



**RAFAEL LEMOS BASTOS**

**MÉTODOS PARA SELEÇÃO DE MODELOS DE  
SEMIVARIOGRAMA EM CAMPOS  
ALEATÓRIOS GAUSSIANOS**

**LAVRAS - MG  
2017**

**RAFAEL LEMOS BASTOS**

**MÉTODOS PARA SELEÇÃO DE MODELOS DE SEMIVARIOGRAMA  
EM CAMPOS ALEATÓRIOS GAUSSIANOS**

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 Espacial, para a obtenção do título de Doutor.

Orientador

Dr. Marcelo Silva de Oliveira

Co-orientador

Dr. Daniel Furtado Ferreira

**LAVRAS - MG**

**2017**

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

Bastos, Rafael Lemos.

Métodos para seleção de modelos de semivariograma em  
campos aleatórios gaussianos / Rafael Lemos Bastos. – 2017.  
136 p. : il.

Orientador: Marcelo Silva de Oliveira.

Coorientador: Daniel Furtado Ferreira.

Tese (doutorado) – Universidade Federal de Lavras, 2017.

Bibliografia.

1. Espacial. 2. Geoestatística. 3. Validação. I. Oliveira, Marcelo  
Silva de. II. Ferreira, Daniel Furtado. III. Título.

**RAFAEL LEMOS BASTOS**

**MÉTODOS PARA SELEÇÃO DE MODELOS DE SEMIVARIOGRAMA  
EM CAMPOS ALEATÓRIOS GAUSSIANOS**

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 Espacial, para a obtenção do título de Doutor.

APROVADO em 18 de agosto de 2017.

Dr. Eric Batista Ferreira	UNIFAL-MG
Dr. Fortunato Silva de Menezes	UFLA
Dr. Renato Ribeiro de Lima	UFLA
Dr. João Domingos Scalon	UFLA
Dr. Deive Ciro de Oliveira	UNIFAL-MG

Dr. Marcelo Silva de Oliveira  
Orientador

Dr. Daniel Furtado Ferreira  
Co-orientador

**LAVRAS - MG  
2017**

## DEDICO

À memória do meu irmão Guilherme Tadeu Lemos Bastos.

A morte é o grande mistério que ronda a humanidade,  
bate em nossa porta quando menos esperamos.

Faz muitos sofrerem, especialmente pela ausência de quem tanto amamos,  
de quem jamais sentiremos novamente o perfume, o abraço carinhoso,  
a voz a nos chamar e tudo o que faz diferença em nossas vidas.

Todos nós perdemos pessoas queridas, mas em nosso íntimo,  
ainda acreditamos que podemos encontrá-las em cada esquina,  
que vamos abrir a porta e dar de cara com elas.

Porém, elas apenas se mostram em nossos sonhos  
e na saudade que cresce a cada dia.

A quem mais amei, amo e amarei.

*Quem nunca saiu em busca de um grande sonho  
Quem nunca pensou em desistir a certo ponto  
As lágrimas verdadeiras expressam a dor  
Elas expõem meu interior ferido e abatido  
Nelas também encontro alívio e  
Retorno a ver uma luz ao fim do túnel  
Quanto vale a sua fé?  
Você tem que acreditar e ser quem você é  
Tenha paciência tudo tem o tempo certo  
Tá faltando pouco você tá chegando perto  
E se você cair?  
Se levante, eu sei que você vai conseguir  
E se prepare, para vitória  
Respire fundo chegou a hora  
Sua estrela vai brilhar  
Você tá pronto pra voar  
Nunca desista, dos seus sonhos  
Nunca desista, dos seus planos  
Nunca desista, nunca desista.  
**Munhoz e Mariano***

## AGRADECIMENTOS

Primeiramente, a Deus pelo dom da vida, por iluminar-me sempre.

Aos meus pais, Geraldo Tadeu Bastos e Teresinha da Penha Lemos Bastos, aos meus irmãos, Guilherme Tadeu Lemos Bastos (*in memoriam*), Cristiano Lemos Bastos e Luana Lemos Bastos, pelo carinho e apoio em todos os momentos.

