



EDEMAR APPEL NETO

**PROPOSTA E COMPARAÇÃO DE MEDIDAS DE
DEPENDÊNCIA ESPACIAL EM GEOESTATÍSTICA**

LAVRAS - MG

2019

EDEMAR APPEL NETO

**PROPOSTA E COMPARAÇÃO DE MEDIDAS DE DEPENDÊNCIA ESPACIAL EM
GEOESTATÍSTICA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária para a obtenção do título de Mestre.

Prof. Dr. Marcelo Silva de Oliveira

Orientador

Prof. Dr. Enio Júnior Seidel

Coorientador

LAVRAS - MG

2019

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Appel Neto, Edegar

Proposta e comparação de medidas de dependência espacial
em geoestatística / Edegar Appel Neto. – Lavras : UFLA,
2019.

39 p. : il.

Dissertação (mestrado acadêmico) - Universidade Federal
de Lavras–Universidade Federal de Lavras, 2019.

Orientador: Prof. Dr. Marcelo Silva de Oliveira.

Bibliografia.

1. Geoestatística. 2. Semivariograma. 3. Índice. I.
Oliveira, Marcelo Silva de. II. Seidel, Enio Júnior. III. Título .

EDEMAR APPEL NETO

**PROPOSTA E COMPARAÇÃO DE MEDIDAS DE DEPENDÊNCIA ESPACIAL EM
GEOESTATÍSTICA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária para a obtenção do título de Mestre.

APROVADA em 28 de junho de 2019.

Prof. Dr. Paulo Henrique Sales Guimarães UFLA
Prof. Dr. Peter de Matos Campos UFSJ
Prof. Dr. Renato Ribeiro de Lima UFLA

Prof. Dr. Marcelo Silva de Oliveira
Orientador

Prof. Dr. Enio Júnior Seidel
Co-Orientador

**LAVRAS - MG
2019**

*Aos meus pais, Tadeu e Claudete;
minhas irmãs, Márcia, Karen, Ivi e Nádia;
meus cunhados Jeancarlo e Evandro;
meus sobrinhos, Manoela, João, Otávio e Rafael;
DEDICO*

AGRADECIMENTOS

Aos meus pais, Tadeu e Claudete; as minhas irmãs, Márcia, Karen, Ivi e Nádia; aos cunhados, Jean e Evandro, por todo o apoio e incentivo em todos os momentos.

Aos meus sobrinhos, Manoela, João, Otávio e Rafael, por todo o carinho e alegria.

Aos professores orientadores Marcelo Silva de Oliveira e Enio Junior Seidel, pelas discussões, sugestões e tempo dedicado a me auxiliar.

Aos membros da Banca Examinadora, por terem aceito o meu convite, contribuindo e enriquecendo a minha dissertação.

Aos professores e demais funcionários do Departamento de Estatística da UFLA.

À Universidade Federal de Lavras (UFLA) por oportunizar estudo gratuito e de qualidade.

Aos colegas de apartamento Gustavo, Daniel e Wendel pelos momentos de distração que, realmente, precisei nos momentos difíceis.

Aos demais amigos que, mesmo longe, de alguma forma contribuíram com incentivo.

A todos os amigos que fiz em Lavras e levarei para a vida toda.

O presente trabalho foi realizado com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG).

RESUMO

A geoestatística é uma técnica bastante utilizada para descrever a dependência espacial de atributos do solo, em especial, o semivariograma é a principal ferramenta na geoestatística para descrever tal dependência. O semivariograma possibilita a interpretação da dependência espacial em função da distância entre as amostras, por meio dos parâmetros efeito pepita (C_0), contribuição (C_1), patamar ($C_0 + C_1$) e alcance (a). Pelo fato do semivariograma ser uma ferramenta visual para interpretação da dependência espacial e apresentar vários parâmetros, surge a necessidade de uma medida que represente as informações contidas no semivariograma de maneira sucinta e objetiva. Na literatura, existem índices que são calculados a partir dos parâmetros do semivariograma e permitem classificar a dependência espacial em escalas de forte, moderada e fraca. Apesar de existirem tais índices, em alguns casos não é possível encontrar coerência entre o que se verifica visualmente no semivariograma e a classificação da dependência espacial dada pelos índices. Com isso, foi desenvolvida neste estudo, uma medida alternativa para descrever a dependência espacial, levando em consideração os parâmetros do semivariograma e a metade da maior distância entre os pontos amostrados. Foi desenvolvida também uma classificação para essa nova medida para mensurar a dependência espacial. Tendo a nova medida e sua respectiva classificação em fraca, moderada e forte dependência espacial, foram realizadas aplicações a dados provenientes de atributos de solo. Por fim, foram realizadas comparações com dois índices já existentes na literatura.

Palavras-chave: Geoestatística, Dependência Espacial, Índice, Semivariograma.

ABSTRACT

Geostatistics is a widely used technique to describe the spatial dependence of soil attributes. The semivariogram is the main tool used to describe such dependence and allows the interpretation of the spatial dependence regarding the distance between the samples using the nugget effect (C_0), contribution (C_1), sill ($C_0 + C_1$) and range (a). Because the semivariogram is a visual tool for interpreting spatial dependence and indicating parameters, a measure that represents the information contained in the semivariogram in a brief and objective manner is required. In the literature, there are indices that are calculated from the semivariogram parameters and that allow the classification of spatial dependence into strong, moderate, and weak. Despite the existence of such indices, in some cases, it is impossible to find coherence between what is visually verified in the semivariogram and the classification of spatial dependence given by the indices. Thus, an alternative measure was developed to describe spatial dependence, considering the semivariogram parameters and half of the largest distance between the sampled points. A classification was also developed to measure spatial dependence. The data derived from the soil attributes were applied considering this new measure and its respective classification into weak, moderate, and strong spatial dependence. Finally, comparisons were made between the two indices found in the literature.

Keywords: Geostatistics, Spatial Dependence, Index, Semivariogram.

LISTA DE FIGURAS

Figura 2.1 – Semivariogramas com diferentes alcances e mesmos patamares.	11
Figura 2.2 – Semivariograma experimental (semivariâncias $\hat{\gamma}(h)$ versus distâncias h). . .	12
Figura 4.1 – Densidades dos 401 resultados que foram gerados através das combinações entre os 20 valores atribuídos a cada uma das duas componentes do índice de dependência espacial <i>MDE</i> (mais o valor zero incluído), para os seis modelos estudados.	32
Figura 4.2 – Histogramas suavizados de índices de dependência espacial aplicados à dados oriundos de atributos físicos, químicos e mineralógicos de solo. . . .	33

LISTA DE TABELAS

Tabela 3.1 – Cenários com os parâmetros utilizados para simulação dos dados com dependência espacial.	23
Tabela 4.1 – Medidas descritivas dos 401 valores gerados pelas combinações das duas componentes do índice <i>MDE</i> (mais o valor zero incluído) para os seis modelos de semivariograma explorados.	31
Tabela 4.2 – Classificação proposta neste trabalho para o índice <i>MDE</i> utilizando a metodologia de Seidel e Oliveira (2016).	32
Tabela 4.3 – Classificação da dependência espacial de atributos físicos, químicos e mineralógicos de solo utilizando a classificação de Cambardella et al. (1994) para o <i>SPD</i> , a classificação de Seidel e Oliveira (2016) para o <i>IDE</i> e a classificação proposta neste estudo para o <i>MDE</i>	34
Tabela 4.4 – Correlações calculadas entre os índices <i>SPD</i> , <i>IDE</i> , <i>MDE</i> e entre o <i>EQM</i> e a variância de krigagem (<i>VAR</i>) da validação cruzada.	34

SUMÁRIO

1	INTRODUÇÃO	9
2	REFERENCIAL TEÓRICO	11
2.1	Dependência espacial	11
2.1.1	Semivariograma	12
2.1.2	Índices de dependência espacial	15
3	MATERIAL E MÉTODOS	22
3.0.3	Dados	22
3.1	Métodos	22
3.1.1	Medida de correlação espacial	22
3.1.2	Classificação do novo índice	22
3.1.3	Comparações dos índices de dependência espacial	23
4	RESULTADOS	25
5	CONCLUSÃO	36
6	REFERÊNCIAS	37

1 INTRODUÇÃO

A geoestatística oferece ferramentas que possibilitam a análise e interpretação de dados com dependência espacial. O semivariograma é a principal ferramenta utilizada na geoestatística para verificação da dependência espacial, pois mostra a medida do grau de dependência espacial entre amostras ao longo de um suporte específico (LANDIM, 2006).

Segundo Seidel e Oliveira (2014a), o semivariograma descreve bem a estrutura de dependência espacial do fenômeno em estudo, apresentando várias informações em seu conjunto gráfico, fato que pode levar a necessidade de haver uma medida única que represente a dependência espacial de forma resumida. Siqueira, Vieira e Ceddia (2008) discutem resultados da variabilidade espacial de atributos físicos de solo baseando-se nos parâmetros do semivariograma, afirmando que, para melhor interpretação, é preciso considerar uma interação entre os parâmetros contribuição, efeito pepita e o alcance. Com base nos parâmetros estimados no semivariograma, é possível construir índices de dependência espacial que possibilitam uma interpretação direta em apenas um número, resumindo as informações contidas no semivariograma.

