



ENIO JÚNIOR SEIDEL

**NOVAS CONTRIBUIÇÕES PARA AVALIAÇÃO
E DESCRIÇÃO DA ESTRUTURA DE
DEPENDÊNCIA ESPACIAL EM
GEOESTATÍSTICA**

LAVRAS – MG

2013

ENIO JÚNIOR SEIDEL

**NOVAS CONTRIBUIÇÕES PARA AVALIAÇÃO E DESCRIÇÃO DA
ESTRUTURA DE DEPENDÊNCIA ESPACIAL EM GEOESTATÍSTICA**

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

Orientador

Prof. Dr. Marcelo Silva de Oliveira

LAVRAS – MG

2012

**Ficha Catalográfica Elaborada pela Divisão de Processos Técnicos da
Biblioteca da UFLA**

Seidel, Enio Junior.

Novas contribuições para avaliação e descrição da estrutura de dependência espacial em geoestatística / Enio Junior Seidel. – Lavras : UFLA, 2013.

145 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2012.

Orientador: Marcelo Silva de Oliveira.

Bibliografia.

1. Simulação de processos estocásticos. 2. Análise variográfica.
3. Modelos de semivariograma. 4. Índices de dependência espacial.
5. Testes de hipótese. I. Universidade Federal de Lavras. II. Título.

CDD – 519.23

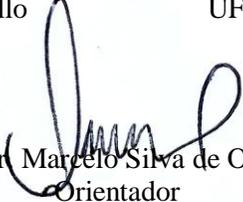
ENIO JÚNIOR SEIDEL

**NOVAS CONTRIBUIÇÕES PARA AVALIAÇÃO E DESCRIÇÃO DA
ESTRUTURA DE DEPENDÊNCIA ESPACIAL EM GEOESTATÍSTICA**

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

APROVADA em 19 de dezembro de 2012.

Prof. Dr. João Domingos Scalon	UFLA
Prof. Dr. Ednaldo Carvalho Guimarães	UFU
Prof. Dr. José Marcio de Mello	UFLA
Prof. Dr. Renato Ribeiro de Lima	UFLA
Prof. Dr. Marcelo Angelo Cirillo	UFLA


Prof. Dr. Marcelo Silva de Oliveira
Orientador

LAVRAS - MG

2012

AGRADECIMENTOS

À Universidade Federal de Lavras (UFLA) e ao Departamento de Ciências Exatas (DEX), pela oportunidade concedida de cursar o doutorado;

À CAPES, pela concessão de bolsa de estudos em parte do desenvolvimento desta tese;

Ao meu orientador, o Prof. Dr. Marcelo Silva de Oliveira, pelos ensinamentos transmitidos, incentivo, atenção e contribuições para o desenvolvimento deste trabalho;

Ao Prof. João Domingos Scalon, pelas contribuições a este trabalho;

Aos professores do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, pelos conhecimentos transmitidos;

À Universidade Federal do Pampa (UNIPAMPA), Campus de Itaqui – RS, pela compreensão e apoio na conclusão desta tese;

Aos membros da banca examinadora desta tese, pela disponibilidade de avaliar e contribuir com o trabalho;

Aos colegas, João Marcos e Gerson, pela atenção em minha iniciação aos estudos em Geoestatística;

Aos meus pais, Enio Selito Seidel e Marli Luft Seidel e à minha irmã Eveline, pelo apoio e confiança dedicados;

À Juliana, pelo apoio, carinho, companheirismo, dedicação e compreensão recebidos;

À Dona “Itinha”, ao “Seu Zé”, à Maria, à Dona Ana, à Dona “Lelé”, à Kelly e todos os amigos da pensão, pelo acolhimento e convivência;

Aos meus grandes amigos e colegas, Tiago, Diana, Danilo, Manoel, Diogo, Jair P., Jair “Wizi”, Marcelo “Brow”, Siomara, Felipe, Crysttian, pelo apoio, parceria, companheirismo e convivência;

À Josi, secretária da pós-graduação, pela amizade, pela atenção, presteza e colaboração durante todo o período do meu Doutorado, e, em especial na qualificação e na defesa desta tese;

À Josiane e à Mirian da secretaria do DEX, à Selminha e à Edila, pela ajuda nos procedimentos da qualificação e da defesa desta tese e pela colaboração durante meu Doutorado;

Aos colegas de doutorado, Moisés, Ana Paula, Tânia, Augusto, Leandro e Ana Lúcia, com quem tive os primeiros contatos no início do doutorado, pela convivência;

Aos amigos, Gibran, Cleber e Amauri, pelo apoio e convivência;

A todos os demais colegas do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, pela convivência;

A Deus por tudo.

RESUMO

Objetivou-se com esta tese desenvolver um estudo sobre aspectos da avaliação e descrição da estrutura de dependência espacial em dados geoestatísticos. Primeiramente é proposta uma correção, nos modelos exponencial e gaussiano, com base no percentual de explicação do parâmetro contribuição. Para possibilitar uma descrição do grau de dependência espacial foi proposta a construção de três índices a partir do conceito de área de dependência espacial do semivariograma. Em seguida, realizou-se um estudo comparativo entre os índices. Para a avaliação da dependência espacial foram construídos dois testes de hipótese, o primeiro a partir de simulação, o segundo por permutação. Após, foi realizado estudo sobre o nível nominal e o poder dos dois testes. A correção realizada nos modelos de semivariogramas exponencial e gaussiano permitiu gerar uma nova classe de modelos exponenciais e gaussianos. Foram formulados três novos índices para descrição da dependência espacial, denominados de *IDE1*, *IDE2* e *IDE3*. Ao realizar uma aplicação comparativa dos três índices, observou-se que, de forma geral, o índice *IDE3* apresentou melhor desempenho. Em relação à avaliação da estrutura de dependência espacial, foram construídos dois testes, denominados de *TDE1* e *TDE2*, independentes entre si, para a hipótese de existência de independência espacial. Os dois testes apresentaram resultados bem semelhantes em relação aos erros Tipo I e Tipo II e ao poder. Concluiu-se que o índice *IDE3* é o melhor, e, por isso, sugere-se a sua utilização para a descrição da dependência espacial. Além disso, os dois testes de hipótese, para avaliar a dependência espacial, tiveram desempenho semelhante, de forma que se sugere a utilização de qualquer um deles.

Palavras-chave: Análise variográfica. Modelos de semivariograma. Testes de hipótese. Índices de dependência espacial. Simulação de processos estocásticos.

ABSTRACT

This thesis aimed at developing a study on aspects of evaluation and description of spatial dependence structure in geostatistical data. First, a correction is proposed, in the exponential and gaussian models, based on the explanation percentage of the contribution parameter. In order to allow a description of the degree of spatial dependence, the construction of three indexes was proposed from the concept of spatial dependence area of the semivariogram. Second, a comparative study between the indexes was done. For the evaluation of spatial dependence, two hypothesis tests were constructed; the first with a simulation, the second by permutation. Subsequently, a study on the nominal level and the power of the two tests was done. The correlation performed on the exponential and gaussian semivariogram models allowed the generation of a new class of exponential and gaussian models. Three new indexes were formulated for the description of the spatial dependence, denominated *IDE1*, *IDE2* and *IDE3*. In carrying out a comparative application of the three indexes, it was observed, in a general way, that the *IDE3* index presented a better performance. Regarding the evaluation of the spatial dependence structure, two tests were constructed, denominated *TDE1* and *TDE2*, independent of each other, for the hypothesis of the existence of spatial independence. Both tests presented similar results in relation to Type I and Type II errors and to the power. It is concluded that the *IDE3* index is the best and, therefore, its use is suggested for the description of spatial dependence. In addition, both hypothesis tests, for evaluating spatial dependence, presented similar performance, which suggests the use of either one.

Key-words: Variographic analysis. Semivariogram models. Hypothesis testing. Spatial dependence indexes. Stochastic process simulation.

LISTA DE FIGURAS

Figura 1	Esquema representativo de um processo estocástico espacial, em que cada ponto s_i do domínio D gera uma variável aleatória $Z(s_i)$ com distribuição de probabilidade específica	21
Figura 2	Exemplo de covariograma	34
Figura 3	Exemplo de correlograma.....	34
Figura 4	Exemplo de semivariograma.....	35
Figura 5	Exemplo de um semivariograma experimental.....	36
Figura 6	Exemplo típico de um semivariograma com seus parâmetros: Contribuição; Efeito pepita; Patamar; e Alcance	39
Figura 7	Exemplos de semivariogramas com envelopes simulados	45
Figura 8	Comportamento dos modelos esférico (ESF), exponencial (EXP) e gaussiano (GAUS) em relação ao alcance AL	59
Figura 9	Comparação entre os comportamentos dos modelos exponencial e gaussiano em relação ao alcance a ser atingido a 95% da contribuição (a) e a 99,99% da contribuição (b)	61
Figura 10	Representação da área de dependência espacial em semivariograma com ajuste de modelo esférico	67
Figura 11	Área de dependência espacial em semivariograma com ajuste de modelo exponencial.....	67
Figura 12	Área de dependência espacial em semivariograma com ajuste de modelo gaussiano	68
Figura 13	Área de dependência espacial em semivariograma com ajuste de modelo exponencial com correção $p=9$	68
Figura 14	Área de dependência espacial em semivariograma com ajuste de modelo gaussiano com correção $p=9$	69
Figura 15	Representação gráfica do modelo linear com patamar	79

Figura 16	Representação gráfica, aproximada, do modelo máximo proposto	80
Figura 17	Área de dependência espacial observada (a) e área de dependência espacial possível (b), para o modelo esférico	83
Figura 18	Área de dependência espacial observada (a) e área de dependência espacial possível (b), para o modelo exponencial ...	88
Figura 19	Área de dependência espacial observada (a) e área de dependência espacial possível (b), para o modelo gaussiano	88
Figura 20	Áreas de dependência e não-dependência espaciais para o modelo esférico	91
Figura 21	Áreas de dependência e não-dependência espaciais para o modelo exponencial.....	96
Figura 22	Áreas de dependência e não-dependência espaciais para o modelo gaussiano	97
Figura 23	Representação de um semivariograma com dependência espacial nula (a), com dependência espacial maior que zero e menor que 100% (b) e com dependência espacial, virtualmente, igual a 100% (c)	99
Figura 24	Representação geral de um correlograma, seus parâmetros e a área de correlação espacial correspondente	101
Figura 25	Representação do alcance corrigido	108
Figura 26	Relacionamento entre os valores de EQM_e e os cenários propostos na simulação, com ajuste de linha de tendência linear, no modelo esférico	113
Figura 27	Frequência de melhores correlações entre cada índice e o EQM_e , considerando, o intervalo de correlações entre -0,1999 e	

	0,0000 (a), entre -0,3999 e -0,2000 (b), entre -0,5999 e -0,4000 (c) e entre -1,0000 e -0,6000 (d), para o modelo esférico.....	114
Figura 28	Frequência de melhores correlações entre cada índice e o EQM_e , considerando, o intervalo de correlações entre -0,1999 e 0,0000 (a), entre -0,3999 e -0,2000 (b), entre -0,5999 e -0,4000 (c) e entre -1,0000 e -0,6000 (d), para o modelo exponencial com $p = 3$	115
Figura 29	Frequência de melhores correlações entre cada índice e o EQM_e , considerando, o intervalo de correlações entre -0,1999 e 0,0000 (a), entre -0,3999 e -0,2000 (b) e entre -0,5999 e -0,4000 (c), para o modelo exponencial com $p = 9$	116
Figura 30	Frequência de melhores correlações entre cada índice e o EQM_e , considerando, o intervalo de correlações entre -0,1999 e 0,0000 (a), entre -0,3999 e -0,2000 (b), entre -0,5999 e -0,4000 (c) e entre -1,0000 e -0,6000 (d), para o modelo gaussiano com $p = 3$	117
Figura 31	Frequência de melhores correlações entre cada índice e o EQM_e , considerando, o intervalo de correlações entre -0,1999 e 0,0000 (a), entre -0,3999 e -0,2000 (b), entre -0,5999 e -0,4000 (c) e entre -1,0000 e -0,6000 (d), para o modelo gaussiano com $p = 9$	118
Figura 32	Representação de um semivariograma de efeito pepita puro, caracterizando a não existência de dependência espacial	122
Figura 33	Esquema representativo do procedimento do teste $TDE1$	124
Figura 34	Esquema com o procedimento do teste $TDE2$	127
Figura 35	Representação do <i>grid</i> na simulação	129
Figura 36	Poder dos testes $TDE1$ e $TDE2$, considerando o modelo esférico	133

Figura 37 Poder dos testes <i>TDE1</i> e <i>TDE2</i> , considerando o modelo exponencial ($p = 3$).....	133
Figura 38 Poder dos testes <i>TDE1</i> e <i>TDE2</i> , considerando o modelo exponencial ($p = 9$).....	134
Figura 39 Poder dos testes <i>TDE1</i> e <i>TDE2</i> , considerando o modelo gaussiano, com ($p = 3$).....	134
Figura 40 Poder do teste <i>TDE1</i> e <i>TDE2</i> , considerando o modelo gaussiano, com ($p = 9$).....	135

LISTA DE TABELAS

Tabela 1	Relação entre os modelos e o alcance de dependência espacial...	54
Tabela 2	Relação entre a correção p , dos modelos exponencial e gaussiano, e o percentual da contribuição que se deseja atingir ..	65
Tabela 3	Esquema com alguns modelos e seus respectivos valores para o fator de modelo	81
Tabela 4	Cenários propostos para avaliar o desempenho dos índices	109
Tabela 5	Ocorrência das melhores correlações entre os índices e o EQM_e , considerando diferentes intervalos de correlações, nas 100 replicações.....	111
Tabela 6	Resultados da aplicação do teste de Qui-Quadrado	119
Tabela 7	Cenários simulados para avaliar o poder dos testes	130
Tabela 8	Valores do $IDE3$ para os modelos esférico, exponencial e gaussiano	132

SUMÁRIO

1	INTRODUÇÃO	15
2	REFERENCIAL TEÓRICO.....	19
2.1	Geoestatística e aplicações.....	19
2.2	Geoestatística geral.....	26
2.3	Geoestatística estacionária (hipóteses de estacionariedade).....	31
2.4	Estimação do semivariograma	35
2.5	Envelopes simulados.....	44
2.6	Modelos teóricos de semivariograma	46
3	METODOLOGIA	49
3.1	Desenvolvimento da proposta de uma correção para os modelos de semivariogramas exponencial e gaussiano.....	49
3.2	Desenvolvimento da proposta de índices para descrever a dependência espacial	50
3.3	Desenvolvimento da proposta de testes de hipótese para a dependência espacial	51
4	RESULTADOS E DISCUSSÃO	53
4.1	Proposta de correção para os modelos teóricos exponencial e gaussiano.....	53
4.1.1	Demonstrações de que o alcance é obtido a 95% da contribuição nos modelos exponencial e gaussiano	56
4.1.2	Proposta de correção para o modelo exponencial	60
4.1.3	Proposta de correção para o modelo gaussiano.....	63
4.2	Proposta de índices de dependência espacial para semivariogramas com patamar	65
4.2.1	Definição da área de dependência espacial no semivariograma	66
4.2.2	Cálculo da estatística <i>ADE</i>	70
4.2.3	Proposta do primeiro índice de dependência espacial.....	82
4.2.3.1	<i>IDE1</i> para o modelo esférico	82
4.2.3.2	Cálculo do índice <i>IDE1</i>	84
4.2.3.3	Generalização do <i>IDE1</i>	88
4.2.4	Proposta do segundo índice de dependência espacial.....	90
4.2.4.1	<i>IDE2</i> para o modelo esférico	90
4.2.4.2	Cálculo do índice <i>IDE2</i>	92
4.2.4.3	Generalização do <i>IDE2</i>	95

4.2.5	Proposta do terceiro índice de dependência espacial	100
4.2.5.1	<i>IDE3</i> para o modelo esférico	100
4.2.5.2	Cálculo da área <i>ACE</i>	102
4.2.5.3	Generalização do <i>IDE3</i>	105
4.2.6	Aplicação comparativa dos índices <i>IDE1</i> , <i>IDE2</i> e <i>IDE3</i>	108
4.3	Proposta de testes de dependência espacial com base em semivariogramas com patamar	121
4.3.1	Procedimento do primeiro teste de dependência espacial (<i>TDE1</i>)	121
4.3.2	Procedimento do segundo teste de dependência espacial (<i>TDE2</i>)	124
4.3.3	Nível nominal dos testes <i>TDE1</i> e <i>TDE2</i>	127
4.3.4	Poder dos testes <i>TDE1</i> e <i>TDE2</i>	130
5	CONCLUSÃO	137
	REFERÊNCIAS	140

1 INTRODUÇÃO

A Geoestatística é um conjunto de modelos usados para descrever a variação de variáveis aleatórias distribuídas no espaço, no tempo, ou em ambos (MINGOTI; LEITE; ROSA, 2006). Assim, em essência, a Geoestatística permite o estudo adequado da estrutura de dependência espacial, possibilitando a mensuração da magnitude e da forma da variabilidade dos atributos de interesse.

A variabilidade de qualquer variável espacial avaliada em um fenômeno pode ser dividida em dois tipos básicos, a saber, aquela de cunho não-estocástico, frequentemente chamada de tendência, e outra de cunho estocástico, frequentemente denominada de dependência espacial. A Geoestatística considera os dois tipos de variabilidade para explicar a continuidade observada em dados espaciais, reservando a média para modelar a estrutura da tendência, e a covariância, correlação e semivariância, para modelar a estrutura de dependência espacial.

Tomando em particular a semivariância, por meio desta é possível construir um gráfico denominado de semivariograma amostral, ou simplesmente semivariograma, que relaciona as semivariâncias com as distâncias entre pontos amostrados e com as suas direções, assemelhando-se a um gráfico de dispersão (nuvem de pontos) da análise de regressão. Sobre esse semivariograma amostral (sobre essa nuvem de pontos) é possível ajustar curvas relativas a modelos matemáticos que permitem a descrição do comportamento da dependência espacial. Contudo, os modelos exponencial e gaussiano possuem uma imposição ao parâmetro alcance, que é devida ao fato de que, nestes modelos o parâmetro patamar é atingido assintoticamente. Contudo, esta imposição é apenas uma convenção, não sendo atribuída por estudo teórico.

Dessa forma, pretende-se, nesta tese, propor uma correção nestes dois modelos matemáticos de semivariograma, permitindo gerar uma nova classe de modelos exponenciais e gaussianos.

Outra questão a ser abordada, nesta tese, está relacionada à avaliação da significância da dependência espacial modelada a partir do semivariograma. Esta avaliação da significância da dependência espacial, quando feita, é realizada através da construção de envelopes simulados no semivariograma, como, por exemplo, quando se utiliza o pacote *geoR* (RIBEIRO JÚNIOR; DIGGLE, 2001) do *software* R (R DEVELOPMENT CORE TEAM, 2012). Esta avaliação gráfica, gerada pelos envelopes, pode, em alguns casos, gerar dúvida quanto à significância da dependência espacial, fazendo com que o usuário não tenha certeza sobre seguir ou não o estudo da estrutura de dependência espacial. Para tentar minimizar as dúvidas quanto à interpretação visual dos envelopes simulados, a proposição de testes de hipótese, para a avaliação da significância da dependência espacial, é realizada nesta tese.

Propor um teste de hipótese, para tal avaliação, possibilita a geração de um valor *p* que permite uma avaliação objetiva da significância da dependência espacial, e, portanto, uma melhor interpretação e, conseqüentemente, uma melhor decisão, sobre a existência ou não, de dependência espacial em dados geoestatísticos.

Outro ponto vulnerável na análise da dependência espacial é a sua descrição através de índices. Os índices, atualmente utilizados, apresentados em Biondi, Myers e Avery (1994) e Cambardella et al. (1994), levam em consideração apenas as estimativas dos parâmetros contribuição, efeito pepita e patamar, não contemplando, dessa forma, todos os aspectos do semivariograma.

Nenhum desses índices considera a influência da estimativa do parâmetro alcance dado pelo semivariograma, de modo que perdem em eficiência, sendo incompletos na descrição da dependência espacial, já que esta

não depende apenas da contribuição, do efeito pepita e do patamar, mas sim depende também do alcance.

Além disso, os índices atualmente utilizados não levam em consideração nenhum aspecto da geometria do semivariograma, fato que deixa estes índices fracos em interpretação. Assim, nesta tese, são propostos novos índices para descrição da estrutura de dependência espacial, que buscam sanar a deficiência dos índices atuais.

As considerações acima suscitam questões pertinentes, que norteiam o desenvolvimento desta tese, as quais estão expostas a seguir, numa ordem adequada para o desenvolvimento lógico da tese:

- a) 1ª - Será possível realizar uma correção nos modelos exponencial e gaussiano de semivariograma, possibilitando criar uma nova classe de modelos?
- b) 2ª - É possível construir índices mais completos para a descrição da dependência espacial, derivados de semivariogramas, que levem em consideração a dependência espacial gerada pelo parâmetro Alcance?
- c) 3ª - É possível fazer inferência sobre a estrutura da dependência espacial por meio da construção de um teste para a hipótese de ausência de dependência espacial?

Assim, postas as questões de pesquisa acima, tem-se como objetivo geral desta tese desenvolver um estudo sobre aspectos da avaliação e descrição da estrutura de dependência espacial em dados geoestatísticos. Como objetivos específicos têm-se:

- a) Propor uma correção nos modelos exponencial e gaussiano de semivariograma;
- b) Construir novos índices para a descrição da dependência espacial;
- c) Construir testes de significância para a hipótese nula de ausência de dependência espacial.

Todo o desenvolvimento dos resultados será realizado para semivariogramas de processos estocásticos que atendam a hipótese de estacionariedade de 2ª ordem, sem tendência (média constante) e isotrópicos (semivariogramas que não dependem das direções, somente das distâncias). Além disso, a predição por Krigagem não será abordada.

2 REFERENCIAL TEÓRICO

Neste capítulo apresentam-se os aspectos teóricos referentes à análise geoestatística, com ênfase na variografia.

2.1 Geoestatística e aplicações

Uma premissa fundamental, que inspira a utilização da Geoestatística, é evidenciada de forma muito clara pela Primeira Lei da Geografia, apresentada por Tobler (1969 citado por LOUZADA, 2011, p. 24), que diz que “coisas mais próximas se parecem mais que coisas mais distantes”.

Essa Lei expressa o aspecto de que atributos variam no decorrer do espaço, de modo que em locais próximos se têm características similares, e em locais afastados por grandes distâncias se têm características menos semelhantes. As características tendem a se tornarem menos semelhantes à medida que as distâncias aumentam.

Desse modo, fica evidente a necessidade de considerar as localizações ou as distâncias no estudo de fenômenos físicos. A ideia é a de que a localização no espaço é importante para entender e até explicar a variabilidade dos atributos em estudo.

Em essência espera-se que os atributos variem continuamente no espaço de modo que as localizações e as distâncias sejam fundamentais para se entender tais variações. Não se tem interesse apenas no tamanho da amostra considerada e na magnitude de cada valor observado. Tem-se interesse no tamanho da amostra, magnitude dos valores observados e também nas localizações de tais valores coletados.

Tal essência já havia sido entendida por Daniel G. Krige, um engenheiro de minas sul africano, que percebeu a necessidade de considerar o aspecto regionalizado nas predições.

Wackernagel (2003) comenta que os primeiros passos rumo à Geoestatística foram dados na África do Sul, com o trabalho de Daniel G. Krige e do estatístico Herbert S. Sichel.

Tendo contato com as ideias iniciais de Krige, o engenheiro de minas francês George Matheron, na década de 60, desenvolveu uma teoria que, de forma revolucionária, iniciou os estudos estatísticos de atributos no espaço. Essa teoria recebeu o nome de Teoria das Variáveis Regionalizadas.

O termo Geoestatística foi derivado da utilização da Estatística em estudos iniciais em Engenharia de Minas e Geologia. Em essência, a Geoestatística era entendida como a Estatística aplicada em dados geológicos.

Ao se considerar dados dependendo das localizações ou distâncias, se está considerando que existe dependência espacial. Segundo Olea (2006), a diferença fundamental entre Geoestatística e Estatística Clássica é que na primeira se supõe a existência de autocorrelação espacial.

De acordo com Resende (2007), a continuidade geográfica da variável regionalizada se manifesta pela tendência de a variável apresentar valores muito próximos (dependentes) em dois pontos vizinhos e muito diferentes em pontos distantes, dando o aspecto de correlação.

Aqui, considera-se a Estatística Clássica como aquela em que as amostras aleatórias são independentes. Já, na Geoestatística, se tem dependência entre as observações.

Assim, tal dependência espacial se deve ao fato de que observações próximas são dependentes, e essa dependência é maior para distâncias menores. A autocorrelação espacial pode ser entendida como uma medida dessa

dependência espacial. O prefixo auto se deve ao fato de que se estuda a correlação da mesma variável em várias localizações.

Portanto, uma diferenciação básica que se deve fazer entre a Geoestatística e a Estatística Clássica, está no que se refere à definição das amostras. A diferenciação é a seguinte: “Na Estatística Clássica se tem uma amostra de tamanho n sendo investigada; já na Geoestatística se têm n amostras de tamanho 1 sendo investigadas”.

A Geoestatística considera que em cada local do espaço é possível considerar a existência de uma variável aleatória, conforme exemplificado na figura 1. Assim, em cada local se coleta uma amostra de tamanho 1 dessa variável aleatória. Esse aspecto diferenciado no processo de amostragem fica claro na abordagem da Geoestatística pela Teoria dos Processos Estocásticos.

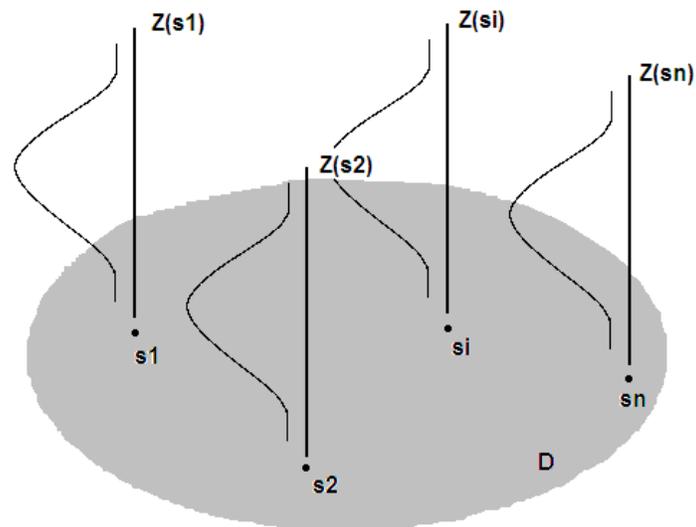


Figura 1 Esquema representativo de um processo estocástico espacial, em que cada ponto s_i do domínio D gera uma variável aleatória $Z(s_i)$ com distribuição de probabilidade específica

Quando uma variável é distribuída no espaço, ela pode ser dita como sendo regionalizada (ou indexada no espaço) e o fenômeno que é representado por tal variável pode ser chamado de regionalização (JOURNEL; HUIJBREGTS, 2003). Assim, um fenômeno natural pode ser caracterizado pela distribuição no espaço, de uma ou mais variáveis, também denominadas, por alguns pesquisadores, de regionalizadas. Segundo Journal e Huijbregts (2003), as variáveis estudadas no espaço possuem duas características importantes:

- a) um aspecto estocástico, relacionado à noção de variável aleatória;
- b) um aspecto estrutural, relacionado com uma representação funcional.

Essas duas características fazem com que a variável dependa de uma componente de efeito fixo, que pode ser modelada por uma função matemática e dependa de uma componente aleatória que expressa a variabilidade espacial. Nesta tese não será utilizado o conceito de variável regionalizada como tratado por Journal e Huijbregts (2003), mas sim, apenas o conceito de variável.

Em essência, uma variável estudada levando em conta o aspecto espacial (indexada em determinadas localizações) pode ser entendida como sendo um processo estocástico. Essa igualdade é evidente pelo fato de que em cada local considerado do espaço é possível se definir uma variável aleatória. Assim, o processo estocástico, é indexado pelas coordenadas do local.

Um fenômeno espacial avaliado em uma variável Z qualquer (discreta ou contínua) pode ser modelado fundamentalmente por um processo estocástico $\{Z(s) : s \in D \subset R^p\}$ que é uma coleção de “clones” da variável Z avaliados nas diferentes posições $s \in D$, constituindo assim o processo estocástico como uma coleção de variáveis aleatórias indexadas por um conjunto D de índices s contidos no R^p , com $p=2$ neste capítulo, contemplando as duas coordenadas

esaciais de uma superfície. O processo acima é dito ser espacial porque a variável Z tem uma indexação de uma coordenada s que a localiza no espaço D . Para todos os fins práticos, espera-se que este processo estocástico seja de 2ª ordem, isto é, exista o segundo momento, ou $E\{Z(s)^2\} < \infty, \forall s \in D$, visando a definição de medidas de dependência espacial adequadas. Caso algum fenômeno espacial em especial não possa ser modelado por um processo estocástico onde o segundo momento exista (processos onde a distribuição de Cauchy está presente, por exemplo), então a dependência espacial deve ser modelada diretamente nas distribuições conjuntas dos vetores aleatórios, se não nos vetores n -variados $[Z(s_1), \dots, Z(s_n)]$, $\forall s_i \in D, n=1,2,3, \dots$, pelo menos nos vetores bivariados $[Z(s_1), Z(s_2)]$. Apesar de essa possibilidade teórica existir, a chance de encontrá-la na prática é muito remota, fazendo com que a exigência da existência de momentos de 2ª ordem seja atendida por quase todos os fenômenos encontrados na prática.

