha\beta} - 2 \sum_{\alpha=1}^n \omega_{\alpha} \tilde{C}_{\alpha 0} \quad (12)$$

Lembrando-se de que o objetivo é minimizar a expressão (12), que é o resultado das derivadas parciais de todos os termos ali presentes. Uma forma flexível para determinar tais derivadas está baseada no multiplicador de Lagrange (JOURNEL, A. G.; HUIJBREGTS, C. J., 1978; ISAAKS; SRIVASTAVA, 1989; WACKERNAGEL, 2003; YAMAMOTO; LAMDIM, 2013) e, ainda, sob a hipótese de que o erro do preditor da krigagem não enviesado, isto é,  $\sum_{\alpha=1}^n \omega_{\alpha} = 1$ , para  $i = 1, \dots, n$ , a equação (12) pode ser reescrita da seguinte forma:

$$\begin{aligned} L(\omega_1, \dots, \omega_n, \lambda) \\ = \tilde{\sigma}^2 + \sum_{\alpha=1}^n \sum_{\beta=1}^n \omega_{\alpha} \omega_{\beta} \tilde{C}_{\alpha\beta} - 2 \sum_{\alpha=1}^n \omega_{\alpha} \tilde{C}_{\alpha 0} + 2\lambda \underbrace{\left( \sum_{\alpha=1}^n \omega_{\alpha} - 1 \right)}_0 \end{aligned} \quad (13)$$

Calculando as derivadas parciais em relação aos pesos e igualando a zero em seguida, obtemos as seguintes equações:

$$\begin{aligned} \frac{\partial(L(\omega_1, \dots, \omega_n, \lambda))}{\partial \omega_1} &= 2 \sum_{\beta=1}^n \omega_{\beta} \tilde{C}_{1\beta} - 2\tilde{C}_{10} + 2\lambda = 0 \Rightarrow \sum_{\beta=1}^n \omega_{\beta} \tilde{C}_{1\beta} + \lambda = \tilde{C}_{10} \\ &\vdots \\ \frac{\partial(L(\omega_1, \dots, \omega_n, \lambda))}{\partial \omega_{\alpha}} &= 2 \sum_{\beta=1}^n \omega_{\beta} \tilde{C}_{\alpha\beta} - 2\tilde{C}_{\alpha 0} + 2\lambda = 0 \Rightarrow \sum_{\beta=1}^n \omega_{\beta} \tilde{C}_{\alpha\beta} + \lambda = \tilde{C}_{\alpha 0} \end{aligned} \quad (14)$$

⋮

$$\frac{\partial(L(\omega_1, \dots, \omega_n, \lambda))}{\partial \omega_n} = 2 \sum_{\beta=1}^n \omega_\beta \tilde{C}_{\alpha\beta} - 2\tilde{C}_{n0} + 2\lambda = 0 \Rightarrow \sum_{\beta=1}^n \omega_\beta \tilde{C}_{n\beta} + \lambda = \tilde{C}_{n0}$$

Derivando em função o multiplicador de Lagrange, temos:

$$\frac{\partial(L(\omega_1, \dots, \omega_n, \lambda))}{\partial \omega_1} = 2 \sum_{\alpha=1}^n \omega_\alpha - 2 = 0 \Rightarrow \sum_{\alpha=1}^n \omega_\alpha = 1 \quad (15)$$

Esse conjunto de equações pode ser resumidamente apresentado nas equações a seguir:

$$\sum_{\beta=1}^n \omega_\beta \tilde{C}_{\alpha\beta} + \lambda = \tilde{C}_{\alpha 0}, \quad \forall \alpha = 1, \dots, n \quad (16)$$

$$\sum_{\alpha=1}^n \omega_\alpha = 1 \quad (17)$$

As equações (16) e (17) juntas produzem o chamado sistema da krigagem ordinária e, sob a notação matricial, podem ser representadas como:

$$\underbrace{\begin{bmatrix} \tilde{C}_{11} & \cdots & \tilde{C}_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \tilde{C}_{n1} & \cdots & \tilde{C}_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}}_{(n+1) \times (n+1)} \cdot \underbrace{\begin{bmatrix} \omega_1 \\ \vdots \\ \omega_n \\ \lambda \end{bmatrix}}_{(n+1) \times 1} = \underbrace{\begin{bmatrix} \tilde{C}_{10} \\ \vdots \\ \tilde{C}_{n0} \\ 1 \end{bmatrix}}_{(n+1) \times 1} \quad (18)$$

$$\text{Definindo } \mathbf{C} = \begin{bmatrix} \tilde{C}_{11} & \cdots & \tilde{C}_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \tilde{C}_{n1} & \cdots & \tilde{C}_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}; \mathbf{w} = \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_n \\ \lambda \end{bmatrix} \text{ e } \mathbf{d} = \begin{bmatrix} \tilde{C}_{10} \\ \vdots \\ \tilde{C}_{n0} \\ 1 \end{bmatrix}$$

A relação (12) pode ser reescrita conforme a equação a seguir:

$$\mathbf{C} \cdot \mathbf{w} = \mathbf{d} \Rightarrow \mathbf{w} = \mathbf{C}^{-1} \cdot \mathbf{d} \quad (19)$$

em que  $\mathbf{C}$  é a matriz de variância-covariância entre os pontos amostrados,  $\mathbf{w}$  contém os vetores de pesos atribuídos aos valores dos pontos amostrados e o multiplicador de Lagrange e  $\mathbf{d}$  é o vetor de covariâncias entre os pontos amostrados e os não amostrados.

O procedimento é similar ao construir o sistema de krigagem ordinária via semivariância ( $\gamma(h)$ ), isto é:

$$\begin{aligned} \sigma_R^2 &= E[Z^*(x_0) - Z(x_0)]^2 \\ &= -\gamma(x_0 - x_0) - \sum_{\alpha=1}^n \sum_{\beta=1}^n \omega_\alpha \omega_\beta \gamma(x_\alpha - x_\beta) + 2 \sum_{\alpha=1}^n \omega_\alpha \gamma(x_\alpha - x_0) \end{aligned} \quad (20)$$

Ao minimizar a variância do erro do preditor pela equação (20), após a aplicação do procedimento de Lagrange, temos:

$$\begin{bmatrix} \gamma(x_1 - x_1) & \cdots & \gamma(x_1 - x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n - x_1) & \cdots & \gamma(x_n - x_n) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_n \\ -\lambda \end{bmatrix} = \begin{bmatrix} \gamma(x_1 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \\ 1 \end{bmatrix} \quad (21)$$

Definindo a matriz:

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma(x_1 - x_1) & \cdots & \gamma(x_1 - x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n - x_1) & \cdots & \gamma(x_n - x_n) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \text{ e os vetores } \mathbf{w} = \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_n \\ -\lambda \end{bmatrix}, \mathbf{g} = \begin{bmatrix} \gamma(x_1 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \\ 1 \end{bmatrix}$$

a relação da equação (21) pode ser assim reescrita (ISAAKS; SRIVASTAVA, 1989; WACKERNAGEL, 2003):

$$\mathbf{\Gamma} \mathbf{w} = \mathbf{g} \Rightarrow \mathbf{w} = \mathbf{\Gamma}^{-1} \mathbf{g} \quad (22)$$

em que  $\mathbf{\Gamma}$  denota a matriz de semivariâncias entre os pontos amostrados,  $\mathbf{w}$  denota o vetor de pesos atribuídos aos valores dos pontos amostrados e  $\mathbf{g}$  denota o vetor de semivariâncias entre os pontos amostrados e os não amostrados.

## 2.2. Abordagem da geoestatística multivariada

Os procedimentos de modelagem geoestatística apresentados anteriormente levam em consideração uma variável regionalizada, no entanto é comum o fato de tais variáveis

não apresentarem estimativas consistentes, sendo necessária a adoção de um mecanismo que permite a melhoria da consistência das estimativas. Um desses mecanismos se baseia na incorporação ao modelo mais de uma variável regionalizada, que é designada na modelagem por variável secundária ou covariável. Sendo assim, sob a hipótese de estacionariedade intrínseca conjunta, isto é, se para todos os pares de variáveis regionalizadas  $U$  e  $V$ :

$$E[Z_u(x+h) - Z_u(x)] = 0 \text{ e } cov[(Z_u(x+h) - Z_u(x)), (Z_v(x+h) - Z_v(x))] = 2\gamma_{u,v}(h)$$

A geoestatística multivariada, uma extensão da geoestatística univariada, surge com o objetivo de incorporar uma variável secundária na modelagem, com base na correlação espacial entre tal variável e a variável primária (ISAAKS; SRIVASTAVA, 1989; WACKERNAGEL, 2003; WEBSTER; OLIVER, 2007; CASTRIGNANÒ *et al.*, 2009; BUTTAFUOCO; GABRIELE *et al.*, 2010; YAMAMOTO; LAMDIM, 2013; LARK *et al.*, 2014).

### 2.2.1. Modelo Linear de Corregionalização

A definição de modelo linear de corregionalização constitui uma das principais fases para o prosseguimento com a análise da geoestatística multivariada, pois é a partir desta fase que se determina o modelo a ser utilizado para a interpolação e mapeamento dos valores interpolados do atributo de interesse.

O modelo linear de corregionalização é utilizado na modelagem do semivariograma univariado e do semivariograma cruzado.

Tal como visto na geoestatística univariada, a caracterização de cada variável regionalizada é feita a partir do respectivo semivariograma amostral. Assim, para cada par de variáveis regionalizadas é caracterizado pelo respectivo semivariograma cruzado amostral (ISAAKS; SRIVASTAVA, 1989; De IACO *et al.*, 2003; LARK; PAPRITZ, 2003; WACKERNAGEL, 2003; MACNAB, 2016).

O modelo de semivariograma cruzado experimental entre  $U$  e  $V$  espacialmente correlacionados, também conhecido como modelo linear de corregionalização, é definido, segundo Webster e Oliver (2007), por:

$$\hat{\gamma}_{u,v}(h) = \frac{1}{2N(h)} \sum_{j=1}^{N(h)} [\{Z_u(x_j) - Z_u(x_j + h)\}\{Z_v(x_j) - Z_v(x_j + h)\}] \quad (23)$$

em que  $N(h)$  representa o número de pares de pontos separados pela distância  $h$ .