Os índices mais usuais na literatura para calcular o grau de dependência espacial são, o *RD* (dependence ratio) (CAMBARDELLA et al., 1994) que considera apenas a razão entre efeito pepita (C_0) e o patamar ($C_0 + C_1$) e ainda o *SPD* (spatial dependence) (BIONDI; MYERS; AVERY, 1994) que faz uma relação inversa ao anterior, sendo a razão entre a contribuição (C_1) e o patamar ($C_0 + C_1$). Ambos os índices possuem uma classificação construída, com base nos quartís, em que, *RD* menor que 25% do patamar a dependência espacial é considerada forte, entre 25% e 75% moderada e acima de 75% é considerada fraca. A proposta de classificação do *SPD* é exatamente o inverso, em que, *SPD* menor que 25% do patamar a dependência espacial é considerada fraca, entre 25% e 75% moderada dependência espacial e acima de 75% é considerada forte dependência espacial. .

Schabenberger e Gotway (2005) fazem uma crítica aos índices que consideram apenas o efeito pepita e o patamar para resumir a dependência espacial apresentada no semivariograma, deixando de contemplar o alcance. Após fazer essa crítica, os autores comentam sobre duas medidas I_1 e I_2 , calculadas a partir da função de correlação denominadas escala integral. Essas medidas são interpretadas como a maior distância e maior distância média, respectivamente, em que as observações são fortemente correlacionadas.

Recentemente, Appel Neto et al. (2018) e Seidel e Oliveira (2014a) propuseram o índice de dependência espacial (*IDE*) baseado em área de correlação espacial, tendo como inspiração a medida I_1 . Para sua construção, o índice *IDE* considera os seguintes parâmetros do semivariograma: contribuição (C_1), patamar ($C_0 + C_1$), alcance (a), além de uma constante inerente ao ajuste do semivariograma denominada fator de modelo (FM) e uma quantidade da máxima distância (qMD) entre pontos amostrados.

Appel Neto et al. (2018) e Seidel e Oliveira (2016) apresentam uma classificação para o *IDE*, considerando intervalos definidos como primeiro quartil, mediana e terceiro quartil de uma distribuição empírica proposta para o índice. Diferente da classificação do *SPD* e *RD*, a classificação do *IDE* tem intervalos diferentes para cada modelo de semivariograma utilizado. O *IDE*, calculado a partir do modelo esférico é classificado como fraca dependência espacial quando $\leq 7\%$, moderada entre 7% e $\leq 15\%$ e forte quando $>15\%$, o modelo exponencial é classificado como fraca dependência espacial quando $\leq 6\%$, moderada entre 6% e $\leq 13\%$ e forte quando $>13\%$ e o modelo gaussiano é classificado como fraca dependência espacial quando $\leq 9\%$, moderada entre 9% e $\leq 20\%$ e forte quando $>20\%$ (SEIDEL; OLIVEIRA, 2016). O *IDE* calculado a partir do modelo cúbico é classificado como fraca dependência espacial quando $\leq 7\%$, moderada entre 7% e $\leq 15\%$ e forte quando $>15\%$, o modelo pentaesférico é classificado como fraca dependência espacial quando $\leq 6\%$, moderada entre 6% e $\leq 12\%$ e forte quando $>12\%$ e o modelo wave é classificado como fraca dependência espacial quando $\leq 11\%$, moderada entre 11% e $\leq 24\%$ e forte quando $>24\%$ (APPEL NETO et al., 2018).

Objetivou-se, com este trabalho, desenvolver um novo índice de dependência espacial a partir da medida I_2 apresentada por Schabenberger e Gotway (2005) e compará-lo computacionalmente via simulações e aplicações a dados reais com os demais índices que já existem na literatura, de modo que foi possível realizar uma discussão sobre diferenças dos índices. Além disso, feito o estudo exploratório do comportamento do novo índice foi proposta uma classificação em intervalos para o índice em fraca, moderada e forte dependência espacial. Com novo índice e sua respectiva classificação foi possível comparar sobre o desempenho dos índices apresentados em relação à mensuração e classificação da dependência espacial.

2 REFERENCIAL TEÓRICO

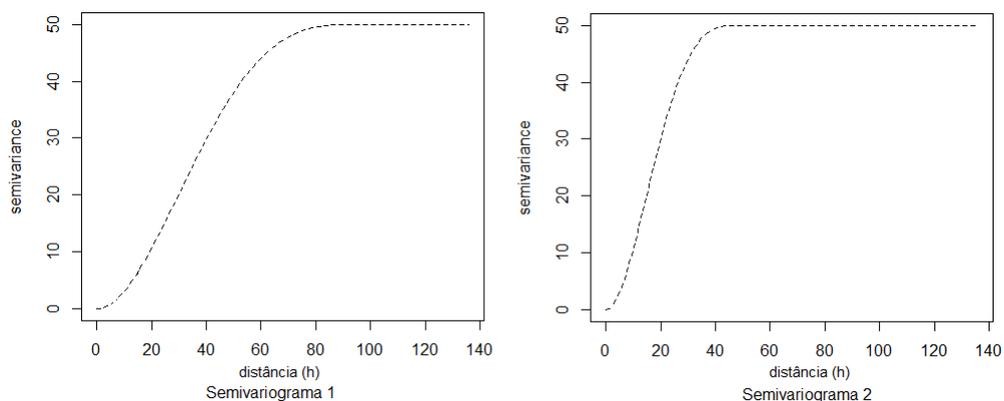
2.1 Dependência espacial

Analisar a dependência espacial tornou-se uma necessidade indispensável quando o fenômeno em estudo trata-se de atributos físicos e químicos do solo. Grego e Vieira (2005) afirmam que, mesmo em experimentos em áreas pequenas, considerar homogeneidade dos atributos de solo pode levar a interpretações errôneas sobre os resultados obtidos, pois a hipótese de dependência espacial estará sendo ignorada.

Segundo Vieira (2000), um dos métodos mais antigos para estimar a dependência no espaço ou no tempo, de amostras vizinhas é a autocorrelação. Porém, quando as amostras são coletadas em duas dimensões, é necessária uma abordagem mais adequada para medir a dependência espacial. A geoestatística surgiu justamente para tratar problemas desse tipo. Mais especificamente tem-se o semivariograma como uma ferramenta para descrever a dependência espacial de atributos medidos em duas ou mais dimensões.

Na Figura 2.1, apresentada em Appel Neto et al. (2018) mostram-se dois semivariogramas, sendo que o semivariograma 1 apresenta o mesmo patamar que semivariograma 2, porém o alcance em que o patamar é atingido é maior no semivariograma 1. Isso mostra a importância de se avaliar a estrutura da dependência espacial levando em consideração todas as informações disponíveis no semivariograma.

Figura 2.1 – Semivariogramas com diferentes alcances e mesmos patamares.



Fonte: Appel Neto et al. (2018).

2.1.1 Semivariograma

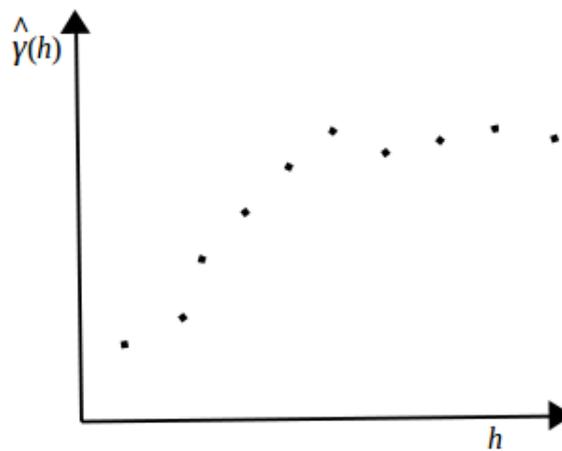
O estimador clássico da semivariância, proposto por Matheron (1962), utilizando o método dos momentos é dado da seguinte forma:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(\mathbf{s}_i) - Z(\mathbf{s}_i + h)]^2, \quad (2.1)$$

em que, $\hat{\gamma}(h)$ é a semivariância estimada na distância h ; $N(h)$ é o número de pares de amostras das variáveis " $Z(\mathbf{s}_i), Z(\mathbf{s}_i + h)$ ", separadas pela mesma distância h ; e Z representa o valor medido do atributo em estudo nas localizações \mathbf{s}_i e $\mathbf{s}_i + h$ (BERTOLANI; VIEIRA, 2001).

Na Figura 2.2, apresenta-se um semivariograma experimental considerando a existência de estacionariedade de segunda ordem (apresenta patamar).

Figura 2.2 – Semivariograma experimental (semivariâncias $\hat{\gamma}(h)$ versus distâncias h).



Fonte: Do autor (2019).

Com o semivariograma experimental construído, ajusta-se um modelo teórico de semivariograma para que seja possível “explicar” a continuidade da semivariância entre os pontos do semivariograma experimental. Também é, a partir do ajuste dos modelos teóricos que se obtém as estimativas dos parâmetros alcance, efeito pepita e contribuição. Há uma restrição quanto aos modelos teóricos que podem ser ajustados, os quais são chamados de “modelos autorizados”. Os modelos teóricos são divididos em dois grupos: modelos com patamar e modelos sem patamar. Os parâmetros do semivariograma são definidos da seguinte forma:

- O alcance (a) é entendido como a distância h em que a semivariância $\gamma(h)$ estabiliza. Para distâncias $h > a$, pode-se afirmar que as amostras não apresentam correlação espacial.