A seguir são apresentados alguns exemplos de aplicações da Geoestatística. Atualmente a Geoestatística é aplicada em várias áreas das ciências onde existe a necessidade de se estudar fenômenos no espaço, como, por exemplo, nas Ciências do Solo, na Agricultura, na Engenharia Agrícola, nas Ciências Florestais, na Epidemiologia, etc.

Em Ciências do Solo, por exemplo, Ceddia et al. (2009) avaliaram a relação entre topografia e variabilidade espacial de algumas propriedades físicas do solo; Vieira et al. (2009) tiveram objetivo de caracterizar a variabilidade espacial dos atributos químicos de um latossolo vermelho após remoção de cafezal; Miguel, Vieira e Grego (2009) avaliaram o efeito da intensidade de pisoteio na variabilidade espacial da infiltração de água em solo sob pastagem; Camargo, Marques e Pereira (2010) caracterizaram a variabilidade espacial dos atributos físicos de argissolos desenvolvidos de arenito e cultivados com cana-

de-açúcar, bem como a relação desses com as curvaturas do relevo; Wu et al. (2008) lançaram objetivo de investigar as diferenças das variabilidades espaciais de quatro metais pesados entre três unidades de mapeamento de solo, e estudar as funções do tipo de solo na previsão espacial da poluição por metais pesados em um local contaminado; Sampaio et al. (2010) estudaram a variabilidade espacial da umidade do solo de latossolo vermelho-amarelo cultivado com milho sob plantio direto, em duas datas de amostragem, e sua relação com o carbono total do solo; Vidal-Vázquez et al. (2009) tiveram por objetivo investigar a adequação da Geoestatística na compreensão da evolução da rugosidade da superfície como uma função da precipitação acumulada.

Em Agricultura e Meio Ambiente, por exemplo, Chiba, Guedes Filho e Vieira (2010), partindo da hipótese de que existe continuidade espacial na população de plantas daninhas em área de cultivo sob semeadura direta, objetivaram quantificar e descrever a sua dependência espacial, visando seu mapeamento; Silva et al. (2010) objetivaram caracterizar a estrutura, a magnitude de dependência espacial e mapear a variabilidade da produção de café e da desfolha das plantas; Bognola et al. (2008) tiveram o objetivo de avaliar a variabilidade espacial do rendimento produtivo de *P. Taeda* em função de propriedades do solo; Alves et al. (2011) objetivaram caracterizar a estrutura espacial e mapear a variação espacial do dano causado pela broca e pelo bicho-mineiro em um agroecossistema de café; Alves et al. (2008) visaram caracterizar a estrutura e magnitude de dependência espacial de algumas variáveis ambientais.

Em Engenharia Agrícola, por exemplo, Rosalen et al. (2011) objetivaram comparar dois métodos de georreferenciamento para caracterizar a variabilidade espacial de atributos de solo, no contexto da agricultura de precisão; Rodrigues, Corá e Fernandes (2012) estudaram atributos de solo e de produtividade de milho, com foco em agricultura de precisão; Vitoria et al.

(2012) estudaram produtividade de planta forrageira em função da densidade do solo e porosidade total, em dois sistemas de manejo; Armindo et al. (2012) estudaram a variabilidade espacial de macro e micronutrientes foliares em pomar sob irrigação localizada; Yanagi Junior et al. (2012) objetivaram caracterizar a variabilidade espacial do nível de ruído gerados por máquinas agrícolas; Konopatzki et al. (2012) visaram mapear a produtividade de pés de pêra, relacionando com a variabilidade espacial de atributos de solo e planta, numa menção á fruticultura de precisão.

Em Ciências Florestais, por exemplo, Mello et al. (2005) avaliaram a estrutura de dependência espacial de quatro características dendrométricas; Mello et al. (2006), dentre alguns objetivos, avaliaram a estrutura de dependência espacial do volume de madeira por hectare; Kanegae Júnior et al. (2007) avaliaram a estrutura da continuidade espacial de características dendrométricas; Lima et al. (2008) avaliaram a resistência do solo a penetração em seção transversal à trilha de tráfego de tratores na colheita de madeira; Mello et al. (2009) estimaram o número de fustes e o volume de madeira por unidade de manejo, por meio de estimadores de amostragem aleatória simples e interpolador geoestatístico; Wojciechowski et al. (2009) utilizaram métodos geoestatísticos na identificação do tamanho e da estrutura da variabilidade espacial de alguns atributos físico-químicos do solo em áreas de Floresta Estacional Decidual.

Em Epidemiologia, por exemplo, Medronho et al. (2003) realizaram análise espacial da soroprevalência da hepatite A em crianças; Fortes et al. (2004) estimaram mapa de risco da ocorrência de infecção por *A. lumbricóides*; Opromolla, Dalben e Cardim (2005) avaliaram a distribuição espacial da hanseníase no Estado de São Paulo de 1991 a 2002; Mingoti, Leite e Rosa (2006) estudaram o número total de casos diagnosticados de AIDS entre 1996 e 1999 em municípios de Minas Gerais.

Esses são apenas alguns exemplos de áreas e situações em que a Geoestatística é aplicada. Existem muitas outras situações onde se necessita de sua aplicação.

2.2 Geoestatística geral

A teoria geral da Geoestatística baseia-se na teoria de processos estocásticos. O entendimento de que o processo estocástico modela fundamentalmente o fenômeno espacial faz com que, neste capítulo, os termos “processo” e “fenômeno” sejam intercambiáveis, equivalendo-se no discurso, mesmo permanecendo a sutil diferença de que o fenômeno representa a realidade estudada enquanto o processo representa apenas o modelo desta realidade.

Uma Geoestatística de fundamentação geral pode ser construída valendo-se de considerações sobre os momentos de primeira e segunda ordem das distribuições bivariadas dentro de $\{Z(s): s \in D \subset R^D\}$. O uso apenas das bivariadas é adequado devido à conveniência matemática e estatística da redução das $2^n - 1 - n$ relações bi, tri, ..., multi, ..., n -variadas, entre as variáveis $Z(s_1)$, ..., $Z(s_n)$ do processo, para apenas $comb(n,2) = \frac{n(n-1)}{2}$ relações bivariadas.

Utilizando apenas as relações bivariadas se consegue expedientes matemáticos e estatísticos que “amarram” todas as n variáveis em suas relações entre si. As relações bivariadas aparecem na Geoestatística geral por meio da abordagem pelas conjuntas $[Z(s), Z(t)]$ e pelas diferenças $Z(s) - Z(t)$. Os tipos de momentos tratados na Geoestatística geral são:

Momentos de primeira ordem:

Médias ou valores esperados $E\{Z(s)\}, \forall s \in D.$

Momentos de segunda ordem:

Variâncias $Var\{Z(s)\} = E\{[Z(s) - \mu(s)]^2\}, \forall s \in D.$

Covariâncias $Cov\{Z(s), Z(t)\} = E\{Z(s)Z(t)\} - E\{Z(s)\}E\{Z(t)\},$

$\forall s, t \in D.$

Correlações $Cor\{Z(s), Z(t)\} = \frac{Cov\{Z(s), Z(t)\}}{\sqrt{Var\{Z(s)\}Var\{Z(t)\}}}, \forall s, t \in D.$

Semivariâncias $\frac{1}{2}Var\{Z(s) - Z(t)\} = \frac{1}{2}E\{[Z(s) - Z(t)]^2\}, \forall s, t \in D.$

Esta abordagem não faz exigências quanto ao conhecimento das formas distribucionais, nem exige conhecer propriedades das distribuições conjuntas n -variadas, com $n > 2$, pois trabalha somente com o primeiro e os segundos momentos, o que a torna uma abordagem mais robusta. A motivação estatística por trás do estabelecimento de declarações apenas sobre o primeiro e segundos momentos também passa pela propriedade única da distribuição Normal, que é caracterizada completamente por estes dois tipos de momentos. Fazendo assim, mesmo que a Geoestatística geral não necessite de normalidade de nenhuma espécie, ela está preparada para conjugar-se perfeitamente com uma modelagem mais rica (mesmo que eventualmente menos robusta) que pressuponha normalidade em alguma instância (nas conjuntas ou nas condicionais) do processo estocástico fundamental. De fato, até mesmo outras distribuições de probabilidade (Lognormal, de Valores extremos, Poisson etc.) podem vir se acoplar a esta modelagem robusta, bem acomodadas na simplicidade dos dois tipos de momentos num mundo bivariado.

A representação de um processo estocástico como sendo $\{Z(s) : s \in D \subset R^p\}$ é geral e, talvez, pouco explicativa em relação a estrutura do processo estocástico no domínio do espaço.

Para Schabenberger e Gotway (2005), de forma geral, os modelos estatísticos são uma decomposição da variável resposta em uma estrutura matemática descrevendo a média e em uma estrutura estocástica adicional que descreve a variação e a covariação entre as respostas. Assim, de forma geral se teria:

$$Dado = \begin{pmatrix} \text{Componente} \\ \text{Estrutural} \end{pmatrix} + \begin{pmatrix} \text{Componente} \\ \text{de Erro} \end{pmatrix} \quad (1)$$

Esta é uma abordagem, inteligente e robusta, que, conjugada com a existência do primeiro e segundos momentos nas bivariadas, realiza uma decomposição clássica fundamental (geral), que a Estatística faz de variáveis aleatórias. Nesta abordagem, a variável aleatória Z , qualquer, pode ser decomposta em:

$$Z = \mu + e \quad (2)$$

em que μ representa a componente determinística do fenômeno (isto é, μ não é variável aleatória) e e representa a componente estocástica (isto é, e é variável aleatória). Pode-se estender imediatamente a Estatística chamada “clássica” para a variável Z que está no processo estocástico espacial, decompondo-a como:

$$Z(s) = \mu(s) + e(s) \quad (3)$$

A componente estocástica $e(s)$ possui média $E\{e(s)\}=0$ e variância $Var\{e(s)\}=\sigma^2$.

Porém, esta representação com a componente estocástica não possibilita destacar a dependência espacial entre as observações. Assim, tem-se a representação do processo estocástico dada por:

$$Z(s) = \mu(s) + \varepsilon'(s) + \varepsilon'' \quad (4)$$

em que $\varepsilon'(s)$ é a componente regionalizada do processo estocástico, e ε'' é a componente aleatória. Percebe-se que $e(s) = \varepsilon'(s) + \varepsilon''$, ou seja, a componente estocástica foi decomposta em duas componentes. As componentes $\mu(s)$, $\varepsilon'(s)$ e ε'' são independentes.

A componente aleatória ε'' segue uma dada distribuição (desejadamente Normal) com média $E\{\varepsilon''\}=0$ e variância $Var\{\varepsilon''\}=\tau^2$.

A covariância entre ε''_i e ε''_j é $Cov\{\varepsilon''_i, \varepsilon''_j\}=0$, ou seja, os erros aleatórios ε'' são independentes.

A média e a variância da componente regionalizada $\varepsilon'(s)$ são, respectivamente $E\{\varepsilon'(s)\}=0$ e $Var\{\varepsilon'(s)\}=Modelo$. O *Modelo* é dado por uma função das localizações ou das distâncias.

Além disso, tem-se que a covariância entre $\varepsilon'(s_i)$ e $\varepsilon'(s_j)$ é dada por:

$$Cov\{\varepsilon'(s_i), \varepsilon'(s_j)\} = f(s_i - s_j), \quad \forall i \text{ e } j \quad (5)$$

Desse modo, a covariância entre $\varepsilon'(s_i)$ e $\varepsilon'(s_j)$ é uma função da distância entre s_i e s_j .

A média do processo $Z(s)$ é dada por:

$$\begin{aligned} E\{Z(s)\} &= E\{\mu(s) + \varepsilon'(s) + \varepsilon''\} = \\ &= E\{\mu(s)\} + E\{\varepsilon'(s)\} + E\{\varepsilon''\} = \mu(s) \end{aligned} \quad (6)$$

A variância do processo $Z(s)$ é dada por

$$\begin{aligned} \text{Var}\{Z(s)\} &= \text{Var}\{\mu(s) + \varepsilon'(s) + \varepsilon''\} = \\ &= \text{Var}\{\mu(s)\} + \text{Var}\{\varepsilon'(s)\} + \text{Var}\{\varepsilon''\} = \text{Modelo} + \tau^2. \end{aligned} \quad (7)$$

Assim, percebe-se que $e(s) = \varepsilon'(s) + \varepsilon''$, ou seja, a componente estocástica do modelo clássico estatístico foi decomposta em duas componentes no modelo geoestatístico. A modelagem acima pode ser denominada de Geoestatística geral. Este é o começo da diferenciação da Geoestatística da assim chamada Estatística Clássica.

Uma diferenciação da Geoestatística em relação à Estatística Clássica é no tocante a existência de limitações na realização do processo estocástico $\{Z(s) : s \in D \subset R^p\}$, que quase sempre tem apenas uma única realização nos fenômenos estudados, exigindo, portanto, alguma hipótese de estacionariedade que permita a análise dos dados. Estas hipóteses tiram a teorização de uma abordagem geral, e introduzem-na numa abordagem de Geoestatística estacionária, desenvolvida na próxima seção.

2.3 Geoestatística estacionária (hipóteses de estacionariedade)

Considerando a abordagem de processos estocásticos espaciais, podem ser consideradas duas hipóteses de estacionariedade a se investigar. A primeira, e mais restritiva, é a hipótese de estacionariedade de segunda ordem. A segunda, e menos restritiva, é a hipótese intrínseca.

A modelagem mais restritiva utilizando hipóteses de estacionariedade pode ser chamada de Geoestatística estacionária, e constitui-se na abordagem geoestatística majoritariamente utilizada.

A hipótese de estacionariedade de primeira ordem está contemplada nas duas hipóteses relatadas anteriormente. Isto é, se o processo estocástico tem estacionariedade de segunda ordem, ele possui estacionariedade de primeira ordem.

O estudo das hipóteses de estacionariedade é fundamental em Geoestatística, pois com base nisso é que se pode desenvolver a investigação do atributo, com apenas uma realização, ou observação, em uma área particular do espaço. Conforme Corstanje, Grunwald e Lark (2008) e Lark (2009), a pressuposição de estacionariedade deve ser assumida para permitir inferências a partir de uma única realização do processo $\{Z(s) : s \in D \subset R^P\}$, que constitui os dados disponíveis. Nesta tese, a hipótese de estacionariedade de segunda ordem é a hipótese adotada.

A estacionariedade de segunda ordem existe se a média e a variância do processo estocástico são independentes da localização e a covariância existe e depende somente da distância h .

Em Babak e Deutsch (2009) tem-se que, a estacionariedade de primeira ordem assume que a média é constante em todo o domínio D . Já na estacionariedade de segunda ordem, assume-se que:

$$\text{Var}\{Z(s)\} = E\{[Z(s) - \mu(s)]^2\} = \sigma^2, \quad \forall s \in D \quad (8)$$

Além disso, tem-se que $\text{Cov}\{Z(s_1), Z(s_2)\} = \kappa\{|s_1 - s_2|\} = \kappa\{h\}$, ou seja, a covariância depende somente das distâncias h entre os pontos s_1 e s_2 .

Esta hipótese tem implicações sobre a função $\kappa(h)$, pois esta modela uma covariância, que deve obedecer a várias condições de contorno de natureza estatística. Algumas destas implicações são:

$$\text{Var}[Z(s)] = \text{Cov}[Z(s), Z(s)] = \kappa(s, s) = \kappa(s - s) = \kappa(0) = \sigma^2,$$

uma constante, não dependente de s , $\forall s \in D$.

$\kappa(h) = \kappa(-h)$, isto é, $\kappa(h)$ é uma função par.

$\kappa(h)$ é uma função não-negativa definida, isto é,

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j \kappa(h_{ij}) = \sum_{i=1}^m \sum_{j=1}^m a_i a_j \kappa(s_i - s_j) \geq 0,$$

onde $h_{ij} = s_i - s_j$, $\forall s_1, \dots, s_m \in D$, $\forall a_1, \dots, a_m$ real e $\forall m = 1, 2, 3, \dots$

Outra possibilidade é assumir a estacionariedade dos dois primeiros momentos da diferença do par de valores em dois pontos ($Z(s) - Z(s+h)$), isto é, a estacionariedade dos dois primeiros momentos do incremento, que lida com a noção de variograma, ou semivariograma (WACKERNAGEL, 2003). De acordo com Corstanje, Grunwald e Lark (2008), se a suposição de

estacionariedade é feita, não sobre o processo estocástico $Z(s)$, mas sobre a diferença de primeira ordem ($Z(s) - Z(s+h)$).

Então, além do processo estocástico $Z(s)$, é possível considerar um segundo processo estocástico, determinado por cada distância h , da forma $Z(s) - Z(s+h)$. Este processo estocástico adjacente é definido como o incremento.

A ideia intuitiva da utilização do incremento é a comparação entre as variáveis $Z(s)$ e $Z(s+h)$, ou seja, avaliar o relacionamento entre as variáveis separadas pela distância h . Por exemplo, avaliar a relação entre as variáveis $Z(s)$ e $Z(s+1)$, que são separadas a 1 unidade de distância.

O incremento pode ser percebido, então, como uma “variável diferença” e, portanto possui média e variância. A média é definida como:

$$E\{Z(s) - Z(s+h)\} = 0 \quad (9)$$

A semivariância é dada por:

$$\frac{1}{2} \text{Var}\{Z(s) - Z(s+h)\} = [\kappa(0) - \kappa(h)] = \gamma(h) \quad (10)$$

em que $\kappa(0) = \text{Var}\{Z(s)\} = \sigma^2$ e $\kappa(h) = \text{Cov}\{Z(s), Z(s+h)\}$.

As figuras 2, 3 e 4 apresentam exemplos de covariograma, correlograma e semivariograma, respectivamente.

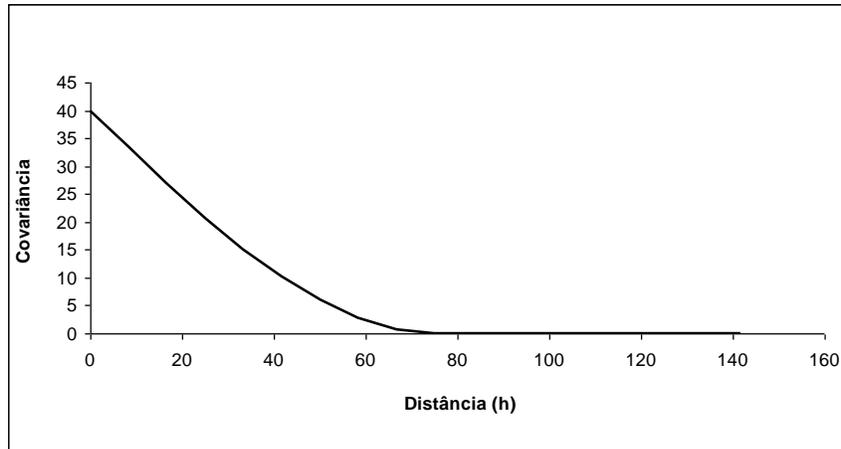


Figura 2 Exemplo de covariograma

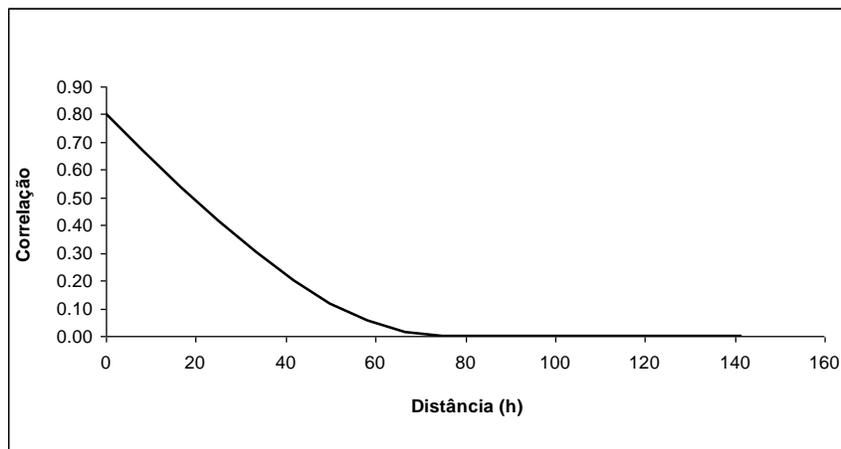


Figura 3 Exemplo de correlograma

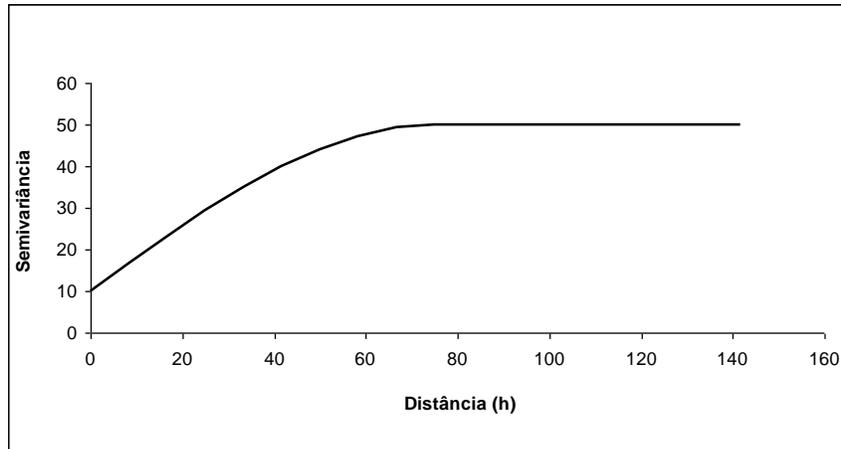


Figura 4 Exemplo de semivariograma

2.4 Estimação do semivariograma

Quando se relacionam as semivariâncias $\gamma(h)$ com as distâncias h em um gráfico, se tem a construção do semivariograma. De acordo com Carvalho, Silveira e Vieira (2002), o gráfico de $\gamma(h)$ versus os valores correspondentes de h é uma função do vetor h , e, portanto depende de ambos, magnitude e direção de h .

Em complemento, para Paz-González et al. (2000), a estimação de semivariâncias pode depender de parâmetros tais como distância entre *lags*, número de *lags*, etc. Os *lags* são distâncias entre os valores de h considerados no gráfico do semivariograma.

Para uma melhor observação do comportamento do semivariograma, pode-se considerar apenas a semivariância média em cada distância h , gerando um gráfico mais claro e fácil de interpretar.

Na realidade, esta semivariância média para cada distância h é definida como o estimador clássico de semivariograma, proposto por Matheron (1963). Este estimador é dado por:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{Z(x_i + h) - Z(x_i)\}^2, \quad (11)$$

em que $\{Z(x_i + h) - Z(x_i)\}$ é o incremento. Assim, o semivariograma experimental é apresentado na figura 5.

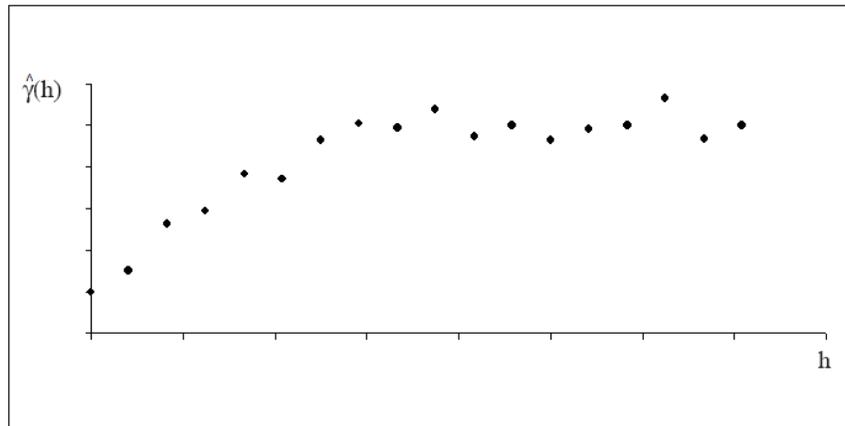


Figura 5 Exemplo de um semivariograma experimental

Com base na figura 5 é possível perceber que as semivariâncias médias crescem conforme se aumentam as distâncias h . Assim, pode-se inferir que quanto maiores forem as distâncias h entre as observações, menores se tornam as relações entre $Z(s+h)$ e $Z(s)$, ou seja, menor se torna a dependência espacial.

Outro estimador do semivariograma, adequado quando os dados apresentam *outliers*, é o estimador robusto, proposto por Cressie e Hawkins (1980). Esse estimador é dado por:

$$\tilde{\gamma}(h) = \left\{ \frac{1}{|2N(h)|} \sum \left\{ \left| Z(x_i + h) - Z(x_i) \right| \frac{1}{2} \right\}^4 \right\} / \left\{ 0,457 + \frac{0,494}{|2N(h)|} \right\} \quad (12)$$

Pelo fato de que primeiramente é calculada a média das raízes quadradas das diferenças e depois o resultado é então elevado à quarta potência, o primeiro termo do estimador robusto é muito pouco afetado por valores extremos se comparado com a média das diferenças quadráticas do estimador de Matheron (SCHABENBERGER; GOTWAY, 2005). Mingoti e Rosa (2008) complementam que este estimador, além de ser robusto à presença de *outliers*, é também robusto quanto à ausência de normalidade dos dados. Além disso, o termo no denominador serve para obter um não-viés aproximado (SCHABENBERGER; GOTWAY, 2005).

Existem outros estimadores possíveis para o semivariograma, tais como os citados no trabalho de Mingoti e Rosa (2008): o estimador da Mediana; o estimador de Haslett; e o estimador de Genton. Porém, a discussão desses estimadores não é o foco desta tese.

Em Olea (2006) podem ser observadas duas regras importantes no cálculo do semivariograma:

- a) O número mínimo de pares na estimação do semivariograma para um dado *lag h* deve ser 30;

- b) Deve-se limitar o número de *lags* do semivariograma experimental à metade da maior distância no domínio amostrado numa direção de interesse.

Observando esses dois aspectos é possível construir um semivariograma experimental capaz de expor a estrutura de variabilidade do atributo em estudo.

Essas indicações dadas em Olea (2006), não necessariamente devem ser seguidas, sendo apenas sugestões. Claro que se deve evitar que o número de pares por *lag* seja pequeno. Até porque, um número pequeno de pares pode gerar um alto erro-padrão para a estimativa.

Em relação à indicação da metade da maior distância do domínio, tem-se que essa observação é feita com base em dois aspectos. O primeiro aspecto é o de que em maiores distâncias obtêm-se poucos pontos, fato que prejudica as estimativas de semivariâncias, o que pode fazer com que não se consiga definir adequadamente o padrão de dependência espacial. O segundo aspecto se refere ao sentido de que em maiores distâncias talvez não seja possível garantir que os pontos respeitem a hipótese de estacionariedade de 2ª ordem. Contudo, são apenas recomendações, de modo que o pesquisador pode escolher que percentual da maior distância deseja utilizar para construir o semivariograma.

Deve-se chamar a atenção para o fato de que o semivariograma não comporta a modelagem de tendências, isto é, deve-se admitir que os dados não apresentem tendência ou que já se providenciou expedientes de “eliminação de tendência dos dados” para que tal tendência, quando presente no fenômeno, não esteja contaminando sua estimação.

Após a construção do semivariograma é necessário ajustar um modelo teórico que explique o comportamento do semivariograma em termos de modelar a variabilidade observada.

Os modelos teóricos de semivariogramas, na sua forma tradicional, possuem 4 parâmetros, que são o alcance, o patamar, o efeito pepita e a contribuição. A figura 6 apresenta um modelo de semivariograma, com patamar, destacando seus quatro parâmetros.

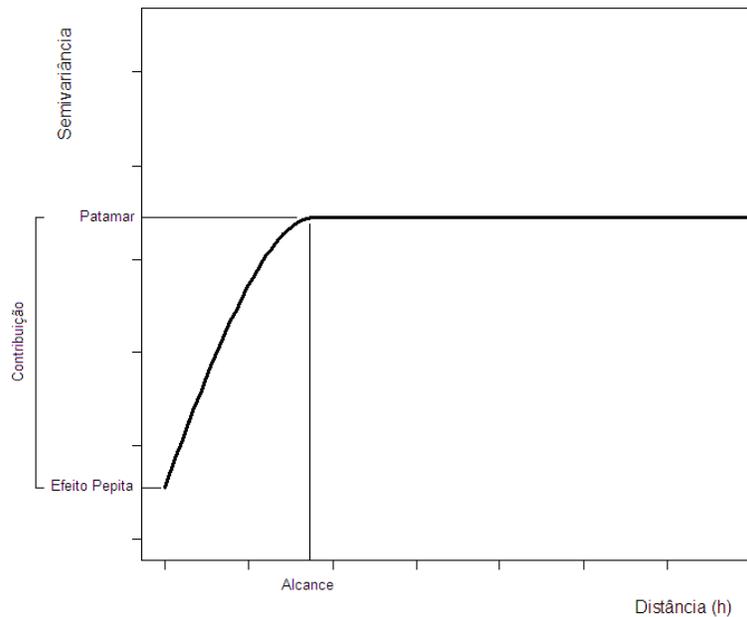


Figura 6 Exemplo típico de um semivariograma com seus parâmetros: Contribuição; Efeito pepita; Patamar; e Alcance

O alcance (a) é compreendido como a distância dentro da qual as amostras apresentam-se correlacionadas espacialmente. Em geral, conforme aumenta o valor no vetor de distância, também aumenta o valor da semivariância média no semivariograma, até um ponto em que as semivariâncias médias atingem um patamar e, a partir daí, se mantêm constantes. O valor de distância em que o semivariograma atinge o patamar é chamado de alcance.