O semivariograma cruzado pode assumir valores negativos, o que representa a correlação negativa entre as variáveis regionalizadas (JOURNAL, A. G.; HUIJBREGTS, C. J., 1978), isto é:

$$-\infty < \hat{\gamma}_{u,v}(h) < \infty \quad (24)$$

### 2.2.2. Cokrigagem ordinária

Uma vez definido o modelo linear de correogionalização a ser utilizado para predição de valores de variáveis em pontos não amostrados, segue-se para a fase de definição do método de predição a ser implementado. A cokrigagem ordinária é popularmente mais utilizada quando se trabalha com a técnica da geoestatística multivariada, segundo a abordagem feita por diferentes autores (JOURNAL, A. G.; HUIJBREGTS, C. J., 1978; FREUND, 1986; ISAAKS; SRIVASTAVA, 1989; GOOVAERTS, 1997; WACKERNAGEL, 2003; YALÇIN, 2005; WEBSTER; OLIVER, 2007; SILVA, A. *et al.*, 2010; CHILÈS; DELFINER, 2012; ADHIKARY *et al.*, 2017).

Este método apresenta vantagens em relação ao método da krigagem ordinária, por exemplo, permite o uso de uma segunda variável correlacionada com uma variável primária de interesse. Seu uso é recomendado quando a covariável é densamente distribuída, o que contribui para a obtenção dos resultados com maior precisão (CHICA-OLMO, 2007).

Sejam  $U$  e  $V$  dois atributos espacialmente distribuídos em determinada área, seja esta  $D$  e definindo  $U$  como atributo primário; então, o preditor de cokrigagem para  $U$  é definido (ISAAKS; SRIVASTAVA, 1989; WACKERNAGEL, 2003; WEBSTER; OLIVER, 2007) como:

$$\begin{aligned} \hat{Z}_u(x_0) &= \sum_{i=1}^n \omega_{1i} Z_u(x_{1i}) + \sum_{j=1}^m \omega_{2j} Z_v(x_{2j}) \\ &= \mathbf{w}_1^T \mathbf{z}_u + \mathbf{w}_2^T \mathbf{z}_v \end{aligned} \quad (25)$$

em que  $\hat{Z}_u(x_0)$  representa a estimativa da cokrigagem;  $\omega_{1i}$  e  $\omega_{2j}$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, m$  são os pesos associados aos valores dos atributos  $U$  e  $V$ , respectivamente;  $Z_u(x_{1i})$  e  $Z_v(x_{2j})$  são os valores dos atributos  $U$  e  $V$ , nas posições  $i$  e  $j$ , respectivamente;  $x_0$  é a posição para a qual se realiza a predição;  $n$  e  $m$  são números de vizinhos dos atributos  $U$  e  $V$ , respectivamente, ao ponto  $x_0$ .

Para que o erro do estimador de predição não seja viesado,  $\omega_{1i}$  e  $\omega_{2j}$  são definidos de modo que:

$$\mathbf{1}^T \mathbf{w}_1 = 1 \text{ e } \mathbf{1}^T \mathbf{w}_2 = 0 \quad (26)$$

Seja  $\mathbf{\Gamma}_{u,v}$  a matriz de semivariâncias (diretas e cruzadas) entre  $U$  e  $V$  amostrados em  $n$  e  $m$  pontos amostrais, respectivamente.  $\mathbf{\Gamma}_{u,v}$  é uma matriz de dimensão  $n \times m$  e é representada (WACKERNAGEL, 2003; WEBSTER; OLIVER, 2007) como:

$$\mathbf{\Gamma}_{u,v} = \begin{bmatrix} \gamma_{u,v}(x_{11}, x_{21}) & \cdots & \gamma_{u,v}(x_{11}, x_{2m}) \\ \vdots & \ddots & \vdots \\ \gamma_{u,v}(x_{1n}, x_{21}) & \cdots & \gamma_{u,v}(x_{1n}, x_{2m}) \end{bmatrix} \quad (27)$$

Sejam também  $\boldsymbol{\gamma}_{u,u/x_0}$  e  $\boldsymbol{\gamma}_{u,v/x_0}$  dois vetores de semivariâncias diretas e cruzadas, respectivamente, determinados entre os pontos amostrados e os não amostrados e definidos como a seguir (WACKERNAGEL, 2003; CHICA-OLMO, 2007; WEBSTER; OLIVER, 2007):

$$\boldsymbol{\gamma}_{u,u/x_0} = \begin{bmatrix} \gamma_{u,u}(x_{10}, x_{11}) \\ \cdots \\ \gamma_{u,u}(x_{10}, x_{1n}) \end{bmatrix} \quad (28)$$

e

$$\boldsymbol{\gamma}_{u,v/x_0} = \begin{bmatrix} \gamma_{u,v}(x_{10}, x_{21}) \\ \cdots \\ \gamma_{u,v}(x_{10}, x_{2m}) \end{bmatrix} \quad (29)$$

Das equações (27) a (29) é obtido o sistema de cokrigagem, que permite o cálculo dos pesos associados aos valores dos atributos  $U$  e  $V$ , em cada ponto amostrado ( $\omega_{1i}$  e

$\omega_{2j}, i = 1, \dots, n; j = 1, \dots, m$ ), respectivamente, que atendam às restrições impostas na equação (26), isto é:

$$\underbrace{\begin{bmatrix} & & & 1 & 0 \\ & \Gamma_{uu} & \Gamma_{uv} & \vdots & \vdots \\ & & & 1 & 0 \\ & & & 0 & 1 \\ & \Gamma_{vu} & \Gamma_{vv} & \vdots & \vdots \\ & & & 0 & 1 \\ 1 & \dots & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & 0 \end{bmatrix}}_G \cdot \underbrace{\begin{bmatrix} \omega_{11} \\ \vdots \\ \omega_{1n} \\ \omega_{21} \\ \vdots \\ \omega_{2m} \\ \lambda_u \\ \lambda_v \end{bmatrix}}_{\theta} = \underbrace{\begin{bmatrix} \gamma_{u,u}(x_{10}, x_{11}) \\ \vdots \\ \gamma_{u,u}(x_{10}, x_{1n}) \\ \gamma_{u,v}(x_{10}, x_{21}) \\ \vdots \\ \gamma_{u,v}(x_{10}, x_{2m}) \\ 1 \\ 0 \end{bmatrix}}_{\gamma} \quad (30)$$

em que  $\Gamma_{u,u}$  e  $\Gamma_{v,v}$  denotam matrizes de semivariâncias diretas dos atributos  $U$  e  $V$ , respectivamente;  $\Gamma_{u,v}$  e  $\Gamma_{v,u}$  denotam matrizes de semivariâncias cruzadas;  $\omega_{1i}$  e  $\omega_{2j}, i = 1, \dots, n; j = 1, \dots, m$  denotam os pesos associados aos atributos  $U$  e  $V$ , respectivamente;  $\gamma_{u,u}(x_{10}, x_{1i})$  e  $\gamma_{u,v}(x_{10}, x_{2j}), i = 1, \dots, n; j = 1, \dots, m$  são elementos dos vetores  $\boldsymbol{\gamma}_{u,u/x_0}$  e  $\boldsymbol{\gamma}_{u,v/x_0}$ , respectivamente;  $\lambda_u$  e  $\lambda_v$  são multiplicadores de Lagrange associados aos atributos  $U$  e  $V$ , respectivamente.

Resumidamente, o sistema apresentado na equação (30) pode ser reescrito da seguinte forma:

$$G \cdot \boldsymbol{\theta} = \boldsymbol{\gamma} \Rightarrow \boldsymbol{\theta} = G^{-1} \cdot \boldsymbol{\gamma} \quad (31)$$

em que  $G$  denota matriz aumentada de semivariâncias diretas e cruzadas nos pontos amostrados;  $\boldsymbol{\theta}$  denota o vetor composto por pesos e multiplicadores de Lagrange; e  $\boldsymbol{\gamma}$  denota o vetor de semivariâncias diretas e cruzadas, obtidas entre os pontos amostrados e os não amostrados.

### 2.3. Validação cruzada

A validação cruzada tem sido utilizada desde os anos de 1959, de acordo com Clark (1986). Foi utilizado por Krige (1959) para comparar os valores preditos e observados. Do ponto de vista geoestatístico, a validação cruzada é uma técnica utilizada para examinar a qualidade de ajuste dos modelos de dependência espacial. É uma das formas utilizadas para

verificar a qualidade do ajuste do modelo empregado para interpolação, tanto por krigagem quanto por cokrigagem. Geralmente, esta técnica se baseia na avaliação simultânea dos valores da média e da variância do erro da estimativa da krigagem ou da cokrigagem (WEBSTER; OLIVER, 2007; PLANT, 2018), sob a hipótese de variância mínima do estimador do erro da predição, isto é:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \{Z(x_i) - \hat{Z}(x_i)\}^2} \quad (32)$$

em que RMSE representa a *raiz quadrada do erro quadrado médio* da interpolação por krigagem ou cokrigagem,  $N$  é o número de valores preditos,  $Z(x_i)$  denota os verdadeiros valores dos atributos nos locais preditos e  $\hat{Z}(x_i)$  denota os valores preditos.

Para considerar um modelo de maior performance, espera-se que o estimador do erro de krigagem ou cokrigagem tenha variância mínima. Assim, valor de  $RMSE \cong 0$  é indicativo de resultado satisfatório para a validação cruzada.

Adicionalmente à medida da RMSE, a validação cruzada pode ser realizada segundo Vieira (2000), analisando os valores dos coeficientes lineares alpha ( $\hat{\beta}_0 \cong 0$ ) e beta ( $\hat{\beta}_1 \cong 1$ ) da regressão linear simples entre os valores observados e os valores preditos pelo método da krigagem ou da cokrigagem ordinária.

#### 2.4. Delineamento de zonas de manejo

O conceito de zonas de manejo no âmbito da agricultura de precisão refere-se às sub-regiões do campo de produção, onde as atribuições sobre as condições da produtividade são homogêneas (LI *et al.*, 2007; MORAL GARCÍA *et al.*, 2010; RODRIGUES *et al.*, 2015). A definição de zonas de manejo não só torna o processo de manejo facilitado, como também economiza os insumos agrícolas.

Em diversos trabalhos são apresentadas diferentes formas de delineamento de zonas de manejo, a exemplo do estudo de Molin *et al.* (2015), em que, entre os vários métodos abordados, consta o método de normalização dos dados pela média. Uma versão

metodológica de agrupamento encontrada na literatura e considerada o melhor método de clusterização é o algoritmo *Fuzzy c-means*. Trata-se de um método não supervisionado de aprendizado de máquina, inicialmente idealizado por Ruspini (1969), que se tornou bastante competitivo no delineamento de zonas de manejo (BEZDEK *et al.*, 1984; MCBRATNEY; DEGRUIJTER, 1992; BURROUGH *et al.*, 2001; GORSEVSKI *et al.*, 2003; LI *et al.*, 2007; RODRIGUES; CORÁ, 2015).