- A contribuição (C_1) é a quantidade de variação do patamar que pode ser explicada pelo modelo teórico.
- O efeito pepita (C_0) é a variação atribuída ao acaso. A semivariância deveria ser igual a zero quando $h = 0$, mas nem sempre isso acontece. Esse fato pode ser explicado por erros de medição ou, até mesmo, por não amostragem de distâncias menores em que o fenômeno ainda apresente uma certa quantidade considerável de variação, caracterizando assim o efeito pepita.
- O patamar (C) é a soma do efeito pepita e da contribuição (variação total), sendo que, alguns modelos atingem o patamar exatamente no alcance. Já outros modelos atingem o patamar assintoticamente. Nesse caso, o patamar é dado aproximadamente pelo efeito pepita mais 95% da contribuição.

Os modelos teóricos são ajustados por algoritmos de aproximação considerando valores iniciais para os parâmetros. Para isso, é utilizado, em geral, algum dos métodos de estimação como, mínimos quadrados (ordinários, ponderados e generalizados), máxima verossimilhança e máxima verossimilhança restrita. Segundo Cressie (1991), pelo fato do método da máxima verossimilhança apresentar alguns pressupostos, que nem sempre são atendidos, muitas vezes obtêm-se estimativas viciadas dos parâmetros populacionais. O método de mínimos quadrados consiste em minimizar a soma de quadrados entre a diferença dos valores observados e estimados da semivariância e, por não apresentar pressuposição quanto a distribuição de probabilidade das amostras, esse método tem obtido a preferência de muitos pesquisadores na estimação de semivariogramas. Olea (2006) apresenta uma lista com os modelos teóricos de semivariograma com patamar mais destacáveis. Esses modelos são:

Modelo exponencial:

$$\gamma(h) = \begin{cases} 0 & , h = 0 \\ C_0 + C_1 \left(1 - e^{-\frac{3h}{a}}\right) & , h > 0 \end{cases} ; \quad (2.2)$$

Modelo gaussiano:

$$\gamma(h) = \begin{cases} 0 & , h = 0 \\ C_0 + C_1 \left(1 - e^{-3\left(\frac{h}{a}\right)^2}\right) & , h > 0 \end{cases} ; \quad (2.3)$$

Modelo esférico:

$$\gamma(h) = \begin{cases} 0 & , h = 0 \\ C_0 + C_1 \left(\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a}\right)^3\right) & , 0 < h < a \\ C_0 + C_1 & , h \geq a \end{cases} ; \quad (2.4)$$

Modelo pentaesférico:

$$\gamma(h) = \begin{cases} 0 & , h = 0 \\ C_0 + C_1 \left(\frac{15h}{8a} - \frac{5}{4} \left(\frac{h}{a}\right)^3 + \frac{3}{8} \left(\frac{h}{a}\right)^5\right) & , 0 < h < a \\ C_0 + C_1 & , h \geq a \end{cases} ; \quad (2.5)$$

Modelo cúbico:

$$\gamma(h) = \begin{cases} 0 & , h = 0 \\ C_0 + C_1 \left(7 \left(\frac{h}{a}\right)^2 - \frac{35}{4} \left(\frac{h}{a}\right)^3 + \frac{7}{2} \left(\frac{h}{a}\right)^5 - \frac{3}{4} \left(\frac{h}{a}\right)^7\right) & , 0 < h < a \\ C_0 + C_1 & , h \geq a \end{cases} ; \quad (2.6)$$

Modelo senoidal ou "Wave":

$$\gamma(h) = \begin{cases} 0 & , h = 0 \\ C_0 + C_1 \left(1 - \frac{\text{sen}\left(\frac{\pi h}{a}\right)}{\pi \frac{h}{a}} \right) & , h > 0 \end{cases} ; \quad (2.7)$$

em que, C_0 é o efeito pepita, C_1 é a contribuição, a é o alcance e h é a distância entre os pontos.

2.1.2 Índices de dependência espacial

A razão do efeito pepita sobre o patamar (CAMBARDELLA et al., 1994), fornece a proporção de variação ao acaso, ou seja, não explicada pelo modelo. Pode-se definir “razão de dependência” (RD) pela expressão:

$$RD = \frac{C_0}{C_0 + C_1} 100 \quad , \quad (2.8)$$

em que, C_0 é o efeito pepita e C_1 é a contribuição.

No estudo de Cambardella et al. (1994) é apresentada uma classificação considerando que: se o índice RD for menor que 25%, a dependência espacial pode ser classificada como forte, no intervalo entre 25% e 75% é classificada como moderada e acima de 75% como fraca dependência espacial.

Biondi, Myers e Avery (1994) apresentaram o inverso do índice RD ao fazer a razão da contribuição sobre o patamar (SPD). No mesmo estudo, os autores classificam o resultado como fraca dependência espacial se o SPD for menor ou igual a 25%, moderada no intervalo entre 25% e 75% e forte dependência espacial acima de 75%.

$$SPD = \left(\frac{C_1}{C_0 + C_1} \right) 100 \quad , \quad (2.9)$$

em que, C_0 é o efeito pepita e C_1 é a contribuição.

Seidel e Oliveira (2015) apresentam uma medida de dependência espacial baseada em áreas de dependência espacial do semivariograma. As áreas de dependência espacial (ADE) foram propostas inicialmente por Seidel e Oliveira (2014b) como estatísticas de testes de hipóteses para a existência de dependência espacial.

O cálculo das $ADE_{observada}$ e da ADE_{maxima} é realizado respectivamente por (2.10) e (2.11):

$$ADE_{observada} = \int_0^a [C - \gamma(h)] dh, \quad (2.10)$$

$$ADE_{maxima} = \int_0^a [C - \gamma(h)_{adaptado}] dh, \quad (2.11)$$

em que,

C é o patamar;

a é o alcance;

$\gamma(h)$ é o modelo teórico de semivariograma;

$\gamma(h)_{adaptado}$ é o modelo teórico de semivariograma considerando $C_0 = 0$ ($C_1 = C$).

Em Seidel e Oliveira (2015) essas áreas foram calculadas para os modelos esférico, exponencial e gaussiano. Para os três modelos, o resultado da razão das áreas de dependência espacial $ADE_{observada}$ e ADE_{maxima} coincidem com o índice SPD . Seidel e Oliveira (2015) argumentam que o índice SPD é apenas apresentado por Biondi, Myers e Avery (1994), como a relação da variação estrutural (C_1) com o patamar (C), com isso, o desenvolvimento da razão das áreas de dependência espacial passa a ser uma justificativa geométrica baseada em áreas de dependência espacial para o índice SPD .

Considerar áreas do semivariograma com o intuito de construir medidas para descrever a dependência espacial é bastante razoável. Schabenberger e Pierce (2001) afirmam que a noção de estrutura espacial de um campo aleatório está relacionada ao seu grau de continuidade, isto é, quanto mais lento o aumento do semivariograma próximo da origem, mais o processo é estruturado espacialmente. Os autores exemplificam essas diferenças de comportamentos com os modelos exponencial e esférico, com o mesmo patamar e alcance, uma vez que o primeiro aumenta mais rapidamente, à medida que a distância aumenta, caracterizando-se, portanto menos estruturado.

A partir da função de correlação de um processo $Z(s)$ estacionário de segunda ordem Lumley e Panofsky (1964) propõem duas medidas (I_1 e I_2) que têm a finalidade de caracterizar tal função com escalas de comprimento, chamadas escalas integrais. Para fenômenos avaliados no espaço bidimensional, Russo e Bresler (1981) comentam que a medida I_2 pode ser interpre-

tada como a maior distância média para a qual $z(\mathbf{s}_1)$ e $z(\mathbf{s}_2)$ são fortemente correlacionados. As escalas integrais são calculadas da seguinte forma:

$$I_1 = \int_0^{\infty} \rho(h)dh \quad , \quad (2.12)$$

$$I_2 = \left\{ 2 \int_0^{\infty} \rho(h)h dh \right\}^{\frac{1}{2}} \quad , \quad (2.13)$$

em que,

$\rho(h)$ é o correlograma;

h é a distância.

Seidel e Oliveira (2014a), consideraram uma forma similar à medida I_1 , ao propor um índice de dependência espacial baseado em área de correlação espacial. O método de construção do índice proposto difere da escala integral, basicamente, por duas alterações: a primeira é a modificação do limite superior de integração, em vez de $+\infty$, os autores consideraram o alcance prático, justificando que para o índice deveriam ser consideradas somente distâncias em que há correlação entre as observações. A diferença do limite superior de integração pode ser verificada a seguir para os modelos exponencial, esférico e gaussiano.