Amostras separadas por distâncias menores que o alcance são correlacionadas umas às outras, o que permite que se façam interpolações para espaçamentos menores do que os amostrados (VIEIRA, 2000). Segundo Silva et al. (2007), quanto menor o alcance, mais rapidamente é obtida a independência entre as amostras (pontos), uma vez que o alcance é a distância limite da dependência espacial.

Segundo Isaaks e Srivastava (1989), o patamar ($C = C_0 + C_1$) é o valor de semivariância média em que se atinge o alcance no semivariograma.

O efeito pepita (C_0) é o valor que representa a descontinuidade do semivariograma, no sentido de que, à medida que h tende a zero, $\gamma(h)$ tende a um valor positivo C_0 . Para Vieira (2000), o valor do efeito pepita revela a descontinuidade do semivariograma para distâncias menores do que a menor distância entre as amostras. Wu et al. (2008) definem que o efeito pepita representa o erro experimental e a variação no espaço dentro de distâncias mínimas de amostragem.

Assim, entende-se o efeito pepita como uma espécie de variância aleatória.

Em geral, o efeito pepita é devido à (CHILÈS; DELFINER, 2006):

- a) uma microestrutura ou “ruído geológico”, ou seja, um componente do fenômeno com uma escala menor do que o suporte de amostragem (efeito pepita puro);
- b) uma estrutura com um alcance menor do que a menor distância entre os pontos;
- c) erros de medição ou posicionamento.

Corroborando com o exposto acima, Gallardo e Paramá (2007) comentam que um alto efeito pepita também pode ocorrer devido a erros de amostragem ou análise, ou ainda por falta de estrutura na escala considerada.

A contribuição (C_1) é a diferença entre o patamar e o efeito pepita, sendo o segmento gráfico nesse intervalo que, praticamente, caracteriza a dependência espacial de processos estocásticos contínuos. A contribuição pode ser entendida como uma variância estruturada que é devida à estrutura espacial do fenômeno em estudo.

Como o patamar é a soma da contribuição, que é a variância estrutural, com o efeito pepita, que é a variância ao acaso, então se tem que o patamar pode ser entendido como a variância total, que, como observado em Abreu et al. (2003) e Vieira (2000), é aproximadamente igual à variância dos dados.

Sendo os semivariogramas funções, e não números singulares, fica difícil a comparação da estrutura de variabilidade espacial representada por dois semivariogramas, visto que conceitos como “maior” e “menor” não são imediatamente aplicados à funções como o são para números. Por causa desta dificuldade, autores de áreas aplicadas têm-se entregado à tarefa de construir índices de dependência espacial (medidas numéricas únicas da dependência espacial), a partir dos parâmetros do modelo ajustado de semivariograma, pelos quais seria possível comparar as estruturas de dependência espacial. A ideia é que estes índices resumam, em um número apenas, toda a informação presente na função semivariograma, seguindo a mesma lógica usada na construção de medidas de dispersão da Estatística Descritiva. Um exemplo é a variância, que resume em um número apenas toda a variabilidade presente numa distribuição de frequências, que é uma função. Neste sentido, Biondi, Myers e Avery (1994) e Cambardella et al. (1994) apresentam seus índices. Segundo Biondi, Myers e Avery (1994), esses índices são úteis para fazer comparações entre diferentes estruturas de dependência espacial.

Uma medida interessante, que é utilizada para descrever, ou resumir, a dependência espacial, é a razão de dependência espacial, ou grau de dependência espacial. Na realidade existem duas expressões para este índice. A primeira dada por (CAMBARDELLA et al., 1994):

$$RD = \left(\frac{C_0}{C_0 + C_1} \right) * 100 \quad (13)$$

A segunda expressa como (BIONDI; Myers; Avery, 1994):

$$DE = \left(\frac{C_1}{C_0 + C_1} \right) * 100 \quad (14)$$

Assim, tem-se uma espécie de estatística descritiva da dependência espacial.

Ambas as expressões (13) e (14) assumem valores entre 0 e 100%. Na expressão (13), quanto maior o valor do índice, menor é a dependência espacial. Já na expressão (14), quanto maior o valor, maior pode ser considerada a dependência espacial.

A literatura mostra que os índices propostos por Biondi, Myers e Avery (1994) e Cambardella et al. (1994) estão sendo utilizados em diversas áreas aplicadas, tais como, por exemplo, ciências florestais (MELLO et al., 2005, 2006), ciências do solo (GALLARDO; PARAMÁ, 2007; MIGUEL; VIEIRA; GREGO, 2009; SAMPAIO et al., 2010; SIQUEIRA; MARQUES JÚNIOR; PEREIRA, 2010; VIDAL-VÁZQUEZ et al., 2009; VIEIRA et al., 2009; WU et al., 2008) e agricultura (CHIBA; GUEDES FILHO; VIEIRA, 2010).

Com base na magnitude da dependência espacial, expressa em (13), quanto menor a proporção do efeito pepita em relação ao patamar, maior a continuidade do fenômeno e maior a dependência espacial apresentada pela

variável (NOVAES FILHO et al., 2007). Para Vidal-Vázquez et al. (2009) a razão de dependência pode gerar uma indicação da proporção da variabilidade que é estruturada, e conseqüentemente, a proporção que é aleatória.

Cambardella et al. (1994) lançaram uma espécie de classificação do índice apresentado na expressão (13). Se o índice assume valor de até 25%, tem-se dependência espacial forte; índice entre 25% e 75% descreve uma dependência espacial moderada; e valores a partir de 75% caracterizam dependência espacial fraca. Nesse sentido, para Wu et al. (2008), considerar a variabilidade espacial das variáveis é importante, para realizar predições, somente quando elas possuem moderada ou forte dependência espacial.

Observa-se que o fato de avaliar o grau de dependência espacial é importante para que se possa comparar fenômenos ou comparar fenômenos no decorrer do tempo, por exemplo. Assim, medidas que permitem quantificar este grau de dependência são essenciais. Essas medidas são consideradas como valores que caracterizam a dependência espacial, da mesma forma que o coeficiente de variação (CV) faz para caracterizar a precisão em experimentação, por exemplo.

Contudo, as medidas sugeridas por Biondi et al. (1994) e Cambardella et al. (1994) não contemplam aspectos mais profundos da geometria do semivariograma. O que pode ser considerada uma fragilidade dessas duas medidas.

Essa fragilidade fica evidente pelo fato de que esses dois índices só consideram os parâmetros contribuição e efeito pepita. Essa é uma deficiência que pode levar o usuário a equivocadas interpretações sobre o grau de dependência espacial de um fenômeno. Essa interpretação equivocada pode acontecer, pois dois fenômenos sob comparação podem ter valores de contribuição e efeito pepita muito semelhante o que faz, por conseqüência, com que os dois índices assumam valores semelhantes. Contudo, podem ter alcances

de dependências espaciais totalmente diferentes. Dessa forma, o usuário, com base no índice, poderia supor que os fenômenos são semelhantes em termos de dependência espacial, quando na verdade não são.

Schabenberger e Gotway (2005) comentam que a medida DE , que relaciona a contribuição com o patamar, também entendida como a variabilidade estrutural relativa, é uma medida bastante grosseira da estrutura de um processo estocástico. Os autores ainda complementam que muitos aspectos do processo são desconsiderados por essa medida, dentre esses o alcance do semivariograma, que muitas vezes, em alguns estudos, é um parâmetro importante.

2.5 Envelopes simulados

Além das medidas sugeridas por Biondi et al. (1994) e Cambardella et al. (1994) para quantificar a dependência espacial e possibilitar a avaliação do grau dessa dependência, existe a possibilidade de avaliar essa dependência utilizando uma técnica denominada de envelopes simulados.

Os envelopes simulados são gerados a partir de permutações dos dados observados nas coordenadas do *grid* amostral. Isto é, mantêm-se inalteradas as coordenadas geográficas, e permutam-se os valores da variável (atributo) observada. O princípio da realização das permutações é tentar quebrar a estrutura de dependência espacial dos dados, gerando uma espécie de dados independentes.

Ao se realizar as permutações, gerando, a princípio, dados aproximadamente sem dependência espacial, têm-se condições de construir envelopes simulados no semivariograma dos dados originais. A figura 7 mostra dois exemplos de semivariogramas com envelopes simulados, com base em 99 permutações, realizados no pacote geoR (RIBEIRO JÚNIOR; DIGGLE, 2001).

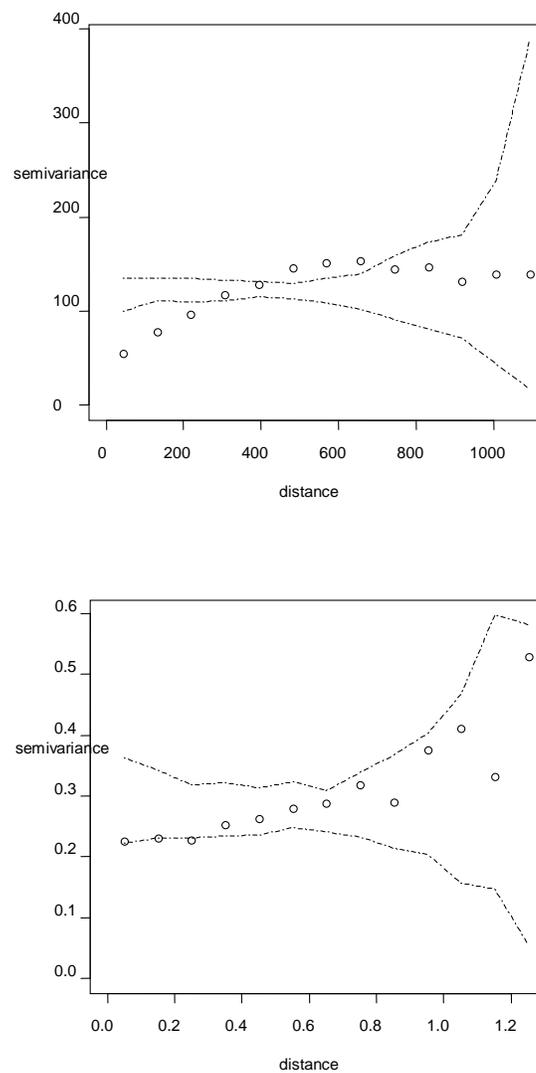


Figura 7 Exemplos de semivariogramas com envelopes simulados

No *geoR* (RIBEIRO JÚNIOR; DIGGLE, 2001), o comando que gera as permutações para construir os envelopes simulados é o *variog.mc.env*.

Contudo, apesar dos envelopes simulados gerarem uma indicação sobre a dependência espacial, é apenas uma análise gráfica, e não um teste sobre tal dependência, pois não gera um valor p para interpretação mais objetiva. Portanto, pode-se considerar que os envelopes simulados são uma espécie de ferramenta de avaliação visual.

Em alguns casos, como, por exemplo, no segundo semivariograma da figura 7, o pesquisador pode ficar em dúvida se a dependência espacial é significativa ou não, pois existem três pontos praticamente coincidindo com o limite inferior do envelope. Além disso, os pontos do semivariograma parecem indicar um comportamento aproximadamente linear, não sendo assim possível, concluir que poderia haver um ajuste por modelo de efeito pepita puro. Este fato mostra a fragilidade da utilização dos envelopes simulados.

Assim, o mais próximo de uma inferência, é a aplicação de envelopes simulados, que, de forma exploratória visual, avaliam a existência ou não de dependência espacial. Ou seja, apesar de realizar-se a simulação dos envelopes, não se tem interesse em testar a significância da continuidade espacial.

Assim, o emprego dos envelopes simulados não é suficiente para realizar uma inferência formal sobre a hipótese nula de não existência de dependência espacial em Geoestatística.

2.6 Modelos teóricos de semivariograma

Após a construção do semivariograma é necessário ajustar um modelo teórico que explique o comportamento do semivariograma em termos de modelar a variabilidade observada.

Segundo Alves et al. (2011), a ideia de ajustar um modelo teórico aos dados experimentais é sumarizar as relações espaciais nos dados. Para Carvalho,

Silveira e Vieira (2002), os modelos matemáticos ajustados permitem visualizar a natureza da variação espacial das variáveis estudadas.

De acordo com Carvalho, Silveira e Vieira (2002), além de permitir a visualização da natureza da variação espacial, os modelos matemáticos são necessários para outras aplicações, como, por exemplo, realizar a krigagem. Da mesma forma, para Silva et al. (2010), a análise do modelo de semivariograma é uma etapa importante, dado que tal modelo é a interpretação da estrutura de correlação espacial a ser utilizada nos procedimentos inferenciais da krigagem.

Os modelos teóricos mais comumente usados são classificados como (JOURNAL; HUIJBREGTS, 2003):

- a) Modelos com patamar (ou modelos de transição) e um relacionamento linear na origem:
 - modelo esférico;
 - modelo exponencial.
- b) Modelos com patamar e um relacionamento parabólico na origem:
 - modelo gaussiano.
- c) Modelos sem patamar (o correspondente a um processo estocástico que é então somente intrínseco e não tem nem covariância nem variância finita *a priori*).
- d) Modelo efeito pepita puro.

Nesta tese são discutidos os modelos esférico, exponencial, gaussiano e modelo de efeito pepita puro. Considerando os parâmetros C_0 , C_1 e a , é possível definir os modelos teóricos.

O modelo esférico é dado pela seguinte expressão (ISAACS; SRIVASTAVA, 1989):

$$\gamma(h)_{esf} = \begin{cases} 0 & , h = 0 \\ C_0 + C_1 \left[\frac{3}{2} \left(\frac{h}{a} \right) - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right] & , 0 < h \leq a . \\ C_0 + C_1 & , h > a \end{cases} \quad (15)$$

O modelo exponencial é dado por (ISAACS; SRIVASTAVA, 1989):

$$\gamma(h)_{exp} = \begin{cases} 0 & , h = 0 \\ C_0 + C_1 \left[1 - \exp\left(-\frac{3h}{a}\right) \right] & , h \neq 0 . \end{cases} \quad (16)$$

O modelo gaussiano é expresso da forma (ISAACS; SRIVASTAVA, 1989):

$$\gamma(h)_{gau} = \begin{cases} 0 & , h = 0 \\ C_0 + C_1 \left\{ 1 - \exp\left[-3\left(\frac{h}{a}\right)^2\right] \right\} & , h \neq 0 . \end{cases} \quad (17)$$

O modelo de efeito pepita puro é dado por (ISAACS; SRIVASTAVA, 1989):

$$\gamma(h)_{epp} = \begin{cases} 0 & , h = 0 \\ C_0 & , h \neq 0 . \end{cases} \quad (18)$$

Resumidamente, tem-se que, uma vez identificada a dependência espacial pelo semivariograma, o objetivo da análise espacial é modelar essa dependência a fim de fazer predições sobre os dados (OLIVEIRA et al., 2006).

3 METODOLOGIA

Em consonância com os objetivos propostos, a tese é desenvolvida em duas etapas metodológicas básicas. A primeira, envolvendo pesquisa bibliográfica, a fim de consolidar o referencial teórico. Já a segunda etapa, com a metodologia empregada para obtenção dos resultados, é formada por:

- a) 1º - Desenvolvimento de uma proposta de correção para os modelos de semivariogramas exponencial e gaussiano, apresentada na seção 3.1;
- b) 2º - Desenvolvimento de uma proposta de novos índices para descrição da estrutura de dependência espacial, apresentada na seção 3.2;
- c) 3º - Desenvolvimento de uma proposta de novos testes de hipótese de dependência espacial, apresentada na seção 3.3.

Esses desenvolvimentos são feitos utilizando-se da teoria geoestatística e do recurso de simulações estocásticas. Todo o desenvolvimento foi realizado no *software* R (R DEVELOPMENT CORE TEAM, 2012).

3.1 Desenvolvimento da proposta de uma correção para os modelos de semivariogramas exponencial e gaussiano

O procedimento metodológico consiste em calcular uma correção, denominada de p , para os modelos exponencial e gaussiano, com base no percentual de explicação da contribuição que se deseja atingir.

A principal motivação para a construção dessa correção se dá no princípio de que, quanto mais próximo do valor real da contribuição (C_1), mais próximo da descrição verdadeira do comportamento do fenômeno se está chegando. E, assim, mais adequada é a forma de escrita da equação dos modelos.

Antes da construção da correção, é realizado um estudo de demonstração para mostrar que, com base na forma escrita dos modelos exponencial e gaussiano apresentada em Isaaks e Srivastava (1989), o alcance prático é obtido em 95% da contribuição.

A correção realizada, inicialmente, se baseia na explicação de 99,99% da contribuição. Após, define-se o procedimento, de forma geral, para qualquer percentual de explicação da contribuição que se deseja atingir.

3.2 Desenvolvimento da proposta de índices para descrever a dependência espacial

Propõe-se a construção de três índices para descrever a dependência espacial, com base no conceito de áreas de dependência espacial observadas no semivariograma. Com base nesse conceito, inicialmente é definida uma medida que é dada pelo cálculo dessa área de dependência espacial.

O primeiro índice a ser proposto é baseado no conceito de áreas de dependência observadas no semivariograma. Este índice relaciona a área de dependência observada com a área de dependência espacial possível de ser atingida. Esta última seria, conceitualmente, a área de dependência espacial máxima para um fenômeno atingir.

O segundo índice baseia-se no fato de que o semivariograma é composto por duas áreas distintas, uma sendo a área de dependência espacial, e outra

sendo a área de não-dependência espacial do semivariograma. Assim, a composição do índice é dada pela relação entre a área de dependência espacial com a área total do semivariograma. Em que a área total é a soma das áreas de dependência e não-dependência espaciais.

O terceiro índice é construído a partir do conceito de correlograma. A ideia é construir uma medida que derive da área de correlação espacial dada pelo correlograma.

Também, é realizada uma aplicação comparativa entre os índices. Para realizar a comparação são simulados 25 cenários com diferentes graus de dependência espacial. Em cada cenário são calculados os índices e, também, o erro quadrático médio dos erros de predição ($EQMe$). O $EQMe$ é obtido através do procedimento de validação conforme pode ser visto, com mais detalhes, em Webster e Oliver (2007). A intenção é relacionar cada índice com o $EQMe$ através do cálculo de correlações. Por fim, para verificar, estatisticamente, se um índice se destaca aos demais, utiliza-se o teste de Qui-Quadrado. Para mais detalhes sobre o teste de Qui-quadrado ver Siegel e Castellan Júnior (2006).

3.3 Desenvolvimento da proposta de testes de hipótese para a dependência espacial

Para a construção dos testes de hipótese, inicialmente, é definida a hipótese nula de ausência de dependência espacial. Após, define-se uma estatística de teste que é gerada a partir do conceito de área de dependência espacial dada pelo semivariograma.

O primeiro teste é construído com base em simulações de fenômenos apresentando característica de efeito pepita puro, ou seja, fenômenos em que a hipótese nula de ausência de dependência espacial é verdadeira.

O segundo teste é construído com base em permutações dos dados originais. Isto é feito pelo fato de que a permutação dos dados nas coordenadas gera independência espacial, pois se quebra a estrutura original. Assim, nas permutações se tem que a hipótese nula de ausência de dependência espacial é verdadeira.

Por fim, é estudado o poder dos dois testes. Utiliza-se o índice de melhor desempenho, conforme procedimento explicado na seção anterior, para gerar uma medida de dependência espacial a partir de diferentes cenários, e com base nisso avaliar o poder dos testes em relação aos valores do índice.

4 RESULTADOS E DISCUSSÃO

Nesta seção são apresentados os resultados obtidos a partir da aplicação da metodologia proposta, a fim de alcançar os objetivos lançados para esta tese. Também, são realizadas as discussões a partir dos resultados obtidos.

4.1 Proposta de correção para os modelos teóricos exponencial e gaussiano

Nesta seção é proposta uma correção para os modelos de semivariogramas exponencial e gaussiano.

É necessário, inicialmente, distinguir os parâmetros da dependência espacial do fenômeno em estudo daqueles parâmetros dos modelos ajustados para explicar tal dependência espacial. Essa distinção, muitas vezes não é clara, e pode ser confundida pelos pesquisadores. Assim, o primeiro produto desta tese é gerar um esclarecimento sobre a distinção que existe entre os parâmetros da dependência espacial de qualquer fenômeno natural, que pode ser avaliado no espaço e os parâmetros dos modelos que podem ser estimados para explicar a dependência espacial de tais fenômenos.

Todo fenômeno que apresenta dependência espacial, em Geoestatística, com estacionariedade de 2ª ordem, possui quatro parâmetros: contribuição (*CT*), efeito pepita (*EP*), patamar (*PT*) e alcance (*AL*).

Contudo, quando se trata dos modelos teóricos de dependência espacial (neste caso, esférico, exponencial e gaussiano), deve-se fazer uma diferenciação necessária nos parâmetros. Isto é, deve-se observar que um parâmetro da dependência espacial de um fenômeno é diferente de um parâmetro do modelo teórico de dependência espacial.

Nos três modelos (esférico, exponencial e gaussiano), tem-se que o parâmetro contribuição é igual à contribuição do fenômeno ($C_1 = CT$), o efeito pepita, também, é o mesmo do fenômeno ($C_0 = EP$), e o patamar, por consequência, também é igual ao patamar da dependência espacial do fenômeno ($C = C_0 + C_1 = PT$).

Contudo, em relação ao parâmetro alcance dos modelos, deve-se fazer uma observação importante. Usa-se a seguinte simbologia para o alcance prático: $AL = a$. Nos fenômenos em que se ajusta modelo esférico, o alcance real é finito. Já nos fenômenos em que se ajustam modelo exponencial ou gaussiano, o alcance real é infinito. Contudo, para fins de operações matemáticas, convencionou-se que o alcance é virtualmente finito. A tabela 1 apresenta as relações entre os modelos e o alcance de dependência espacial.

Tabela 1 Relação entre os modelos e o alcance de dependência espacial

Modelo	Alcance real	Alcance prático	Parâmetro do modelo (alcance teórico)
Esférico	AL	AL=a	a
Exponencial	∞	AL=a	a/3*
Gaussiano	∞	AL=a	a/ $\sqrt{3}$ *

* Nos modelos exponencial e gaussiano, o alcance prático é obtido a 95% da contribuição.

Para o modelo esférico de dependência espacial tem-se que o parâmetro é igual ao alcance prático. Contudo, para o modelo exponencial, tem-se que o parâmetro é igual a $AL/3 = a/3$. No modelo gaussiano, tem-se o parâmetro $AL/\sqrt{3} = a/\sqrt{3}$. Dessa forma, tem-se uma parametrização mais adequada, no sentido de que permite uma interpretação mais facilitada da diferença que existe entre o alcance prático e o parâmetro nos modelos exponencial e gaussiano.

O modelo esférico fica escrito da mesma forma dada em Isaaks e Srivastava (1989), como:

$$\gamma(h)_{esf} = \begin{cases} 0 & , h = 0 \\ C_0 + C_1 \left[1,5 \left(\frac{h}{a} \right) - 0,5 \left(\frac{h}{a} \right)^3 \right] & , 0 < h \leq a \\ C_0 + C_1 & , h > a \end{cases} \quad (19)$$

Já o modelo exponencial, com a parametrização $a/3$, fica escrito como:

$$\gamma(h)_{exp} = \begin{cases} 0 & , h = 0 \\ C_0 + C_1 \left[1 - e^{\left(-\frac{h}{a/3} \right)} \right] & , h \neq 0 \end{cases} \quad (20)$$

O modelo gaussiano, com a parametrização $a/\sqrt{3}$, é escrito como:

$$\gamma(h)_{gau} = \begin{cases} 0 & , h = 0 \\ C_0 + C_1 \left\{ 1 - e^{\left[-\left(\frac{h}{a/\sqrt{3}} \right)^2 \right]} \right\} & , h \neq 0 \end{cases} \quad (21)$$

Esta forma de escrever os modelos exponencial e gaussiano, dada em (20) e (21), facilita o entendimento do que é o alcance teórico, se comparada com a forma de escrever os modelos utilizados por Isaaks e Srivastava (1989).

Além disso, os modelos em (20) e (21), estão apresentados de forma semelhante àquela implementada no pacote geoR (RIBEIRO JÚNIOR; DIGGLE, 2001).

Nos modelos exponencial (20) e gaussiano (21), o alcance AL (alcance prático) é obtido em 95% da contribuição. Em Biondi et al. (1994) também há uma indicação de que o alcance, nos modelos exponencial e gaussiano, é obtido na distância em que o semivariograma atinge 95% da contribuição. Assim, os resultados dados nesta tese e em Biondi et al. (1994), sobre a relação entre contribuição e alcance, são comprovados, através do estudo de demonstração, na próxima seção.

4.1.1 Demonstrações de que o alcance é obtido a 95% da contribuição nos modelos exponencial e gaussiano

A demonstração para o modelo exponencial:

$$\gamma(h)_{\text{exp}} = C_0 + C_1 \left[1 - \exp\left(-\frac{h}{a/3}\right) \right], h \neq 0$$

Fazendo $\gamma(h)_{\text{exp}} = C_0 + 0,95(C_1)$, tem-se

$$C_0 + C_1 \left[1 - \exp\left(-\frac{h}{a/3}\right) \right] = C_0 + 0,95(C_1)$$

$$C_1 \left[1 - \exp\left(-\frac{h}{a/3}\right) \right] = C_0 - C_0 + 0,95(C_1)$$

$$C_1 \left[1 - \exp\left(-\frac{h}{a/3}\right) \right] = 0,95(C_1)$$

$$1 - \exp\left(-\frac{h}{a/3}\right) = \frac{0,95(C_1)}{C_1}$$

$$1 - \exp\left(-\frac{h}{a/3}\right) = 0,95$$

$$-\exp\left(-\frac{h}{a/3}\right) = 0,95 - 1$$

$$-\exp\left(-\frac{h}{a/3}\right) = -0,05$$

Aplicando o \ln nos dois lados da equação, tem-se

$$\frac{h}{a/3} = \ln(0,05)$$

$$\frac{h}{a/3} = -2,9957 \cong -3$$

$$h \cong 3(a/3)$$

$$h \cong a$$

A demonstração para o modelo gaussiano:

$$\gamma(h)_{gau} = C_0 + C_1 \left\{ 1 - \exp\left[-\left(\frac{h}{a/\sqrt{3}} \right)^2 \right] \right\}, h \neq 0$$

Fazendo $\gamma(h)_{gau} = C_0 + 0,95(C_1)$, tem-se

$$C_0 + C_1 \left\{ 1 - \exp \left[- \left(\frac{h}{a/\sqrt{3}} \right)^2 \right] \right\} = C_0 + 0,95(C_1)$$

$$C_1 \left\{ 1 - \exp \left[- \left(\frac{h}{a/\sqrt{3}} \right)^2 \right] \right\} = C_0 - C_0 + 0,95(C_1)$$

$$C_1 \left\{ 1 - \exp \left[- \left(\frac{h}{a/\sqrt{3}} \right)^2 \right] \right\} = 0,95(C_1)$$

$$1 - \exp \left[- \left(\frac{h}{a/\sqrt{3}} \right)^2 \right] = \frac{0,95(C_1)}{C_1}$$

$$1 - \exp \left[- \left(\frac{h}{a/\sqrt{3}} \right)^2 \right] = 0,95$$

$$- \exp \left[- \left(\frac{h}{a/\sqrt{3}} \right)^2 \right] = 0,95 - 1$$

$$- \exp \left[- \left(\frac{h}{a/\sqrt{3}} \right)^2 \right] = -0,05$$

Aplicando o \ln nos dois lados da equação, tem-se:

$$\left(\frac{h}{a/\sqrt{3}} \right)^2 = \ln(0,05)$$

$$\frac{h^2}{a^2/3} = -2,9957 \cong 3$$

$$h^2 \cong 3(a^2/3)$$

$$h \cong a$$

Na figura 8 tem-se o comportamento dos modelos esférico, exponencial e gaussiano em relação ao alcance AL .