O algoritmo *Fuzzy c-means* baseia-se na minimização da chamada função de mínimos quadrados generalizada dos erros,  $J_m$ , que é uma medida ponderada da distância do valor a clusterizar e o centroide da classe (GORSEVSKI *et al.*, 2003; LI *et al.*, 2007), isto é:

$$J_m = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m (d_{ik})^2 \quad (33)$$

em que  $u_{ik} \in \{0; 1\}$  são valores das entradas da matriz  $\mathbf{U}$  de associação *Fuzzy* que indica associação do  $i$ -ésimo objeto, na  $k$ -ésima classe;  $m \in (1; \infty)$  é o expoente de ponderação *Fuzzy*, conhecido também por coeficiente de nebulosidades, e sua função é controlar o peso atribuído a cada quadrado de erro;  $(d_{ik})^2$  é o quadrado da medida de distância considerada, que pode ser euclidiana, de Mahalanobis ou distância diagonal e  $n$  número de objetos que se pretendem classificar; e  $c$  é o número de classes considerado na clusterização.

A maior dificuldade encontrada com o emprego desse método está relacionada à forma de como determinar o valor ótimo de número  $c$  a considerar no algoritmo. Essa é uma das vantagens do algoritmo *Fuzzy c-means* em relação aos outros métodos de clusterização, pois *Fuzzy c-means* utiliza valor que minimiza o *Fuzzyness*, obtido através da avaliação simultânea de duas funções: *Índice de Performance Fuzzy* (FPI) e *Entropia de Partição Modificada* (MPE), definidos pelas equações a seguir (ODEH *et al.*, 1992; LARK; STAFFORD, 1997; GORSEVSKI *et al.*, 2003; LI *et al.*, 2007):

$$FPI = 1 - \frac{c}{c-1} \left[ 1 - \frac{\sum_{k=1}^n \sum_{i=1}^c (u_{ik})^2}{n} \right] \quad (34)$$

e

$$MPE = \frac{n}{n-c} \left[ - \frac{\sum_{k=1}^n \sum_{i=1}^c u_{ik} \log(u_{ik})}{n} \right] \quad (35)$$

A função do índice de desempenho *Fuzzy* estima o grau de imprecisão gerado por um número específico de classes, enquanto a entropia de partição modificada estima o grau de desorganização criado por um número de classes. Assim, o número ótimo de classes é o  $c$ , para o qual FPI e MPE atingem valores mínimos.

### **3. Geoestatística multivariada como ferramenta alternativa no delineamento de zonas de manejo**

#### **3.1. Introdução**

A prática da Agricultura de Precisão vem cada vez mais ganhando espaço no seio de agricultores, devido às numerosas vantagens relacionadas com a sustentabilidade econômica e ambiental. A busca pelo conhecimento sobre a variabilidade espacial dos atributos do solo em determinado campo de produção agrícola tornou-se um dos fatores-chave para a boa prática da agricultura de precisão, pois permite a identificação das áreas com menor e maior variabilidade de atributos, o que torna possível a prática do gerenciamento diferenciado do solo.

Em diversos campos de produção agrícola, é comum a mensuração de vários atributos do solo, o que demanda métodos multivariados, tal como a cokrigagem para melhor analisar o padrão de variabilidade espacial conjunta de tais atributos (ISAACS; SRIVASTAVA, 1989; WACKERNAGEL, 2003; GOOVAERTS, 1997; YAMAMOTO; LAMDIM, 2013).

A cokrigagem ordinária, um método de interpolação encontrado dentro da geoestatística multivariada, tem sido utilizada para a predição dos valores dos atributos do solo em áreas não amostradas, uma vez que leva em consideração a estrutura conjunta da dependência espacial. O método permite também a construção de mapas de clusters, que possibilitam a identificação de zonas com padrão espacial homogêneo dentro do campo de produção agrícola, segundo Lima *et al.* (2016).

O uso da cokrigagem ordinária permite também a identificação dos grupos de atributos do solo que apresentam padrão espacial semelhante, tornando flexível o manejo adequado do solo. Sendo assim, o objetivo principal deste capítulo foi avaliar se a geoestatística multivariada pode auxiliar na definição das zonas de manejo. O objetivo secundário foi avaliar a redução da densidade amostral do atributo primário ao utilizar a cokrigagem.

## 3.2. Material e métodos

### 3.2.1. Descrição da área de estudo e coleta de dados

Os dados considerados nesta pesquisa foram obtidos numa fazenda localizada no município de Goianápolis, no estado de Goiás, Brasil, e originalmente utilizados no trabalho de Costa e Queiroz (2011). A fazenda situa-se nas coordenadas médias de 16° 28' 20" S e 49° 00' 32" N, de acordo com Costa e Queiroz (2011).

A coleta de dados foi feita numa grade de amostragem regular com 150 pontos, em uma área de 75 hectares (Figura 1). O receptor Global Navigation Satellite System (GNSS), o GPS topográfico da Marca Magellan, modelo Promark 3, foram utilizados para o georreferenciamento dos pontos amostrais.

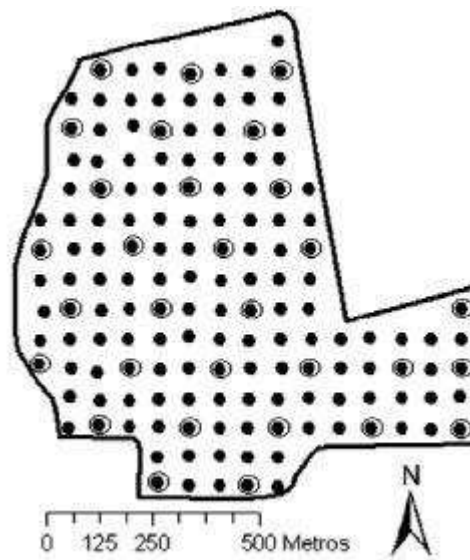


Figura 1 – Grade de amostragem dos 150 pontos amostrados numa área de 75 hectares localizada numa fazenda no município de Goianápolis-GO – Brasil.

Fonte: COSTA; QUEIROZ, 2011.

Em cada um dos 150 pontos amostrados foram mensurados quatro atributos químicos do solo, nomeadamente magnésio ( $Mg^{+2}$  ( $cmol_c dm^{-3}$ )), acidez potencial ( $H+Al$  ( $cmol_c dm^{-3}$ )), capacidade de troca de cátions pH 7,0 ( $T$  ( $cmol_c dm^{-3}$ )) e matéria orgânica ( $MO$  ( $dag kg^{-1}$ )) e dois atributos físicos do solo, nomeadamente argila (%) e areia (%).

### 3.2.2. Análise estatística

#### 3.2.2.1. Análise univariada dos atributos

A análise descritiva de dados é uma das primeiras etapas, uma vez que permite o conhecimento de estatísticas, bem como da distribuição dos dados. Assim, inicialmente foi realizada a análise descritiva, com o objetivo de descrever estatisticamente as distribuições dos valores mensurados dos atributos (variáveis), através da análise dos respectivos histogramas e dos valores da média, mediana, desvio-padrão e coeficiente de variação. Determinou-se também o grau de relacionamento entre os atributos físico-químicos do solo. A existência de relacionamento entre atributos é uma das pressuposições da geoestatística multivariada.

### 3.2.2.2. Abordagem da Geoestatística Multivariada

A geoestatística multivariada, uma extensão da geoestatística univariada segundo Goovaerts (1997) e Wackernagel (2003), foi desenvolvida através do método de predição espacial denominado cokrigagem. Uma das premissas para a utilização da cokrigagem, de acordo com Yamamoto e Lamdim (2013), consiste na existência de altas correlações entre os atributos primário e secundário para que as predições sejam consistentes. Assim, neste trabalho foram utilizados pares de atributos que apresentaram correlação linear de Pearson ( $r$ ), tal que  $r \geq 0,5$ , de acordo com a recomendação de Cohen (1988).

O método de predição da cokrigagem, comumente chamado de cokrigagem ordinária, baseia-se em dois tipos de funções de interesse: (i) função de semivariograma univariado, que é utilizada para descrever a variabilidade espacial de um atributo de interesse; e (ii) função de semivariograma cruzado, que descreve a variabilidade espacial conjunta de dois atributos (ISAAKS; SRIVASTAVA, 1989; WEBSTER; OLIVER, 2007).

Sejam  $Z_u(x)$  e  $Z_v(x)$  duas variáveis regionalizadas que representam, respectivamente, os possíveis valores dos atributos  $U$  e  $V$  na posição  $x$ . Supondo que  $U$  e  $V$  sejam espacialmente altamente correlacionados, sob a hipótese de estacionariedade intrínseca conjunta das  $P$  funções aleatórias, tal que  $\forall x, x + h \in D$ , em que  $D$  denota a área em estudo e  $h$  a distância entre dois pontos amostrados, tem-se, segundo Wackernagel (2003), que:

$$E[Z_u(x) - Z_u(x + h)] = 0 \quad (36)$$

e

$$cov[(Z_u(x) - Z_u(x + h))(Z_v(x) - Z_v(x + h))] = 2\gamma_{u,v}(h) \quad (37)$$

A partir das equações (36) e (37) é definido o modelo de função de semivariograma univariado para o atributo  $U$ , segundo Webster e Oliver (2007), como:

$$\gamma_{u,u}(h) = \frac{1}{2}E[\{Z_u(x) - Z_u(x+h)\}^2] \quad (38)$$

em que  $h$  é a distância que separa dois pontos que contêm valores de  $U$ , e o modelo para o atributo  $V$  é definido de modo semelhante ao apresentado na equação (38).

O modelo de semivariograma cruzado que descreve a relação espacial de dois atributos é também definido como na equação (38), porém leva em conta dois atributos diferentes, isto é:

$$\gamma_{u,v}(h) = \frac{1}{2}E[\{Z_u(x) - Z_u(x+h)\}\{Z_v(x) - Z_v(x+h)\}] \quad (39)$$

O modelo de semivariograma cruzado experimental entre  $U$  e  $V$ , conhecido também como modelo linear de correção regionalização, é definido, segundo Webster e Oliver (2007), por:

$$\hat{\gamma}_{u,v}(h) = \frac{1}{2N(h)} \sum_{j=1}^{N(h)} [\{Z_u(x_j) - Z_u(x_j+h)\}\{Z_v(x_j) - Z_v(x_j+h)\}] \quad (40)$$

em que  $N(h)$  representa o número de pares de pontos separados pela distância  $h$ . Na literatura de geoestatística, a variável  $U$  na equação (40) é comumente chamada de variável primária e  $V$ , variável secundária ou covariável.