À Universidade Federal de Lavras (UFLA), ao Departamento de Estatística e ao Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, pela oportunidade concedida para realização do Doutorado.

Ao Orientador, Professor Dr. Marcelo Silva de Oliveira, pela amizade, conhecimentos ensinados e pelas mãos estendidas quando mais precisei.

Ao Coorientador, Professor Dr. Daniel Furtado Ferreira, pelo apoio científico e *fitness*, amizade e companheirismo.

Ao Eterno Orientador Eric Batista Ferreira, por ter acreditado em mim, pelos conhecimentos ensinados, dedicação e paciência, proporcionando-me uma excelente base estatística.

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

Aos professores da Universidade Federal de Lavras e de Alfenas, da Escola Estadual Secretário Tristão da Cunha e de outras instituições que estudei.

Aos meus familiares, afilhados, vizinhos e amigos.

À Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) pela bolsa concedida no Doutorado e ao Grupo de Pesquisa em Estatística Espacial (GPS-UFLA) pelo apoio científico.

Enfim, a todos que fizeram parte da minha carreira acadêmica.

## RESUMO

Vários fenômenos apresentam certa estruturação na variação entre vizinhos, apresentando dependência espacial. A Geoestatística engloba um conjunto de ferramentas que permitem o estudo da dependência espacial, sendo o semivariograma uma de suas principais ferramentas, no qual expressa a dependência espacial entre as amostras. O ajuste de um modelo ao semivariograma é fundamental, pois toda predição Geoestatística depende desse modelo. Dentre vários modelos que podem ser ajustados, deve-se selecionar o que mais bem se ajusta, para que o mapa de krigagem a ser gerado após a seleção, seja o mais coerente com a realidade. Como método de seleção para modelos de semivariograma, os pesquisadores utilizam diferentes estatísticas, ocorrendo a seleção de diferentes modelos para o mesmo semivariograma. A seleção de diferentes modelos reflete uma decisão subjetiva dos pesquisadores, tendo como resultado diferentes informações num mesmo experimento. O objetivo do presente trabalho enfoca nesse problema de pesquisa, buscando recomendar um método para proceder a seleção de um modelo que mais bem se ajusta ao semivariograma. Foram encontrados oito métodos usuais de seleção e sete foram propostos, sendo todos eles programados no *software* R. Os métodos foram aplicados em milhares de dados simulados em diferentes cenários, tornando-se possível recomendar aquele que apresentar um melhor comportamento em relação as proporções de acerto. Diante da classificação dos métodos usuais e propostos em relação as proporções de acerto, destaca-se que o método proposto 7 esteve entre as primeiras posições em todos cenários adotados. Esse método possui uma grande vantagem, pois é composto por outros métodos que estiveram entre as primeiras posições. Portanto, recomenda-se utilizá-lo na seleção do modelo que mais bem se ajusta ao semivariograma, em campos aleatórios gaussianos. Espera-se que este trabalho seja fundamental para uma análise Geoestatística mais eficiente e informativa.

Palavras-Chave: Espacial, Geoestatística, Validação.

## ABSTRACT

There is a wide range of phenomena that present some level of structuring and spatial dependence in the variation between observations. Geostatistics include a set of tools that allow the study of spatial dependence. For example, the Semivariogram, which is considered one of the main tools, expresses the spatial dependence between two observations as a function of the distance between them. The adjustment of a model to the semivariogram is fundamental since all geostatistical prediction depends on this model. It is of great importance to adjust data to models that fit the data well so that the kriging map that is generated after the selection can present a great level of coherence with reality. As a method of selection for semivariogram models, researchers use different statistics by selecting different models for the same semivariogram. However, the selection of different models reflects on a subjective decision of the researchers, resulting in different information in the same experiment. Therefore, focusing on this research problem, this work seeks to recommend a method to better select a model that best fits the semivariogram. A total of eight usual methods of selection were found and seven were proposed using the R software environment for statistical computing and graphics. The methods were applied in thousands of simulated data in different scenarios, and it was possible to recommend one that presented the best behavior in relation to the correctness rate. Considering the classification of the usual and proposed methods in relation to the correctness rates, it is worth noting that the proposed method number 7 was in the first position in all scenarios adopted. This method is advantageous because it is composed by other methods that were also close to the first positions. Therefore, it is recommended to use it in selecting the model that best fits the semivariogram in Gaussian random fields. It is expected that this work will be fundamental for a more efficient and informative geostatistical analysis.