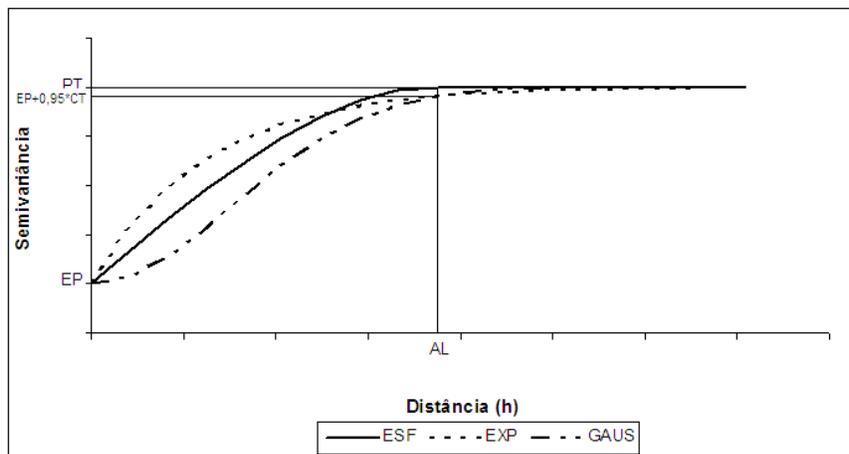


Figura 8 Comportamento dos modelos esférico (ESF), exponencial (EXP) e gaussiano (GAUS) em relação ao alcance AL

4.1.2 Proposta de correção para o modelo exponencial

Como pode ser observado nas equações (20) e (21), os modelos exponencial e gaussiano são modelados com base nos valores 3 e $\sqrt{3}$, respectivamente, pois consideram que o alcance prático é obtido a 95% do valor da contribuição ($0,95 * C_1$).

Contudo, esta definição de que o alcance prático é obtido a 95% do valor da contribuição, é apenas uma convenção, como já dito, para operacionalização matemática, já que, nos modelos exponencial e gaussiano, o alcance real é infinito.

Na literatura clássica (ISAAKS; SRIVASTAVA, 1989; JOURNAL; HUIJBREGTS, 2003; WACKERNAGEL, 2003; WEBSTER; OLIVER, 2007) não se tem uma descrição ou demonstração do porque utilizar-se o fator 95%. Isto é comprovado em Webster e Oliver (2007), ao comentarem que, para aplicações práticas, considera-se que o alcance efetivo (alcance finito) é, usualmente, a distância na qual o semivariograma é igual a 95% da contribuição, assumindo aproximadamente o valor de $3a$.

Dessa forma, o segundo produto desta tese é a proposta de uma correção para os modelos exponencial e gaussiano. Tal correção é gerada a partir do princípio de que, quanto mais próximo do valor real do alcance ($AL = \infty$), mais próximo da descrição verdadeira de comportamento do fenômeno se está chegando.

A ideia é propor que a correção seja tal que o alcance prático seja atingido a 99,99% da contribuição ($0,9999 * C_1$). Com este novo conceito, consegue-se atingir maior proximidade do alcance real ($AL = \infty$), se comparada com a modelagem realizada atualmente com 95% da contribuição ($0,95 * C_1$). A

figura 9 apresenta a comparação do comportamento dos modelos exponencial e gaussiano entre considerar 95% da contribuição ou 99,99% da contribuição.

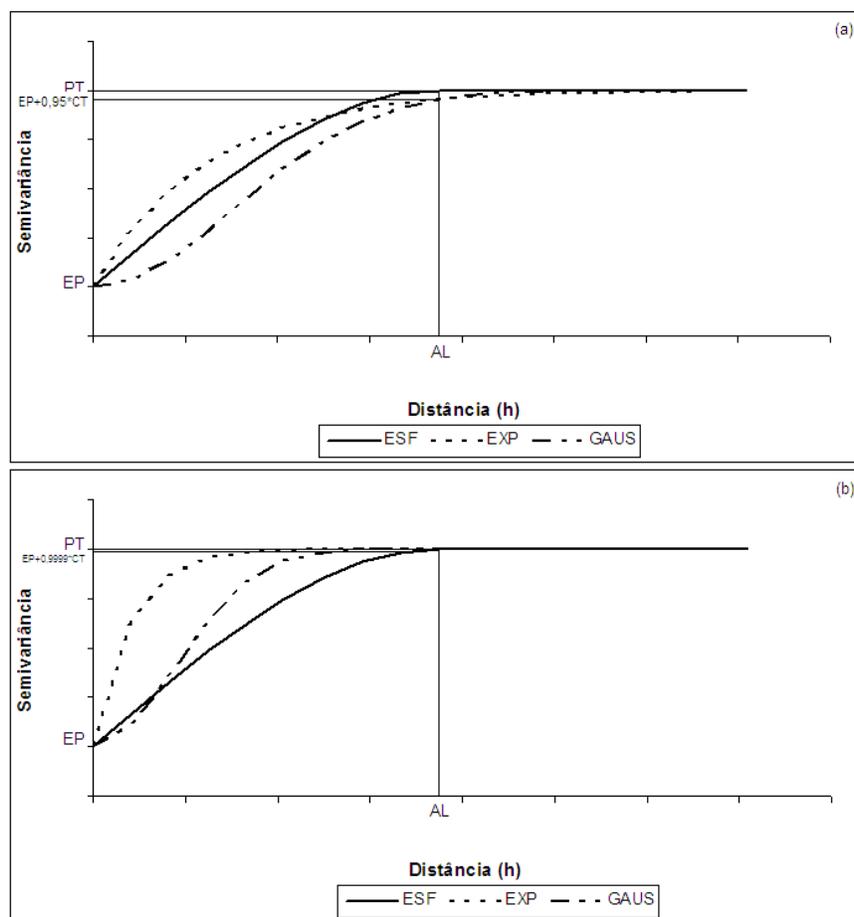


Figura 9 Comparação entre os comportamentos dos modelos exponencial e gaussiano em relação ao alcance a ser atingido a 95% da contribuição (a) e a 99,99% da contribuição (b)

Assim, com base na figura 9, observa-se que o alcance, no caso de se considerar 99,99% da contribuição, tem maior aproximação do alcance real (que é infinito). Portanto, a correção sugerida nesta tese, possui características mais

adequadas para representar melhor o comportamento verdadeiro da dependência espacial do fenômeno em estudo.

Ao considerar 99,99% da contribuição como sendo o valor para o qual se obtém o valor do alcance prático para o modelo exponencial, tem-se que a correção (p) é dada por:

$$\begin{aligned} \Rightarrow \gamma_{\text{exp}}(h) &= C_0 + 0,9999(C_1) \\ \Rightarrow C_0 + C_1 \left[1 - e^{\left(-\frac{a}{a/p} \right)} \right] &= C_0 + 0,9999(C_1) \\ \Rightarrow C_1 \left[1 - e^{\left(-\frac{a}{a/p} \right)} \right] &= 0,9999(C_1) \\ \Rightarrow 1 - e^{\left(-\frac{a}{a/p} \right)} &= 0,9999 \\ \Rightarrow e^{(-p)} &= 0,0001 \\ \Rightarrow -p &= \ln(0,0001) \\ \Rightarrow p &= 9,2 \\ \Rightarrow p &\cong 9 \end{aligned}$$

O valor da correção, para 99,99% da contribuição, é 9. Assim, de posse deste novo valor, é possível reescrever a equação do modelo exponencial como:

$$\gamma_{\text{exp}}(h) = C_0 + C_1 \left[1 - e^{\left(-\frac{h}{a/9} \right)} \right], \quad h \neq 0 \quad (22)$$

4.1.3 Proposta de correção para o modelo gaussiano

No caso de considerar-se o modelo gaussiano e adotar 99,99% da contribuição como sendo o valor para o qual se obtém o valor do alcance prático, tem-se que a correção (p) é dada por:

$$\begin{aligned} \Rightarrow \gamma_{gaus}(h) &= C_0 + 0,9999(C_1) \\ \Rightarrow C_0 + C_1 \left[1 - e^{-\left[\left(\frac{a}{\sqrt{p}} \right)^2 \right]} \right] &= C_0 + 0,9999(C_1) \\ \Rightarrow C_1 \left[1 - e^{-\left[\left(\frac{a}{\sqrt{p}} \right)^2 \right]} \right] &= 0,9999(C_1) \\ \Rightarrow 1 - e^{-\left[\left(\frac{a}{\sqrt{p}} \right)^2 \right]} &= 0,9999 \\ \Rightarrow e^{(-p)} &= 0,0001 \\ \Rightarrow -p &= \ln(0,0001) \\ \Rightarrow p &= 9,2 \\ \Rightarrow p &\cong 9 \end{aligned}$$

O valor da correção, para 99,99% da contribuição, é 9. Assim, de posse deste valor, é possível reescrever a equação do modelo gaussiano como:

$$\gamma_{gaus}(h) = C_0 + C_1 \left[1 - e^{\left[-\left(\frac{h}{a/\sqrt{9}} \right)^2 \right]} \right], \quad h \neq 0 \quad (23)$$

Generalizando os modelos exponencial e gaussiano, pode-se escrever as equações, respectivamente, na forma:

$$\gamma_{exp}(h) = C_0 + C_1 \left[1 - e^{\left(\frac{h}{a/p} \right)} \right] \quad (24)$$

e

$$\gamma_{gaus}(h) = C_0 + C_1 \left[1 - e^{\left[-\left(\frac{h}{a/\sqrt{p}} \right)^2 \right]} \right] \quad (25)$$

em que a correção p está relacionada com o percentual de explicação do comportamento do fenômeno em estudo, ou seja, relacionada com o percentual da contribuição para o qual se define o alcance prático. Na tabela 2 apresenta-se a relação entre o percentual da contribuição que se define e o valor da correção p .

Tabela 2 Relação entre a correção p , dos modelos exponencial e gaussiano, e o percentual da contribuição que se deseja atingir

Percentual da contribuição	Correção p
100%	∞
99,99%	9
⋮	⋮
95%	3
⋮	⋮
0%	0

Vê-se que a determinação de uma correção a partir de um percentual da contribuição revela a verdadeira natureza dos modelos exponencial e gaussiano. De fato, a consideração desse percentual permite criar uma classe de modelos novos para semivariogramas, permitindo ampliar o leque de modelagem de fenômenos exponenciais ou gaussianos.

O pesquisador tem, então, a opção de definir o valor de p a partir do percentual da contribuição que deseja atingir, ou explicar. Nesta tese, são utilizados $p = 3$ e $p = 9$, para fins de apresentação dos resultados, pois se opta em utilizar 95% e 99,99%, da contribuição, como percentual de explicação, respectivamente.

4.2 Proposta de índices de dependência espacial para semivariogramas com patamar

Nesta seção são propostos três índices de dependência espacial (*IDE*) que têm suas construções baseadas na geometria do semivariograma ou do correlograma. A ideia da construção desses índices é devida ao desenvolvimento do conceito de área de dependência espacial que é abordado na próxima seção.

4.2.1 Definição da área de dependência espacial no semivariograma

Nesta seção é proposta a definição de uma área, no semivariograma, que permite compreender ou quantificar a dependência espacial. Essa área, que caracteriza a dependência espacial, é calculada com base na geometria do semivariograma, sendo definida como a área de dependência espacial. Esta área está relacionada com os índices de dependência espacial que são propostos nesta tese.

Então, no semivariograma, essa área está entre o patamar e o modelo teórico e entre a origem e o alcance prático. Ela permite interpretar a dependência espacial em termos de área, facilitando, dessa forma, por exemplo, a comparação de semivariogramas. No modelo esférico, a área de dependência espacial, está representada na Figura 10. No modelo exponencial, a área de dependência espacial, está representada na figura 11. No modelo gaussiano, a área de dependência espacial, está representada na figura 12. A figura 13 apresenta a área de dependência espacial no modelo exponencial com correção $p=9$. E a figura 14 apresenta a área de dependência espacial no modelo gaussiano com correção $p=9$.

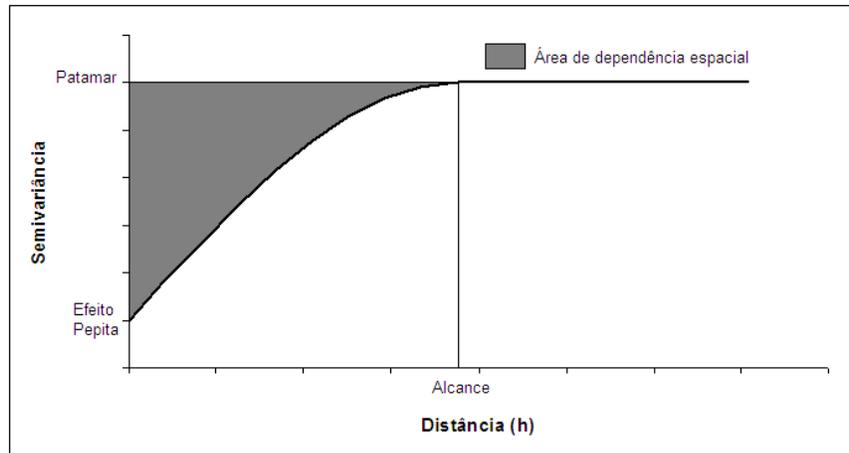


Figura 10 Representação da área de dependência espacial em semivariograma com ajuste de modelo esférico

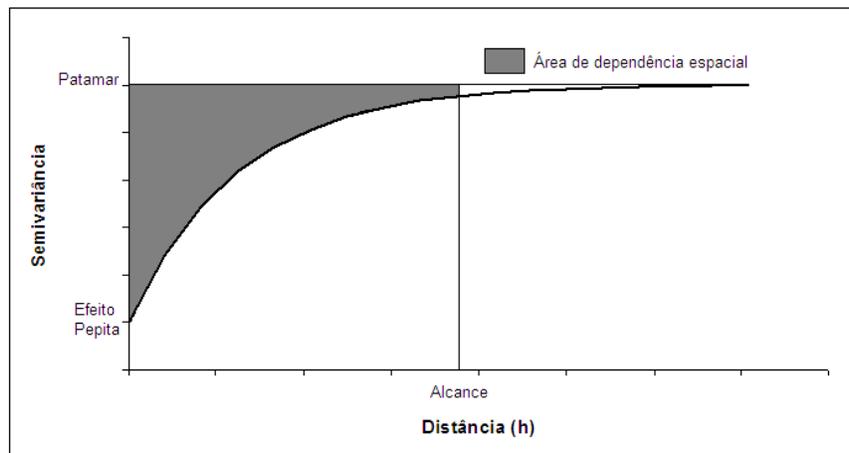


Figura 11 Área de dependência espacial em semivariograma com ajuste de modelo exponencial

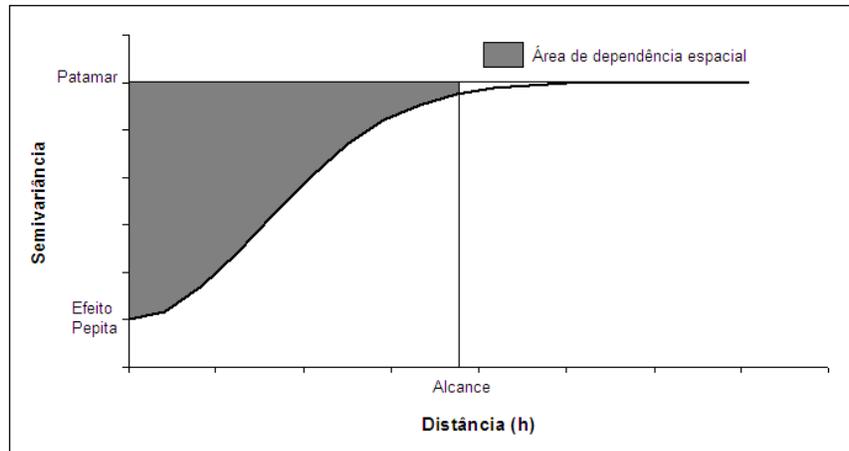


Figura 12 Área de dependência espacial em semivariograma com ajuste de modelo gaussiano

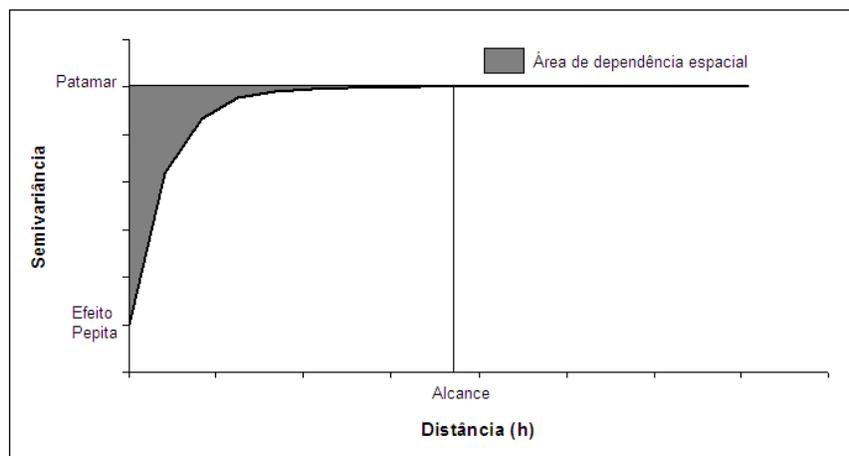


Figura 13 Área de dependência espacial em semivariograma com ajuste de modelo exponencial com correção $p=9$

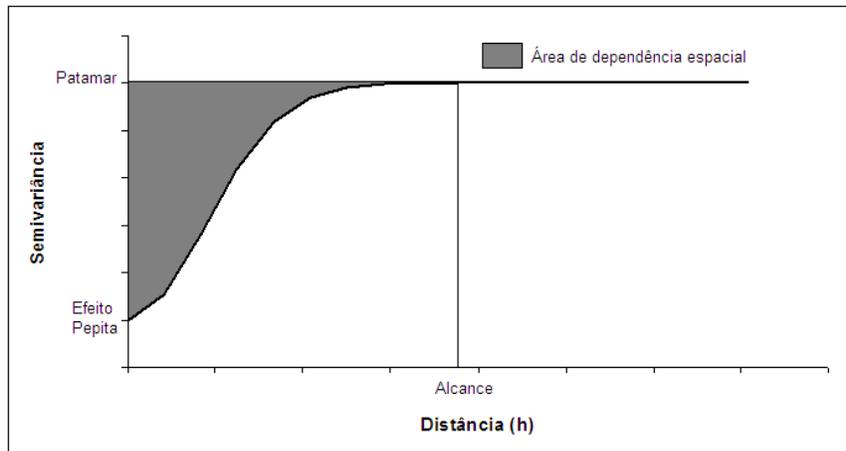


Figura 14 Área de dependência espacial em semivariograma com ajuste de modelo gaussiano com correção $p=9$

Com base nas áreas de dependência espacial, definidas nas figuras 10 a 14, é possível construir, por integração, uma estatística que reflète a dependência espacial, pois ela é a própria área de dependência espacial. Como essa estatística é a área de dependência espacial, sua denominação é “*ADE*”, que é a abreviatura de “Área de Dependência Espacial”. Os cálculos completos da estatística *ADE*, para os modelos estudados nesta tese, estão apresentados na seção a seguir.

4.2.2 Cálculo da estatística ADE

Cálculo da estatística ADE para o modelo esférico:

$$\begin{aligned}
 ADE_{esf} &= \text{área de dependência} \\
 &= \int_0^a (C - \gamma_{esf}(h)) dh \\
 &= \int_0^a \left(C - \left(C_0 + C_1 \left[\frac{3}{2} \left(\frac{h}{a} \right) - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right] \right) \right) dh \\
 &= a \cdot \left[C - C_0 - \left(\frac{3C_1}{4} \right) + \left(\frac{C_1}{8} \right) \right] \\
 &= a \cdot \left[C - C_0 - C_1 \left(\frac{5}{8} \right) \right] \\
 &= a \cdot [C - C_0 - 0,625 \cdot (C_1)]
 \end{aligned}$$

Considerando $C = C_0 + C_1$, tem-se

$$\begin{aligned}
 ADE_{esf} &= a \cdot [C_0 + C_1 - C_0 - 0,625 \cdot (C_1)] \\
 &= a \cdot [C_1 - 0,625 \cdot (C_1)] \\
 &= [0,375 \cdot (C_1)] a
 \end{aligned}$$

Cálculo da estatística ADE para o modelo exponencial com $p=3$:

$$\begin{aligned}
 ADE_{\text{exp}} &= \text{área de dependência} \\
 &= \int_0^a (C - \gamma_{\text{exp}}(h)) dh \\
 &= \int_0^a \left(C - \left(C_0 + C_1 \left[1 - e^{-\left(\frac{h}{a/3}\right)} \right] \right) \right) dh \\
 &= a \cdot [C - C_0 - C_1] + a C_1 \cdot \left[\frac{e^3 - 1}{3e^3} \right] \\
 &= a \cdot [C - C_0 - C_1] + a C_1 \cdot \left[\frac{1 - e^{-3}}{3} \right] \\
 &= a \cdot \left[C - C_0 - C_1 + C_1 \cdot \left(\frac{1 - e^{-3}}{3} \right) \right]
 \end{aligned}$$

Considerando $C = C_0 + C_1$, tem-se

$$\begin{aligned}
 ADE_{\text{exp}} &= a \cdot \left[C_0 + C_1 - C_0 - C_1 + C_1 \cdot \left(\frac{1 - e^{-3}}{3} \right) \right] \\
 &= \left[C_1 \cdot \left(\frac{1 - e^{-3}}{3} \right) \right] \cdot a \\
 &= \left[C_1 \cdot \left(\frac{1 - 0,05}{3} \right) \right] \cdot a \\
 &= [C_1 \cdot (0,317)] \cdot a \\
 &= [0,317 \cdot (C_1)] \cdot a
 \end{aligned}$$

Cálculo da estatística ADE para o modelo exponencial com $p=9$:

$$\begin{aligned}
 ADE_{\text{exp}} &= \text{área de dependência} \\
 &= \int_0^a (C - \gamma_{\text{exp}}(h)) dh \\
 &= \int_0^a \left(C - \left(C_0 + C_1 \left[1 - e^{\left(-\frac{h}{a/9} \right)} \right] \right) \right) dh \\
 &= a \cdot [C - C_0 - C_1] + a C_1 \cdot \left[\frac{e^9 - 1}{9e^9} \right] \\
 &= a \cdot [C - C_0 - C_1] + a C_1 \cdot \left[\frac{1 - e^{-9}}{9} \right] \\
 &= a \cdot \left[C - C_0 - C_1 + C_1 \cdot \left(\frac{1 - e^{-9}}{9} \right) \right]
 \end{aligned}$$

Considerando $C = C_0 + C_1$, tem-se

$$\begin{aligned}
 ADE_{\text{exp}} &= a \cdot \left[C_0 + C_1 - C_0 - C_1 + C_1 \cdot \left(\frac{1 - e^{-9}}{9} \right) \right] \\
 &= \left[C_1 \cdot \left(\frac{1 - e^{-9}}{9} \right) \right] \cdot a \\
 &= \left[C_1 \cdot \left(\frac{1 - 0,00012}{9} \right) \right] \cdot a \\
 &= [C_1 \cdot (0,111)] \cdot a \\
 &= [0,111 \cdot (C_1)] \cdot a
 \end{aligned}$$

Cálculo da estatística ADE para o modelo gaussiano com $p=3$:

$$\begin{aligned}
 ADE_{gaus} &= \text{área de dependência} \\
 &= \int_0^a (C - \gamma_{gaus}(h)) dh \\
 &= \int_0^a \left(C - \left(C_0 + C_1 \left[1 - e^{-\left(\frac{h}{a/\sqrt{3}} \right)^2} \right] \right) \right) dh \\
 &= a \cdot [C - C_0 - C_1] + C_1 \cdot \int_0^a e^{-\left(\frac{h}{a/\sqrt{3}} \right)^2} dh
 \end{aligned}$$

$$\begin{aligned}
 \int_0^a e^{-\left(\frac{h}{a/\sqrt{3}} \right)^2} dh &= \sqrt{2\pi} \cdot \frac{a}{\sqrt{6}} \cdot \int_0^a \frac{e^{-\frac{1}{2} \left(\frac{h}{a/\sqrt{6}} \right)^2}}{\sqrt{2\pi} \cdot a/\sqrt{6}} dh \\
 &= \sqrt{2\pi} \cdot \frac{a}{\sqrt{6}} \cdot \int_0^{b\sqrt{6}} \frac{e^{-\frac{1}{2} \left(\frac{h}{b} \right)^2}}{\sqrt{2\pi} \cdot b} dh \\
 &= \sqrt{2\pi} \cdot \frac{a}{\sqrt{6}} \cdot (0,4928471) \\
 &= 0,50434356a
 \end{aligned}$$

Como $\int_0^{b\sqrt{6}} \frac{e^{-\frac{1}{2} \left(\frac{h}{b} \right)^2}}{\sqrt{2\pi} \cdot b} dh$ é uma distribuição normal padronizada

acumulada de zero até $\sqrt{6}$, tem como resultado numérico o valor 0,4928471.

Dessa forma se obtém

$$\begin{aligned} & \sqrt{2\pi} \cdot \frac{a}{\sqrt{6}} \cdot \int_0^{b \cdot \sqrt{6}} \frac{e^{-\frac{1}{2} \left(\frac{h}{b} \right)^2}}{\sqrt{2\pi} \cdot b} dh \\ &= \sqrt{2\pi} \cdot \frac{a}{\sqrt{6}} \cdot (0,4928471) \\ &= 0,50434356a \end{aligned}$$

Assim, tem-se

$$\begin{aligned} ADE_{gaus} &= a \cdot [C - C_0 - C_1] + C_1 \cdot \int_0^a e^{\left(-\left(\frac{h}{a/\sqrt{3}} \right)^2 \right)} dh \\ &= a \cdot [C - C_0 - C_1] + C_1 \cdot [0,504 \cdot a] \end{aligned}$$

Considerando $C = C_0 + C_1$, tem-se

$$\begin{aligned} ADE_{gaus} &= a \cdot [C_0 + C_1 - C_0 - C_1] + a \cdot C_1 \cdot [0,504] \\ &= a \cdot C_1 \cdot (0,504) \\ &= [0,504 \cdot (C_1)] \cdot a \end{aligned}$$

Cálculo da estatística ADE para o modelo gaussiano com $p=9$:

$$\begin{aligned}
 ADE_{gaus} &= \text{área de dependência} \\
 &= \int_0^a (C - \gamma_{gaus}(h)) dh \\
 &= \int_0^a \left(C - \left(C_0 + C_1 \left[1 - e^{-\left(\frac{h}{a/\sqrt{9}} \right)^2} \right] \right) \right) dh \\
 &= a \cdot [C - C_0 - C_1] + C_1 \cdot \int_0^a e^{-\left(\frac{h}{a/\sqrt{9}} \right)^2} dh
 \end{aligned}$$

Pode-se escrever a expressão $\int_0^a e^{-\left(\frac{h}{a/\sqrt{9}} \right)^2} dh$ da forma “gaussiana”:

$$\int_0^a e^{-\left(\frac{h}{a/\sqrt{9}} \right)^2} dh = \sqrt{2\pi} \cdot \frac{a}{\sqrt{18}} \cdot \int_0^a \frac{e^{-\frac{1}{2} \left(\frac{h}{a/\sqrt{18}} \right)^2}}{\sqrt{2\pi} \cdot a/\sqrt{18}} dh$$

Fazendo $b = a/\sqrt{18}$ ser o desvio padrão, tem-se:

$$\begin{aligned} \int_0^a e^{-\left(\frac{h}{a/\sqrt{9}}\right)^2} dh &= \sqrt{2\pi} \cdot \frac{a}{\sqrt{18}} \cdot \int_0^a \frac{e^{-\frac{1}{2}\left(\frac{h}{a/\sqrt{18}}\right)^2}}{\sqrt{2\pi} \cdot a/\sqrt{18}} dh \\ &= \sqrt{2\pi} \cdot \frac{a}{\sqrt{18}} \cdot \int_0^{b \cdot \sqrt{18}} \frac{e^{-\frac{1}{2}\left(\frac{h}{b}\right)^2}}{\sqrt{2\pi} \cdot b} dh \\ &= \sqrt{2\pi} \cdot \frac{a}{\sqrt{18}} \cdot (0,499989) \\ &= 0,29540248a \end{aligned}$$

Como $\int_0^{b \cdot \sqrt{18}} \frac{e^{-\frac{1}{2}\left(\frac{h}{b}\right)^2}}{\sqrt{2\pi} \cdot b} dh$ é uma distribuição normal padronizada

acumulada de zero até $\sqrt{18}$, tem como resultado numérico o valor 0,499989.

Dessa forma se obtém

$$\begin{aligned} \sqrt{2\pi} \cdot \frac{a}{\sqrt{18}} \cdot \int_0^{b \cdot \sqrt{18}} \frac{e^{-\frac{1}{2}\left(\frac{h}{b}\right)^2}}{\sqrt{2\pi} \cdot b} dh \\ &= \sqrt{2\pi} \cdot \frac{a}{\sqrt{18}} \cdot (0,499989) \\ &= 0,29540248a \end{aligned}$$

Assim, tem-se

$$\begin{aligned} ADE_{gaus} &= a.[C - C_0 - C_1] + C_1 \cdot \int_0^a e^{\left(-\left(\frac{h}{a/\sqrt{9}}\right)^2\right)} dh \\ &= a.[C - C_0 - C_1] + C_1.[0,295.a] \end{aligned}$$

Considerando $C = C_0 + C_1$, tem-se

$$\begin{aligned} ADE_{gaus} &= a.[C_0 + C_1 - C_0 - C_1] + a.C_1.[0,295] \\ &= a.C_1.(0,295) \\ &= [0,295.(C_1)]a \end{aligned}$$

Portanto, para o modelo esférico, a estatística ADE é dada por:

$$ADE_{esf} = \text{área de dependência} = 0,375.(C_1.a), \quad (26)$$

em que C_1 é a contribuição e a é o alcance prático.