A predição por cokrigagem resulta da combinação linear dos valores dos atributos primário e secundário e é definida, segundo Isaaks e Srivastava (1989) e Silva, S. A. *et al.* (2010), por:

$$\hat{Z}_u(x_0) = \sum_{i=1}^n \omega_{1i} Z_u(x_{1i}) + \sum_{j=1}^m \omega_{2j} Z_v(x_{2j}) \quad (41)$$

em que  $\hat{Z}_u(x_0)$  representa a estimativa da cokrigagem;  $\omega_{1i}$  e  $\omega_{2j}$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, m$  são os pesos associados aos valores dos atributos  $U$  e  $V$ , respectivamente;  $Z_u(x_{1i})$  e  $Z_v(x_{2j})$  são os valores dos atributos  $U$  e  $V$  nas posições  $i$  e  $j$ , respectivamente;  $x_0$  é a posição para a qual se realiza a predição; e  $n$  e  $m$  são números de vizinhos dos atributos  $U$  e  $V$ , respectivamente.

Os valores dos pesos  $\omega_{1i}$  e  $\omega_{2j}$  são determinados de tal modo que a estimativa do erro da cokrigagem não seja tendenciosa e a variância da estimativa do erro da cokrigagem seja mínima. Assim, para observação das duas condições, são definidas as seguintes restrições sobre os pesos  $\omega_{1i}$  e  $\omega_{2j}$  (ISAAKS; SRIVASTAVA, 1989; WACKERNAGEL, 2003 ; SILVA, S. A. *et al.*, 2010):

$$\sum_{i=1}^n \omega_{1i} = 1 \quad (42)$$

e

$$\sum_{j=1}^m \omega_{2j} = 0 \quad (43)$$

Existem trabalhos da área de agricultura de precisão que se basearam no procedimento geoestatístico de cokrigagem como método de interpolação de valores dos atributos do solo, a exemplo os trabalhos de Wu *et al.* (2009), Silva, A. *et al.* (2010) e Adhikary *et al.* (2017).

### 3.2.2.3. Delineamento de zonas de manejo

Com o objetivo de avaliar o adensamento amostral do atributo primário nos resultados obtidos por cokrigagem, foram criados três cenários que representam diferentes proporções dos pontos amostrados na área. Em um dos cenários, todos os pontos amostrados, ou seja, 100% dos pontos amostrados em que foram observados o atributo primário, foram selecionados. Outro cenário foi aquele em que 75% dos pontos foram selecionados. Em outro cenário, 50% dos pontos foram selecionados. Menores densidades de amostragem não foram utilizadas porque o número de pontos não seria suficiente para a modelagem da dependência espacial, de acordo com Webster e Oliver (1992). Segundo esses autores, variogramas calculados em menos de 50 observações não apresentam boas estimativas pela krigagem e pelo menos 100 a 150 observações são necessárias. Após essa seleção foram ajustados modelos de semivariância-cruzado para cada um daqueles cenários, e os mapas de interpolação por cokrigagem foram obtidos.

A partir dos mapas de interpolação, foi utilizado o método de agrupamento *Fuzzy c-means* para delinear as zonas de manejo. Este método não supervisionado particionou todas

as observações em  $c$  clusters, em que cada observação foi automaticamente atribuída a um grupo mais próximo da média, como na distância euclidiana, segundo o trabalho de Gorsevski *et al.* (2003). O número  $c$  ótimo de partições foi definido a partir da avaliação conjunta de duas funções, índice de performance *Fuzzy* (FPI), que representa o grau de *fuzziness* (imprecisão) gerado por um número específico de partições; e a entropia de partição modificada (MPE), que representa o grau de desorganização criada também por um número específico de partições. Assim, de acordo com Gorsevski *et al.* (2003), o número ótimo de partições foi estabelecido para  $c$ , em que as duas funções apresentaram valores mínimos.

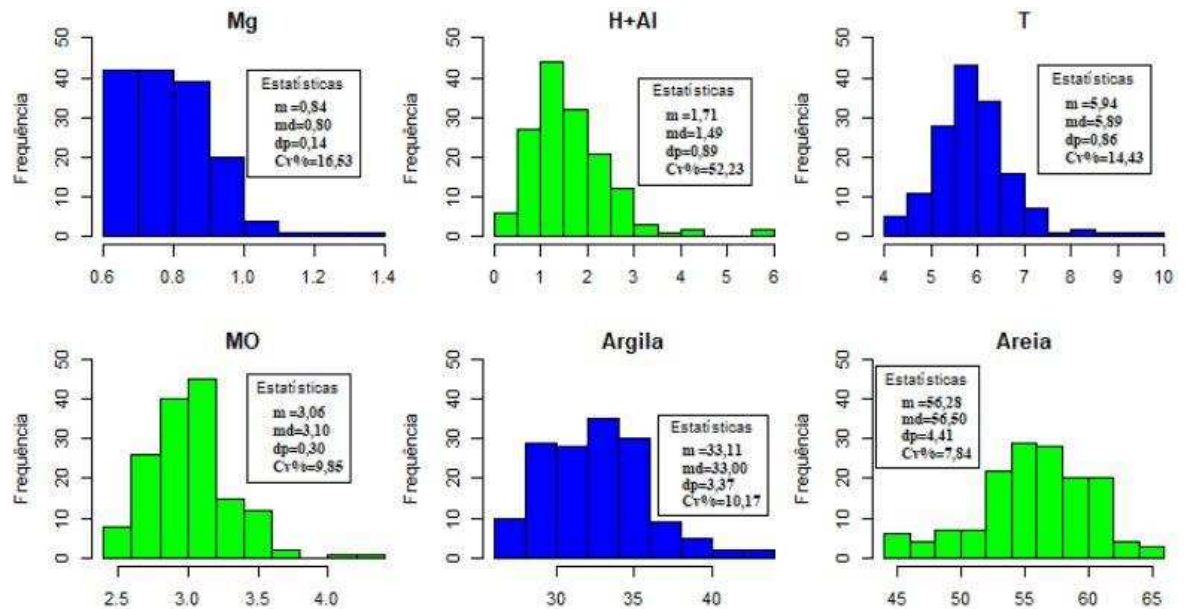
Após o delineamento das zonas de manejo foi avaliada a concordância daquelas zonas obtidas nos cenários com 100% dos pontos amostrados para o atributo primário com aquelas zonas de manejo conseguidas nos cenários com 75% e 50% dos pontos amostrados para o atributo primário. A avaliação da concordância foi realizada a partir da obtenção do coeficiente Kappa (CONGALTON; MEAD, 1986), bem como para o teste de hipóteses para conhecer a sua significância. Tal análise de concordância teve por objetivo verificar se as zonas de manejo obtidas com dados subamostrados (incompletos no atributo primário) apresentam padrão de variabilidade espacial similar ao padrão espacial das zonas de manejo obtidas com dados completos e dentro do mesmo par de atributos. As medidas FPI, MPE e Kappa são comumente utilizadas em trabalhos de agricultura de precisão (FRIDGEN *et al.*, 2003; GORSEVSKI *et al.*, 2003; LI *et al.*, 2007; COSTA; QUEIROZ, 2011; MÁQUINA *et al.*, 2020).

### **3.3. Resultados e discussão**

#### **3.3.1. Estatística descritiva dos atributos**

Os atributos magnésio, acidez potencial, capacidade de troca catiônica, matéria orgânica e argila tenderam a apresentar distribuição assimétrica positiva, enquanto o atributo físico areia mostrou tendência para a distribuição assimétrica negativa (Figura 2). O tipo de distribuição dos atributos do solo pode estar relacionado ao solo da área em estudo. Utilizando as classes de dispersão propostas por Warrick e Nielsen (1980), verificou-se que, de acordo com os valores do coeficiente de variação observados, os atributos químicos tenderam a apresentar maior dispersão que os atributos físicos.

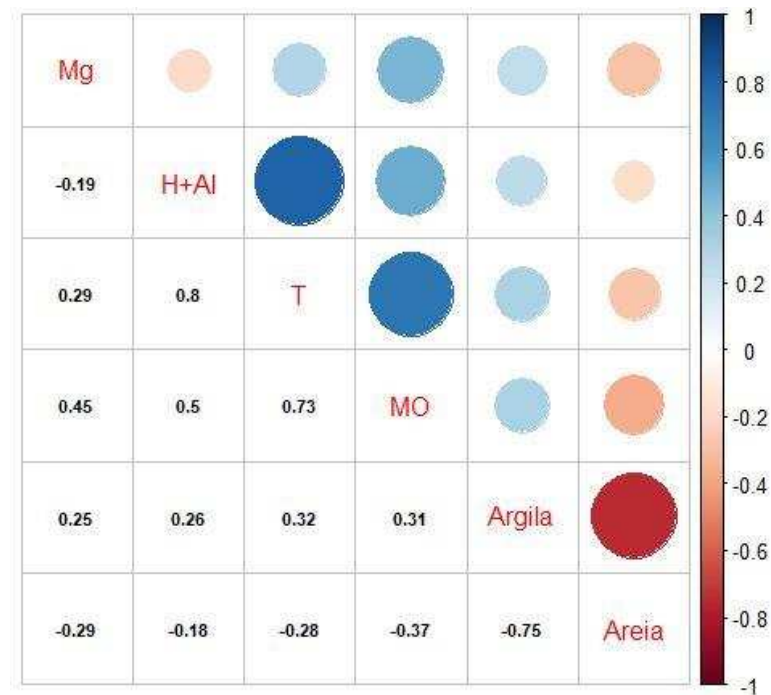
De acordo com Santos *et al.* (2017) e Máquina *et al.* (2020), distribuições assimétricas são características comuns dos atributos do solo em condições tropicais, principalmente em Latossolos Vermelhos.



m = Média; md = Mediana; dp = Desvio-padrão; e Cv = Coeficiente de variação (%).

Figura 2 – Estatísticas descritivas dos atributos químicos Mg ( $\text{cmol}_c \text{dm}^{-3}$ ) = magnésio; H+Al ( $\text{cmol}_c \text{dm}^{-3}$ ) = acidez potencial; T ( $\text{cmol}_c \text{dm}^{-3}$ ) = capacidade de troca de cátions  $\text{pH } 7,0$ ; e MO ( $\text{dag kg}^{-1}$ ) = matéria orgânica; e atributos físicos argila (%) e areia (%).

São apresentadas na Figura 3 as correlações entre todos os pares possíveis dos seis atributos físico-químicos considerados nesta pesquisa, com o objetivo de extrair os pares de atributos para a realização da geoestatística multivariada.



Mg = Magnésio; H+Al = Acidez potencial; T = Capacidade de troca de cátions pH 7,0; e MO = Matéria orgânica.

Figura 3 – Matriz de correlações dos seis atributos físico-químicos do solo considerados na pesquisa.