$$\begin{aligned} I_{1exponencial} &= \int_0^{\infty} \rho(h)dh \\ &= \int_0^{\infty} \frac{[C - \gamma(h)]}{C} dh \\ &= \int_0^{\infty} \frac{[C - C_0 - C_1 (1 - e^{-\frac{3h}{a}})]}{C} dh \\ &= \frac{C_1}{C_0 + C_1} \int_0^{\infty} e^{-\frac{3h}{a}} dh \\ &= \frac{C_1}{C_0 + C_1} \lim_{h \rightarrow \infty} \left[\left(-\frac{a}{3} e^{-\frac{3h}{a}} \right) - \left(-\frac{a}{3} e^{\frac{0}{a}} \right) \right] \\ &= \frac{C_1}{C_0 + C_1} \frac{a}{3} \\ I_{1exponencial} &= 0,333a \frac{C_1}{C_0 + C_1} \end{aligned} \quad (2.14)$$

$$\begin{aligned}
I_{1esferico} &= \int_0^{\infty} \rho(h)dh \\
&= \int_0^{\infty} \frac{[C - \gamma(h)]}{C} dh \\
&= \int_0^{\infty} \frac{\left[C - C_0 - C_1 \left(\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right) \right]}{C} dh \\
&= \int_0^{\infty} \frac{C_1}{C_0 + C_1} - \frac{C_1}{C_0 + C_1} \left(\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right) dh \\
&= \frac{C_1}{C_0 + C_1} a - \frac{C_1}{C_0 + C_1} \left(\frac{3a^2}{4} - \frac{1a^4}{8a^3} \right) \\
&= \frac{C_1}{C_0 + C_1} a - \frac{C_1}{C_0 + C_1} \frac{3}{4} a + \frac{C_1}{C_0 + C_1} \frac{1}{8} a \\
I_{1esferico} &= 0,375a \frac{C_1}{C_0 + C_1} \tag{2.15}
\end{aligned}$$

$$\begin{aligned}
I_{1gaussiano} &= \int_0^{\infty} \rho(h)dh \\
&= \int_0^{\infty} \frac{[C - \gamma(h)]}{C} dh \\
&= \int_0^{\infty} \frac{\left[C - C_0 - C_1 \left(1 - e^{-\frac{3h^2}{a^2}} \right) \right]}{C} dh \\
&= \frac{C_1}{C_0 + C_1} \int_0^{\infty} e^{-\frac{3h^2}{a^2}} dh \\
I_{1gaussiano} &= 0,512a \frac{C_1}{C_0 + C_1} \tag{2.16}
\end{aligned}$$

Integrando a área de correlação espacial (*ACE*) até o alcance, desenvolvido em Seidel e Oliveira (2014a).

$$\begin{aligned}
ACE_{\text{exponencial}} &= \int_0^a \rho(h) dh \\
&= \int_0^a \frac{[C - \gamma(h)]}{C} dh \\
&= \int_0^a \frac{\left[C - C_0 - C_1 \left(1 - e^{-\frac{3h}{a}} \right) \right]}{C} dh \\
&= \frac{C_1}{C_0 + C_1} \int_0^a e^{-\frac{3h}{a}} dh \\
&= \frac{C_1}{C_0 + C_1} \left[\left(-\frac{a}{3} e^{-\frac{3a}{a}} \right) - \left(-\frac{a}{3} e^{\frac{0}{a}} \right) \right] \\
&= \frac{C_1}{C_0 + C_1} \left[-\frac{a}{3} e^{-3} + \frac{a}{3} \right] \\
ACE_{\text{exponencial}} &= 0,317a \frac{C_1}{C_0 + C_1} \tag{2.17}
\end{aligned}$$

$$\begin{aligned}
ACE_{\text{esferico}} &= \int_0^a \rho(h) dh \\
&= \int_0^a \frac{[C - \gamma(h)]}{C} dh \\
&= \int_0^a \frac{\left[C - C_0 - C_1 \left(\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right) \right]}{C} dh \\
&= \int_0^a \frac{C_1}{C_0 + C_1} - \frac{C_1}{C_0 + C_1} \left(\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right) dh \\
&= \frac{C_1}{C_0 + C_1} a - \frac{C_1}{C_0 + C_1} \left(\frac{3a^2}{4a} - \frac{1a^4}{8a^3} \right) \\
&= \frac{C_1}{C_0 + C_1} a - \frac{C_1}{C_0 + C_1} \frac{3}{4} a + \frac{C_1}{C_0 + C_1} \frac{1}{8} a \\
ACE_{\text{esferico}} &= 0,375a \frac{C_1}{C_0 + C_1} \tag{2.18}
\end{aligned}$$

$$\begin{aligned}
ACE_{gaussiano} &= \int_0^{\infty} \rho(h) dh \\
&= \int_0^a \frac{[C - \gamma(h)]}{C} dh \\
&= \int_0^a \frac{\left[C - C_0 - C_1 \left(1 - e^{-\frac{3h^2}{a^2}} \right) \right]}{C} dh \\
&= \frac{C_1}{C_0 + C_1} \int_0^{\infty} e^{-\frac{3h^2}{a^2}} dh \\
ACE_{gaussiano} &= 0,504a \frac{C_1}{C_0 + C_1} \tag{2.19}
\end{aligned}$$

A segunda alteração realizada por Seidel e Oliveira (2014a) foi a divisão do alcance prático (resultante da integral) por uma proporção da máxima distância entre os pontos amostrados ($q.MD$), proporção esta definida pelo próprio pesquisador usuário do índice. Os autores justificam que tal alteração foi feita para tornar o índice adimensional, facilitando assim sua interpretação, diferentemente da escala integral I_1 , que é dada na unidade de medida do alcance. Em outro estudo Seidel e Oliveira (2016) recomendam que, ao invés de utilizar a proporção q , seja utilizada a metade da maior distância entre os pontos amostrados. A justificativa dos autores para utilizar $0,5MD$ vem da literatura clássica da Geoestatística que recomenda usar metade da máxima distância entre pontos para se obter o semivariograma experimental (JOURNEL; HUIJBREGTS, 1978; OLEA, 2006). Para casos em que o alcance prático seja maior que a metade da maior distância entre os pontos amostrados, o valor da razão $\frac{a}{0,5MD}$ deve ser truncado em 1 para que essa componente tenha variação somente no intervalo (0,1). Com isso a razão $\frac{a}{0,5MD}$ permite ter a mesma ideia de comparação da contribuição com o patamar $\frac{C_1}{C_0+C_1}$. O índice de dependência espacial (IDE) é diferente para cada modelo de semivariograma utilizado em função do fator de modelo. Em Appel Neto et al. (2018) o IDE é apresentado também para os modelos cúbico, pentaesférico e wave. Assim, os índices para os seis modelos são dados por:

$$IDE_{Esferico}\% = 0,375 \left(\frac{C_1}{C_0 + C_1} \right) \left(\frac{a}{0,5MD} \right) 100 \quad , \tag{2.20}$$

$$IDE_{Exponencial}\% = 0,317 \left(\frac{C_1}{C_0 + C_1} \right) \left(\frac{a}{0,5MD} \right) 100 \quad , \tag{2.21}$$

$$IDE_{Gaussiano} \% = 0,504 \left(\frac{C_1}{C_0 + C_1} \right) \left(\frac{a}{0,5MD} \right) 100 \quad , \quad (2.22)$$

$$IDE_{Cubico} \% = 0,365 \left(\frac{C_1}{C_0 + C_1} \right) \left(\frac{a}{0,5MD} \right) 100 \quad , \quad (2.23)$$

$$IDE_{Pentaesferico} \% = 0,312 \left(\frac{C_1}{C_0 + C_1} \right) \left(\frac{a}{0,5MD} \right) 100 \quad , \quad (2.24)$$

$$IDE_{Wave} \% = 0,589 \left(\frac{C_1}{C_0 + C_1} \right) \left(\frac{a}{0,5MD} \right) 100 \quad , \quad (2.25)$$

em que, os valores constantes são os FM , a componente $0,5MD$ é o valor que corresponde à metade da máxima distância MD entre os pontos amostrados.

Em Seidel e Oliveira (2016) é apresentada uma classificação do IDE para os modelos esférico, exponencial e gaussiano baseada nos quartis de uma distribuição empírica decorrente de 169 valores obtidos com a combinação de possíveis valores que as componentes $\frac{C_1}{C_0+C_1}$ e $\frac{a}{0,5MD}$ do índice podem assumir. Cada modelo teve classificações de fraca, moderada e forte dependência espacial, sendo que os valores limites para essas classificações foram diferentes para cada modelo, pelo fato de não possuírem o mesmo fator de modelo. Assim como em Seidel e Oliveira (2016), no estudo de Appel Neto et al. (2018) podem ser verificadas classificações do IDE para os modelos pentaesférico, cúbico e wave.

3 MATERIAL E MÉTODOS

3.0.3 Dados

Foram utilizados os dados de Santos et al. (2018) que são compostos por 500 ajustes de semivariogramas relacionados a atributos de solos, coletados de 25 artigos da área de Ciência do Solo publicados em periódicos indexados no portal SciELO Brasil (*Scientific Electronic Library Online*), no período de 2006 a 2015.

3.1 Métodos

3.1.1 Medida de correlação espacial

A proposta da nova medida de mensuração da dependência espacial (*MDE*) foi desenvolvida com base na escala integral I_2 , de forma similar a Seidel e Oliverira (2014a), considerando o mesmo entendimento de que o limite superior de integração adotado será o alcance, para que sejam consideradas apenas distâncias em que há correlação espacial. Desse modo, os cálculos foram desenvolvidos a partir da medida de correlação espacial (*MCE*):

$$MCE = \left\{ 2 \int_0^a \rho(h)hdh \right\}^{\frac{1}{2}} \quad (3.1)$$

Considerando que o processo $\{Z(\mathbf{s})\}$ apresente estacionariedade de segunda ordem a correlação pode ser estimada por meio da relação entre variância e semivariância como visto em (2.1), a abordagem sobre área de correlação espacial apresentada em Seidel e Oliveira (2014a) tem como particularidade de que cada modelo de semivariograma apresenta um fator de modelo diferente do outro, com isso, seguindo a afirmação de Olea (2006) sobre os modelos de semivariograma mais usualmente utilizados, foram explorados os modelos de semivariograma esférico, exponencial, gaussiano, pentaesférico, cúbico e wave.