Para o modelo exponencial com correção $p=3$, a estatística ADE é dada por:

$$ADE_{exp} = 0,317.(C_1.a) \quad (27)$$

Para o modelo gaussiano com correção $p=3$, a estatística ADE é dada por:

$$ADE_{gaus} = 0,504.(C_1.a) \quad (28)$$

Para o modelo exponencial com correção $p=9$, a estatística ADE é dada por:

$$ADE_{exp} = 0,111.(C_1.a) \quad (29)$$

Para o modelo gaussiano com correção $p=9$, a estatística ADE é dada por:

$$ADE_{gaus} = 0,295.(C_1.a) \quad (30)$$

É possível observar nas expressões (26) a (30) que cada modelo, com sua respectiva correção p , apresenta uma constante em sua respectiva estatística ADE . Essa constante, inerente a cada modelo, pode ser entendida como um fator de modelo que reflete a força da dependência espacial. Definindo-se essa constante como “fator de modelo” (FM), pode-se, então, generalizar a estatística ADE , para os modelos exponencial e gaussiano com qualquer correção p . A estatística ADE , em sua forma geral, é dada por:

$$ADE_{Modelo} = FM.(C_1.a), \quad (31)$$

em que, FM é o fator do modelo, C_1 é a contribuição e a é o alcance prático.

A estatística ADE é uma variável aleatória, pois é composta por duas variáveis aleatórias: a contribuição (C_1) e o alcance prático (a). Portanto, a estatística ADE assume valores de forma aleatória (cada amostra gera um valor para a estatística ADE).

O fator de modelo (FM) pode ser entendido como um valor que expressa a força da dependência espacial que o modelo pode atingir. O FM , teoricamente, pode atingir valores entre 0 e 1. Quanto mais próximo de 1 for o valor de FM , maior é a força da dependência espacial do modelo. O modelo que apresenta o menor valor de fator de modelo é o modelo de efeito pepita puro, que possui FM igual a zero. Já o modelo linear com patamar, não discutido nesta tese, é o que apresenta o valor mediano para o fator de modelo, com $FM = 0,5$. Uma representação gráfica do modelo linear com patamar está apresentada na figura 15.

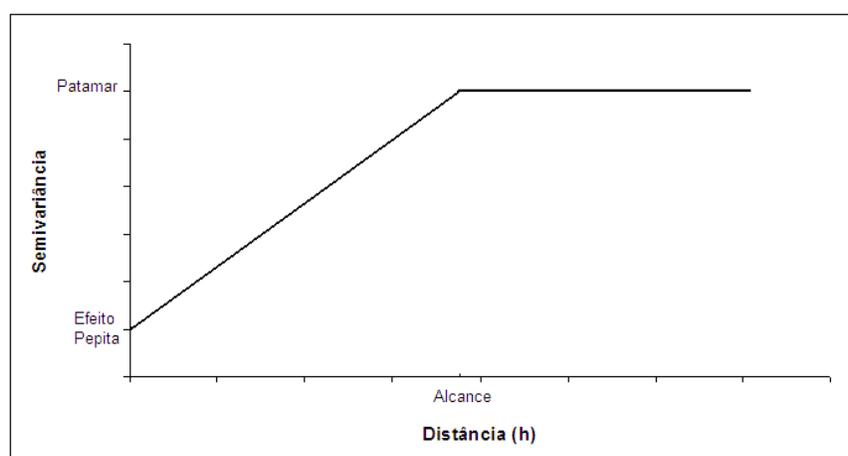


Figura 15 Representação gráfica do modelo linear com patamar

Contudo, na prática, até então não se têm nenhum modelo, conhecido na literatura, que apresente fator de modelo máximo, ou seja, que possua $FM = 1$. Assim, apresenta-se, nesta tese, a proposta de um modelo teórico, com dependência espacial máxima, que possui $FM = 1$. A forma matemática deste modelo máximo proposto é dada por:

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ C_0, & 0 < h \leq a \\ C_0 + C_1, & h > a \end{cases} \quad (32)$$

A forma gráfica, aproximada, do modelo proposto em (32), com $FM = 1$, é apresentada na figura 16. Esse modelo máximo, apesar de teórico, pode ser mais desenvolvido no futuro, ou ser aproximado por algum outro modelo ainda não desenvolvido na Geoestatística. Por enquanto, esse modelo é utilizado, nesta tese, como característico da máxima dependência espacial, caracterizando o máximo valor de FM .

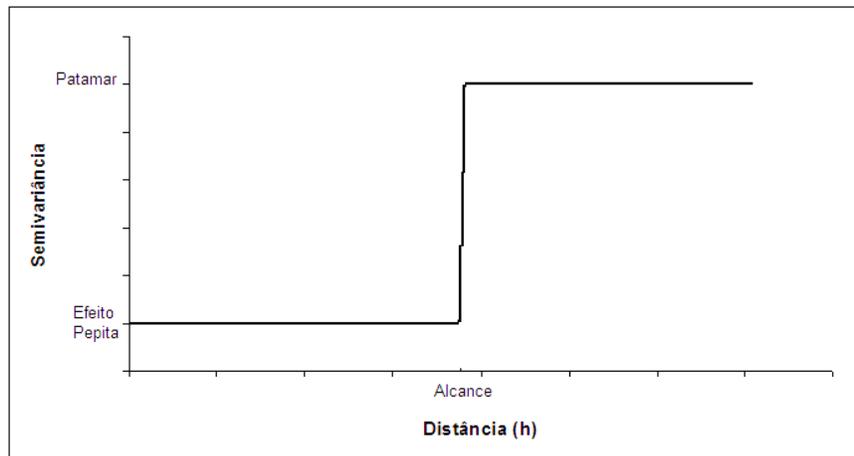


Figura 16 Representação gráfica, aproximada, do modelo máximo proposto

A tabela 3 mostra um esquema com alguns modelos e seus respectivos valores para o fator de modelo.

Tabela 3 Esquema com alguns modelos e seus respectivos valores para o fator de modelo

Modelo teórico	Fator de modelo (FM)
Modelo proposto em (4.2.3)	1,000
Gaussiano com $p = 3$	0,504
Linear com patamar	0,500
Esférico	0,375
Exponencial com $p = 3$	0,317
Gaussiano com $p = 9$	0,295
Exponencial com $p = 9$	0,111
Efeito pepita puro	0,000

Observa-se, na tabela 3, que os modelos estudados nesta tese apresentam fator de modelo de mediano a baixo, indicando que apresentam baixas forças de dependência espacial. O modelo gaussiano com correção $p = 3$, dentre os estudados, é o que apresenta maior força, com valor mediano de 0,504.

Uma observação importante deve ser feita em relação ao comportamento conjunto das correções (p) para os modelos exponencial e gaussiano e o fator de modelo (FM). Quanto menor for o percentual de explicação que se deseja atingir da contribuição, menor é o valor da correção (p) e, em consequência, maior é o valor do fator de modelo (FM). Esse fato, particular dessa nova classe de modelos exponenciais e gaussianos, é devido à definição da correção (p), pois, a área de dependência espacial (ADE) fica maior conforme diminui o percentual de explicação que se deseja da contribuição, o que permite modelar mais tipos de fenômenos exponenciais e gaussianos.

4.2.3 Proposta do primeiro índice de dependência espacial

O primeiro índice de dependência espacial (denominado de *IDEI*), que se propõe nesta tese, é gerado do estudo da geometria do semivariograma, através da estatística *ADE* apresentada anteriormente. A construção do índice se dá pela relação entre a área de dependência espacial observada no semivariograma e a área de dependência espacial possível de ser atingida por um fenômeno no semivariograma.

Cada modelo teórico ajustado (esférico, exponencial e gaussiano) ao semivariograma possui seu próprio índice.

4.2.3.1 *IDEI* para o modelo esférico

Com base na visualização da figura 17 (a), é possível perceber que a área de dependência espacial observada (ADE_{obs}) para um fenômeno qualquer, em estudo, está definida entre o patamar ($C = C_0 + C_1$) e o modelo esférico ($\gamma_{esf}(h)$). Já, com a observação da figura 17 (b), tem-se que a área de dependência espacial possível de ser atingida ($ADE_{possível}$), que seria a área de dependência espacial máxima, está definida entre o patamar ($C = C_0 + C_1$) e o modelo esférico adaptado ($\gamma_{esf}(h)_{Adaptado}$).

O modelo adaptado é um modelo obtido por meio do conceito de área de dependência máxima possível de ser atingida quando o parâmetro efeito pepita é nulo. Esse modelo é aquele que gera a máxima área de dependência espacial possível de ser atingida. O modelo esférico adaptado é dado por:

$$\gamma_{esf}(h)_{Adaptado} = C \left[1,5 \left(\frac{h}{a} \right) - 0,5 \left(\frac{h}{a} \right)^3 \right], \quad 0 < h < a \quad (33)$$

em que C é o patamar, a é o alcance prático e h é a distância entre pontos.

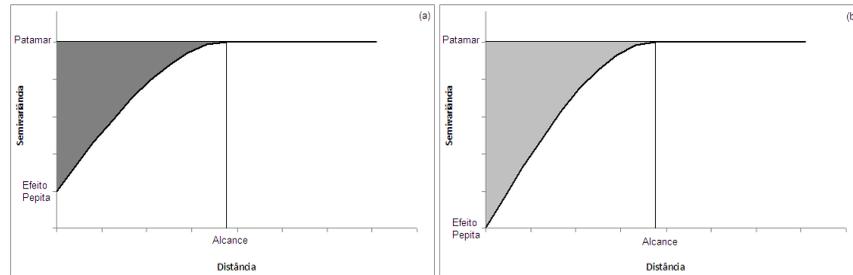


Figura 17 Área de dependência espacial observada (a) e área de dependência espacial possível (b), para o modelo esférico

Para a construção do índice $IDE1_{esf}$ faz-se, então, a razão entre a área de dependência espacial observada, apresentada na figura 17 (a), e a área de dependência espacial possível de ser atingida, apresentada na figura 17 (b). Ou seja, calcula-se a área definida na figura 17 (a) e a área definida na figura 17 (b) e, então, calcula-se a razão entre elas. Dessa forma, tem-se que o índice $IDE1$ para o modelo esférico é dado por:

$$\begin{aligned}
 IDE1_{esf} &= \frac{ADE_{obs}}{ADE_{possível}} \\
 &= \frac{\int_0^a (C - \gamma_{esf}(h)) dh}{\int_0^a (C - \gamma_{esf}(h)_{Adaptado}) dh} \\
 &= \frac{C_1}{C_0 + C_1}
 \end{aligned}$$

Assim, verifica-se que o $IDE1_{esf}$, para semivariogramas com ajuste de modelo esférico, é:

$$IDEI_{esf} (\%) = \left(\frac{C_1}{C_0 + C_1} \right) \cdot 100. \quad (34)$$

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

Pode-se observar que a expressão (34) é igual à expressão do índice dado em Biondi et al. (1994), ou seja, percebe-se que a expressão do índice derivado das áreas de dependência espacial do semivariograma esférico é idêntico ao índice DE .

Os cálculos completos do $IDEI$, para os modelos estudados nesta tese, estão apresentados na seção a seguir.

4.2.3.2 Cálculo do índice $IDEI$

Cálculo do índice $IDEI$ para o modelo esférico:

$$\begin{aligned} IDEI_{esf} &= \frac{ADE_{obs}}{ADE_{possível}} \\ &= \frac{\int_0^a (C - \gamma_{esf}(h)) dh}{\int_0^a (C - \gamma_{esf}(h)_{Adaptado}) dh} \\ &= \frac{\int_0^a \left(C - \left(C_0 + C_1 \left[\frac{3}{2} \left(\frac{h}{a} \right) - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right] \right) \right) dh}{\int_0^a \left(C - \left(C \left[\frac{3}{2} \left(\frac{h}{a} \right) - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right] \right) \right) dh} \\ &= \frac{[0,375.(C_1)]_a}{[0,375.(C)]_a} \\ &= \frac{C_1}{C} \\ &= \frac{C_1}{C_0 + C_1} \end{aligned}$$

Cálculo do índice *IDEI* para o modelo exponencial com $p=3$:

$$\begin{aligned}
 IDEI_{\text{exp}} &= \frac{ADE_{\text{obs}}}{ADE_{\text{possível}}} \\
 &= \frac{\int_0^a (C - \gamma_{\text{exp}}(h)) dh}{\int_0^a (C - \gamma_{\text{exp}}(h)_{\text{Adaptado}}) dh} \\
 &= \frac{\int_0^a \left(C - \left(C_0 + C_1 \left[1 - e^{\left(\frac{-h}{a/3} \right)} \right] \right) \right) dh}{\int_0^a \left(C - \left(C \left[1 - e^{\left(\frac{-h}{a/3} \right)} \right] \right) \right) dh} \\
 &= \frac{[0,317.(C_1)]a}{[0,317.(C)]a} \\
 &= \frac{C_1}{C} \\
 &= \frac{C_1}{C_0 + C_1}
 \end{aligned}$$

Cálculo do índice *IDEI* para o modelo exponencial com $p=9$:

$$\begin{aligned}
 IDEI_{\text{exp}} &= \frac{ADE_{\text{obs}}}{ADE_{\text{possível}}} \\
 &= \frac{\int_0^a (C - \gamma_{\text{exp}}(h)) dh}{\int_0^a (C - \gamma_{\text{exp}}(h)_{\text{Adaptado}}) dh} \\
 &= \frac{\int_0^a \left(C - \left(C_0 + C_1 \left[1 - e^{\left(-\frac{h}{a/9} \right)} \right] \right) \right) dh}{\int_0^a \left(C - \left(C \left[1 - e^{\left(-\frac{h}{a/9} \right)} \right] \right) \right) dh} \\
 &= \frac{[0,111.(C_1)]_a}{[0,111.(C)]_a} \\
 &= \frac{C_1}{C} \\
 &= \frac{C_1}{C_0 + C_1}
 \end{aligned}$$

Cálculo do índice *IDEI* para o modelo gaussiano com $p=3$:

$$\begin{aligned}
 IDEI_{gaus} &= \frac{ADE_{obs}}{ADE_{possível}} \\
 &= \frac{\int_0^a (C - \gamma_{gaus}(h)) dh}{\int_0^a (C - \gamma_{gaus}(h)_{Adaptado}) dh} \\
 &= \frac{[0,504.(C_1)]a}{[0,504.(C)]a} \\
 &= \frac{C_1}{C} \\
 &= \frac{C_1}{C_0 + C_1}
 \end{aligned}$$

Cálculo do índice *IDEI* para o modelo gaussiano com $p=9$:

$$\begin{aligned}
 IDEI_{gaus} &= \frac{ADE_{obs}}{ADE_{possível}} \\
 &= \frac{\int_0^a (C - \gamma_{gaus}(h)) dh}{\int_0^a (C - \gamma_{gaus}(h)_{Adaptado}) dh} \\
 &= \frac{[0,295.(C_1)]a}{[0,295.(C)]a} \\
 &= \frac{C_1}{C} \\
 &= \frac{C_1}{C_0 + C_1}
 \end{aligned}$$

4.2.3.3 Generalização do *IDEI*

Com base no resultado obtido para o modelo esférico, e com base nas figuras 18 e 19, pode-se construir uma generalização para os modelos exponencial e gaussiano. Assim, o índice *IDEI* é apresentado, de forma geral, como:

$$IDEI_{Modelo} (\%) = \left(\frac{C_1}{C_0 + C_1} \right) \cdot 100 \quad (35)$$

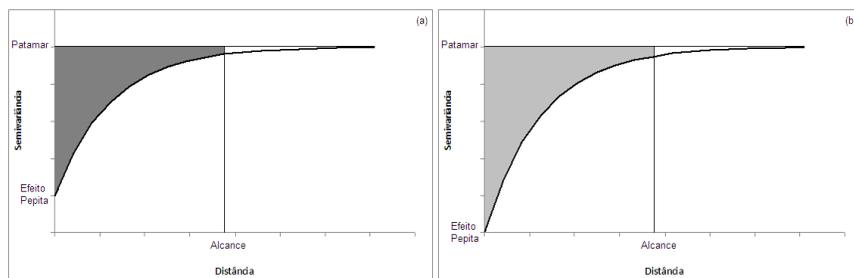


Figura 18 Área de dependência espacial observada (a) e área de dependência espacial possível (b), para o modelo exponencial

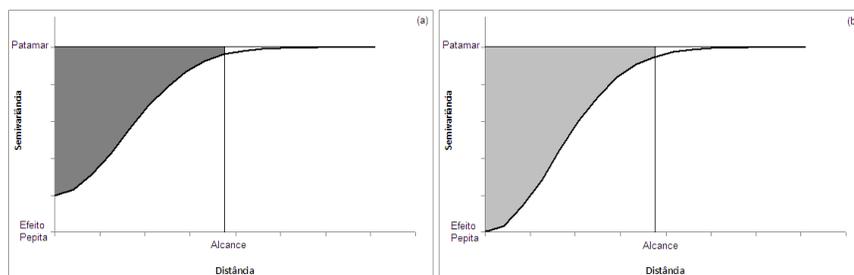


Figura 19 Área de dependência espacial observada (a) e área de dependência espacial possível (b), para o modelo gaussiano

O *IDEI*, ao relacionar a área de dependência espacial observada (ADE_{obs}) com a área de dependência espacial possível de ser atingida ($ADE_{possível}$), cria uma razão que pode assumir valores no intervalo entre 0 e 100%. Assim, a interpretação do *IDEI* se refere a quanto da área de dependência espacial possível se atingiu, no fenômeno em estudo, com a área de dependência espacial observada.

Além disso, verifica-se que a expressão do *IDEI* é a mesma, independentemente do modelo de ajuste considerado (esférico, exponencial ou gaussiano). Ao se construir um índice para descrição da dependência espacial, nos mesmos moldes de um coeficiente de variação (*CV*), observa-se que apenas a contribuição (C_1) e o efeito pepita (C_0) compõem a expressão do índice.

No trabalho de Biondi et al. (1994), não há nenhuma indicação da forma como foi gerado o índice *DE*, apenas há a definição de que o *DE* entende-se como o percentual da variância espacial que é explicada pela dependência espacial. Os autores parecem, apenas, estudar a variabilidade gerada no eixo das semivariâncias no semivariograma, relacionando a variância estruturada (dada pela contribuição) com a variância total (dada pelo patamar).

Assim, nesta tese, o índice *IDEI*, proposto, mostra, por meio do estudo geométrico do semivariograma, envolvendo áreas de dependência espacial, que pode ser utilizado para descrever a dependência espacial de semivariogramas com ajuste de modelos esférico, exponencial e gaussiano, quando se admite o conceito de comparação de áreas de dependências observada e possível de ser atingida.

Contudo, deve-se observar que esse índice apresenta a mesma fragilidade dos índices *RD* e *DE*, que é o fato de só considerar os parâmetros contribuição e efeito pepita, e não o alcance e/ou o modelo ajustado.

4.2.4 Proposta do segundo índice de dependência espacial

A construção do segundo índice (denominado de *IDE2*) se dá pela relação entre a área de dependência espacial observada e a área de dependência espacial total do semivariograma. A definição da área de dependência espacial (*ADE*) observada já foi apresentada anteriormente.

Se for observado o semivariograma, dado na figura 20, pode-se visualizar a área de dependência espacial em destaque. Toda a área restante do semivariograma pode ser definida como sendo uma área de não-dependência espacial (*ANDE*). Ou seja, a área de não-dependência espacial pode ser dividida em duas. Uma estando definida entre o modelo e o eixo das distâncias, e entre a origem e o alcance prático, denominada de *ANDE1*. A outra definida entre o patamar e o eixo das distâncias, e entre o alcance prático e o máximo valor do eixo das distâncias, denominada de *ANDE2*.

Contudo, convencionou-se, nesta tese, que as áreas utilizadas estão definidas entre o patamar e o eixo das distâncias, e entre a origem e o alcance prático. Isto é, utilizam-se as áreas *ADE* e *ANDE1*. Dessa forma, define-se a área de dependência espacial total (*AT*) como sendo a soma das áreas *ADE* e *ANDE1*, que forma um retângulo entre o patamar e o alcance prático.

4.2.4.1 *IDE2* para o modelo esférico

Com base na visualização da figura 20, é possível perceber que a área de dependência espacial observada (ADE_{obs}) para um fenômeno qualquer, em estudo, está definida entre o patamar ($C = C_0 + C_1$) e o modelo esférico ($\gamma_{esf}(h)$). Já a área de não-dependência espacial observada ($ANDE1_{obs}$) que se considera, nesta tese, para fins de cálculo do índice, está definida entre o modelo

esférico ($\gamma_{esf}(h)$) e o eixo das distâncias e entre a origem e o alcance prático. A soma das áreas de dependência espacial observada e não-dependência espacial observada resultam na área total observada (AT_{obs}). A área total pode ser escrita por:

$$AT_{obs} = C.a \quad (36)$$

em que C é o patamar e a é o alcance prático.

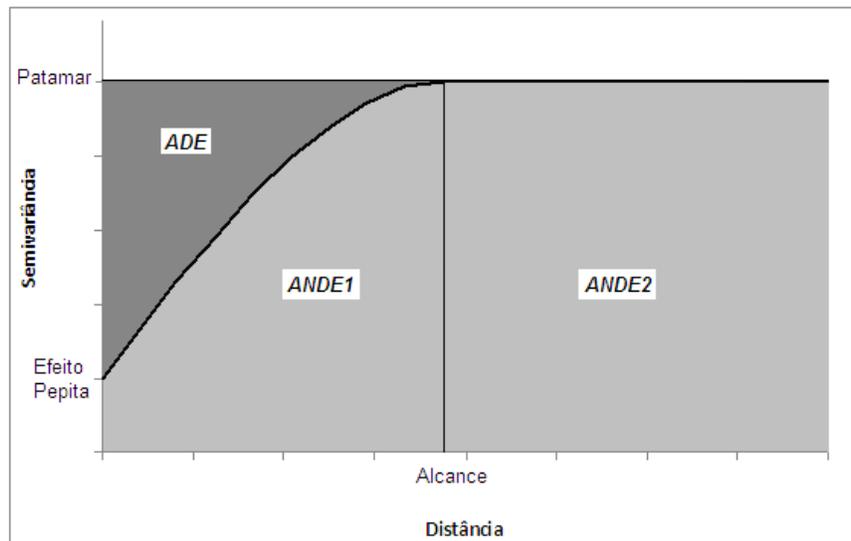


Figura 20 Áreas de dependência e não-dependência espaciais para o modelo esférico

A intenção é relacionar a área de dependência espacial ADE_{obs} com a área total AT_{obs} . Assim, define-se o valor do $IDE2_{esf}$ como sendo a razão entre a área de dependência espacial observada e a área total observada. Tem-se:

$$\begin{aligned}
 IDE2_{esf} &= \frac{ADE_{obs}}{AT_{obs}} \\
 &= \frac{\int_0^a (C - \gamma_{esf}(h)) dh}{C.a} \\
 &= \frac{0,375.(C_1)}{C_0 + C_1}
 \end{aligned}$$

Assim, verifica-se que o $IDE2_{esf}$, para semivariogramas com ajuste de modelo esférico, é dado por:

$$IDE2_{esf} (\%) = 0,375 \left(\frac{C_1}{C_0 + C_1} \right) . 100 \quad (37)$$

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

O índice proposto, aqui, assume valores no intervalo entre 0 e 37,5%. Isto é devido a se utilizar a constante 0,375 que é o fator de modelo para o modelo esférico.

Apesar de esse índice variar entre 0 e 37,5%, e não entre 0 e 100%, ele é eficiente na interpretação da dependência espacial. A única diferença é que a escala de valores possíveis para esse índice é reduzida.

Os cálculos completos do $IDE2$, para os modelos estudados nesta tese, estão apresentados na próxima seção.

4.2.4.2 Cálculo do índice $IDE2$

Cálculo do índice $IDE2$ para o modelo esférico:

$$\begin{aligned}
 IDE2_{esf} &= \frac{ADE_{obs}}{AT_{obs}} \\
 &= \frac{\int_0^a (C - \gamma_{esf}(h)) dh}{C.a} \\
 &= \frac{[0,375.(C_1)].a}{C.a} \\
 &= \frac{0,375.(C_1)}{C} \\
 &= \frac{0,375.(C_1)}{C_0 + C_1}
 \end{aligned}$$

Cálculo do índice *IDE2* para o modelo exponencial com $p=3$:

$$\begin{aligned}
 IDE2_{exp} &= \frac{ADE_{obs}}{AT_{obs}} \\
 &= \frac{\int_0^a (C - \gamma_{exp}(h)) dh}{C.a} \\
 &= \frac{[0,317.(C_1)].a}{C.a} \\
 &= \frac{0,317.(C_1)}{C} \\
 &= \frac{0,317.(C_1)}{C_0 + C_1}
 \end{aligned}$$

Cálculo do índice *IDE2* para o modelo exponencial com $p=9$:

$$\begin{aligned}
 IDE2_{\text{exp}} &= \frac{ADE_{\text{obs}}}{AT_{\text{obs}}} \\
 &= \frac{\int_0^a (C - \gamma_{\text{exp}}(h)) dh}{C \cdot a} \\
 &= \frac{[0,111 \cdot (C_1)] \cdot a}{C \cdot a} \\
 &= \frac{0,111 \cdot (C_1)}{C} \\
 &= \frac{0,111 \cdot (C_1)}{C_0 + C_1}
 \end{aligned}$$

Cálculo do índice *IDE2* para o modelo gaussiano com $p=3$:

$$\begin{aligned}
 IDE2_{\text{gaus}} &= \frac{ADE_{\text{obs}}}{AT_{\text{obs}}} \\
 &= \frac{\int_0^a (C - \gamma_{\text{gaus}}(h)) dh}{C \cdot a} \\
 &= \frac{[0,504 \cdot (C_1)] \cdot a}{C \cdot a} \\
 &= \frac{0,504 \cdot (C_1)}{C} \\
 &= \frac{0,504 \cdot (C_1)}{C_0 + C_1}
 \end{aligned}$$

Cálculo do índice *IDE2* para o modelo gaussiano com $p=9$:

$$\begin{aligned}
 IDE2_{gaus} &= \frac{ADE_{obs}}{AT_{obs}} \\
 &= \frac{\int_0^a (C - \gamma_{gaus}(h)) dh}{C \cdot a} \\
 &= \frac{[0,295 \cdot (C_1)] \cdot a}{C \cdot a} \\
 &= \frac{0,295 \cdot (C_1)}{C} \\
 &= \frac{0,295 \cdot (C_1)}{C_0 + C_1}
 \end{aligned}$$

4.2.4.3 Generalização do *IDE2*

Com base no resultado obtido para o modelo esférico e a partir da visualização das figuras 21 e 22, é possível generalizar o índice *IDE2* para os modelos exponencial e gaussiano. Uma observação a ser feita, a partir das figuras 21 e 22, para os modelos exponencial e gaussiano, é a de que nesses dois modelos a área de dependência espacial vai além do alcance prático, diferentemente do que ocorre no modelo esférico que tem área de dependência espacial apenas até o alcance prático. Assim, nos modelos exponencial e gaussiano, só é considerada a área de dependência espacial definida até o alcance prático. A expressão do índice *IDE2*, generalizada, para os modelos exponencial e gaussiano, é dada por:

$$IDE2_{Modelo} (\%) = FM \cdot \left(\frac{C_1}{C_0 + C_1} \right) \cdot 100 \quad (38)$$

em que FM é o fator de modelo, C_0 é o efeito pepita e C_1 é a contribuição.

De forma geral, o $IDE2(\%)$ varia no intervalo $(0 \leq IDE2(\%) \leq FM \cdot 100\%)$. Por exemplo, para o modelo gaussiano com correção $p=3$, tem-se que o índice $IDE2_{gaus}(\%)$ varia entre 0 e 50,4%.