Assim, os pares de atributos selecionados que apresentaram fortes correlações ( $r \geq |0,5|$ ), de acordo com a classificação de Cohen (1988), foram: capacidade de troca de cátions pH 7,0 *vs.* matéria orgânica; capacidade de troca de cátions pH 7,0 *vs.* acidez potencial; acidez potencial *vs.* matéria orgânica; e argila *vs.* areia.

Para este trabalho, os atributos físico-químicos matéria orgânica e capacidade de troca de cátions pH 7,0, físico e argila foram considerados como as variáveis primárias para a realização da cokrigagem. Portanto, de acordo com o critério de correlação previamente estabelecido e as variáveis primárias determinadas, a modelagem e interpolação da cokrigagem foram aplicadas aos seguintes pares de atributos: capacidade de troca de cátions pH 7,0 *vs.* acidez potencial; matéria orgânica *vs.* acidez potencial; e argila *vs.* areia.

### 3.3.2. Análise de variabilidade espacial multivariada

Os resultados dos ajustes dos modelos lineares de correlogramas para cada um dos três pares de atributos identificados anteriormente são apresentados na Tabela 1.

Tabela 1 – Modelos teóricos de semivariogramas e respectiva estatística de validação cruzada

Par de atributos	Cenário	Modelo de semivariância	Parâmetro de semivariância			RMS
			$a$	$C_0$	$C_0 + C_1$	
T vs. (H+Al)	50	Exponencial	733,83	0,14	1,67	0,58
	75	Gaussiano	320,90	0,27	0,57	0,51
	100	Gaussiano	291,09	0,12	0,27	0,40
MO vs. (H+Al)	50	Esférico	565,60	0,14	0,27	0,41
	75	Esférico	367,58	0,13	0,21	0,37
	100	Esférico	405,90	0,03	0,17	0,19
Argila vs. Areia	50	Esférico	255,40	10,61	-3,93	1,28
	75	Esférico	288,11	12,52	-9,41	4,01
	100	Esférico	308,32	-1,44	-11,76	2,11

H+Al = acidez potencial; T = capacidade de troca de cátions pH 7,0; MO = matéria orgânica;  $a$  = alcance, em metros;  $C_0$  = efeito pepita;  $C_0 + C_1$  = patamar; *RMSE* = raiz quadrada do erro quadrático médio da validação cruzada; e cenário = % dos pontos amostrados do atributo primário.

Para cada um dos três cenários do par de atributos químicos T vs. (H+Al), foi possível construir um modelo transitivo, isto é, um modelo de semivariância com patamar. Para o cenário com 50% dos pontos do atributo primário foi ajustado o modelo exponencial, o qual, entre todos os modelos ajustados, apresentou matematicamente os maiores valores de alcance e patamar, o que está de acordo com as características gerais do modelo exponencial, como explicam Isaaks e Srivastava (1989). De acordo com estes autores, o modelo de semivariância exponencial atinge o patamar assintoticamente, e o valor do alcance prático é definido como a distância na qual o valor de semivariograma é 95% do patamar.

O modelo esférico foi o que apresentou o melhor ajuste para todos os cenários considerados do par MO vs. (H+Al). O cenário considerando 50% dos pontos amostrados para o atributo primário mostrou maior valor de alcance. Do ponto de vista geoestatístico, maior valor de alcance observado reflete maior distância dentro da qual é observada a dependência espacial entre os atributos que compõem o par.

Os modelos ajustados nos três cenários para cada um dos pares T vs. (H+Al) e MO vs. (H+Al) apresentaram tendência de aumento dos valores de efeito pepita e patamar à medida que o número de pontos amostrados para atributo primário diminui. Essa semelhança observada pode ser justificada pelo mesmo atributo secundário utilizado nos dois pares.

Foi possível também ajustar modelos de semivariância para os três cenários do par de atributos de textura do solo (argila *vs.* areia). Os valores negativos observados nos parâmetros de efeito pepita ( $C_0$ ) e o patamar ( $C_0 + C_1$ ) são devidos à correlação negativa entre os atributos argila e areia. Do ponto de vista geoestatístico, tais atributos tendem a apresentar padrões de variabilidade espacial divergentes, tal como explicam Castrignanò *et al.* (2009).

Os valores da raiz quadrada do erro quadrado médio (*RMSE*) observados na Tabela 1, obtidos em todos os cenários dos pares dos atributos capacidade de troca de cátions pH 7,0 *vs.* acidez potencial e matéria orgânica *vs.* acidez potencial, tendem a apresentar aumento do *RMSE* para cenários com menores pontos amostrados para o atributo primário. Contudo, nos três cenários do par argila *vs.* areia, os valores de *RMSE* observados na Tabela 1 são razoavelmente superiores em relação aos dos cenários dos demais pares de atributos. Resultados do *RMSE* similares foram obtidos por Gozdowski *et al.* (2015) em trabalho para avaliar a acurácia dos métodos de interpolação dos valores da textura do solo.

Os valores de *RMSE* (Tabela 1) próximos de zero para a maioria dos três modelos de correionalização, em cada um dos três pares de atributos de solo, indicam que tais modelos podem ser utilizados para produzir interpolações para os atributos primários capacidade de troca de cátions pH 7,0 e matéria orgânica, usando o atributo acidez potencial como atributo secundário; assim como a produção de um mapa de interpolação para o atributo primário argila, usando o atributo como atributo secundário areia. A performance observada nos modelos de semivariância, em todos os cenários dos três pares de atributos, é também sustentada por Webster e Oliver (1992). Segundo estes autores, modelos de semivariâncias ajustados utilizando pelo menos 50 pontos amostrais tendem a apresentar maior acurácia. Sendo assim, do ponto de vista geoestatístico, o uso desses modelos para a interpolação dos atributos primários pode produzir resultados com maior precisão.

Estudos envolvendo cokrigagem ordinária como o de Bottega *et al.* (2011) corroboram os resultados desta pesquisa. Nesse estudo, um dos objetivos dos autores se baseava na predição do atributo químico cálcio, utilizado como atributo secundário, e no potencial hidrogeniônico, através do método da cokrigagem. Outros estudos que utilizaram a cokrigagem ordinária como método de predição de valores de atributos químicos em pontos não amostrados são de Adhikary *et al.* (2017) e L. King *et al.* (2000).

### 3.3.2.1 Delineamento de zonas de manejo

Nas Figuras 4 a 6 são apresentados os gráficos dos índices de partição *Fuzzy* (FPI) e de entropia de partição modificada (MPE) em função do número de clusteres. Esses gráficos determinam o número ótimo de zonas de manejo a serem delineadas nos mapas de interpolação por cokrigagem ordinária para os pares dos atributos capacidade de troca de cátions pH 7,0 *vs.* acidez potencial; matéria orgânica *vs.* acidez potencial; e argila *vs.* areia.

Para o par capacidade de troca de cátions pH 7,0 *vs.* acidez potencial (Figura 4), o número ótimo foi de dois clusteres para os três cenários avaliados. Entretanto, para o par argila *vs.* areia (Figura 6), o número ótimo foi de três clusteres nos três cenários analisados. A correlação ( $r > 0,7$ ) existente entre os atributos de cada par (Figura 3) pode ter contribuído para a uniformidade no número de zonas de manejo nos três cenários estudados para cada par.

No par de atributos matéria orgânica *vs.* acidez potencial (Figura 5) não houve uniformidade no número de zonas de manejo para os três cenários. O número ótimo de três clusteres foi registrado nos cenários obtidos considerando 100% e 75% de pontos amostrados para o atributo primário; no entanto, para o cenário obtido considerando 50%, o número ótimo foi igual a dois clusteres. A falta de uniformidade no número de clusteres para o par de atributos matéria orgânica *vs.* acidez potencial pode ser devida à correlação observada ( $r = 0,5$ ) entre esses dois atributos (Figura 3), pois, embora esteja dentro do intervalo recomendado pelo Cohen (1988), a correlação entre os atributos não foi mantida para as diferentes porcentagens disponíveis da variável secundária.

Em geral, foi possível definir clusteres para cada um dos três cenários de cada um dos três pares considerados na análise, e os números ótimos de clusteres determinados que variam de dois a três são basicamente os mais utilizados na definição de zonas de manejo em diferentes trabalhos que usam os atributos do solo. Por exemplo, esses números de zonas de manejo foram empregados nos trabalhos de Li *et al.* (2007), Buttafuoco, G. *et al.* (2010), Costa e Queiroz (2011), Khan *et al.* (2020) e Plant (2001).

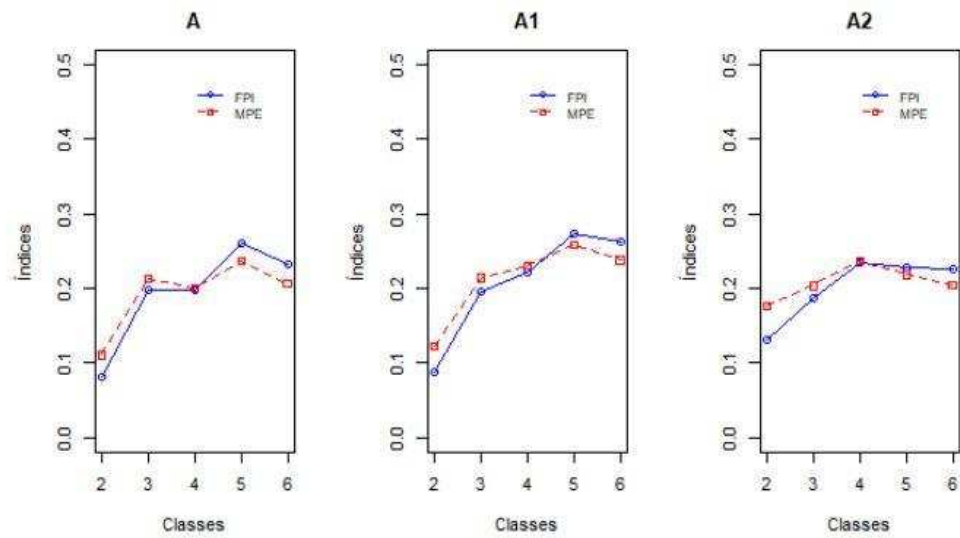


Figura 4 – Índice de Performance *Fuzzy* (FPI) e Entropia de Partição Modificada (MPE) para o par de atributos T vs. (H+Al), considerando três cenários: A = 100% de pontos amostrados do atributo primário; A1 = 75% de pontos amostrados do atributo primário; A2 = 50% de pontos amostrados do atributo primário; T = capacidade de troca de cátions *pH* 7,0 (atributo primário); e H+Al = acidez potencial (atributo secundário).