Keywords: Spatial, Geostatistics, Validation.

## LISTA DE FIGURAS

Figura 1	Modelos Wave, Exponencial, Gaussiano e Esférico ajustados ao semivariograma, considerando dados de teor de zinco . . . . .	19
Figura 2	Mapas de krigagem utilizando diferentes modelos de semivariograma . . . . .	20
Figura 3	Exemplo de semivariograma com suas características . . . . .	29
Figura 4	Esquemática da acurácia versus precisão . . . . .	39
Figura 5	Pontos formados pelos valores de $M2$ versus $M1$ . . . . .	48
Figura 6	Pontos formados pelos valores de $M2 - M1$ versus $(M1 + M2)/2$ . . . . .	49
Figura 7	Pontos formados pelos desvios padrões versus médias dos postos . . . . .	54
Figura 8	Malha dos dados reais de teor de zinco ( $mg/dm^3$ ) no solo . . . . .	57
Figura 9	Modelo esférico considerando $C_0=0$ , $C_1=10$ , $a=30$ e $C=10$ . . . . .	58
Figura 10	Malha de um exemplo de dados simulados com dependência espacial esférica . . . . .	61
Figura 11	Semivariogramas referentes aos dados simulados com dependência espacial esférica, com 100%, 95%, 90%, 80%, 75% e 70% da distância máxima . . . . .	63
Figura 12	Exemplo de intervalo de confiança para $Z(s_i)$ . . . . .	66
Figura 13	Variâncias de krigagem em relação aos valores preditos dos modelos adotados . . . . .	68
Figura 14	Exemplo de intervalo de confiança para $Z(s_i) - \hat{Z}(s_i)$ . . . . .	70
Figura 15	Ajuste perfeito da reta aos pontos $p_i$ 's, $i = 1, 2, \dots, n$ . . . . .	71
Figura 16	Reta ajustada aos $p_i$ 's e a reta ideal acrescentada . . . . .	72
Figura 17	Especificação dos $i$ -ésimo $p_i$ 's e $pm_i$ 's, $i = 1, 2, 3, 4$ . . . . .	73

Figura 18	Pontos formados pelos desvios padrões em relação as médias dos postos, com as distâncias desses pontos em relação ao ponto (0,0)	80
Figura 19	Semivariogramas referentes a dados simulados com dependência espacial esférica de acordo com a seção 3.1.2 . . . . .	82
Figura 20	Semivariogramas estimados referentes a 1000 simulações de dados com dependência espacial exponencial, esférico, gaussiano e wave, considerando 100%, 80% e 70% da distância máxima . . .	83
Figura 21	Semivariogramas estimados referentes a 1000 simulações de dados com dependência espacial exponencial, esférico, gaussiano e wave, condicionados ao $R^2$ acima de 0,7 em relação ao semivariograma paramétrico desejado, considerando 100%, 80% e 70% da distância máxima . . . . .	84
Figura 22	(a) Semivariograma referente a dados simulados com dependência espacial wave com valores paramétricos $C_0=0$ , $C_1=10$ , $a=5$ e $C=10$ , e (b) o erro ocasionado no processo de autovalidação . . .	86
Figura 23	Semivariograma esférico com (a) 100%, (b) 80% e (c) 70% da distância máxima . . . . .	88
Figura 24	Modelo Exponencial, Esférico, Gaussiano e Wave ajustados ao semivariograma esférico com (a) 100%, (b) 80% e (c) 70% da distância máxima . . . . .	89
Figura 25	Valores preditos pela validação do modelo esférico plotados nos intervalos com 95% de confiança para medida individual, considerando 100% da distância máxima . . . . .	95
Figura 26	Valores preditos pela validação do modelo esférico plotados nos intervalos com 95% de confiança para medida individual, considerando 80% da distância máxima . . . . .	96