3.1.2 Classificação do novo índice

Para a obtenção dos níveis de classificação do índice proposto foram consideradas várias combinações de possíveis resultados que o novo índice poderá assumir baseando-se no intervalo de variação de duas componentes necessárias para o cálculo do índice. Foram considerados 20 valores para cada componente variando em intervalos de 0,05 em 0,05 (0,05;0,10;0,15;...;0,95;

1,00), os quais foram combinados e multiplicados, resultando em 400 valores e ainda por fim foi incluído o valor zero, a fim de contemplar uma única situação em que possa haver efeito pepita puro ($C_1 = 0$ e $a = 0$). Esses resultados ainda foram multiplicados pelos *FM* dos modelos de semivariograma em estudo. Obtidos os resultados dessas combinações foram realizadas análises descritivas para explorar o comportamento da distribuição empírica desses dados. Com base nessa distribuição, foram avaliados os intervalos que aparentam ser mais equilibrados para classificar o novo índice em forte, moderada ou fraca dependência espacial.

3.1.3 Comparações dos índices de dependência espacial

Foram realizados alguns cenários de simulações com a finalidade de avaliar quais índices em estudo sintetizam melhor as informações contidas no semivariograma. A ideia consistiu em comparar a correlação de cada índice com o erro quadrático médio (*EQM*) e com a variância de krigagem (*VAR*) da validação cruzada realizada a partir da krigagem ordinária, tendo como entendimento que, o índice que melhor representa a dependência espacial é o que apresenta correlações com *EQM* e *VAR* mais próximas de -1, ou seja, quanto maior for o índice de dependência espacial, menor será o *EQM* e a *VAR* e vice-versa.

Tabela 3.1 – Cenários com os parâmetros utilizados para simulação dos dados com dependência espacial.

Cenário	Valores iniciais		
	C_1	C_0	a
1	5	45	7,07
2	5	45	17,68
3	5	45	35,36
4	5	45	53,03
5	5	45	63,64
6	12,5	37,5	7,07
7	12,5	37,5	17,68
8	12,5	37,5	35,36
9	12,5	37,5	53,03
10	12,5	37,5	63,64
11	25	25	7,07
12	25	25	17,68
13	25	25	35,36
14	25	25	53,03
15	25	25	63,64
16	37,5	12,5	7,07
17	37,5	12,5	17,68
18	37,5	12,5	35,36
19	37,5	12,5	53,03
20	37,5	12,5	63,64
21	45	5	7,07
22	45	5	17,68
23	45	5	35,36
24	45	5	53,03
25	45	5	63,64

Fonte: Do autor (2019).

Os dados com dependência espacial foram simulados, por meio da função *grf* do pacote *geoR* (RIBEIRO JÚNIOR; DIGGLE, 2001) do *Software R* (R CORE TEAM, 2017). Na Tabela 3.1, apresentam-se os 25 cenários de diferentes estruturas de dependência espacial que foram considerados para simulação dos dados com dependência espacial. As 25 diferentes estruturas de dependência espacial foram escolhidas combinando 5 proporções da contribuição sobre o patamar (10%, 25%, 50%, 75% e 90%) e 5 proporções do alcance sobre a metade da maior distância entre os pontos da malha de amostra (10%, 25%, 50%, 75% e 90%). Para todos os cenários, foram consideradas malhas regulares de dimensão 100 x 100m com 169 pontos e média igual a zero. Ao todo, foram realizadas simulações para 5 dos 6 modelos em estudo, com exceção do modelo pentaesférico, pelo fato do mesmo não estar implementado no pacote *geoR*. Cada cenário foi replicado 100 vezes com a finalidade de amenizar possíveis variações do algoritmo de simulação dos dados. A partir dos 100 valores gerados, foram calculadas as médias dos índices, dos erros quadráticos médios (*EQMs*) e das variâncias de krigagem (*VARs*), logo as correlações foram calculadas sobre dois vetores de médias. O ajuste do semivariograma teórico foi realizado, por meio do método de mínimos quadrados ponderados, utilizando a função *optim* para realizar as aproximações das estimativas dos parâmetros. Os valores iniciais para se realizar os ajustes foram os mesmos utilizados para gerar os dados com dependência. Com a finalidade de ser possível replicar os resultados das simulações, foi utilizada a semente “2016”.

Com base nos dados reais coletados, foram calculados os índices de dependência espacial *SPD*, *IDE* e o novo índice proposto (*MDE*) em cada ajuste de semivariograma. Ao todo, foram calculados 500 *IDE's*, *SPD's* e *MDE's* e, com base nesses resultados, foram quantificadas as diferenças entre as suas respectivas classificações.

4 RESULTADOS

A seguir, são apresentados os cálculos das medidas de correlação espacial a partir da medida I_2 . Para a realização dos cálculos, foi feita uma adaptação no limite superior de integração, em que foi considerado o alcance (a).

$$\begin{aligned}
MCE_{esferico} &= \left\{ 2 \int_0^a \rho(h) h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \frac{[C - \gamma(h)]}{C} h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \frac{[C - (C_0 + C_1 (\frac{3}{2} (\frac{h}{a}) - \frac{1}{2} (\frac{h}{a})^3))] }{C} h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \left[\frac{C_1}{C_0 + C_1} - \frac{C_1}{C_0 + C_1} \left(\frac{3}{2} \left(\frac{h}{a} \right) - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right) \right] h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \frac{C_1}{C_0 + C_1} \left[1 - \left(\frac{3}{2} \left(\frac{h}{a} \right) - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right) \right] h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \left[\int_0^a \frac{C_1}{C_0 + C_1} h dh - \int_0^a \frac{C_1}{C_0 + C_1} \left(\frac{3}{2} \left(\frac{h}{a} \right) \right) h dh + \right. \right. \\
&\quad \left. \left. \int_0^a \frac{C_1}{C_0 + C_1} \left(\frac{1}{2} \left(\frac{h}{a} \right)^3 \right) h dh \right] \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \left[\frac{C_1}{C_0 + C_1} \left(\frac{a^2}{2} \right) - \frac{C_1}{C_0 + C_1} \left(\frac{3a^2}{6} \right) + \frac{C_1}{C_0 + C_1} \left(\frac{a^2}{10} \right) \right] \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \left[\left(\frac{C_1}{C_0 + C_1} \right) a^2 \left(\frac{1}{2} - \frac{3}{6} + \frac{1}{10} \right) \right] \right\}^{\frac{1}{2}} \\
MCE_{esferico} &= 0,447 \left(\frac{C_1}{C_0 + C_1} \right)^{\frac{1}{2}} a \tag{4.1}
\end{aligned}$$

$$\begin{aligned}
MCE_{\text{exponencial}} &= \left\{ 2 \int_0^a \rho(h) h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \frac{[C - \gamma(h)]}{C} h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \frac{[C - (C_0 + C_1 (1 - e^{-\frac{3h}{a}}))] }{C} h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \left(\frac{C_1}{C_0 + C_1} \right) e^{-\frac{3h}{a}} h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \left(\frac{C_1}{C_0 + C_1} \right) \int_0^a e^{-\frac{3h}{a}} h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \left[\left(\frac{C_1}{C_0 + C_1} \right) \frac{(-4e^{-3} + 1)a^2}{9} \right] \right\}^{\frac{1}{2}} \\
MCE_{\text{exponencial}} &= 0,422 \left(\frac{C_1}{C_0 + C_1} \right)^{\frac{1}{2}} a \tag{4.2}
\end{aligned}$$

No estudo de Shukla et al. (2017), foram desenvolvidos os cálculos para a escala integral I_2 para os modelos esférico e exponencial. Os resultados encontrados pelos autores são diferentes das equações 4.1 e 4.2 apenas no fator de modelo. Essa diferença é explicada pelo fato dos autores usarem $+\infty$ como limite superior integração, diferentemente, neste estudo foi utilizado o alcance como limite superior.

Tem-se a seguir os cálculos para o modelo gaussiano:

$$\begin{aligned}
MCE_{gaussiano} &= \left\{ 2 \int_0^a \rho(h) h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \frac{[C - \gamma(h)]}{C} h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \frac{\left[C - \left(C_0 + C_1 \left(1 - e^{-3\left(\frac{h}{a}\right)^2} \right) \right) \right]}{C} h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \left(\frac{C_1}{C_0 + C_1} \right) e^{-3\left(\frac{h}{a}\right)^2} h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \left(\frac{C_1}{C_0 + C_1} \right) \int_0^a e^{-3\left(\frac{h}{a}\right)^2} h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \left[\left(\frac{C_1}{C_0 + C_1} \right) \frac{-a^2 (e^{-3} - e^0)}{6} \right] \right\}^{\frac{1}{2}} \\
MCE_{gaussiano} &= 0,563 \left(\frac{C_1}{C_0 + C_1} \right)^{\frac{1}{2}} a \tag{4.3}
\end{aligned}$$