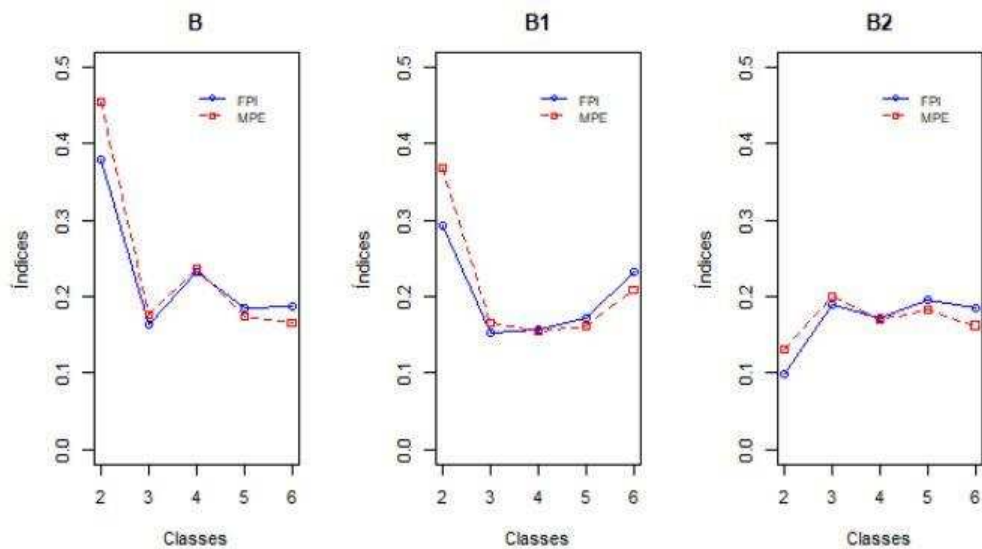


Figura 5 – Índice de Performance *Fuzzy* (FPI) e Entropia de Partição Modificada (MPE) para o par de atributos MO vs. H+Al, considerando três cenários: B = 100% de pontos amostrados do atributo primário; B1 = 75% de pontos amostrados do atributo primário; B2 = 50% de pontos amostrados do atributo primário; MO = matéria orgânica (atributo primário); e H+Al = acidez potencial (atributo secundário).

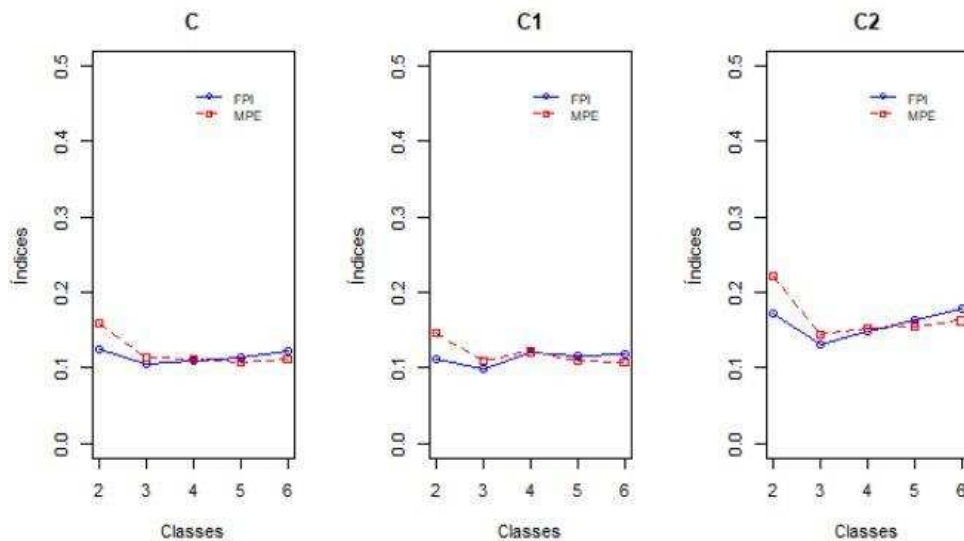


Figura 6 – Índice de Performance *Fuzzy* (FPI) e Entropia de Partição Modificada (MPE) para o par de atributos argila *vs.* areia, considerando três cenários: C = 100% de pontos amostrados do atributo primário; C1 = 75% de pontos amostrados do atributo primário; e C2 = 50% de pontos amostrados dos atributos argila (atributo primário) e areia (atributo secundário).

Os resultados do teste de hipótese do índice Kappa (Tabela 2) para cada par de atributos, com o objetivo de avaliar a concordância das zonas de manejos obtidas para o cenário com 100% dos pontos amostrados do atributo primário com as zonas de manejo obtidas para os cenários com 50% e 75%, são apresentados na Tabela 2. Foi observada concordância significativa dos cenários 50% e 75% com o cenário 100% para o par de atributos T *vs.* (H+Al) e, também, para o par argila *vs.* areia. Para o par MO *vs.* (H + Al) também foi observada concordância significativa entre os cenários 100% e 75% dos pontos amostrados do atributo primário. Não foi avaliada a concordância entre os 100% e 50% dos pontos amostrados do atributo primário para o par MO *vs.* (H + Al), porque o número ótimo de classes obtido em cada um desses dois cenários não foi o mesmo.

A significância dos índices Kappa obtidos indica que resultados satisfatórios podem ser alcançados ao realizar a cokrigagem para o atributo capacidade de troca de cátions *pH* 7,0 utilizando, no mínimo, 50% dos pontos amostrados para esse atributo. Resultados similares podem ser obtidos para o atributo argila. Adicionalmente, podem-se obter resultados satisfatórios para o atributo matéria orgânica usando 75% dos pontos amostrados desse atributo. Do ponto de vista geoestatístico, utilizar menor número de pontos amostrados do atributo primário para a realização da cokrigagem pode contribuir para a redução de custos de amostragem e para a análise dos atributos do solo. Resultados similares foram

obtidos por Bottega *et al.* (2011) em um trabalho que teve por objetivo a predição dos valores de cálcio e magnésio utilizando como atributo secundário o potencial hidrogeniônico.

O uso do teste de hipótese do índice Kappa é comum nos trabalhos direcionados ao delineamento de zonas de manejo, a exemplo do de Máquina *et al.* (2020), que utilizaram o teste para avaliar o grau de concordância entre zonas de manejo obtidas levando em conta apenas um atributo, como aquelas conseguidas considerando um grupo de atributos. Outros autores que usaram esse tipo de teste de hipótese foram Costa e Queiroz (2011) e Carvalho (2015).

Tabela 2 – Índice Kappa relativo à concordância entre as zonas de manejo obtidas do cenário 100% com aquelas conseguidas nos cenários 50% e 75% de pontos amostrados do atributo primário

Atributo	Cenário	
	50%	75%
T <i>vs.</i> (H+Al)	0,66*	0,91*
MO <i>vs.</i> (H+Al)	NA	0,75*
Argila <i>vs.</i> Areia	0,16*	0,53*

T = capacidade de troca catiônica pH 7,0, H+Al; MO = matéria orgânica; \* = significativo a 5%; e NA = não aplicável para cálculo do valor de índice Kappa.

Os mapas das zonas de manejo (Figura 7) obtidos para todos os cenários dos pares capacidade de troca de cátions pH 7,0 *vs.* acidez potencial e argila *vs.* areia apresentaram similaridade no padrão de variabilidade espacial. Padrão de variabilidade espacial similar foi também observado nos mapas obtidos entre os cenários 75% e 100% para o par de atributos matéria orgânica *vs.* acidez potencial. Esses resultados corroboram os do teste de hipótese de índice Kappa, que indicaram grau de concordância significativa entre diferentes cenários dentro do mesmo par de atributos.

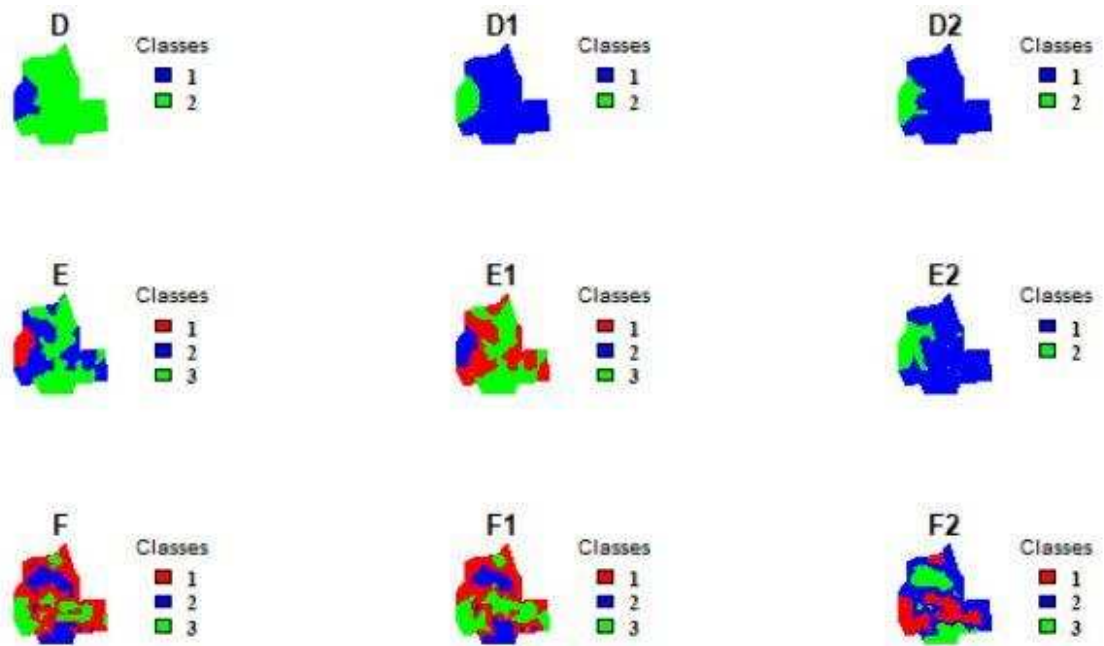


Figura 7 – Mapas de zonas de manejo obtidas pelo método *Fuzzy c-means* para: D = 100% de pontos amostrados do atributo primário do par T vs. (H+Al), D1 = 75% dos pontos amostrados do atributo primário do par T vs. (H+Al) e D2 = 50% dos pontos amostrados do atributo primário do par T vs. (H+Al); T = capacidade de troca de cátions  $pH$  7,0 (atributo primário), H+Al = acidez potencial; E = 100% dos pontos amostrados do atributo primário do par MO vs. (H+Al), E1 = 75% dos pontos amostrados do atributo primário do par MO vs. (H+Al) e E2 = 50% dos pontos amostrados do atributo primário do par MO vs. (H+Al); e F = 100% dos pontos amostrados do atributo primário do par argila vs. areia, F1 = 75% dos pontos amostrados do atributo primário do par argila vs. areia e F2 = 50% dos pares amostrados do atributo primário do par argila vs. areia.