Figura 27	Valores preditos pela validação do modelo esférico plotados nos intervalos com 95% de confiança para medida individual, considerando 70% da distância máxima . . . . .	96
Figura 28	Intervalo de confiança para o erro de predição em relação a validação do modelo esférico, considerando 100% da distância máxima	98
Figura 29	Intervalo de confiança para o erro de predição em relação a validação do modelo esférico, considerando 80% da distância máxima	98
Figura 30	Intervalo de confiança para o erro de predição em relação a validação do modelo esférico, considerando 70% da distância máxima	99
Figura 31	Pontos formados pelos desvios padrões em relação as médias dos ranks, com as distâncias desses pontos em relação ao ponto (0,0), considerando (a) 100% , (b) 80% e (c) 70% da distância máxima	105
Figura 32	Semivariograma exponencial com (a) 100%, (b) 80% e (c) 70% da distância máxima . . . . .	109
Figura 33	Semivariograma gaussiano com (a) 100%, (b) 80% e (c) 70% da distância máxima . . . . .	113
Figura 34	Semivariograma wave com (a) 100%, (b) 80% e (c) 70% da distância máxima . . . . .	117

## LISTA DE TABELAS

Tabela 1	Decisões possíveis na seleção de um modelo . . . . .	38
Tabela 2	Estimativas classificadas dos métodos selecionados, com as médias e os desvios padrões dos postos calculados . . . . .	53
Tabela 3	Processo de seleção das $i$ -ésima variâncias de krigagem entre os modelos . . . . .	69
Tabela 4	Variâncias de krigagem selecionadas . . . . .	69
Tabela 5	Estimativas classificadas dos $n$ métodos selecionados, com as médias e os desvios padrões dos ranks calculados . . . . .	79
Tabela 6	Nomes das funções dos métodos usuais e propostos de seleção .	81
Tabela 7	Cenários adotados na verificação do comportamento dos métodos	87
Tabela 8	Estimativas dos parâmetros dos modelos ajustados, considerando 100%, 80% e 70% da distância máxima do semivariograma esférico	90
Tabela 9	Resultados das aplicações dos métodos usuais em dados simulados com dependência espacial esférica, considerando 100% da distância máxima . . . . .	92
Tabela 10	Resultados das aplicações dos métodos usuais em dados simulados com dependência espacial esférica, considerando 80% da distância máxima . . . . .	93
Tabela 11	Resultados das aplicações dos métodos usuais em dados simulados com dependência espacial esférica, considerando 70% da distância máxima . . . . .	94
Tabela 12	Estimativas das proporções do Método 1 . . . . .	97
Tabela 13	Estimativas das proporções do Método 2 . . . . .	99

Tabela 14	Valores- $p$ do teste $t$ e as distâncias médias entre os valores observados e preditos, considerando 100%, 80% e 70% da distância máxima . . . . .	100
Tabela 15	Valores- $p$ do teste $F$ e as distâncias médias entre os valores observados e preditos, considerando 100%, 80% e 70% da distância máxima . . . . .	101
Tabela 16	Valores- $p$ referentes ao teste da correlação e teste do viés, e estimativas dos limites de concordância, considerando 100%, 80% e 70% da distância máxima . . . . .	102
Tabela 17	$DQM$ , $QV$ , $DQDP$ e $FCPDP$ , em relação aos modelos adotados, considerando 100%, 80% e 70% da distância máxima . . .	103
Tabela 18	Estimativas ranqueadas, com as médias e os desvios padrões dos ranks, considerando 100%, 80% e 70% da distância máxima . . .	104
Tabela 19	Distâncias referentes a Figura 31 . . . . .	105
Tabela 20	Resultados das aplicações dos métodos propostos em dados simulados com dependência espacial esférica, considerando 100% da distância máxima . . . . .	106
Tabela 21	Resultados das aplicações dos métodos propostos em dados simulados com dependência espacial esférica, considerando 80% da distância máxima . . . . .	107
Tabela 22	Resultados das aplicações dos métodos propostos em dados simulados com dependência espacial esférica, considerando 70% da distância máxima . . . . .	108
Tabela 23	Resultados das aplicações dos métodos usuais e propostos em dados simulados com dependência espacial exponencial, considerando 100% da distância máxima . . . . .	110