Para o modelo cúbico tem-se:

$$\begin{aligned}
MCE_{cubico} &= \left\{ 2 \int_0^a \rho(h) h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \frac{[C - \gamma(h)]}{C} h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \frac{\left[C - \left(C_0 + C_1 \left(7 \left(\frac{h}{a} \right)^2 - \frac{35}{4} \left(\frac{h}{a} \right)^3 + \frac{7}{2} \left(\frac{h}{a} \right)^5 - \frac{3}{4} \left(\frac{h}{a} \right)^7 \right) \right]}{C} h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \left[\frac{C_1}{C_0 + C_1} - \frac{C_1}{C_0 + C_1} \left(7 \left(\frac{h}{a} \right)^2 - \frac{35}{4} \left(\frac{h}{a} \right)^3 + \frac{7}{2} \left(\frac{h}{a} \right)^5 - \frac{3}{4} \left(\frac{h}{a} \right)^7 \right) \right] h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \frac{C_1}{C_0 + C_1} \left[1 - \left(7 \left(\frac{h}{a} \right)^2 - \frac{35}{4} \left(\frac{h}{a} \right)^3 + \frac{7}{2} \left(\frac{h}{a} \right)^5 - \frac{3}{4} \left(\frac{h}{a} \right)^7 \right) \right] h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \left[\int_0^a \frac{C_1}{C_0 + C_1} h dh - \int_0^a \frac{C_1}{C_0 + C_1} \left(7 \left(\frac{h}{a} \right)^2 \right) h dh + \int_0^a \frac{C_1}{C_0 + C_1} \left(\frac{35}{4} \left(\frac{h}{a} \right)^3 \right) h dh \right. \right. \\
&\quad \left. \left. - \int_0^a \frac{C_1}{C_0 + C_1} \left(\frac{7}{2} \left(\frac{h}{a} \right)^5 \right) h dh + \int_0^a \frac{C_1}{C_0 + C_1} \left(\frac{3}{4} \left(\frac{h}{a} \right)^7 \right) h dh \right] \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \frac{C_1}{C_0 + C_1} \left[\frac{h^2}{2} - \frac{7 h^4}{4 a^2} + \frac{35 h^5}{20 a^3} - \frac{7 h^7}{14 a^5} - \frac{3 h^9}{36 a^7} \right]_0^a \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \frac{C_1}{C_0 + C_1} \left[\frac{a^2}{2} - \frac{7 a^4}{4 a^2} + \frac{35 a^5}{20 a^3} - \frac{7 a^7}{14 a^5} - \frac{3 a^9}{36 a^7} \right] \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \left[\left(\frac{C_1}{C_0 + C_1} \right) a^2 \left(\frac{1}{2} - \frac{7}{4} + \frac{35}{20} - \frac{7}{14} + \frac{3}{36} \right) \right] \right\}^{\frac{1}{2}} \\
MCE_{cubico} &= 0,408 \left(\frac{C_1}{C_0 + C_1} \right)^{\frac{1}{2}} a \tag{4.4}
\end{aligned}$$

Para o modelo pentaesférico tem-se:

$$\begin{aligned}
MCE_{pentaesferico} &= \left\{ 2 \int_0^a \rho(h) h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \frac{[C - \gamma(h)]}{C} h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \frac{[C - (C_0 + C_1 (\frac{15}{8} (\frac{h}{a}) - \frac{5}{4} (\frac{h}{a})^3 + \frac{3}{8} (\frac{h}{a})^5))] }{C} h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \left[\frac{C_1}{C_0 + C_1} - \frac{C_1}{C_0 + C_1} \left(\frac{15}{8} \left(\frac{h}{a} \right) - \frac{5}{4} \left(\frac{h}{a} \right)^3 + \frac{3}{8} \left(\frac{h}{a} \right)^5 \right) \right] h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \frac{C_1}{C_0 + C_1} \left[1 - \left(\frac{15}{8} \left(\frac{h}{a} \right) - \frac{5}{4} \left(\frac{h}{a} \right)^3 + \frac{3}{8} \left(\frac{h}{a} \right)^5 \right) \right] h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \left[\int_0^a \frac{C_1}{C_0 + C_1} h dh - \int_0^a \frac{C_1}{C_0 + C_1} \left(\frac{15}{8} \left(\frac{h}{a} \right) \right) h dh \right. \right. \\
&\quad \left. \left. + \int_0^a \frac{C_1}{C_0 + C_1} \left(\frac{5}{4} \left(\frac{h}{a} \right)^3 \right) h dh - \int_0^a \frac{C_1}{C_0 + C_1} \left(\frac{3}{8} \left(\frac{h}{a} \right)^5 \right) h dh \right] \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \frac{C_1}{C_0 + C_1} \left[\frac{h^2}{2} - \frac{15 h^3}{24 a} + \frac{5 h^5}{20 a^3} - \frac{3 h^7}{56 a^5} \right]_0^a \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \left[\left(\frac{C_1}{C_0 + C_1} \right) a^2 \left(\frac{1}{2} - \frac{15}{24} + \frac{5}{20} - \frac{3}{56} \right) \right] \right\}^{\frac{1}{2}} \\
MCE_{pentaesferico} &= 0,378 \left(\frac{C_1}{C_0 + C_1} \right)^{\frac{1}{2}} a \tag{4.5}
\end{aligned}$$

Por fim, para o modelo wave.

$$\begin{aligned}
MCE_{wave} &= \left\{ 2 \int_0^a \rho(h) h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \frac{[C - \gamma(h)]}{C} h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \int_0^a \frac{\left[C - \left(C_0 + C_1 \left(1 - \frac{\sin(\frac{\pi h}{a})}{\pi \frac{h}{a}} \right) \right) \right]}{C} h dh \right\}^{\frac{1}{2}} \\
&= \left\{ 2 \left(\frac{C_1}{C_0 + C_1} \right) \int_0^a \frac{\sin(\frac{\pi h}{a})}{\pi \frac{h}{a}} h dh \right\}^{\frac{1}{2}} \\
MCE_{wave} &= 0,637 \left(\frac{C_1}{C_0 + C_1} \right)^{\frac{1}{2}} a
\end{aligned}$$

Seguindo a mesmo método de Seidel e Oliveira (2014a), o índice *MDE* foi criado a partir dos resultados das *MCEs* adaptando apenas a divisão do alcance pela metade da maior distância entre os pontos amostrados. Essa adaptação foi realizada para garantir que o índice seja adimensional e o alcance em relação à metade da maior distância tenha a mesma ideia de proporção como a contribuição tem do patamar.

$$MDE_{esferico} = 0,447 \left(\frac{C_1}{C_0 + C_1} \right)^{\frac{1}{2}} \frac{a}{0,5MD} 100 \quad , \quad (4.6)$$

$$MDE_{exponencial} = 0,422 \left(\frac{C_1}{C_0 + C_1} \right)^{\frac{1}{2}} \frac{a}{0,5MD} 100 \quad , \quad (4.7)$$

$$MDE_{gaussiano} = 0,563 \left(\frac{C_1}{C_0 + C_1} \right)^{\frac{1}{2}} \frac{a}{0,5MD} 100 \quad , \quad (4.8)$$

$$MDE_{cubico} = 0,408 \left(\frac{C_1}{C_0 + C_1} \right)^{\frac{1}{2}} \frac{a}{0,5MD} 100 \quad , \quad (4.9)$$

$$MDE_{pentaesferico} = 0,378 \left(\frac{C_1}{C_0 + C_1} \right)^{\frac{1}{2}} \frac{a}{0,5MD} 100 \quad , \quad (4.10)$$

$$MDE_{wave} = 0,637 \left(\frac{C_1}{C_0 + C_1} \right)^{\frac{1}{2}} \frac{a}{0,5MD} 100 \quad , \quad (4.11)$$

Nas equações 4.6 a 4.11 são apresentados os índices de dependência espacial desenvolvidos a partir das medidas de correlação espacial (*MCEs*). Percebe-se que esse novo índice proposto difere basicamente do *IDE* no fator de modelo e também uma raiz quadrada que está sobre a componente $\frac{C_1}{C_0+C_1}$.

Uma generalização do índice *MDE* é apresentada na equação 4.12.

$$MDE_{modelo} = FM \left(\frac{C_1}{C_0 + C_1} \right)^{\frac{1}{2}} \frac{a}{0,5MD} 100 \quad , \quad (4.12)$$

em que,

FM é o fator de modelo;

C_1 é a contribuição;

C_0 é o efeito pepita;

a é o alcance.

Vale ressaltar que na equação 4.12 a componente $\frac{a}{0,5MD}$ deve ser truncada em 1, em situações que o alcance seja maior que a metade da maior distância.

Tabela 4.1 – Medidas descritivas dos 401 valores gerados pelas combinações das duas componentes do índice *MDE* (mais o valor zero incluído) para os seis modelos de semivariograma em estudo.

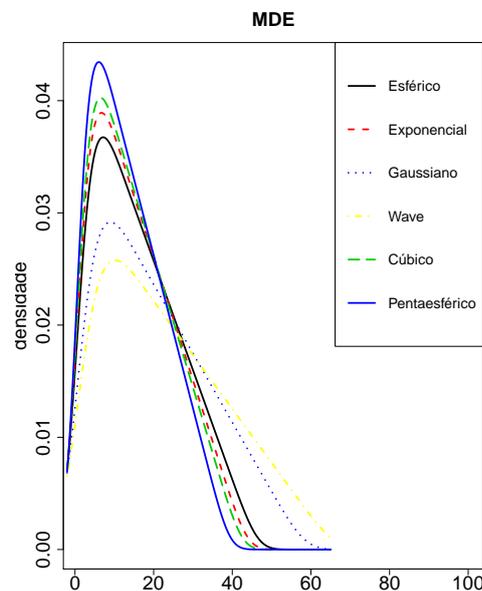
	Esférico	Exponencial	Gaussiano	Wave	Cúbico	Pentaesférico
Mínimo	0,00	0,00	0,00	0,00	0,00	0,00
Máximo	44,70	42,20	56,30	63,70	40,80	37,80
1º Quartil	7,07	6,67	8,90	10,07	6,45	5,98
Mediana	14,42	13,61	18,16	20,54	13,16	12,19
3º Quartil	23,80	22,47	29,98	33,92	21,72	20,13
Média	16,14	15,24	20,33	23,00	14,73	13,65
Desvio Padrão	10,73	10,13	13,52	15,30	9,80	9,08
Coef. Assimetria	0,55	0,55	0,55	0,55	0,55	0,55
Coef. Curtose	-0,62	-0,62	-0,62	-0,62	-0,62	-0,62

Fonte: Do autor (2019).