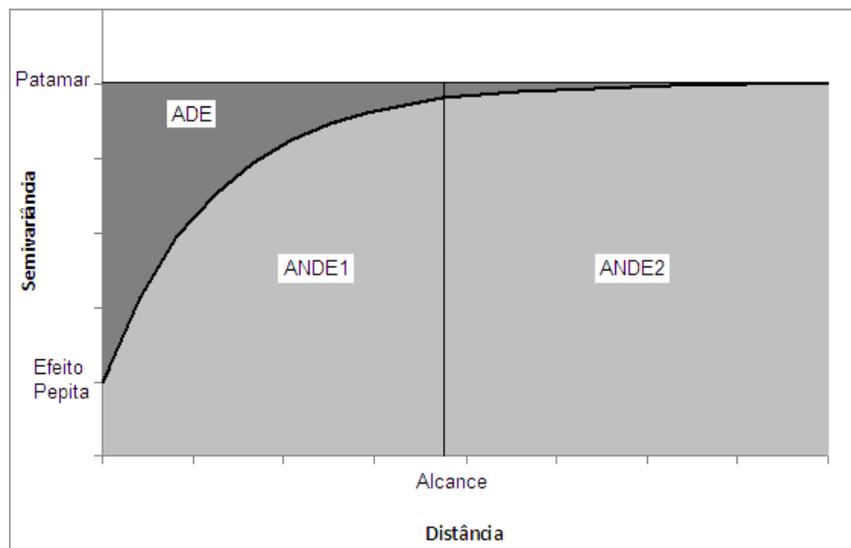


Figura 21 Áreas de dependência e não-dependência espaciais para o modelo exponencial

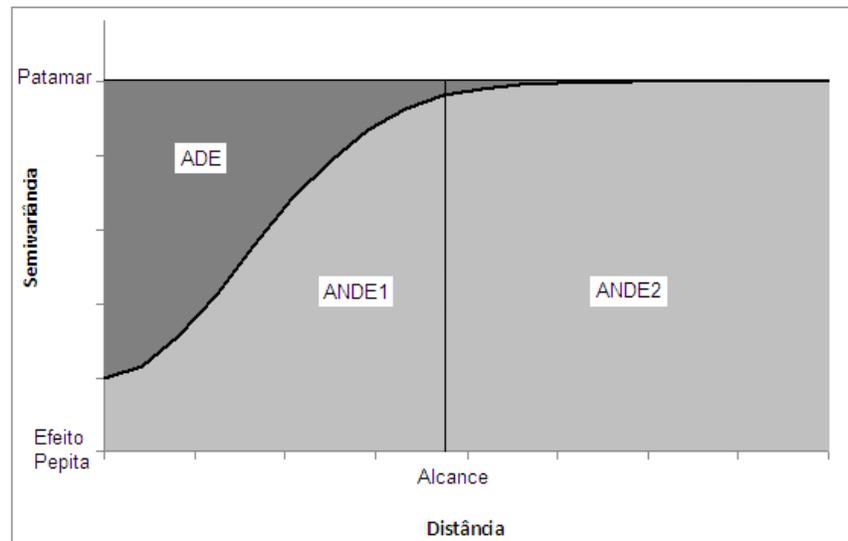


Figura 22 Áreas de dependência e não-dependência espaciais para o modelo gaussiano

O índice *IDE2* parece descrever de forma mais adequada a dependência espacial dos fenômenos, já que, representa o percentual da área total (soma entre as áreas de dependência e não-dependência espaciais) do semivariograma, que é ocupada pela área de dependência espacial. De acordo com este conceito, representado na figura 23, tem-se que nenhum fenômeno real pode possuir dependência espacial de 100%. Um fenômeno só teria 100% de dependência espacial se sua área de dependência espacial fosse igual à área total (ver figura 23(b)), o que não se observa na realidade. Assim, a definição de 100% de dependência espacial é apenas virtual.

Já o índice *IDE1*, que é igual ao índice *DE*, dado em Biondi et al. (1994), não traz este conceito, mas sim, um conceito de que a máxima dependência espacial observada é obtida quando a contribuição é igual ao patamar (efeito pepita nulo), como mostrado nas figuras 17(b), 18(b) e 19(b).

O índice *IDE2* é mais completo que o índice *IDE1*, pois considera o fator de modelo e é gerado a partir do conceito de áreas de dependência e não-dependência espaciais, como pode ser observado na figura 23. Contudo, esse índice também apresenta a fragilidade de só considerar, além do modelo ajustado, a contribuição e o efeito pepita em sua composição, não contemplando o alcance prático.

A figura 23(a) apresenta um semivariograma sem área de dependência espacial, pois é estimado o modelo de efeito pepita puro, indicando a existência de independência espacial. Esse é o caso onde o semivariograma é formado apenas pela área de não-dependência espacial.

Na figura 23(b), tem-se um semivariograma onde se define uma área de dependência espacial entre o patamar e o modelo teórico estimado, e entre a origem e o alcance prático. Nesse semivariograma pode-se fazer uma comparação entre a área de dependência espacial e a área total, pois a área total é composta por duas áreas, a área de dependência e a área de não-dependência espacial.

Por fim, na figura 23(c) apresenta-se um semivariograma com a máxima área de dependência espacial, pois a área total é composta somente pela área de dependência espacial.

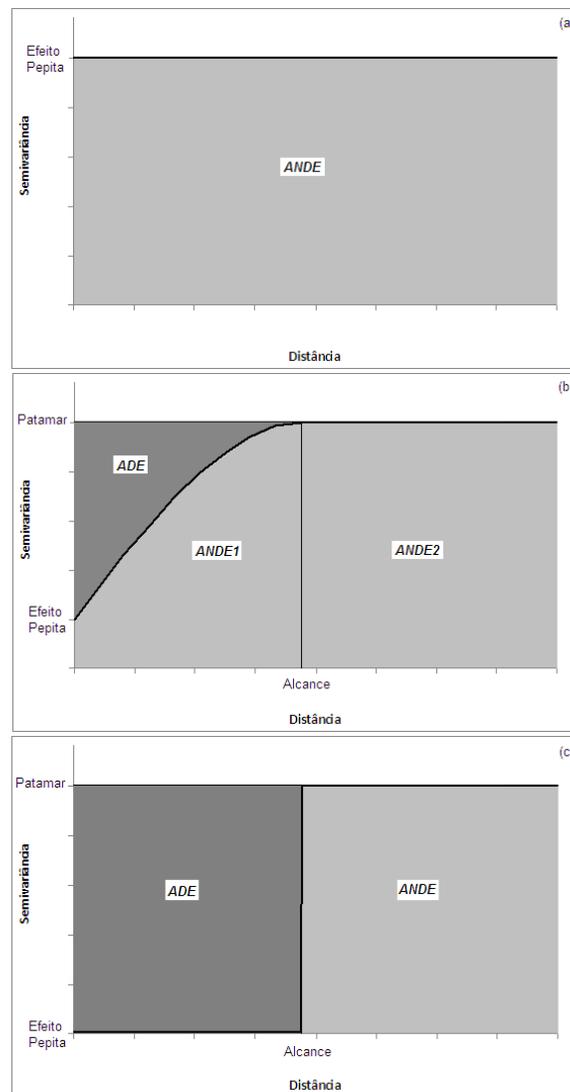


Figura 23 Representação de um semivariograma com dependência espacial nula (a), com dependência espacial maior que zero e menor que 100% (b) e com dependência espacial, virtualmente, igual a 100% (c)

4.2.5 Proposta do terceiro índice de dependência espacial

O índice *IDE3* é gerado a partir da definição do correlograma como forma de avaliar a correlação espacial existente no fenômeno em estudo. Assim, este terceiro índice descreve, em essência, o grau de correlação espacial.

4.2.5.1 *IDE3* para o modelo esférico

Na Geoestatística, quando se observa a existência de estacionariedade de segunda ordem, uma relação entre o semivariograma e o correlograma é válida. Essa relação diz que o correlograma pode ser escrito como:

$$\rho(h) = 1 - \frac{\gamma(h)}{C} \quad (39)$$

Com base nesta relação existente entre o correlograma e o semivariograma, é possível construir a figura 24, que apresenta uma representação geral de um correlograma com seus respectivos parâmetros e a área de correlação espacial correspondente.

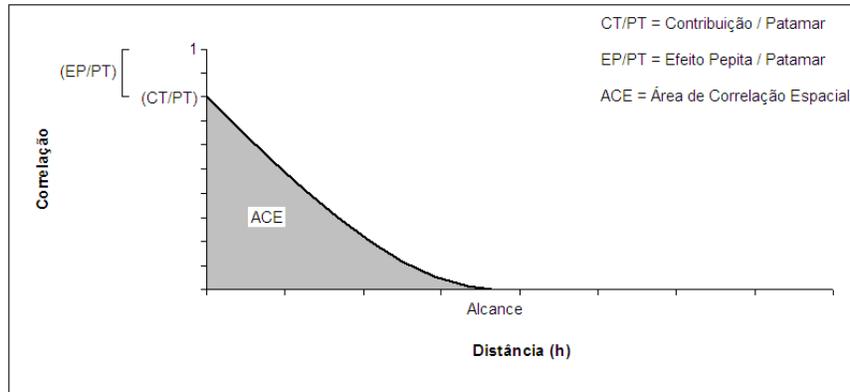


Figura 24 Representação geral de um correlograma, seus parâmetros e a área de correlação espacial correspondente

Percebe-se, com base na figura 24, que o correlograma gera uma área de correlação espacial, que é análoga a área de dependência espacial gerada no semivariograma. A partir da visualização do correlograma, é possível obter a área de correlação espacial (*ACE*), por meio da integração, definida entre zero e a , da expressão dada em (4.21). A *ACE* para o modelo esférico é dada por:

$$\begin{aligned}
 ACE_{ESF} &= \int_0^a \left(1 - \frac{\gamma_{esf}(h)}{C} \right) dh \\
 &= 0,375 \cdot \left(\frac{C_1}{C_0 + C_1} \right) \cdot a
 \end{aligned}
 \tag{40}$$

Como é possível observar na expressão (40), a *ACE*, para o modelo esférico, assume valores no intervalo entre 0 e $0,375 \cdot a$, e é dada na unidade de medida do alcance prático, que, em geral, é dada em metros.

Para facilitar a interpretação desta medida, tornando-a livre de uma unidade de medida, realiza-se uma modificação em sua expressão, introduzindo

o termo $\frac{1}{q.MD}$. Assim, com esta modificação, constrói-se o terceiro índice, dado como:

$$IDE3_{Esf} (\%) = 0,375 \cdot \left(\frac{C_1}{C_0 + C_1} \right) \cdot \left(\frac{a}{q.MD} \right) \cdot 100, \quad (41)$$

em que, C_0 é o efeito pepita, C_1 é a contribuição, a é o alcance prático e $q.MD$ é o valor que corresponde à proporção (q) que se deseja atingir da máxima distância (MD) entre pontos amostrados. Caso a relação $\frac{a}{q.MD}$ resulte em valor maior que 1, esta razão é, então, truncada em 1, para que assumam apenas valores entre zero e 1.

Esse índice ($IDE3$), baseado no correlograma, é mais completo do que os índices $IDE1$ e $IDE2$, no sentido de que já contempla, em sua essência, uma medida de correlação espacial, além de contemplar todos os parâmetros dos modelos da estrutura de dependência espacial (efeito pepita, contribuição, alcance, fator de modelo).

Os cálculos completos da ACE , para os modelos estudados nesta tese, estão apresentados na próxima seção.

4.2.5.2 Cálculo da área ACE

Cálculo da área ACE para o modelo esférico:

$$\begin{aligned}
ACE_{esf} &= \int_0^a \left(1 - \frac{\gamma_{esf}(h)}{C} \right) dh \\
&= \int_0^a \left(\frac{C - \gamma_{esf}(h)}{C} \right) dh \\
&= \frac{1}{C} \int_0^a (C - \gamma_{esf}(h)) dh \\
&= \frac{[0,375.(C_1)].a}{C} \\
&= 0,375 \cdot \left(\frac{C_1}{C_0 + C_1} \right) \cdot a
\end{aligned}$$

Cálculo da área ACE para o modelo exponencial com p=3:

$$\begin{aligned}
ACE_{exp} &= \int_0^a \left(1 - \frac{\gamma_{exp}(h)}{C} \right) dh \\
&= \int_0^a \left(\frac{C - \gamma_{exp}(h)}{C} \right) dh \\
&= \frac{1}{C} \int_0^a (C - \gamma_{exp}(h)) dh \\
&= \frac{[0,317.(C_1)].a}{C} \\
&= 0,317 \cdot \left(\frac{C_1}{C_0 + C_1} \right) \cdot a
\end{aligned}$$

Cálculo da área ACE para o modelo exponencial com $p=9$:

$$\begin{aligned}
 ACE_{\text{exp}} &= \int_0^a \left(1 - \frac{\gamma_{\text{exp}}(h)}{C} \right) dh \\
 &= \int_0^a \left(\frac{C - \gamma_{\text{exp}}(h)}{C} \right) dh \\
 &= \frac{1}{C} \int_0^a (C - \gamma_{\text{exp}}(h)) dh \\
 &= \frac{[0,111.(C_1)]a}{C} \\
 &= 0,111. \left(\frac{C_1}{C_0 + C_1} \right) a
 \end{aligned}$$

Cálculo da área ACE para o modelo gaussiano com $p=3$:

$$\begin{aligned}
 ACE_{\text{gaus}} &= \int_0^a \left(1 - \frac{\gamma_{\text{gaus}}(h)}{C} \right) dh \\
 &= \int_0^a \left(\frac{C - \gamma_{\text{gaus}}(h)}{C} \right) dh \\
 &= \frac{1}{C} \int_0^a (C - \gamma_{\text{gaus}}(h)) dh \\
 &= \frac{[0,504.(C_1)]a}{C} \\
 &= 0,504. \left(\frac{C_1}{C_0 + C_1} \right) a
 \end{aligned}$$

Cálculo da área ACE para o modelo gaussiano com $p=9$:

$$\begin{aligned}
 ACE_{gaus} &= \int_0^a \left(1 - \frac{\gamma_{gaus}(h)}{C} \right) dh \\
 &= \int_0^a \left(\frac{C - \gamma_{gaus}(h)}{C} \right) dh \\
 &= \frac{1}{C} \int_0^a (C - \gamma_{gaus}(h)) dh \\
 &= \frac{[0,295.(C_1)].a}{C} \\
 &= 0,295 \cdot \left(\frac{C_1}{C_0 + C_1} \right) \cdot a
 \end{aligned}$$

4.2.5.3 Generalização do $IDE3$

A partir do resultado obtido para o modelo esférico, pode-se generalizar o índice $IDE3$ para os modelos exponencial e gaussiano. Assim, de forma geral, a expressão do índice $IDE3$ pode ser escrita como:

$$IDE3_{Modelo}(\%) = FM \cdot \left(\frac{C_1}{C_0 + C_1} \right) \cdot \left(\frac{a}{q.MD} \right) \cdot 100 \quad (42)$$

em que, C_0 é o efeito pepita, C_1 é a contribuição, a é o alcance prático e $q.MD$ é o valor que corresponde à proporção (q) que se deseja atingir da máxima distância (MD) entre pontos amostrados.

Esse índice pode assumir valores no intervalo entre 0 e $FM.100(\%)$. O índice apresenta uma vantagem em relação aos outros índices já discutidos nesta tese que é o fato de contemplar em sua expressão tanto a contribuição e o efeito

pepita, quanto o alcance prático. Isso mostra que esse índice contempla mais aspectos do semivariograma se comparado aos índices anteriores.

Em Schabenberger e Gotway (2005) existe uma indicação de duas medidas, denominadas de Escala Integral, que, no R^1 e no R^2 , são dada por, respectivamente:

$$I_1 = \int_0^{\infty} R(h)dh \quad (43)$$

e

$$I_2 = \left(2 \int_0^{\infty} R(h)hdh \right)^{1/2} \quad (44)$$

em que $R(h)$ é a função correlograma e h é a distância entre as observações. Segundo Schabenberger e Gotway (2005), a ideia é que a escala integral considera distâncias em que as observações são altamente correlacionadas, e não considera distâncias em que as observações não são mais correlacionadas.

Processos com alta continuidade espacial obtêm maiores valores de escala integral, pois as correlações enfraquecem mais lentamente conforme as distâncias aumentam (SCHABENBERGER; GOTWAY, 2005).

A expressão para o cálculo da área *ACE* é semelhante á medida apresentada em (43), contudo, não é definida como uma medida de escala integral, mas sim, é construída pelo cálculo da área de correlação do correlograma. Além disso, o índice *IDE3* é valido, tanto para o R^1 , quanto para o R^2 , pois é originado do correlograma. Já a escala integral tem diferentes expressões para cada um dos domínios R^1 ou R^2 .

Outra observação importante é a de que as escalas integrais dadas em (43) e (44) são dadas na escala de medida do alcance, por exemplo, em metros. Já o índice *IDE3* é livre de escalas, o que o torna mais fácil de interpretar.

Porém, mais estudos devem ser realizados para o entendimento das escalas integrais e possíveis comparações com o índice *IDE3*.

Quando se considera o *IDE3* sem o fator $\frac{1}{q.MD}$, se tem uma espécie de valor de alcance corrigido pela força do modelo, dando ideia de um alcance ideal (ou alcance efetivo). Isto é possível, pois os valores de *FM* e $\frac{C_1}{C_0 + C_1}$ podem corrigir o valor do alcance prático. Considerando que o produto $FM \cdot \left(\frac{C_1}{C_0 + C_1} \right)$ dá um valor de força do modelo, tem-se que o alcance corrigido varia de 0 a $FM.a$. Dessa forma, o alcance corrigido é dado na unidade de medida do alcance prático. A expressão do alcance corrigido é dada por:

$$a_{corrigido} = FM \cdot \left(\frac{C_1}{C_0 + C_1} \right) \cdot a \quad (45)$$

A figura 25 mostra uma representação do alcance corrigido. Este alcance corrigido, construído a partir do *IDE3*, poderia ser considerado um quarto índice, mesmo sendo dado na unidade de medida do alcance prático, mas que não será abordado nesta tese.

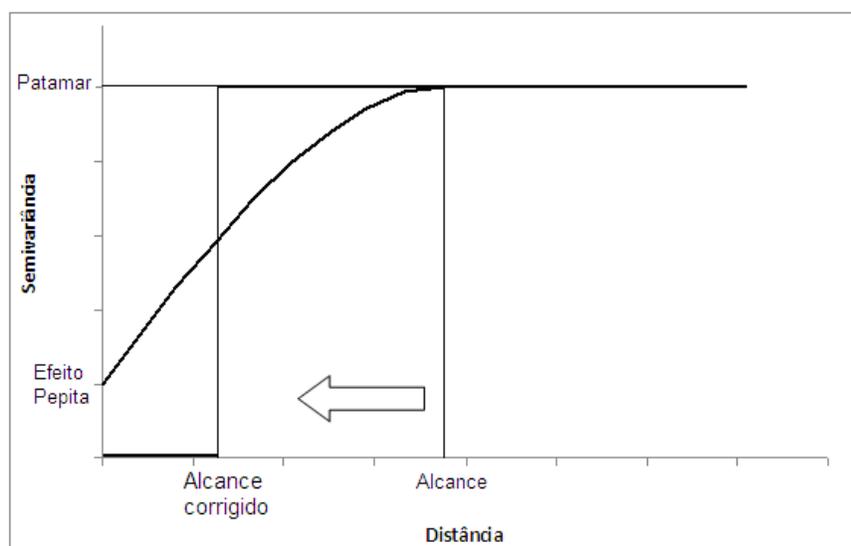


Figura 25 Representação do alcance corrigido

Esse alcance corrigido é o valor do alcance equivalente à transformação do fenômeno atual para seu estado ideal, em que se tem efeito pepita zero e fator de modelo igual a 1.

4.2.6 Aplicação comparativa dos índices *IDE1*, *IDE2* e *IDE3*

Nesta seção, é realizada uma comparação entre os índices propostos nesta tese e o índice dado em Biondi et al. (1994). Para realizar esta comparação são simulados cenários, por meio da função *gwf* do pacote *geoR* (RIBEIRO JÚNIOR; DIGGLE, 2001), que estão apresentados na tabela 4. São cenários que correspondem a dependências espaciais de diferentes graus. As simulações apresentam as seguintes características:

- 1º) média=0;
- 2º) patamar=50;

- 3º) $n=169$;
 4º) *grid* regular de 100x100 m;
 5º) processo estocástico gaussiano
 6º) estimador clássico de semivariograma;

Tabela 4 Cenários propostos para avaliar o desempenho dos índices

Cenário	Parâmetros da dependência espacial			
	<i>EP</i>	<i>CT</i>	<i>AL</i>	<i>DE</i>
1	45	5	10	10%
2	45	5	25	10%
3	45	5	50	10%
4	45	5	75	10%
5	45	5	100	10%
6	37,5	12,5	10	25%
7	37,5	12,5	25	25%
8	37,5	12,5	50	25%
9	37,5	12,5	75	25%
10	37,5	12,5	100	25%
11	25	25	10	50%
12	25	25	25	50%
13	25	25	50	50%
14	25	25	75	50%
15	25	25	100	50%
16	12,5	37,5	10	75%
17	12,5	37,5	25	75%
18	12,5	37,5	50	75%
19	12,5	37,5	75	75%
20	12,5	37,5	100	75%
21	5	45	10	90%
22	5	45	25	90%
23	5	45	50	90%
24	5	45	75	90%
25	5	45	100	90%

Ao simular os 25 cenários, geram-se 25 *IDE1=DE*, 25 *IDE2*, 25 *IDE3* e gera-se, também, 25 erros quadráticos médios dos erros de predição (EQM_e),

uma dessas medidas para cada cenário. A partir disso, é possível calcular a correlação entre cada índice e o EQM_e .

A ideia é de que o melhor índice é aquele que apresenta melhor correlação (mais próxima de -1) com o EQM_e , pois esta última mede o erro das predições (mapas), isto é, mede a qualidade dos mapas. Como uma única correlação é pouco expressiva para possibilitar uma adequada conclusão, a respeito do comportamento dos índices, são realizadas 100 replicações da simulação dos 25 cenários, e, por consequência, 100 correlações entre cada um dos índices e o EQM_e , para que se tenha como observar qual índice apresenta maior frequência de melhores correlações (mais próximas de -1) com o EQM_e .

Pelo fato de que para os cenários com maiores graus de dependência espacial tem-se que; o EQM_e diminui seus valores, e os índices aumentam seus valores, percebe-se que as correlações entre os índices e o EQM_e são negativas. Isto é, menores valores de EQM_e estão relacionados com maiores valores dos índices. Dessa forma, tem-se que as correlações variam de -1 até zero.

São considerados quatro intervalos de correlações para avaliar o desempenho dos índices. Esses intervalos estão apresentados na tabela 5.

Tabela 5 Ocorrência das melhores correlações entre os índices e o EQM_e , considerando diferentes intervalos de correlações, nas 100 replicações

Intervalo de correlações	Nº de ocorrências
Modelo esférico	
(-0,1999 a 0,0000)	1
(-0,3999 a -0,2000)	18
(-0,5999 a -0,4000)	44
(-1,0000 a -0,6000)	37
Total	100
Modelo exponencial (p=3)	
(-0,1999 a 0,0000)	8
(-0,3999 a -0,2000)	42
(-0,5999 a -0,4000)	43
(-1,0000 a -0,6000)	7
Total	100
Modelo exponencial (p=9)	
(-0,1999 a 0,0000)	62
(-0,3999 a -0,2000)	33
(-0,5999 a -0,4000)	5
(-1,0000 a -0,6000)	0
Total	100
Modelo gaussiano (p=3)	
(-0,1999 a 0,0000)	15
(-0,3999 a -0,2000)	47
(-0,5999 a -0,4000)	35
(-1,0000 a -0,6000)	3
Total	100
Modelo gaussiano (p=9)	
(-0,1999 a 0,0000)	23
(-0,3999 a -0,2000)	50
(-0,5999 a -0,4000)	23
(-1,0000 a -0,6000)	4
Total	100

Observa-se, a partir da tabela 5, para o modelo esférico, que a maior frequência de melhores correlações ocorreu entre -1,0000 e -0,4000, com 81% das 100 replicações nesse intervalo. Se fossem consideradas apenas correlações moderadas a fortes, no intervalo de -1,0000 a -0,5000, este percentual atinge 63%. A maioria das correlações se classifica como de moderada a forte.

Ainda, na tabela 5, para o modelo exponencial com correção $p=3$, observa-se maior frequência de melhores correlações entre -0,5999 e -0,2000, com 85% das 100 replicações nesse intervalo. A maioria das correlações é classificada como de moderada a fraca.

Observa-se, também pela tabela 5, para o modelo exponencial com correção $p=9$, que a maior frequência de melhores correlações ocorreu no intervalo entre -0,3999 e 0,0000, com 95% das 100 replicações nesse intervalo, classificando a maioria das correlações como fracas.

Em relação ao modelo gaussiano com correção $p=3$, verifica-se, na tabela 5, que a maior frequência de melhores correlações ocorreu entre -0,5999 e -0,2000, com 82% das 100 replicações nesse intervalo, indicando que a maioria das correlações é classificada como de fraca a moderada.

Verifica-se, na tabela 5, para o modelo gaussiano com correção $p=9$, que a maior frequência de melhores correlações ocorreu entre -0,3999 e -0,2000, com 50% das 100 replicações nesse intervalo. A maioria das correlações é classificada como fraca.

A figura 26 mostra um exemplo de relacionamento existente entre os valores de EQM_e e os cenários propostos para as simulações. À medida que a numeração dada aos cenários (de 1 a 25) aumenta, há a tendência de se aumentar a dependência espacial (ver tabela 4, onde o DE aumenta de 10% a 90% com o aumento dado à numeração dos cenários). Esse relacionamento mostra que, quanto maior o grau de dependência espacial, menor é o erro quadrático médio dos erros de predição (EQM_e). Isso faz com que se verifique que os cenários

simulados são adequados para realizar a avaliação do desempenho dos índices em descrever a dependência espacial, já que os cenários com maior dependência espacial e maior alcance minimizam os valores de EQM_e .

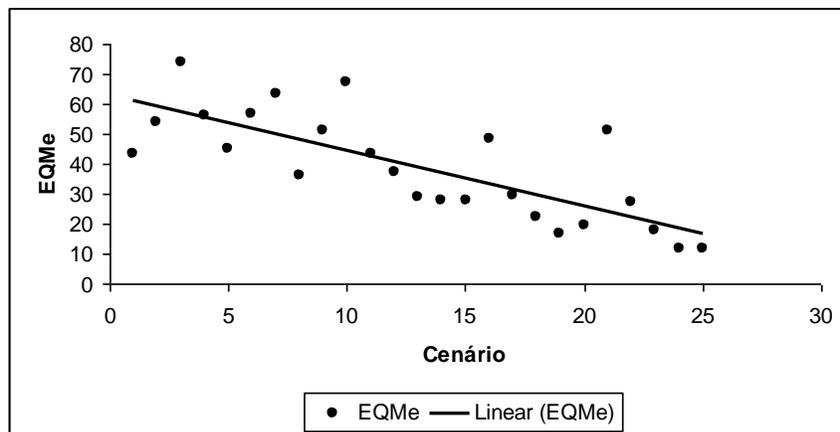


Figura 26 Relacionamento entre os valores de EQM_e e os cenários propostos na simulação, com ajuste de linha de tendência linear, no modelo esférico

A representação dos comportamentos entre os cenários e o EQM_e , observada na figura 26, é praticamente a mesma para todos os modelos considerados nesta tese. Dessa forma, toma-se essa representação, para o modelo esférico, e omitem-se as representações para os modelos exponencial e gaussiano.

As figuras 27, 28, 29, 30 e 31, apresentam a frequência de melhores correlações entre cada índice e o EQM_e , para os modelos esférico, exponencial (p=3), exponencial (p=9), gaussiano (p=3) e gaussiano (p=9), respectivamente.

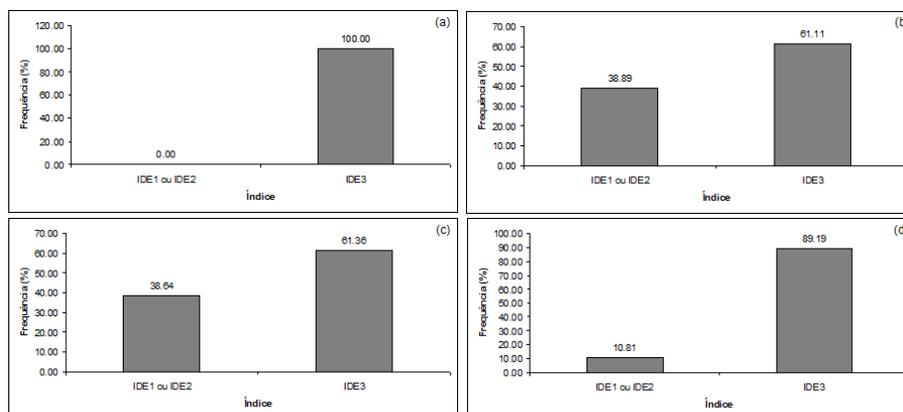


Figura 27 Frequência de melhores correlações entre cada índice e o EQM_e , considerando, o intervalo de correlações entre -0,1999 e 0,0000 (a), entre -0,3999 e -0,2000 (b), entre -0,5999 e -0,4000 (c) e entre -1,0000 e -0,6000 (d), para o modelo esférico

Na figura 27(a), observa-se que o índice IDE3 apresentou a única melhor correlação com o EQM_e , no intervalo de -0,1999 e 0,0000. Já na figura 27(b), verifica-se que o índice IDE3 apresentou maior frequência de melhores correlações (61,11% das 18 replicações), considerando correlações entre -0,3999 e -0,2000. Conforme a figura 27(c), observa-se que, mais uma vez, o índice IDE3 apresentou maior frequência (61,36% das 44 replicações). Por fim, pela figura 27(d), observa-se, também, que a maior frequência foi obtida pelo IDE3, com 89,19% das 37 replicações entre -1,0000 e -0,6000.