## 4. CONCLUSÕES

1. Uma observação criteriosa das pressuposições da geoestatística multivariada culminou com a construção de modelos lineares de correionalização, isto é, modelos que incorporaram na modelagem mais uma variável regionalizada. Essa incorporação contribuiu significativamente para a obtenção de resultados consistentes da modelagem.
2. As zonas de manejo delineadas nos mapas obtidos pela cokrigagem mostraram resultados satisfatórios, nos diferentes cenários dentro do mesmo par de atributos.
3. O uso da técnica de predição por cokrigagem não só poderá minimizar os custos de recolha, como também poderá garantir melhor análise das amostras de solo para os atributos de interesse.

## REFERÊNCIAS

ADHIKARY, S.; MUTTIL, N.; YILMAZ, A. **Cokriging for Enhanced Spatial Interpolation of Rainfall in Two Australian Catchments**. Viena, 2017.

BECKETT, P. H. T. A. W., R.. Soil variability – A review. **Soils and Fertilizers**, v. 34, p. 1-15, 1971.

BEZDEK, J. C.; EHRLICH, R.; FULL, W. FCM: the fuzzy c-means clustering algorithm. **Computers & Geosciences**, v. 10, n. 2, p. 191-203, 1984/01/01/1984. ISSN 0098-3004. Disponível em: <http://www.sciencedirect.com/science/article/pii/0098300484900207>.

BOTTEGA, E. L.; SILVA, S. de A.; COSTA, M. M; BOTTEGA, S. P. Cokrigagem na estimativa dos teores de Ca e Mg em um Latossolo Vermelho distroférico. **Revista Ciência Agronômica**, v. 42, p. 821-828, 2011. ISSN 1806-6690. Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1806-66902011000400001&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-66902011000400001&nrm=iso).

BURROUGH, P. A. Soil variability: a late 20th century view. **Soils and fertilizers**, v. 56, p. 531-62, 1993.

BURROUGH, P. A.; WILSON, J P.; GAANS, P. F. M. V.; HANSEN, A. J. Fuzzy k-means classification of topo-climatic data as an aid to forest mapping in the Greater Yellowstone Area, USA. **Landscape Ecology**, v. 16, n. 6, p. 523-46, 2001/08/01/2001. ISSN 1572-9761. Disponível em: <https://doi.org/10.1023/A:1013167712622>.

BUTTAUFUOCO, G.; CASTRIGNANÒ A., COLECCHIA A.S., RICCA N. Delineation of management zones using soil properties and a multivariate geostatistical approach. **Italian Journal of Agronomy**, v. 5, p. 323-32, 2010.

CAMBARDELLA, C. A.; KARLEN, D. L. Spatial analysis of soil fertility parameters. **Precision Agriculture**, v. 1, n. 1, p. 5-14, January 01, 1999. ISSN 1573-1618. Disponível em: <https://doi.org/10.1023/A:1009925919134>.

CARVALHO, P. S. M. **Agricultura de Precisão no cultivo de cacueiro**: delineamento de zonas de manejo em lavoura do sul da Bahia. 2015. 114 f. Dissertação (Mestrado em Produção Vegetal) – Universidade Estadual de Santa Cruz, Ilhéus, BA, 2015.

CASTRIGNANÒ, A. VENTRELA, D.; GIGLIO, L.; GUASTAFERRO, F **A multivariate geostatistical algorithm on delineation of management zones**. [S.l. : s.n.t.], 2009.

CHICA-OLMO, J. Prediction of housing location price by a multivariate spatial method: cokriging. **Journal of Real Estate Research**, v. 29, n. 1, p. 91-114, 2007/01/01/2007. ISSN

0896-5803. Disponível em:  
<https://www.tandfonline.com/doi/abs/10.1080/10835547.2007.12091188>.

CHILÈS, J.-P.; DELFINER, P. **Geostatistics: modeling spatial uncertainty**. 2<sup>nd</sup>. [S.l. : s.n.t.], 2012.

CLARK, I. **The art of cross validation in geostatistical applications** – Application of computers and operations research in the mineral industry. 19<sup>th</sup>. Littleton: University, T. P. Littleton Colorado, Society of Mining Engineers, Inc., 1986. p. 211-20.

COHEN, J. **Statistical power analysis for the behavioral sciences**. NJ: Erlbaum, 1988.

CONGALTON, R. G.; MEAD, R. A. A review of discrete multivariate analysis techniques used in assessing the accuracy of remotely sensed data from error matrices. **IEEE Transactions on Geoscience and Remote Sensing**, v. 24, n. I, p. 169-74, 1986.

COSTA, M. M.; QUEIROZ, D. M. **Condutibilidade elétrica aparente do solo, como ferramenta para agricultura de precisão em uma área sob cerrado**. Viçosa, MG: Departamento de Engenharia Agrícola/UFV, 2011. 89 p.

DE IACO, S.; MYERS, D.; POSA, D. **The linear coregionalization model and the product-sum space-time variogram**. [S.l. : s.n.t.], 2003. p. 25-38.

FOSFATO, I. D. P. D. **Manual internacional de fertilidade do solo**. Piracicaba, SP, 1998.

FREUND, M. J. Cokriging: multivariable analysis in petroleum exploration. **Computers & Geosciences**, v. 12, n. 4, p. 485-91, 1986/01/01/1986. ISSN 0098-3004. Disponível em: <http://www.sciencedirect.com/science/article/pii/0098300486900634>.

FRIDGEN, J. J.; KITCHEN, N. R.; SUDDUTH, K. A.; DRUMMOND, S. T.; WIEBOLD, W. J. Management Zone Analyst (MZA): software for subfield management zone delineation. [S.l. : s.n.t.], 2003.

GOOVAERTS, P. **Geostatistics for natural resource evaluation**. [S.l. : s.n.t.], 1997. v. 42.

GORSEVSKI, P. V.; GESSLER, P. E.; JANKOWSKI, P. Integrating a fuzzy k-means classification and a Bayesian approach for spatial prediction of landslide hazard. **Journal of Geographical Systems**, v. 5, n. 3, p. 223-51, November 01 2003. ISSN 1435-5930. Disponível em: <https://doi.org/10.1007/s10109-003-0113-0>.

GOZDOWSKI, D.; STĘPIEŃ, M.; SAMBORSKI, S.; DOBERS, E. S.; SZATYŁOWICZ, J. ; CHORMAŃSKI, J. Prediction accuracy of selected spatial interpolation methods for

soil texture at farm field scale. *Journal of Soil Science and Plant Nutrition*, v. 15, p. 639-50, 12/01/2015.

ISAAKS, E. H.; SRIVASTAVA, R. M. **An Introduction to applied geostatistics**. New York: Oxford University Press, 1989.

JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining geostatistics**. London: Academic Press, 1978. 600 p.

KHAN, H.; FAROOQUE, A. A.; ACHARYA, B.; ABBAS, F.; ESAU, T. J.; ZAMAN, Q. U. Delineation of management zones for site-specific information about soil fertility characteristics through proximal sensing of potato fields. *Agronomy*, v. 10, n. 12, p. 1854, 2020. ISSN 2073-4395. Disponível em: <https://www.mdpi.com/2073-4395/10/12/1854>.

KONAK, A. **Estimating path loss in wireless local area networks using ordinary kriging**. [S.l. : s.n.t.], 2011. p. 2888-96.

KRIGE, D. G. A study of the relationship between development values and recovery grades on the South African goldfields. **Journal of the South African Institution of Mining and Metallurgy**, n. 61, p. 317-31, 1959.

L. KING, S.; LISTER, A.; HOPPUS, M. **A comparison of kriging and cokriging for mapping forest volume in connecticut**. [S.l. : s.n.t.], 2000.

LARK, R. M.; ANDER, E.L.; , CAVE , M.R.; KNIGHTS, K.V.; GLENNON, M.M.; SCANLON, R.P. **Mapping trace element deficiency by cokriging from regional geochemical soil data: a case study on cobalt for grazing sheep in Ireland**. 2014. p. 64-78.

LARK, R. M.; PAPRITZ, A. Fitting a linear model of coregionalization for soil properties using simulated annealing. **Geoderma**, v. 115, n. 3, p. 245-60, 2003/08/01/2003. ISSN 0016-7061. Disponível em: <http://www.sciencedirect.com/science/article/pii/S001670610300065X>.

LARK, R. M.; STAFFORD, J. V. Classification as a first step in the interpretation of temporal and spatial variation of crop yield. **Ann. Appl. Biol.**, v. 130, p. 111-21, 1997.

LI, Y.; ZHOW, S.; FENG, L.; LI, H-YI. Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. **Computers and Electronics in Agriculture**, v. 56, n. 2, p. 174-86, 2007/04/01/2007. ISSN 0168-1699. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0168169907000312>.

LIMA, J. S. S.; SILVA, S. A.; OLIVEIRA, R. B.; FONSECA, A. B. Estimativa da produtividade de café conilon utilizando técnicas de cokrigagem. **Revista Ceres**, v. 63, n. 1, p. 54-61, 2016.

MACNAB, Y. C. Linear models of coregionalization for multivariate lattice data: order-dependent and order-free cMCARs. **Statistical Methods in Medical Research**, v. 25, n. 4, p. 1118-44, 2016. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/0962280216660419>.

MÁQUINA, C. M. SANTOS, N.T.; COSTA, M. M; SILVA, S. A. Using multivariate analysis to design management zones. **Científica**, v. 48, n. 1, p. 25-35, 2020. ISSN 1984-5529.

MATHERON, G. **Les variables régionalisées et leur estimation**. Paris: Masson, 1965.

MCBRATNEY, A. B.; DEGRUIJTER, J. J. A continuum approach to soil classification by modified fuzzy k-means with extragrades. **Journal of Soil Science**, v. 43, p. 159-75, 1992.

MCBRATNEY, A. B.; PRINGLE, M. J. Estimating average and proportional variograms of soil properties and their potential use in Precision Agriculture. **Precision Agriculture**, v. 1, n. 2, p. 125-52, September 01 1999. ISSN 1573-1618. Disponível em: <https://doi.org/10.1023/A:1009995404447>.

MOLIN, J. P.; AMARAL, L. R. D.; COLAÇO, A. F. **Agricultura de Precisão**. 1. ed. São Paulo: 2015. ISBN 978-85-7975-213-1.