Tabela 24	Resultados das aplicações dos métodos usuais e propostos em dados simulados com dependência espacial exponencial, considerando 80% da distância máxima . . . . .	111
Tabela 25	Resultados das aplicações dos métodos usuais e propostos em dados simulados com dependência espacial exponencial, considerando 70% da distância máxima . . . . .	112
Tabela 26	Resultados das aplicações dos métodos usuais e propostos em dados simulados com dependência espacial gaussiana, considerando 100% da distância máxima . . . . .	114
Tabela 27	Resultados das aplicações dos métodos usuais e propostos em dados simulados com dependência espacial gaussiana, considerando 80% da distância máxima . . . . .	115
Tabela 28	Resultados das aplicações dos métodos usuais e propostos em dados simulados com dependência espacial gaussiana, considerando 70% da distância máxima . . . . .	116
Tabela 29	Resultados das aplicações dos métodos usuais e propostos em dados simulados com dependência espacial wave, considerando 100% da distância máxima . . . . .	118
Tabela 30	Resultados das aplicações dos métodos usuais e propostos em dados simulados com dependência espacial wave, considerando 80% da distância máxima . . . . .	119
Tabela 31	Resultados das aplicações dos métodos usuais e propostos em dados simulados com dependência espacial wave, considerando 70% da distância máxima . . . . .	120
Tabela 32	Comportamento dos métodos usuais e propostos diante os cenários	121

Tabela 33	Proporções de acerto dos modelos selecionados pelos métodos usuais e propostos, aplicados em dados com dependência espacial esférica . . . . .	124
Tabela 34	Proporções de acerto dos modelos selecionados pelos métodos usuais e propostos, aplicados em dados com dependência espacial exponencial . . . . .	125
Tabela 35	Proporções de acerto dos modelos selecionados pelos métodos usuais e propostos, aplicados em dados com dependência espacial gaussiana . . . . .	126
Tabela 36	Proporções de acerto dos modelos selecionados pelos métodos usuais e propostos, aplicados em dados com dependência espacial wave . . . . .	127
Tabela 37	Classificação dos métodos usuais e propostos em relação as maiores proporções de acerto diante os cenários adotados . . . . .	128
Tabela 38	Estimativas ranqueadas, com as médias e os desvios padrões dos ranks, considerando dados reais de teor de zinco . . . . .	129