Assim, observa-se que o índice *IDE3*, em todas as situações analisadas, apresentou melhor desempenho em relação aos índices *IDE1=DE* e *IDE2*. Ainda, com base na figura 27, percebe-se que os índices *IDE1* e *IDE2* apresentaram mesma correlação com o EQM_e em todas as 100 replicações.

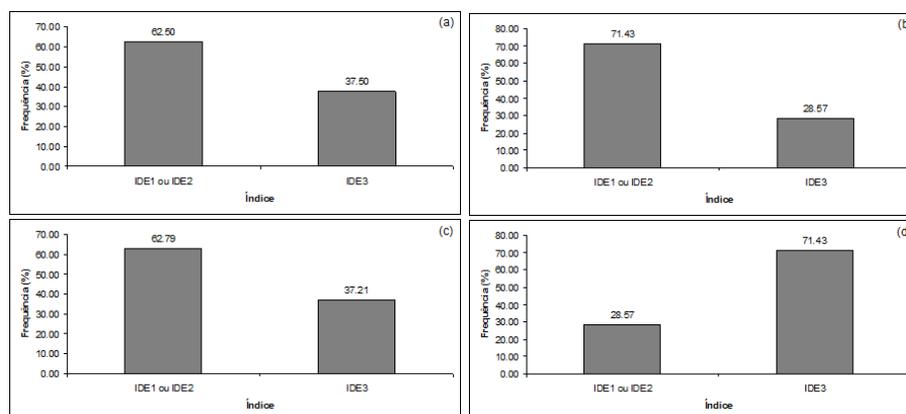


Figura 28 Frequência de melhores correlações entre cada índice e o EQM_e , considerando, o intervalo de correlações entre -0,1999 e 0,0000 (a), entre -0,3999 e -0,2000 (b), entre -0,5999 e -0,4000 (c) e entre -1,0000 e -0,6000 (d), para o modelo exponencial com $p = 3$

Na figura 28(a), observa-se que os índices $IDE1$ e $IDE2$ apresentaram maior frequência de melhores correlações com o EQM_e (62,50% das 8 replicações), no intervalo de -0,1999 e 0,0000. Na figura 28(b), também, os índices $IDE1$ e $IDE2$ apresentaram maior frequência de melhores correlações (71,43% das 42 replicações), considerando correlações entre -0,3999 e -0,2000. Pela figura 28(c), observa-se que, mais uma vez, os dois índices apresentaram maior frequência (62,79% das 43 replicações). Contudo, pela figura 28(d), observa-se que a maior frequência foi obtida pelo $IDE3$, com 71,43% das 7 replicações entre -1,0000 e -0,6000.

Assim, observa-se que os índices $IDE1$ e $IDE2$ apresentaram melhor desempenho, em relação ao índice $IDE3$, apenas para correlações de fracas a moderadas. Enquanto que, para correlações moderadas a fortes, o índice $IDE3$ obteve melhor desempenho. Mais uma vez, com base na figura 28, percebe-se que os índices $IDE1$ e $IDE2$ apresentaram mesma correlação com o EQM_e em todas as 100 replicações.

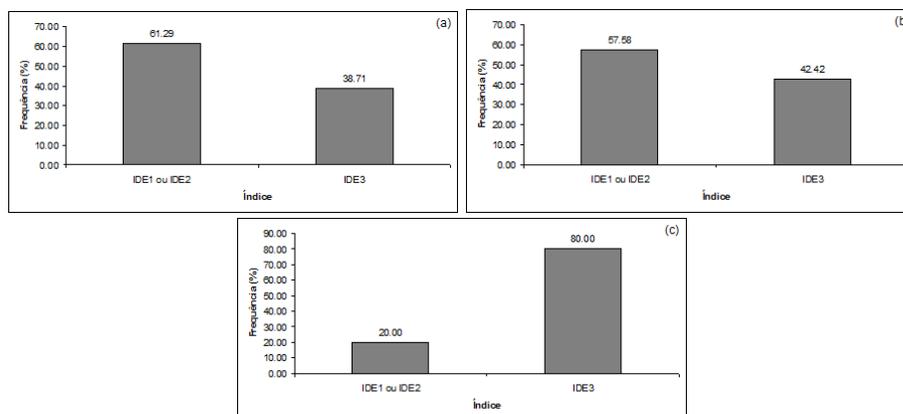


Figura 29 Frequência de melhores correlações entre cada índice e o EQM_e , considerando, o intervalo de correlações entre -0,1999 e 0,0000 (a), entre -0,3999 e -0,2000 (b) e entre -0,5999 e -0,4000 (c), para o modelo exponencial com $p = 9$

Na figura 29(a), observa-se que os índices *IDE1* e *IDE2* apresentaram maior frequência de melhores correlações com o EQM_e (61,29% das 62 replicações), no intervalo de -0,1999 e 0,0000. Na figura 29(b), também, os índices *IDE1* e *IDE2* apresentaram maior frequência de melhores correlações (57,58% das 33 replicações), considerando correlações entre -0,3999 e -0,2000. Contudo, pela figura 29(c), observa-se que a maior frequência foi obtida pelo *IDE3*, com 80,00% das 5 replicações entre -0,5999 e -0,4000.

Assim, observa-se que os índices *IDE1* e *IDE2* apresentaram melhor desempenho, em relação ao índice *IDE3*, apenas para correlações de fracas a moderadas. Enquanto que, para correlações moderadas a fortes, o índice *IDE3* obteve melhor desempenho.

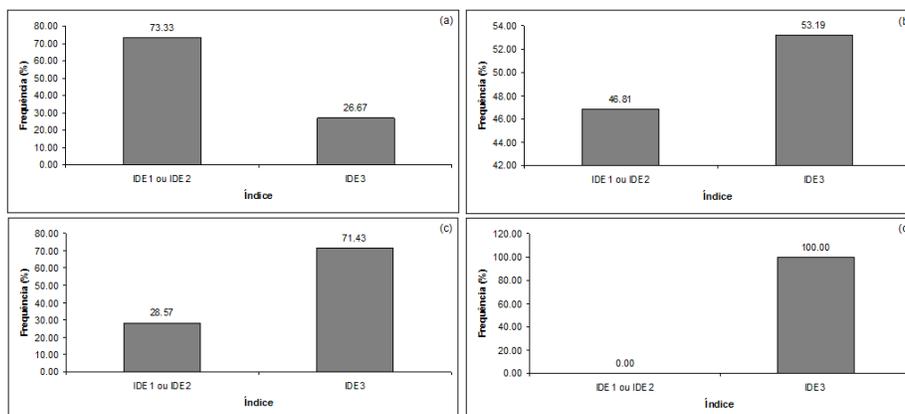


Figura 30 Frequência de melhores correlações entre cada índice e o EQM_e , considerando, o intervalo de correlações entre -0,1999 e 0,0000 (a), entre -0,3999 e -0,2000 (b), entre -0,5999 e -0,4000 (c) e entre -1,0000 e -0,6000 (d), para o modelo gaussiano com $p = 3$

Na figura 30(a), observa-se que os índices $IDE1$ e $IDE2$ apresentam maior frequência de melhores correlações (73,33% das 15 replicações) com o EQM_e , no intervalo de -0,1999 e 0,0000. Já na figura 30(b), verifica-se que o índice $IDE3$ apresentou maior frequência de melhores correlações (53,19% das 47 replicações), considerando correlações entre -0,3999 e -0,2000. Conforme a figura 30(c), observa-se que, mais uma vez, o índice $IDE3$ apresentou maior frequência (71,43% das 35 replicações). Por fim, pela figura 30(d), observa-se, também, que a maior frequência foi obtida pelo $IDE3$, com 100,00% das 3 replicações entre -1,0000 e -0,6000.

Assim, observa-se que o índice $IDE3$, na maioria das situações analisadas, apresentou melhor desempenho em relação aos índices $IDE1=DE$ e $IDE2$.

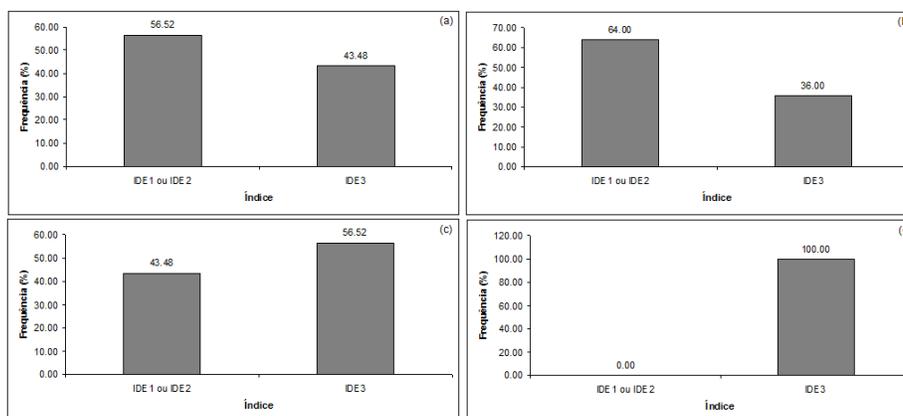


Figura 31 Frequência de melhores correlações entre cada índice e o EQM_e , considerando, o intervalo de correlações entre -0,1999 e 0,0000 (a), entre -0,3999 e -0,2000 (b), entre -0,5999 e -0,4000 (c) e entre -1,0000 e -0,6000 (d), para o modelo gaussiano com $p = 9$

Na figura 31(a), observa-se que os índices $IDE1$ e $IDE2$ apresentam maior frequência de melhores correlações (56,52% das 23 replicações) com o EQM_e , no intervalo de -0,1999 e 0,0000. Na figura 31(b), os índices $IDE1$ e $IDE2$ apresentam maior frequência de melhores correlações (64,00% das 50 replicações), considerando correlações entre -0,3999 e -0,2000. Já na figura 31(c), observa-se que o índice $IDE3$ apresentou maior frequência (56,52% das 23 replicações). Por fim, pela figura 31(d), observa-se, também, que a maior frequência foi obtida pelo $IDE3$, com 100,00% das 4 replicações entre -1,0000 e -0,6000.

Observa-se, então, que os índices $IDE1$ e $IDE2$ foram melhores nas duas primeiras situações, mas o índice $IDE3$ foi melhor nas duas últimas situações. Contudo, o índice $IDE3$ teve melhor desempenho para correlações mais fortes.

Aplica-se o teste de Qui-Quadrado, com a intenção de avaliar a homogeneidade de frequências entre os índices e os intervalos de correlações. A tabela 6 apresenta os resultados da aplicação do teste de Qui-Quadrado.

Tabela 6 Resultados da aplicação do teste de Qui-Quadrado

Faixas de correlação	Índice	
	IDE1 e IDE2	IDE3
Modelo esférico		
(-0,3999 a 0,0000)*	7	12
(-0,5999 a -0,4000)	17	27
(-1,0000 a -0,6000)	4	33
Teste de Qui-Quadrado		p=0,0134
Modelo exponencial (p=3)		
(-0,3999 a 0,0000)*	35	15
(-1,0000 a -0,4000)*	29	21
Teste de Qui-Quadrado		p=0,2112
Modelo exponencial (p=9)		
(-0,3999 a 0,0000)*	38	24
(-1,0000 a -0,4000)*	20	18
Teste de Qui-Quadrado		p=0,3945
Modelo gaussiano (p=3)		
(-0,3999 a 0,0000)*	11	4
(-0,5999 a -0,4000)	22	25
(-1,0000 a -0,6000)	10	28
Teste de Qui-Quadrado		p=0,0060
Modelo gaussiano (p=9)		
(-0,3999 a 0,0000)*	13	10
(-0,5999 a -0,4000)	32	18
(-1,0000 a -0,6000)	10	17
Teste de Qui-Quadrado		p=0,0751

* As faixas de -0,1999 a 0,000 e de -0,3999 a -0,2000 foram unidas. As faixas de -0,5999 a -0,400 e de -1,0000 a -0,6000 foram unidas.

Observa-se, pela tabela 6, que, para o modelo esférico, a relação entre os índices e as faixas de correlação é significativa ($p=0,0134$). Conclui-se que, estatisticamente, o índice *IDE3* possui maior frequência em todas as faixas de correlação.

Para o modelo exponencial com correção $p=3$, na tabela 6, observa-se que a relação entre os índices e as faixas de correlação não é significativa ($p=0,2113$). Estatisticamente, não há um índice que apresente maior frequência na maioria das faixas.

Em relação ao modelo exponencial com correção $p=9$, pela tabela 6, verifica-se que a relação entre os índices e as faixas de correlação não foi significativa ($p=0,3945$). Não houve, estatisticamente, um índice com maior frequência na maioria das faixas.

Ainda, pela tabela 6, para o modelo gaussiano com correção $p=3$, observa-se que a relação entre os índices e as faixas de correlação foi significativa ($p=0,0060$). O índice *IDE3* obteve, estatisticamente, maior frequência na maioria das faixas de correlação.

Já para o modelo gaussiano com correção $p=9$, na tabela 6, observa-se que a relação entre os índices e as faixas de correlação não foi significativa ($p=0,0751$). Estatisticamente, nenhum dos índices obteve maior frequência na maioria das faixas.

De forma geral, considerando todos os modelos simultaneamente, o índice *IDE3* teve melhor desempenho, evidenciando que deve ser preferido em relação aos demais índices. Ou seja, recomenda-se o índice *IDE3* porque ele é mais completo do que os índices de Biondi et al. (1994) ($IDE1=DE$), Cambardella et al. (1994) $\left(RD = \frac{1}{DE}\right)$ e $IDE2 = FM.IDE1 = FM.DE$.

Assim, observa-se a importância de se considerar a proporção da dependência espacial gerada pelo parâmetro alcance, pois foi possível mostrar que o índice que contempla esta dependência dada pelo alcance teve melhor desempenho.

Contudo, a de se frisar que os índices construídos nesta tese valem apenas para os modelos considerados para semivariogramas com patamar. Mais

estudos devem ser realizados para outros modelos e semivariogramas sem patamar.

4.3 Proposta de testes de dependência espacial com base em semivariogramas com patamar

Nesta seção são apresentados dois procedimentos para realização de teste de hipótese de existência ou não de dependência espacial para semivariogramas com patamar, ou seja, para semivariogramas com ajuste de modelos esférico, exponencial ou gaussiano.

A construção dos testes de dependência espacial, na forma que são propostos nesta tese, leva em consideração características dos modelos ajustados ao semivariograma. Assim, o primeiro passo antes da aplicação dos testes, é ajustar o modelo adequado ao semivariograma em estudo.

A estatística de teste a ser utilizada é a estatística *ADE*, já discutida nesta tese. Essa estatística assume seus valores de forma aleatória, caracterizando-se assim, como uma estatística de teste adequada do ponto de vista da inferência estatística.

4.3.1 Procedimento do primeiro teste de dependência espacial (*TDE1*)

O procedimento para o primeiro teste de dependência espacial (*TDE1*) é computacional baseado em simulações. As hipóteses do teste de dependência espacial são as seguintes:

$$\begin{cases} H_0 : \text{Não existe dependência espacial} \\ H_1 : \text{Existe dependência espacial} \end{cases}$$

As simulações utilizadas para o teste são baseadas na hipótese nula, de não existência de dependência espacial. Ou seja, são simulados semivariogramas de efeito pepita puro, caracterizando independência espacial, com base na representação da figura 32.

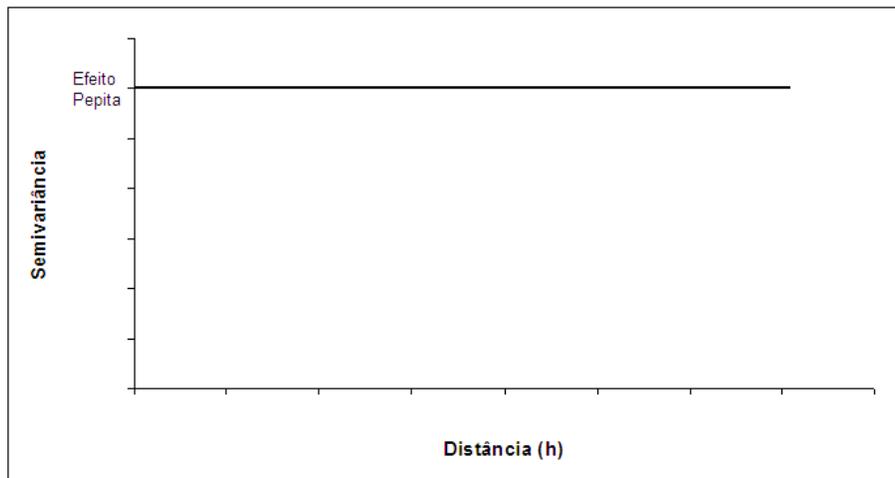


Figura 32 Representação de um semivariograma de efeito pepita puro, caracterizando a não existência de dependência espacial

Para proceder ao teste de dependência espacial, é realizada a seguinte sequência:

- a) Calcular a estatística $ADE_{calculada}$ para o teste de dependência espacial, com base no modelo ajustado ao semivariograma observado nos dados em estudo;
- b) Simular, sob hipótese nula de independência espacial, n semivariogramas com modelo efeito pepita puro;

Nesta tese utiliza-se $n = 99$. As simulações dos semivariogramas com modelo efeito pepita puro são realizadas com a função *grf* do *geoR* (RIBEIRO JÚNIOR; DIGGLE, 2001).

- c) Ajustar o modelo de interesse (esférico, exponencial ou gaussiano) aos n semivariogramas simulados sob hipótese nula;

Para realizar os ajustes, utilizam-se, como valores iniciais, os valores de contribuição, efeito pepita e alcance, gerados pelo ajuste do semivariograma construído a partir dos dados originais em estudo.

Quando o ajuste do modelo ao semivariograma gera valor de alcance maior que o máximo comprimento do vetor de distâncias do semivariograma, atribui-se valor nulo ao alcance. Isto é feito como correção ao ajuste “superestimado”, quando ocorre, aumentando a chance de aceitar a hipótese nula, ou seja, priorizando minimizar o Erro Tipo I.

- d) Calcular o *ADE* em cada uma das n simulações;
- e) Construir a distribuição da estatística *ADE* (valor do $ADE_{calculada}$ + os n valores de *ADE* das simulações);
- f) Calcular um valor p com base na distribuição da estatística *ADE*. O valor p é dado pela proporção de valores de *ADE* que são maiores ou iguais ao $ADE_{calculada}$.

Um esquema com o procedimento para a realização do teste *TDEI* pode ser observado na figura 33.

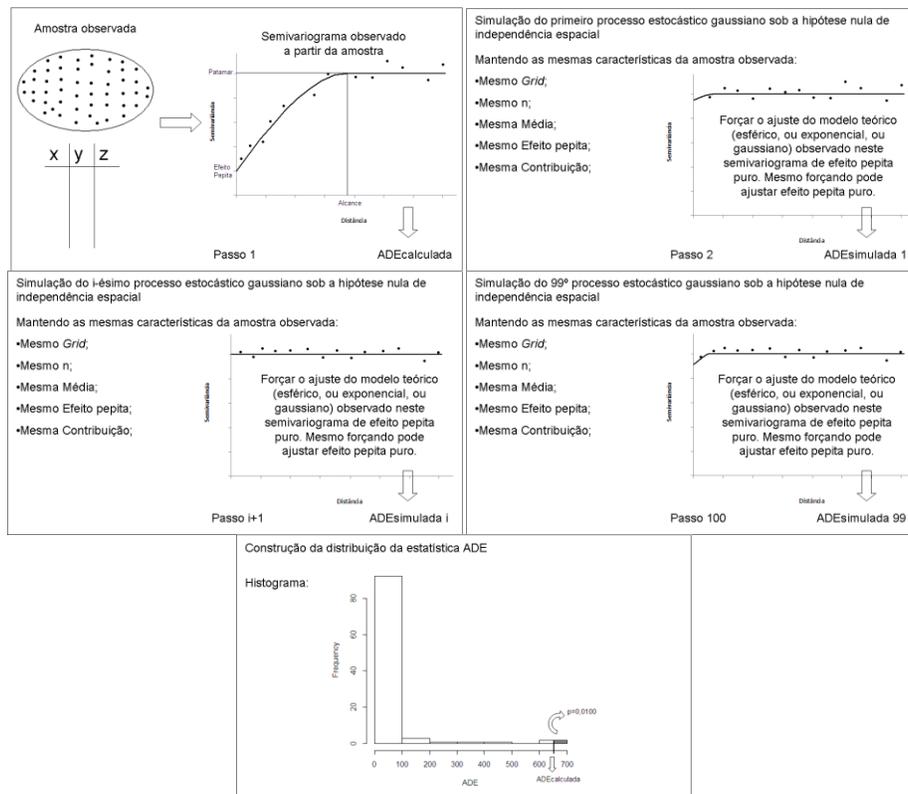


Figura 33 Esquema representativo do procedimento do teste *TDE1*

Seguindo o esquema dado na figura 33, realiza-se o teste e se pode concluir sobre aceitar ou não a hipótese nula (H_0), ou seja, concluir sobre a existência ou não de dependência espacial.

4.3.2 Procedimento do segundo teste de dependência espacial (*TDE2*)

O procedimento para o segundo teste de dependência espacial (*TDE2*) é computacional baseado em permutações. As hipóteses do teste de dependência espacial são as seguintes:

$$\begin{cases} H_0 : \text{N\~{a}o existe depend\~{e}ncia espacial} \\ H_1 : \text{Existe depend\~{e}ncia espacial} \end{cases}$$

Este teste utiliza o princ\xedpio da permuta\~{c}\~{a}o dos valores do atributo nas posi\~{c}\~{o}es geogr\u00e1ficas. Como ferramenta para permutar os dados observados, usa-se a fun\~{c}\~{a}o *sample*, do pacote *base* do *software* R (R DEVELOPMENT CORE TEAM, 2012). Esta fun\~{c}\~{a}o gera permuta\~{c}\~{o}es ou aleatoriza\~{c}\~{o}es a partir de uma amostra. Assim, \u00e9 poss\xedvel realizar simula\~{c}\~{o}es de processos estoc\u00e1sticos quase independentes (com correla\~{c}\~{a}o espacial fraca, ou seja, com baixa depend\~{e}ncia espacial).

Para proceder ao teste de depend\~{e}ncia espacial, \u00e9 realizada a seguinte sequ\~{e}ncia:

- a) Calcular a estat\xedstica $ADE_{calculada}$ para o teste de depend\~{e}ncia espacial, com base no modelo ajustado ao semivariograma observado nos dados em estudo;
- b) Permutar, n vezes, os valores do atributo nas coordenadas geogr\u00e1ficas, gerando, dessa forma, n bases de dados com fraca depend\~{e}ncia espacial;

Nesta tese utiliza-se $n = 99$. A partir das permuta\~{c}\~{o}es n\u00e3o se quebra a depend\~{e}ncia espacial efetivamente, mas se enfraquece essa depend\~{e}ncia de modo que a hip\u00f3tese nula de independ\~{e}ncia espacial \u00e9 quase atingida. Assim, o fato de que a depend\~{e}ncia dos dados \u00e9 enfraquecida ao se permutar os dados viabiliza a realiza\~{c}\~{a}o do teste.

- c) Construir n semivariogramas, um para cada base de dados gerada;

Os semivariogramas respeitam a hipótese nula de independência espacial, pois são construídos a partir de bases de dados geradas pelo processo de permutação.

- d) Ajustar o modelo de interesse (esférico, exponencial ou gaussiano) aos n semivariogramas;

Para realizar os ajustes, utilizam-se, como valores iniciais, os valores de contribuição, efeito pepita e alcance, gerados pelo ajuste do semivariograma construído a partir dos dados originais em estudo.

Quando o ajuste do modelo ao semivariograma gera valor de alcance maior que o máximo comprimento do vetor de distâncias do semivariograma, atribui-se valor nulo ao alcance. Isto é feito como correção ao ajuste “superestimado”, quando ocorre, aumentando a chance de aceitar a hipótese nula, ou seja, priorizando minimizar o Erro Tipo I.

- e) Calcular o ADE em cada um dos n semivariogramas;
- f) Construir a distribuição da estatística ADE (valor do $ADE_{calculada}$ + os n valores de ADE gerados dos semivariogramas construídos com base na permutação);
- g) Calcular um valor p com base na distribuição da estatística ADE . O valor p é dado pela proporção de valores de ADE que são maiores ou iguais ao $ADE_{calculada}$.

A figura 34 apresenta um esquema com o procedimento para realização do teste TDE2.

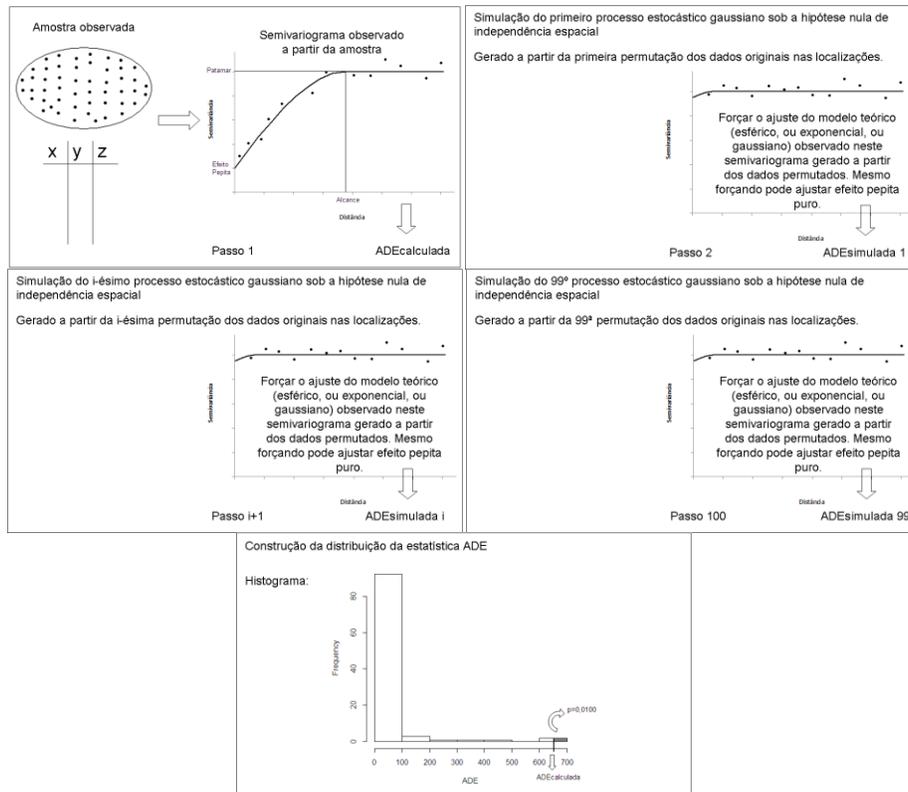


Figura 34 Esquema com o procedimento do teste TDE2

Seguindo o procedimento exposto na figura 34, realiza-se o teste e se pode concluir sobre aceitar ou não a hipótese nula (H_0), ou seja, concluir sobre a existência ou não de dependência espacial.

4.3.3 Nível nominal dos testes *TDE1* e *TDE2*

Para estudar o nível de significância nominal dos testes é simulado um cenário, sob modelo de efeito pepita puro, de modo que a hipótese nula (H_0),

de independência espacial, é verdadeira. O cenário simulado, com o pacote *RandomFields* (SCHLATHER, 2001), apresenta as seguintes características:

- 1º) média=0;
- 2º) alcance=0;
- 3º) efeito pepita=50;
- 4º) contribuição=0;
- 5º) $n=169$;
- 6º) *grid* regular de 100x100 m;
- 7º) processo estocástico gaussiano
- 8º) estimador clássico de semivariograma;

A figura 35 apresenta a representação do *grid* na simulação.

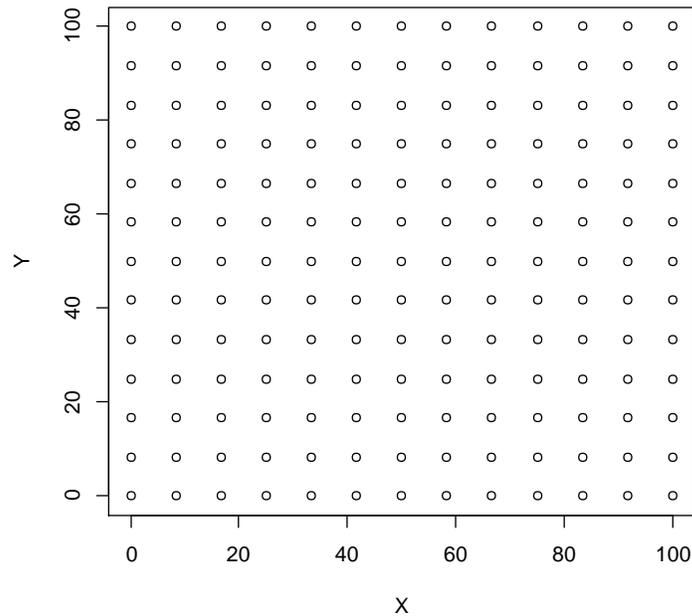


Figura 35 Representação do *grid* na simulação

Realizam-se os dois testes 100 vezes (cada teste é aplicado repetidamente 100 vezes), nesse cenário, para estudar o nível nominal dos testes. Em cada uma das 100 vezes, o valor p é calculado, e o percentual de vezes em que a hipótese nula, de independência espacial, é corretamente rejeitada define o nível nominal dos testes. O nível de significância adotado é de 5% (nível nominal), ou seja, só é considerada rejeitada a hipótese nula, se o valor p é menor que 0,05.