Na Tabela 4.2, é possível verificar que os valores máximos em cada modelo são atingidos pelo índice *MDE*, decorrente do mesmo apresentar maiores *FM*s que o índice *IDE* (APPEL NETO et al., 2018; SEIDEL; OLIVEIRA, 2014a, 2016) para os modelos estudados. Apesar de atingir maiores valores máximos, isso não significa que a dependência espacial seja mais

forte, visto que cada índice e cada modelo tenham os seus próprios limites para classificação. Ainda, por meio da Tabela 4.1 pode-se verificar que o MDE tem comportamento de assimetria positiva de forma semelhante ao IDE (APPEL NETO et al., 2018; SEIDEL; OLIVEIRA, 2016). Contudo, essa assimetria é mais suave, pelo fato de que a componente $frac{C_1 C_0 + C_1$ tem uma raiz quadrada aplicada. Na Figura 4.1 corrobora-se esse comportamento de assimetria positiva do MDE.

Figura 4.1 – Densidades dos 401 resultados que foram gerados através das combinações entre os 20 valores atribuídos a cada uma das duas componentes do índice de dependência espacial *MDE* (mais o valor zero incluído), para os seis modelos estudados.



Fonte: Do autor (2019).

Tabela 4.2 – Classificação proposta neste trabalho para o índice *MDE* utilizando a metodologia de Seidel e Oliveira (2016).

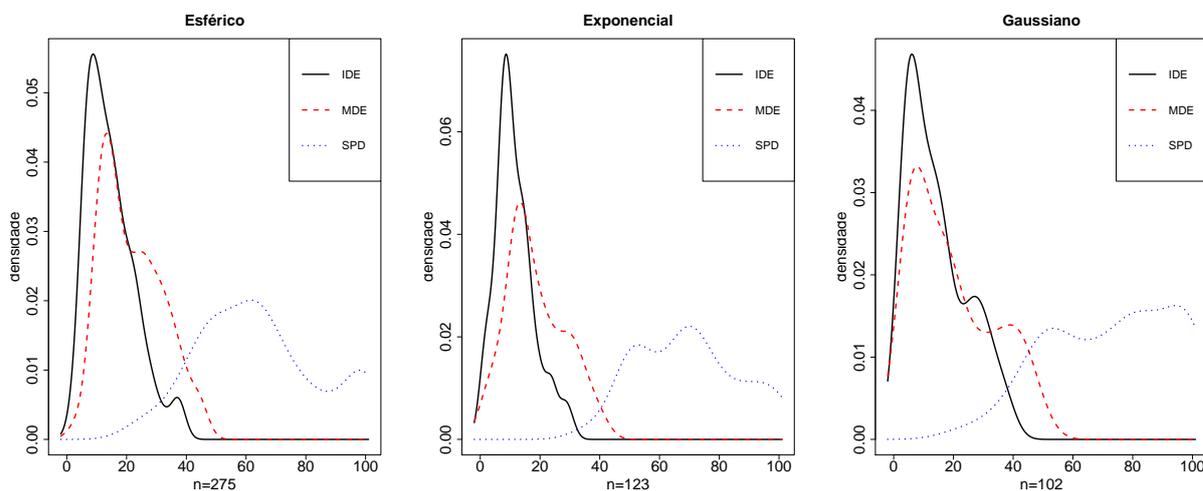
	Fraca	Moderada	Forte
Esférico	$0,00 \leq MDE \leq 14,00$	$14,00 < MDE \leq 24,00$	$24,00 < MDE \leq 44,70$
Exponencial	$0,00 \leq MDE \leq 14,00$	$14,00 < MDE \leq 22,00$	$22,00 < MDE \leq 42,20$
Gaussiano	$0,00 \leq MDE \leq 18,00$	$18,00 < MDE \leq 30,00$	$30,00 < MDE \leq 56,30$
Cúbico	$0,00 \leq MDE \leq 13,00$	$13,00 < MDE \leq 22,00$	$22,00 < MDE \leq 40,80$
Pentaesférico	$0,00 \leq MDE \leq 12,00$	$12,00 < MDE \leq 20,00$	$20,00 < MDE \leq 37,80$
Wave	$0,00 \leq MDE \leq 21,00$	$21,00 < MDE \leq 34,00$	$34,00 < MDE \leq 63,70$

Fonte: Do autor (2019).

Na Tabela 4.2, apresenta-se uma classificação proposta, baseada em Seidel e Oliveira (2016). De forma similar, essa classificação é composta de três níveis de dependência espacial, forte, moderada e fraca. Pelo fato de o índice apresentar uma assimetria positiva, os limites de

cada nível dessa classificação foram definidos com base na mediana e no terceiro quartil da distribuição empírica resultante das combinações dos 20 valores atribuídos para cada componente do índice.

Figura 4.2 – Histogramas suavizados de índices de dependência espacial aplicados à dados oriundos de atributos físicos, químicos e mineralógicos de solo.



Fonte: Do autor (2019).

Na Figura 4.2, são apresentadas as densidades empíricas dos índices *SPD*, *IDE* e *MDE* aplicados a dados de atributos de solo, onde se verifica os índices *IDE* e *MDE* apresentam um comportamento assimétrico, em decorrência do efeito do alcance. Ao estudar os mesmos dados, Santos et al. (2018) verificaram que o *IDE* teve correlações de moderadas a fortes com o alcance, que, por sua vez, apresentou um comportamento de forte assimetria à direita. Percebe-se, então, que o alcance exerce forte influência nos índices *IDE* e *MDE*. Essa aplicação a dados reais corrobora a ideia de que esses índices precisam ter uma classificação diferente da proposta por Cambardella et al. (1994), pelo fato de que há uma assimetria positiva. Ainda, na Figura 4.2, percebe-se que o comportamento do índice *SPD* é diferente para os 3 modelos (esférico, exponencial e gaussiano), em que, para o modelo gaussiano parece haver uma maior concentração de valores próximos de 100. Já para os modelos esférico e exponencial a maior concentração dos valores está em torno de 60 e 70 respectivamente.

Na Tabela 4.3, apresenta-se a classificação dos resultados dos índices *SPD*, *IDE* e *MDE* aplicados a dados de atributos de solo. Para o modelo esférico, pode-se verificar que o *SPD* apresentou uma grande quantidade de atributos classificados como moderada dependência espacial e poucos foram classificados como fraca dependência espacial. Já os índices *IDE* e *MDE* para o modelo esférico, apresentaram maior quantidade de atributos classificados como

Tabela 4.3 – Classificação da dependência espacial de atributos de solo utilizando a classificação de Cambardella et al. (1994) para o *SPD*, a classificação de Seidel e Oliveira (2016) para o *IDE* e a classificação proposta neste estudo para o *MDE*.

	Esférico			Exponencial			Gaussiano			Total		
	<i>SPD</i>	<i>IDE</i>	<i>MDE</i>	<i>SPD</i>	<i>IDE</i>	<i>MDE</i>	<i>SPD</i>	<i>IDE</i>	<i>MDE</i>	<i>SPD</i>	<i>IDE</i>	<i>MDE</i>
Fraca	3	43	80	0	17	49	1	40	54	4	163	183
Moderada	201	115	84	82	60	33	49	34	21	332	157	138
Forte	71	117	111	41	46	41	52	28	27	164	180	179
Total		275			123			102			500	

Fonte: Do autor (2019).

forte dependência espacial e um equilíbrio entre a diferença dos classificados como moderada e fraca dependência espacial. No modelo exponencial, o *SPD* teve resultados parecidos com o modelo esférico, o índice *IDE*, por sua vez, apresentou maior quantidade de atributos classificados como moderada dependência espacial, já o índice *MDE* apresentou maior quantidade de atributos classificados como fraca dependência espacial. No modelo gaussiano, o índice *SPD* apresentou maior número de atributos classificados como forte dependência espacial, em contrapartida os índices *IDE* e *MDE* apresentaram maior número de atributos classificados como fraca dependência espacial.

Tabela 4.4 – Correlações calculadas entre os índices *SPD*, *IDE*, *MDE* e entre o *EQM* e a variância de krigagem (*VAR*) da validação cruzada.