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	17
<b>1.1</b>	<b>Objetivos</b>	22
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	23
<b>2.1</b>	<b>Geoestatística</b>	23
<b>2.1.1</b>	<b>Semivariograma</b>	28
<b>2.1.2</b>	<b>Modelos de semivariograma</b>	33
<b>2.1.3</b>	<b>Importância do modelo de semivariograma</b>	35
<b>2.1.4</b>	<b>Avaliação da adequabilidade do modelo de semivariograma</b>	38
<b>2.1.5</b>	<b>Autovalidação</b>	40
<b>2.1.6</b>	<b>Métodos usuais de seleção do modelo de semivariograma</b>	41
<b>2.1.7</b>	<b>Métodos de seleção exógenos à Geoestatística</b>	48
<b>2.2</b>	<b>Simulação</b>	55
<b>3</b>	<b>MATERIAL E MÉTODOS</b>	57
<b>3.1</b>	<b>Material</b>	57
<b>3.1.1</b>	<b>Dados reais com estrutura de dependência espacial</b>	57
<b>3.1.2</b>	<b>Dados simulados com estrutura de dependência espacial</b>	58
<b>3.2</b>	<b>Métodos</b>	62
<b>3.2.1</b>	<b>Métodos utilizados na análise geoestatística dos dados de teor de zinco</b>	62
<b>3.2.2</b>	<b>Métodos utilizados na análise geoestatística dos dados simulados</b>	62
<b>3.2.3</b>	<b>Proposição de métodos de seleção do modelo de semivariograma</b>	65
<b>3.2.4</b>	<b>Estratégia utilizada para recomendação de um ou mais métodos de seleção</b>	82
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b>	88
<b>4.1</b>	<b>Aplicações dos métodos usuais e propostos em dados simulados com dependência espacial esférica</b>	88
<b>4.2</b>	<b>Aplicações dos métodos usuais e propostos em dados simulados com dependência espacial exponencial, gaussiana e wave</b>	109
<b>4.3</b>	<b>Discussão em relação aos métodos usuais e propostos</b>	122
<b>4.4</b>	<b>Resultados das simulações para identificação dos método recomendado</b>	124
<b>4.5</b>	<b>Aplicação do método recomendado em dados reais</b>	129
<b>5</b>	<b>CONCLUSÕES</b>	130
	<b>REFERÊNCIAS</b>	131

## 1 INTRODUÇÃO

Uma boa análise estatística de dados espaciais não pode ser feita sob a hipótese de independência espacial. Entretanto, este tipo de hipótese não pode simplesmente ser pressuposta antes que se prove a não existência de correlação de amostras entre os pontos georeferenciados.

Com a necessidade do desenvolvimento de ferramentas estatísticas que consideram a estrutura de variabilidade espacial de um determinado fenômeno, surgiu a Geoestatística. A consideração da existência de uma dependência da variação com relação ao espaço de amostragem é o que diferencia a Geoestatística da Estatística Clássica.

A Geoestatística se fundamenta no que foi inicialmente chamado de Teoria das Variáveis Regionalizadas, segundo a qual os seus valores estão relacionados à sua disposição espacial, ou seja, os valores de dois pontos vizinhos são parecidos e a medida que se distanciam, tornam-se menos parecidos. Hoje é dito que se fundamenta na teoria de campos aleatórios (processos estocásticos). A Geoestatística, como metodologia de análise de dados, está difundida em vários ramos da ciência, envolvendo áreas de ciências da Terra, Engenharias, Agrárias, Biológicas, Humanas, Exatas, etc.

As principais ferramentas da Geoestatística são o semivariograma, que é um gráfico de semivariâncias em relação às distâncias, fundamental para descrever a estrutura de variabilidade e desenvolver as predições de valores para locais não amostrados, que é a krigagem, a segunda ferramenta. Com a aplicação da Geoestatística, por meio de uma malha amostral da área em estudo é possível gerar um mapa apresentando valores para qualquer posição dessa área.

O ajuste de um modelo ao semivariograma é um dos aspectos mais im-

portantes dessa aplicação, pois toda predição do mapa gerado depende do valor desse modelo para cada distância especificada. Portanto, se o modelo não estiver bem ajustado, todos os cálculos subsequentes não apresentarão os melhores resultados e o mapa em relação ao fenômeno estudado não será o mais coerente com a realidade.

Diante dos vários modelos possíveis de semivariograma, deve-se selecionar o que mais bem se ajusta. Uma técnica consagrada na Geoestatística para auxiliar na seleção é a validação de cada modelo. No presente trabalho adotou-se a autovalidação, porém os resultados podem ser também estendidos para o caso do uso de validação cruzada. Por meio da autovalidação, retira-se um ponto de cada vez da malha amostral e para a coordenada referente ao ponto retirado é realizada a predição pelo método de krigagem, utilizando os valores observados dos outros pontos e o modelo de semivariograma adotado. Portanto, após realizar a autovalidação de cada modelo, para cada coordenada onde há um valor observado, também haverá um valor predito. Comparando os valores observados e preditos, várias estatísticas podem ser utilizadas como método de seleção.