Após a aplicação de ambos os testes, 100 vezes, observa-se que a hipótese nula foi aceita em todas as aplicações, ou seja, 100% das vezes o valor p foi maior ou igual a 5%. Conseqüentemente, 0% das vezes rejeitou-se H_0 . O que faz com que o Erro Tipo I seja igual a zero.

Assim, obteve-se um nível real de 0% frente a um nível nominal de 5%, indicando um ótimo resultado, pois o nível real foi menor que o nominal. Esse fato de obter nível real menor que o nominal pode ser devido ao baixo número de repetições do teste (100 vezes), ou ser o verdadeiro comportamento dos dois testes.

4.3.4 Poder dos testes *TDE1* e *TDE2*

Para realizar o estudo do poder dos testes, são simulados 15 cenários de dependência espacial para cada um dos modelos teóricos esférico, exponencial e gaussiano. Nestes cenários, diferentes valores de contribuição, efeito pepita e alcance são atribuídos. Estes cenários são considerados, pois são construídos sob a hipótese alternativa (de existência de dependência espacial), isto é, são cenários em que a hipótese alternativa (H_1) é verdadeira. A tabela 7 mostra os cenários a serem simulados.

Tabela 7 Cenários simulados para avaliar o poder dos testes

Cenário	Parâmetros da dependência espacial		
	<i>EP</i>	<i>CT</i>	<i>AL</i>
1	45	5	25
2	45	5	50
3	45	5	75
4	37,5	12,5	25
5	37,5	12,5	50
6	37,5	12,5	75
7	25	25	25
8	25	25	50
9	25	25	75
10	12,5	37,5	25
11	12,5	37,5	50
12	12,5	37,5	75
13	5	45	25
14	5	45	50
15	5	45	75

Com base na tabela 7, realiza-se uma simulação de cada cenário para possibilitar o estudo do poder dos testes. Em cada um desses cenários é calculado o índice *IDE3* para que se possa avaliar o poder dos testes para diferentes graus de dependência espacial. O *IDE3* foi escolhido, pois foi o índice de melhor desempenho na descrição do grau de dependência espacial, já estudado anteriormente.

Os cenários simulados, com o pacote *RandomFields* (SCHLATHER, 2001), apresentam as seguintes características:

- 1º) média=0;
- 2º) patamar=50;
- 3º) $n=169$;
- 4º) *grid* regular de 100x100 m;
- 5º) processo estocástico gaussiano
- 6º) estimador clássico de semivariograma;

A tabela 8 apresenta os valores do *IDE3* para cada modelo considerado, com base nos 15 cenários simulados.

Tabela 8 Valores do *IDE3* para os modelos esférico, exponencial e gaussiano

Cenário	Esf	Exp (p=3)	Exp (p=9)	Gaus (p=3)	Gaus (p=9)
1	1%	2%	0%	1%	1%
2	2%	2%	1%	6%	1%
3	4%	2%	1%	3%	3%
4	2%	2%	1%	5%	2%
5	6%	7%	2%	6%	4%
6	10%	11%	2%	11%	5%
7	7%	7%	1%	8%	2%
8	13%	11%	3%	16%	6%
9	15%	15%	5%	25%	16%
10	7%	9%	3%	12%	3%
11	18%	14%	3%	30%	13%
12	31%	23%	6%	37%	19%
13	9%	11%	2%	17%	4%
14	19%	20%	6%	35%	12%
15	32%	31%	9%	42%	21%

Em cada um dos cenários simulados, apresentados na tabela 7, realizam-se os testes 100 vezes (cada teste é aplicado repetidamente 100 vezes). Isto é feito para avaliar o poder dos dois testes. Em cada uma das 100 vezes, o valor p é calculado, e o percentual de vezes em que a hipótese nula, de independência espacial, é corretamente rejeitada define o poder de cada teste. O nível de significância adotado é de 5% (nível nominal), ou seja, só é considerada rejeitada a hipótese nula, se o valor p é menor que 0,05.

As figuras 36, 37, 38, 39 e 40 apresentam o resultado do estudo do poder dos testes em relação ao índice *IDE3*.

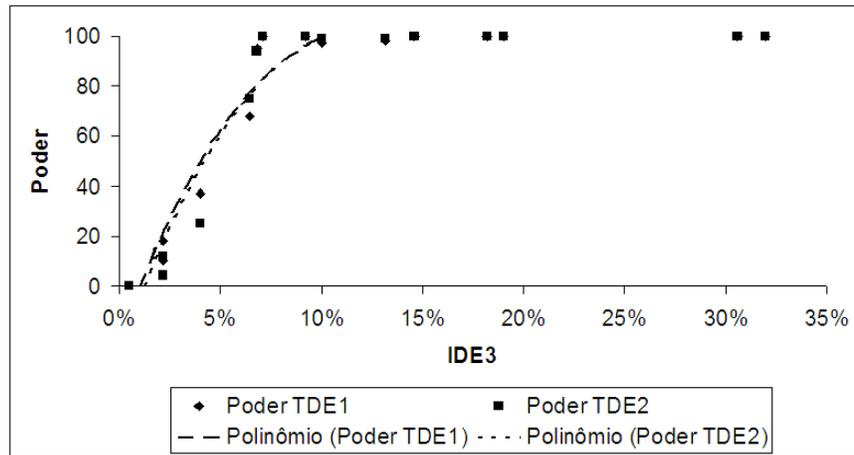


Figura 36 Poder dos testes *TDE1* e *TDE2*, considerando o modelo esférico

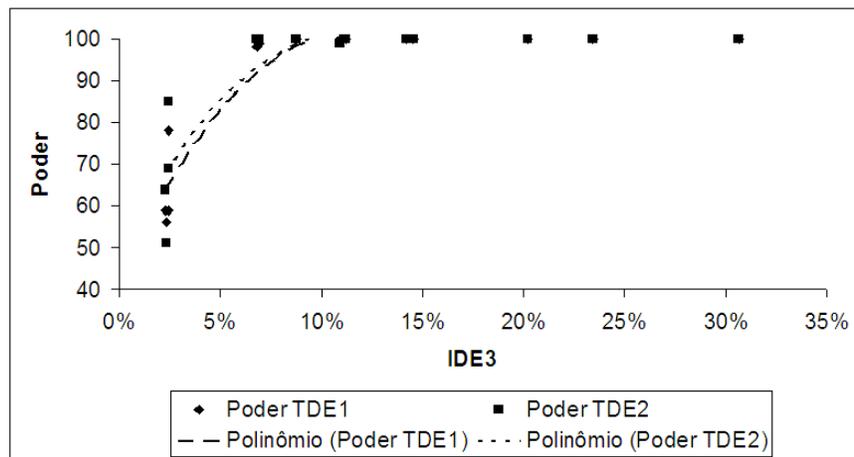


Figura 37 Poder dos testes *TDE1* e *TDE2*, considerando o modelo exponencial ($p = 3$)

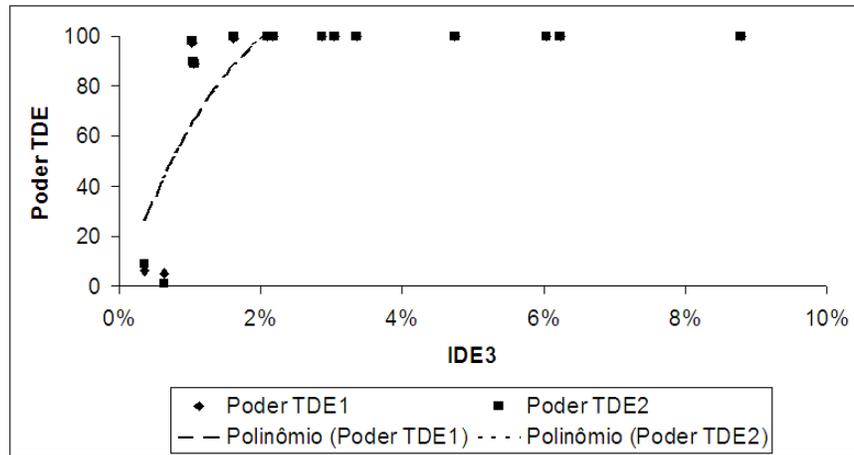


Figura 38 Poder dos testes *TDE1* e *TDE2*, considerando o modelo exponencial ($p = 9$)

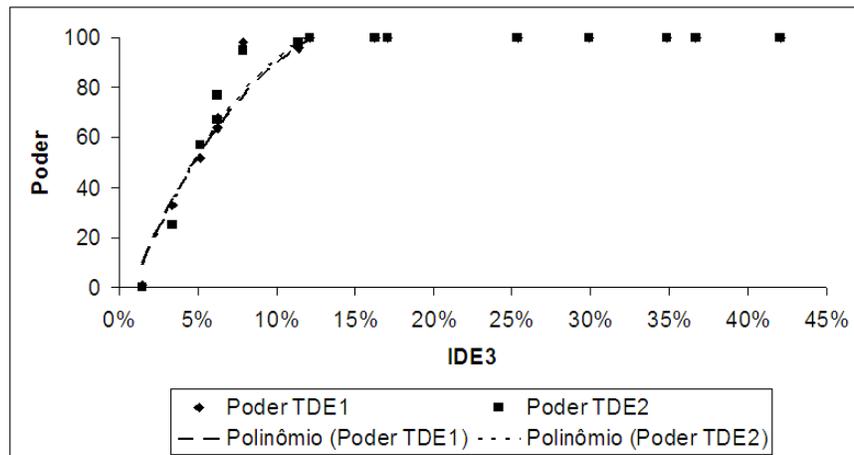


Figura 39 Poder dos testes *TDE1* e *TDE2*, considerando o modelo gaussiano, com ($p = 3$)

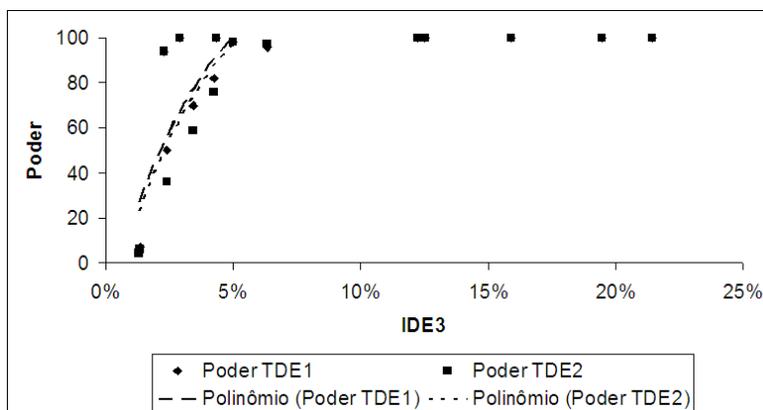


Figura 40 Poder do teste *TDE1* e *TDE2*, considerando o modelo gaussiano, com ($p = 9$)

Com base nas figuras 36, 37, 38, 39 e 40, pode-se observar a curva que representa o comportamento do poder dos testes *TDE1* e *TDE2*. O comportamento do poder dos testes foi ajustado, em todos os modelos, por um modelo polinomial de 3º grau. A curva do polinômio é traçada até atingir o valor 100 no eixo do poder, desconsiderando o restante da curva. O poder cresce rapidamente para menores valores de *IDE3* e se aproxima de 100% para maiores valores de *IDE3*. O ajuste das curvas polinomiais é, nesta tese, apenas para ilustrar o comportamento do poder dos testes em relação aos valores do *IDE3*, não tendo pretensão de prever o poder a partir do *IDE3*. Porém, é uma possibilidade de estudos futuros a utilização dessa relação entre poder e *IDE3* para, por exemplo, a determinação de valores de *IDE3* que classificam o grau de dependência espacial.

O que fica claro é que para os cenários com baixa dependência espacial o poder dos testes é menor, e para aqueles cenários com maior grau de dependência, o poder dos testes é maior. Além disso, as curvas de poder de ambos os testes são muito semelhantes, quase coincidentes, indicando que os dois testes parecem gerar resultados semelhantes, parecendo serem indistintos.

Uma observação deve ser feita: os testes propostos aqui não têm a intenção de substituir o uso dos envelopes simulados já existentes, mas sim, vem a complementar a análise puramente visual feita hoje, a exemplo do que o teste de Shapiro Wilk vem fazer em relação ao histograma ou ao *qqplot*, na avaliação da normalidade de dados.

Ou seja, os testes propostos aqui vêm a ser um complemento à avaliação da dependência espacial. Sendo que podem ser utilizados conjuntamente com os envelopes simulados desenvolvidos, por Ribeiro Júnior e Diggle (2001), no pacote *geoR*.

Nesta tese, a comparação dos testes propostos parece indicar uma grande semelhança entre eles, no que tange aos erros Tipo I e Tipo II, tornando-os igualmente recomendáveis, a princípio. Porém, há a relevância em se considerar outros critérios de desempenho, além dos erros, para a decisão de recomendá-los. Tais outros critérios podem incluir, entre outros:

- a) Custo de computação;
- b) Tipo de procedimento computacional adotado: simulações Monte Carlo podem ser mais confiáveis do que permutações, pois enquanto na simulação Monte Carlo a amostragem é do tipo amostragem aleatória simples, na permutação tem-se amostragem aleatória estratificada, com o risco de se estar amostrando num estrato de comportamento anômalo.

Enfim, pelo fato de os dois testes serem inéditos, e não ser o objetivo desta tese a testagem exaustiva destes novos testes (mas principalmente sua apresentação, descrição, e ilustração de uso), necessita-se realizar mais estudos, incluindo mais cenários, mais simulações e maior investigação das propriedades dos testes.

5 CONCLUSÃO

Retomando os objetivos específicos lançados para esta tese, pode-se concluir por:

1º) Propor uma correção nos modelos exponencial e gaussiano de semivariograma;

Foi possível realizar uma correção nos modelos de semivariogramas exponencial e gaussiano a partir de um novo conceito sobre o percentual de explicação da contribuição. Mostrou-se que a partir deste novo conceito é possível construir diferentes expressões matemáticas para esses dois modelos, criando, dessa forma, uma nova classe de modelos exponenciais e gaussianos.

Para trabalhos futuros sugerem-se mais estudos sobre essa nova classe de modelos para sua possível implantação computacional.

2º) Construir novos índices para a descrição da dependência espacial;

Foram formulados três índices para descrição da dependência espacial de dados geoestatísticos. Os três índices foram construídos com base no conceito de áreas de dependência espacial.

O primeiro índice, denominado de *IDE1*, foi formulado a partir da comparação entre a área de dependência observada e a área de dependência possível de ser atingida. Após a construção deste índice, observou-se que é idêntico ao índice encontrado em Biondi et al. (1994). Ou seja, encontrou-se, a partir da geometria do semivariograma, a demonstração matemática do índice dado em Biondi et al. (1994).

O segundo índice, denominado de *IDE2*, tomou por base a comparação entre a área de dependência espacial e a área total do semivariograma.

A construção do terceiro índice, denominado de IDE3, foi baseada no conceito de área de correlação espacial dada pelo correlograma. Ou seja, o índice *IDE3* é construído a partir da área de correlação espacial.

A partir do índice *IDE3* foi possível definir um conceito de alcance corrigido, que caracteriza, com base na força do modelo, o valor de um alcance ideal, ou efetivo, do modelo.

Ao realizar uma aplicação comparativa dos três índices, observou-se que, de forma geral, o índice *IDE3* apresentou melhor desempenho. Deste modo, recomenda-se seu uso para realizar a descrição da dependência espacial de dados geoestatísticos.

Para futuros estudos sugere-se a construção de índices para outros modelos de semivariogramas, como os que não apresentam o parâmetro patamar.

3º) Construir testes de significância para a hipótese nula de ausência de dependência espacial;

Construíram-se dois testes de hipótese, independentes entre si, para avaliar a existência ou não de dependência espacial. A hipótese nula a ser testada diz que não existe dependência espacial nos dados, ou seja, fala que os dados possuem independência espacial.

A partir do conceito de áreas de dependência espacial, gerados nesta tese, construiu-se uma nova estatística de dependência espacial. Esta estatística, utilizada para o teste de hipótese de dependência espacial, denominada de *ADE*, é a própria área de dependência espacial calculada por integração. Ou seja, esta estatística é formulada a partir da geometria do semivariograma amostral.

O primeiro teste construído, denominado de *TDEI*, utiliza o princípio de simulação, sob hipótese nula de independência espacial, gerado a partir de dados sob comportamento de efeito pepita puro (modelo de efeito pepita puro).

O segundo teste construído, nesta tese, denominado de *TDE2*, se baseia no princípio de simulação, sob hipótese nula de independência espacial, gerado partindo-se de dados permutados no *grid* de amostragem.

Os dois testes apresentaram resultados bem semelhantes em relação aos erros Tipo I e Tipo II. Dessa foram, sugere-se a utilização de qualquer um deles.

Contudo, para futuros estudos necessita-se ampliar o número de cenários e o número de simulações para uma melhor avaliação dos testes. Além disso, o teste *TDE2*, baseado em permutações, precisa de maior investigação já que a permutação dos dados não quebra a dependência espacial dos dados, mas sim, apenas enfraquece esta dependência.

Por fim, avaliando de forma geral o desenvolvimento desta tese, conclui-se que foi alcançado o objetivo principal de propor métodos para avaliar, por meio de testes de hipótese, e descrever, por meio de índices, a estrutura de dependência espacial de dados geoestatísticos, ou seja, espera-se que os métodos desenvolvidos nesta tese contribuam para que os usuários da Geoestatística possam realizar inferências mais formais e descrições mais adequadas sobre a estrutura de dependência espacial.

REFERÊNCIAS

ABREU, S. L. et al. Variabilidade espacial de propriedades físico-hídricas do solo, da produtividade e da qualidade de grãos de trigo em Argissolo Franco Arenoso sob plantio direto. **Ciência Rural**, Santa Maria, v. 33, n. 2, p. 275-282, mar./abr. 2003.

ALVES, M. C. et al. Geostatistical analysis of the spatial variation of the Berry borer and leaf miner in a coffee agroecosystem. **Precision Agriculture**, Dordrecht, v. 12, n. 1, p. 18-31, 2011.

_____. Variabilidade espacial de variáveis geobiofísicas nas nascentes da bacia hidrográfica do Ribeirão Santa Cruz. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 12, n. 5, p. 527-535, 2008.

ARMINDO, R. A. et al. Spatial variability of leaf nutrient contents in a drip irrigated citrus orchard. **Engenharia Agrícola**, Jaboticabal, v. 32, n. 3, p. 479-489, 2012.

BABAK, O.; DEUTSCH, C. V. Statistical approach to inverse distance interpolation. **Stochastic Environmental Research and Risk Assessment**, New York, v. 23, n. 5, p. 543-553, July 2009.

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

BOGNOLA, I. A. et al. Modelagem uni e bivariada da variabilidade espacial de rendimento de *Pinus taeda* L. **Floresta**, Curitiba, v. 38, n. 2, p. 373-385, 2008.

CAMARGO, L. A.; MARQUES JÚNIOR, J.; PEREIRA, G. T. Spatial variability of physical attributes of an alfisol under different hillslope curvatures. **Revista Brasileira de Ciência do Solo**, Viçosa, MG, v. 34, n. 3, p. 617-630, maio/jun. 2010.

CAMBARDELLA, C. A. et al. Field-scale variability of soil properties in Central Iowa soils. **Soil Science Society America Journal**, Madison, v. 58, n. 5, p. 1501-1511, Sept./Oct. 1994.

CARVALHO, J. R. P.; SILVEIRA, P. M.; VIEIRA, S. R. Geoestatística na determinação da variabilidade espacial de características químicas do solo sob diferentes preparos. **Pesquisa Agropecuária Brasileira**, Brasília, v. 37, n. 8, p. 1151-1159, ago. 2002.

CEDDIA, M. B. et al. Topography and spatial variability of soil physical properties. **Scientia Agricola**, Piracicaba, v. 66, n. 3, p. 338-352, 2009.

CHIBA, M. K.; GUEDES FILHO, O.; VIEIRA, S. R. Variabilidade espacial e temporal de plantas daninhas em Latossolo Vermelho argiloso sob semeadura direta. **Acta Scientiarum. Agronomy**, Maringá, v. 32, n. 4, p. 735-742, 2010.

CHILÈS, J. P.; DELFINER, P. **Geostatistics: modeling spatial uncertainty**. New York: J. Wiley, 2006. 695 p.

CORSTANJE, R.; GRUNWALD, S.; LARK, R. M. Inferences from fluctuations in the local variogram about the assumption of stationarity in the variance. **Geoderma**, Amsterdam, v. 143, n. 1/2, p. 123-132, Jan. 2008.

CRESSIE, N. A.; HAWKINS, D. M. Robust estimation of the variogram: I. **Mathematical Geology**, New York, v. 12, n. 2, p. 115-125, 1980.

FORTES, B. P. M. D. et al. Modelagem geoestatística da infecção por *Ascaris lumbricóides*. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 20, n. 3, p. 727-734, 2004.

GALLARDO, A.; PARAMÁ, R. Spatial variability of soil elements in two plant communities of NW Spain. **Geoderma**, Amsterdam, v. 139, n. 1/2, p. 199-208, Jan. 2007.

ISAAKS, E. H.; SRIVASTAVA, R. M. **Applied geostatistics**. New York: Oxford University, 1989. 561 p.

JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining geostatistics**. Caldwell: Blackburn, 2003. 600 p.

KANEGAE JÚNIOR, H. et al. Avaliação da continuidade espacial de características dendométricas em diferentes idades de povoamentos clonais de *Eucalyptus* sp. **Revista Árvore**, Viçosa, MG, v. 31, n. 5, p. 859-866, set./out. 2007.

KONOPATZKI, M. R. S. et al. Spatial variability of yield and other parameters associated with pear trees. **Engenharia Agrícola**, Piracicaba, v. 32, n. 2, p. 381-392, 2012.

LARK, R. M. Kriging a soil variable with a simple nonstationary variance model. **Journal of Agricultural, Biological and Environmental Statistics**, Alexandria, v. 14, n. 3, p. 301-321, 2009.

LIMA, J. S. S. et al. Métodos geoestatísticos no estudo da resistência do solo à penetração em trilha de tráfego de tratores na colheita de madeira. **Revista Árvore**, Viçosa, MG, v. 32, n. 5, p. 931-938, set./out. 2008.

LOUZADA, J. M. **Teste de normalidade multivariada em geoestatística utilizando bootstrap**. 2011. 167 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2011.

MATHERON, G. Principles of geostatistics. **Economic Geology**, Lancaster, v. 58, p. 1246-1266, 1963.

MEDRONHO, R. A. et al. Análise espacial da soroprevalência da hepatite A em crianças de uma região carente de Duque de Caxias, RJ, Brasil. **Revista Brasileira de Epidemiologia**, São Paulo, v. 6, n. 4, p. 328-334, 2003.

MELLO, J. M. et al. Estudo da dependência espacial de características dendrométricas para *Eucalyptus grandis*. **Cerne**, Lavras, v. 11, n. 2, p. 113-126, 2005.

_____. Métodos de amostragem e geoestatística para estimativa do número de fustes e volume em plantios de *Eucalyptus grandis*. **Floresta**, Curitiba, v. 39, n. 1, p. 157-166, 2009.

_____. Uso do estimador geoestatístico para predição volumétrica por talhão. **Floresta**, Curitiba, v. 36, n. 2, p. 251-260, 2006.

MIGUEL, F. R. M.; VIEIRA, S. R.; GREGO, C. R. Variabilidade espacial da infiltração de água em solo sob pastagem em função da intensidade de pisoteio. **Pesquisa Agropecuária Brasileira**, Brasília, v. 44, n. 11, p. 1513-1519, nov. 2009.

MINGOTI, S. A.; LEITE, A. G.; ROSA, G. Describing the total number of diagnosed cases of AIDS by means of geostatistics. **Revista de Matemática e Estatística**, Marília, v. 24, n. 1, p. 61-76, 2006.

MINGOTI, S. A.; ROSA, G. A note on robust and non-robust variogram estimators. **REM: Revista Escola de Minas**, Ouro Preto, v. 61, n. 1, p. 87-95, 2008.

NOVAES FILHO, J. P. et al. Distribuição espacial de carbono em solo sob floresta primária na Amazônia Meridional. **Revista Árvore**, Viçosa, MG, v. 31, n. 1, p. 83-92, jan./fev. 2007.

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

OLIVEIRA, M. C. N. et al. Estimação dos parâmetros de uma função de covariância no estudo da variabilidade espacial do teor de cálcio no solo. **Revista de Matemática e Estatística**, Marília, v. 24, n. 4, p. 117-132, 2006.

OPROMOLLA, P. A.; DALBEN, I.; CARDIM, M. Análise da distribuição espacial da hanseníase no estado de São Paulo, 1991-2002. **Revista Brasileira de Epidemiologia**, São Paulo, v. 8, n. 4, p. 356-364, 2005.

PAZ-GONZÁLEZ, A. et al. The effect of cultivation on the spatial variability of selected properties of an umbric horizon. **Geoderma**, Amsterdam, v. 97, n. 3/4, p. 273-292, Sept. 2000.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2012. Software.

RESENDE, M. D. V. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Colombo: EMBRAPA Florestas, 2007. 362 p.

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.

RODRIGUES, M. S.; CORÁ, J. E.; FERNANDES, C. Soil sampling intensity and spatial distribution pattern of soils attributes and corn yield in no-tillage system. **Engenharia Agrícola**, Jaboticabal, v. 32, n. 5, p. 852-865, 2012.

ROSALEN, D. L. et al. GPS receivers for georeferencing of spatial variability of soil attributes. **Engenharia Agrícola**, Jaboticabal, v. 31, n. 6, p. 1162-1169, 2011.

SAMPAIO, F. M. T. et al. Variabilidade espacial da umidade de um latossolo vermelho-amarelo sob plantio direto. **Engenharia Agrícola**, Jaboticabal, v. 30, n. 5, p. 854-861, 2010.

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

SCHLATHER, M. Simulation and analysis of random fields. **R News**, New York, v. 1, n. 2, p. 18-20, 2001.

SIEGEL, S.; CASTELLAN JÚNIOR, N. J. **Estatística não-paramétrica para ciências do comportamento**. 2. ed. Porto Alegre: Artmed, 2006. 448 p.

SILVA, F. M. et al. Efeitos da colheita manual na bienalidade do cafeeiro em Ijaci, Minas Gerais. **Ciência e Agrotecnologia**, Lavras, v. 34, n. 3, p. 625-632, maio/jun. 2010.

_____. Variabilidade espacial de atributos químicos e de produtividade na cultura do café. **Ciência Rural**, Santa Maria, v. 37, n. 2, p. 401-407, mar./abr. 2007.

SIQUEIRA, D. S.; MARQUES JÚNIOR, J.; PEREIRA, G. T. The use of landforms to predict the variability of soil and orange attributes. **Geoderma**, Amsterdam, v. 155, n. 1/2, p. 55-66, Feb. 2010.

VIDAL-VÁZQUEZ, E. et al. Geostatistical analysis of microrelief of an oxisol as a function of tillage and cumulative rainfall. **Scientia Agricola**, Piracicaba, v. 66, n. 2, p. 225-232, 2009.

VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAES, R. F. et al. (Ed.). **Tópicos em ciência do solo**. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2000. v. 1, p. 1-53.

VIEIRA, S. R. et al. Spatial variability of soil chemical properties after coffee tree removal. **Revista Brasileira de Ciência do Solo**, Viçosa, MG, v. 33, n. 5, p. 1507-1514, set./out. 2009.

VITORIA, E. L. et al. Correlação linear e espacial entre produtividade de *Brachiaria brizantha*, densidade do solo e porosidade total em função do sistema de manejo do solo. **Engenharia Agrícola**, Jaboticabal, v. 32, n. 5, p. 909-919, 2012.

WACKERNAGEL, H. **Multivariate geostatistics**: an introduction with applications. New York: Springer-Verlag, 2003. 387 p.

WEBSTER, R.; OLIVER, M. **Geostatistics for environmental scientists**. 2nd ed. Chichester: J. Wiley, 2007. 315 p.

WOJCIECHOWSKI, J. C. et al. Geoestatística aplicada ao estudo das características físico-químicas do solo em áreas de floresta estacional decidual. **Ciência Florestal**, Santa Maria, v. 19, n. 4, p. 383-391, 2009.

WU, C. et al. Statistical and geoestatistical characterization of heavy metal concentrations in a contaminated area taking into account soil map units. **Geoderma**, Amsterdam, v. 144, n. 1/2, p. 171-179, Mar. 2008.

YANAGI JÚNIOR, T. et al. Spatial variability of noise level in agricultural machines. **Engenharia Agrícola**, Jaboticabal, v. 32, n. 2, p. 217-225, 2012.