MOLIN, J. P.; CASTRO, C. N. D. Establishing management zones using soil electrical conductivity and other soil properties by the fuzzy clustering technique. **Scientia Agricola**, v. 65, p. 567-73, 2008. ISSN 0103-9016.

MORAL GARCÍA, F. J.; TERRÓN, J. M.; MARQUES DA SILVA, J. **Delineation of management zones using mobile measurements of soil electrical conductivity and multivariate geostatistical techniques**. [S.l. : s.n.t.], 2010. p. 335-43.

ODEH, I. O. A.; MCBRATNEY, A. B.; CHITTLEBOROUGH, D. J. Soil pattern recognition with fuzzy-c-means: application to classification and soil-landform interrelationships. **Soil Sci. Soc. Am. J.**, v. 56, p. 505-16, 1992.

OLIVER, M.; WEBSTER, R.; GERRARD, J. Geostatistics in physical geography. Part I: Theory. **Transactions of the Institute of British Geographers**, v. 14, n. 3, p. 259-69, 1989. ISSN 00202754, 14755661. Disponível em: <http://www.jstor.org/stable/622687>.

PHAM, T. G., KAPPAS, M.; HUYNH, C. V.; NGUYEN, L. H. K. Application of ordinary kriging and regression kriging method for soil properties mapping in hilly region of central

vietnam. **ISPRS International Journal of Geo-Information**, v. 8, n. 3, p. 147, 2019. ISSN 2220-9964. Disponível em: <http://www.mdpi.com/2220-9964/8/3/147>

PLANT, R. E. *Spatial data analysis in ecology and agriculture using R*. 2<sup>nd</sup> ed. Florida: CRC Press, 2018. 666p. ISBN 9780815392750.

PLANT, R. E. Site-specific management: the application of information technology to crop production. **Computers and Electronics in Agriculture**, v. 30, p. 9-29, 2001.

RODRIGUES, M. S.; CORÁ, J. E. Management zones using fuzzy clustering based on spatial-temporal variability of soil and corn yield. **Engenharia Agrícola**, v. 35, p. 470-83, 2015. ISSN 0100-6916.

RODRIGUES, M.S.; SANTANA M.C.; UCHÔA A. L. P.; MENEZES A. X.S.M. M.; CAVALCANTE I.H.L, Lima AMN (2015) Delineamento de zonas de manejo baseadas em atributos físicos do solo em uma área de goiabeira irrigada no Semiárido nordestino. Anais do XXXV Congresso Brasileiro de Ciência do Solo. Centro de Convenções Natal: Sociedade Brasileira de Ciências de Solo. 4p

RONQUIM, C. C. **Conceitos de fertilidade do solo e manejo adequado para as regiões tropicais**. Campinas, SP: EMBRAPA, 2010. ISBN 1806-3322.

RUSPINI, E. H. A new approach to clustering. **Inf. Control**, v. 15, p. 22-32, 1969.

SABINO, P. D. S. N.; SANTOS, A. C. dos ; LEITE, R.L. L.; DIM, V. P.; CRUZ , R. S; PEDRICO, A.; NETO, D N. N. **Análise espacial de parâmetros da fertilidade do solo em região de ecótono sob diferentes usos e manejos**. [S.l. : s.n.t.], 2011.

SANTOS, R. O.; FRANCO, L. B.; SILVA, S. A., SODRÉ, G. A.; MENEZES, A.A. Spatial variability of soil fertility and its relation with cocoa yield. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 21, n. 2, p. 88-93, 2017.

SILVA, A.; REGINA LOPES ZIMBACK, C.; OLIVEIRA, R. **Cokrigagem na estimativa da evapotranspiração em Campinas (SP)**. Campinas, SP, 2010.

SILVA, S. A.; LIMA, J. S. S.; SOUZA, G. S. Estudo da fertilidade de um Latossolo Vermelho-Amarelo húmico sob cultivo de café arábica por meio de geoestatística. **Revista Ceres**, v. 57, n. 4, p. 560- 7, 2010.

SILVA, V. M.; COSTA, J. F. C. L. Sensitivity analysis of ordinary kriging to sampling and positional errors and applications in quality control. **REM – International Engineering Journal**, v. 69, p. 491-6, 2016. ISSN 2448-167X. Disponível em:

[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S2448-167X2016000400491&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2448-167X2016000400491&nrm=iso).

VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R. F. *et al.* (ed). **Tópicos em ciência do solo**. [S.l.]: SBGS, 2000. v. 1, p. 1-54.

WACKERNAGEL, H. **Multivariate geostatistics** – An introduction with applications. [S.l. : s.n.t.], 2003. 388 p.

WARRICK, A. W.; NIELSEN, D. R. Spatial variability of soil physical properties in the field. In: HILLEL, D. (ed.). **Applications of soil physics**. [S.l. : s.n.t.], p. 319-44, 1980.

WEBSTER, M.; OLIVER, M. A. **Statistical methods in soil and land resource survey**. [S.l. : s.n.t.], 1990.

WEBSTER, R.; OLIVER, M. A. Sample adequately to estimate variograms of soil properties. **Soil Science**, v. 43, n. 1, p. 117-92, 1992. –

WEBSTER, R.; OLIVER, M. A. **Geostatistics for environmental scientists**. 2<sup>nd</sup>. [S.l. : s.n.t.], 2007. ISBN 9780470028582.

WU, C.; WU, J.; ZHANG, I.; DEGLORIA, S. D. **Spatial prediction of soil organic matter content using cokriging with remotely sensed data**. [S.l. : s.n.t.], 2009.

YALÇIN, E. **Cokriging and its effect on the estimation precision**. [S.l. : s.n.t.], 2005. 228 p.

YAMAMOTO, J. K.; LAMDIM, P. M. B. **Geoestatística: conceitos e aplicações**. São Paulo: Oficina de Textos, 2013.

## APÊNDICES

## APÊNDICE A – RESUMO DO ARTIGO ANÁLISE DE DADOS DE SOBREVIVÊNCIA ESPACIALMENTE CORRELACIONADOS

### ANÁLISE DE DADOS DE SOBREVIVÊNCIA ESPACIALMENTE CORRELACIONADOS

Calisto Manuel MÁQUINA<sup>1</sup>  
 Fábio Nogueira DEMARQUI<sup>2</sup>  
 Marcos Oliveria PRATES<sup>2</sup>  
 Nerilson Terra SANTOS<sup>1</sup>

- **RESUMO:** O ajuste de modelos de fragilidade para dados de sobrevivência espacialmente correlacionados é discutido. Considerando uma abordagem Bayesiana completamente paramétrica, o modelo ICAR (*intrinsically conditional autoregressive*) é assumido para a distribuição das fragilidades, enquanto a distribuição de base dos tempos de sobrevivência é modelada pela distribuição Weibull. Um critério para a seleção de modelos Bayesianos é discutido. Uma descrição detalhada de como utilizar o *software* R em conjunto com o *software* OpenBUGS para realizar inferência em modelos de sobrevivência espacial é fornecida. A análise de um banco de dados reais envolvendo pacientes diagnosticados com câncer de bexiga no estado de Connecticut, EUA, é apresentada.
- **PALAVRAS-CHAVE:** Fragilidade espacial; modelo ICAR, inferência Bayesiana; OpenBUGS.

### 1. Introdução

O avanço computacional observado nas últimas décadas, aliado ao desenvolvimento de novas metodologias estatísticas, tem possibilitado a modelagem de dados com estruturas cada vez mais complexas. Neste contexto, modelos estatísticos que incorporam efeitos aleatórios, isto é, variáveis não observáveis que são inseridas na modelagem a fim de acomodar diferentes tipos de estruturas de dados, têm se tornado cada vez mais frequentes.

<sup>1</sup>Universidade Federal de Viçosa - UFV, Departamento de Estatística, CEP: 36570-977, Viçosa, MG, Brasil. E-mail: calisto.calmaquina.maquina@gmail.com; nsantos@ufv.br

<sup>2</sup>Universidade Federal de Minas Gerais - UFMG, Departamento de Estatística, CEP: 31270-901, Belo Horizonte, MG, Brasil. E-mail: fndemarqui@est.ufmg.br; marcosap@est.ufmg.br

## APÊNDICE B – USO DE ANÁLISE MULTIVARIADA NO DELINEAMENTO DE ZONAS DE MANEJO

ISSN: 1984-5529

# Científica

 Jaboticabal  
 v.48, n.1, p.25–35, 2020

<http://dx.doi.org/10.15381/1984-5529.2020v48n1p25-35>

### Using multivariate analysis to design management zones

### Uso de análise multivariada no delineamento de zonas de manejo

Calisto Manuel MÁQUINA<sup>1</sup>; Nerilson Terra SANTOS<sup>2</sup>; Marcelo Marques COSTA<sup>3</sup>;  
 Samuel de Assis SILVA<sup>4</sup>

<sup>1</sup> *Correspondent Author*, Assistant Professor, Rovuma University, Department of Statistics and Information Management, Nampula, Mozambique. Email: calisto.calmagna.maquina@gmail.com

<sup>2</sup> Full Professor, Federal University of Viçosa, Department of Statistics, Viçosa, MG-Brazil; E-mail: nsantos@ufv.br

<sup>3</sup> Professor, Federal University of Jataí, Special Academic Unit of Agrarian Sciences, Jataí, GO-Brazil; E-mail: marcelo.marques.costa@gmail.com

<sup>4</sup> Professor, Federal University of Espírito Santo, Rural Engineering Department, ES-Brazil; E-mail: samuel.assilva@gmail.com

Recebido em: 12-10-2019; Aceito em: 20-12-2019

#### Abstract

Soil chemical and physical attributes are important in any agricultural cropping system, but in precision agriculture they are more relevant due to the possibility of application using different management practices along a production field. However, the correlation between these attributes has been little explored in the delineation of management zones. This work aims to maximize the use of joint spatial variability for soil attributes. Its secondary objectives were 1) reduction of spatial variability dimensionality among all attributes and 2) assessment of agreement between univariate and multivariate management zones. The management zones resulting from the interpolation of attribute values, as well as from the scores of each of the three main components, were delineated using the Fuzzy c-means algorithm. The fuzzy performance and modified partition entropy indexes were used to determine the optimal number of management zones. The Kappa index was used to evaluate the agreement of management zones obtained from attributes with those obtained from principal components. By using principal component analysis, it was possible to reduce the dimensionality of the number of variables that contribute to the joint spatial variability existing in the study area. There was no complete agreement between the uni- and multivariate management zones outlined, which is why further studies on the subject are needed.

**Additional keywords:** Fuzzy c-means algorithm; geostatistics; principal component; spatial variability.

Fonte: Revista Científica.