	Esférico			Exponencial			Gaussiano			Cúbico			Wave		
	<i>SPD</i>	<i>IDE</i>	<i>MDE</i>	<i>SPD</i>	<i>IDE</i>	<i>MDE</i>	<i>SPD</i>	<i>IDE</i>	<i>MDE</i>	<i>SPD</i>	<i>IDE</i>	<i>MDE</i>	<i>SPD</i>	<i>IDE</i>	<i>MDE</i>
<i>EQM</i>	-0,67	-0,84	-0,80	-0,92	-0,91	-0,90	-0,79	-0,79	-0,72	-0,63	-0,86	-0,81	-0,77	-0,51	-0,45
<i>VAR</i>	-0,44	-0,97	-0,94	-0,70	-0,97	-0,95	-0,77	-0,83	-0,77	-0,34	-0,95	-0,91	-0,78	-0,55	-0,50
<i>SPD</i>	1,00	0,43	0,36	1,00	0,76	0,72	1,00	0,61	0,51	1,00	0,75	0,66	1,00	0,41	0,30
<i>IDE</i>	0,43	1,00	0,99	0,76	1,00	0,99	0,61	1,00	0,98	0,75	1,00	0,99	0,41	1,00	0,98
<i>MDE</i>	0,36	0,99	1,00	0,72	0,99	1,00	0,51	0,98	1,00	0,66	0,99	1,00	0,30	0,98	1,00

Fonte: Do autor (2019).

Na Tabela 4.4 é possível verificar que, para o modelo esférico, a maior correlação negativa dos índices com o *EQM* foi do *IDE*, seguido pelo *MDE* e *SPD* respectivamente. Por meio desse resultado pode-se interpretar que, para o modelo esférico, o índice que melhor representa inversamente o *EQM* é o *IDE*. Já para o modelo exponencial, o índice que obteve a maior correlação negativa com o *EQM* foi o *SPD*, porém a diferença foi pequena em comparação com os demais índices. Para o modelo gaussiano, os índices *SPD* e *IDE* apresentaram uma ligeira vantagem em relação ao índice *MDE*, tratando-se da maior correlação negativa com o *EQM*. O modelo cúbico apresentou resultados similares ao modelo esférico, com o índice *IDE* apresentando maior correlação negativa com o *EQM*, o índice *MDE* com resultado próximo ao *IDE* e o *SPD* com uma correlação negativa menor que os demais índices. O modelo wave foi o

modelo em que os índices *IDE* e *MDE* apresentaram menor correlação negativa com o *EQM*, -0,51 e -0,45 respectivamente, já o índice *SPD* apresentou uma correlação negativa de -0,77 com o *EQM*.

Calculadas as correlações também dos índices com a variância de krigagem da validação cruzada (*VAR*), pode-se observar, na Tabela 4.4, que para o modelo esférico os índices *IDE* e *MDE* apresentaram resultados próximos, com uma correlação negativa de -0,97 e -0,94 respectivamente, já o índice *SPD* apresentou uma correlação negativa com a *VAR* de -0,44. Para o modelo exponencial os resultados foram muito próximos do modelo esférico em relação aos índices *IDE* e *SPD*, -0,97 e -0,95 respectivamente, já o índice *SPD* teve uma correlação maior do que comparada ao modelo esférico, porém, ainda apresentou a menor correlação negativa dos 3 índices, -0,70. No modelo gaussiano, o *IDE* apresentou-se a maior correlação negativa com a *VAR* -0,83, seguido do *MDE* e *SPD*, ambos com uma correlação negativa de -0,77. O modelo cúbico foi o que apresentou resultados semelhantes entre os índices *IDE* e *MDE*, com uma correlação negativa com a *VAR* de -0,95 e -0,91 respectivamente, já o *SPD* apresentou uma correlação negativa com a *VAR* de -0,34, a menor de todos os índices e modelos. Já no modelo wave o *SPD* apresentou a maior correlação negativa com a *VAR*, -0,78. Os índices *IDE* e *MDE* apresentaram uma correlação de -0,55 e -0,50 respectivamente.

Na Tabela 4.4, são apresentadas também as correlações entre os índices *SPD*, *IDE* e *MDE*. É possível verificar que o índice *SPD*, na maioria dos modelos, não apresenta muitas correlações positivas fortes com os índices *IDE* e *MDE*, já estes apresentam altos valores de correlação positiva entre si. Isso pode ser explicado pelo fato do índice *SPD* não levar em consideração em seu cálculo a componente do alcance sobre a metade da maior distância, em contrapartida, os índices *IDE* e *SPD* têm essa componente.

5 CONCLUSÃO

A partir da medida I_2 foi possível construir um novo índice de dependência espacial que possui características muito próximas do índice proposto por Seidel e Oliveira (2014a). O novo índice proposto (*MDE*) apresenta comportamento assimétrico positivo e tem a particularidade de dar menos peso para a influência do alcance.

O número de atributos de solo que tiveram a dependência espacial classificada utilizando os índices *IDE* e *MDE* e suas respectivas classificações foram próximos em cada intervalo de classificação, contudo, comparados com o índice *SPD* e a classificação Cambardella et al. (1994) o número de atributos foi visivelmente maior nos intervalos de moderada e forte dependência espacial para os três modelos que tiveram aplicações a dados reais.

Com os resultados das correlações provenientes das simulações foi possível verificar que não houve um índice que se sobrepôs aos demais, em todos os modelos, em relação à maior correlação negativa com o *EQM* e a *VAR*. Contudo, os índices *IDE* e *MDE* apresentaram mais correlações fortes com *EQM* e *VAR*.

Utilizar um grupo maior de índices para avaliar a estrutura de dependência espacial é fundamental para melhor interpretação do comportamento do fenômeno em estudo.

6 REFERÊNCIAS

APPEL NETO, E. et al. Spatial dependence index for cubic, pentaspherical and wave semivariogram models. **Boletim de Ciências Geodésicas**, v.24, n.1, p. 142-151, 2018.

BERTOLANI, F.; VIEIRA, S. Variabilidade espacial da taxa de infiltração de água e da espessura do horizonte a, em um argissolo vermelho-amarelo, sob diferentes usos. **Revista Brasileira de Ciência do Solo**, v. 25, n. 4, p. 987–995, 2001.

BIONDI, F.; MYERS, D. E.; AVERY, C. C. Geostatistically modeling stem size and increment in an old-growth forest. **Canadian Journal of Forest Research**, v. 24, n. 7, p. 1354–1368, 1994.

CAMBARDELLA, C. et al. Field-scale variability of soil properties in central iowa soils. **Soil Science Society of America journal**, v. 58, n. 5, p. 1501–1511, 1994.

CRESSIE N. **Statistics for Spatial Data**. New York: Wiley, 1991. 900 p.

GREGO, C. R.; VIEIRA, S. R. Variabilidade espacial de propriedades físicas do solo em uma parcela experimental. **Revista Brasileira de Ciência do Solo**, v. 29, n. 2, p. 169-177, 2005.

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

LANDIM, P. M. B. Sobre geoestatística e mapas. **Terrae Didatica**, v. 2, n. 1, p. 19-33, 2006.

LUMLEY, J. L.; PANOFSKY, H. A. **The structure of atmospheric turbulence**. Interscience Monographs and Texts in Physics and Astronomy, New York: Wiley, 1964.

MATHERON, G. **Traité de géostatistique appliquée**. Tome I. Memoires du Bureau de Recherches Geologiques et Minières. n. 14. Paris: Technip. 1962.

OLEA, R. A. A six-step practical approach to semivariogram modeling. **Stochastic Environmental Research and Risk Assessment**, v. 20, n. 5, p. 307–318, 2006.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing. 2017.

RIBEIRO JÚNIOR, P. J.; DIGGLE, P. J. GeoR: a package for geostatistical analysis. **R News**, New York, v. 1, n. 2, p. 15-18, 2001.

RUSSO, D.; BRESLER, E.. Soil Hydraulic Properties as Stochastic Processes: I. An Analysis of Field Spatial Variability 1. **Soil Science Society of America Journal**, v. 45, n. 4, p. 682-687, 1981.

SANTOS, E. L. E. dos et al. Some aspects about the spatial dependence index for variability of soil attributes. **Ciência Rural**, Santa Maria, v. 48, n. 6, p. 1-7, 2018.

SCHABENBERGER, O.; GOTWAY, C.A. **Statistical methods for spatial data analysis**. Boca Raton, Chapman e Hall/CRC, 2005. 488p.

SCHABENBERGER, O.; PIERCE, F. J. **Contemporary statistical models for the plant and soil sciences**. CRC press, 2001. 738p.

SEIDEL, E. J.; OLIVEIRA, M. S. de. New geostatistical index for measuring spatial dependence. **Revista Brasileira de Ciência do Solo**, v. 38, n. 3, p. 699–705, 2014a.

SEIDEL, E. J.; OLIVEIRA, M. S. de. Proposta de um teste de hipótese para a existência de dependência espacial em dados geoestatísticos. **Boletim de Ciências Geodésicas**, v. 20, n. 4, 2014b.

SEIDEL, E. J.; OLIVEIRA, M. S. de. Medidas de dependência espacial baseadas em duas perspectivas do semivariograma paramétrico. **Ciência e Natura**, v. 37, n. 3, 2015.

SEIDEL, E. J.; OLIVEIRA, M. S. de. A classification for a geostatistical index of spatial dependence. **Revista Brasileira de Ciência do Solo**, v. 40, 2016.

SHUKLA, M. K. et al. Comparison of Some Approaches to Determine Spatial Dependence of Soil Properties. **Journal on Agricultural Engineering**, v. 3, n. 1, 2017.

SIQUEIRA, G.M.; VIEIRA, S.R.; CEDDIA, M.B. Variabilidade de atributos físicos do solo determinados por métodos diversos. **Bragantia**, 67:203-211, 2008.

VIEIRA, S. R. Geoestatística aplicada à agricultura de precisão. In: GIS Brasil, 98., 2000, Curitiba. Anais... Curitiba, 1998. p. 1-53.