As análises geoestatísticas estão cada vez mais sendo realizadas, porém, os pesquisadores ainda utilizam diferentes estatísticas da autovalidação como método para seleção de modelos de semivariograma, ocorrendo a seleção de diferentes modelos para um mesmo semivariograma, resultando diferentes conclusões num mesmo experimento.

Com intuito de mostrar que a ausência de um método de seleção do modelo que melhor se ajusta ao semivariograma coloca o pesquisador diante de diferentes mapas, dentre os quais ele não sabe qual utilizar, apresenta-se uma análise geoestatística com dados reais (veja subseção 3.1.1) de teor de zinco ( $mg/dm^3$ ) do solo.

Na Figura 1 são apresentados os modelos wave, exponencial, gaussiano e esférico ajustados ao semivariograma estimado em relação aos dados de teor de zinco.

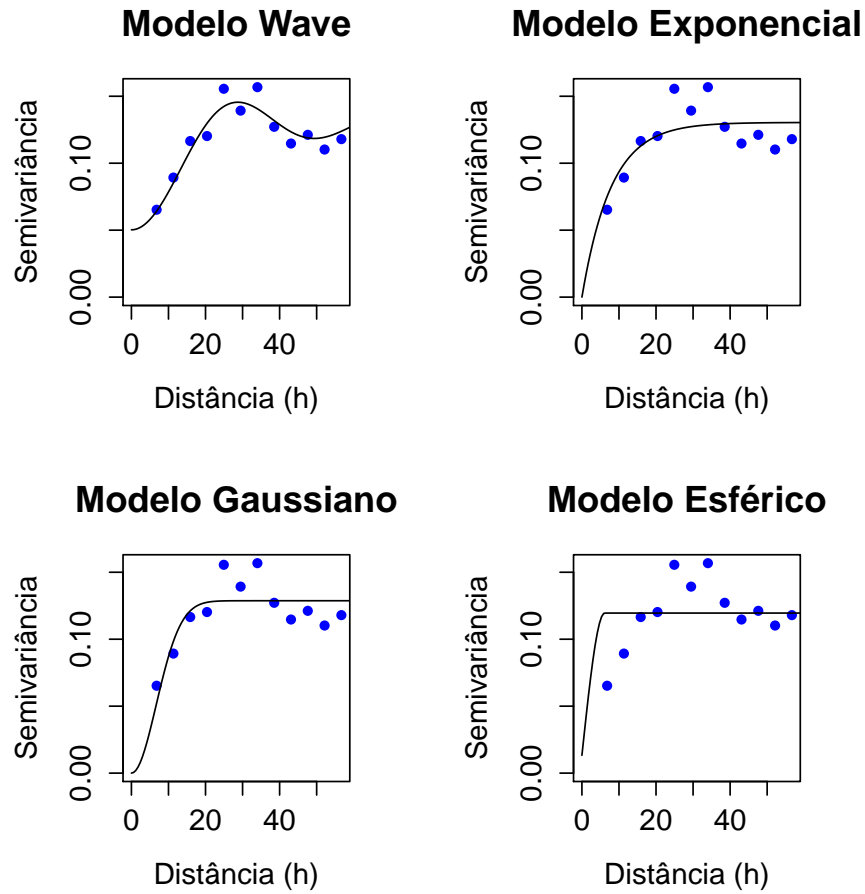


Figura 1 Modelos Wave, Exponencial, Gaussiano e Esférico ajustados ao semivariograma, considerando dados de teor de zinco

Utilizando os modelos ajustados ao semivariograma da Figura 1, foram gerados seus respectivos mapas, utilizando krigagem ordinária. Esses mapas são apresentados na Figura 2.

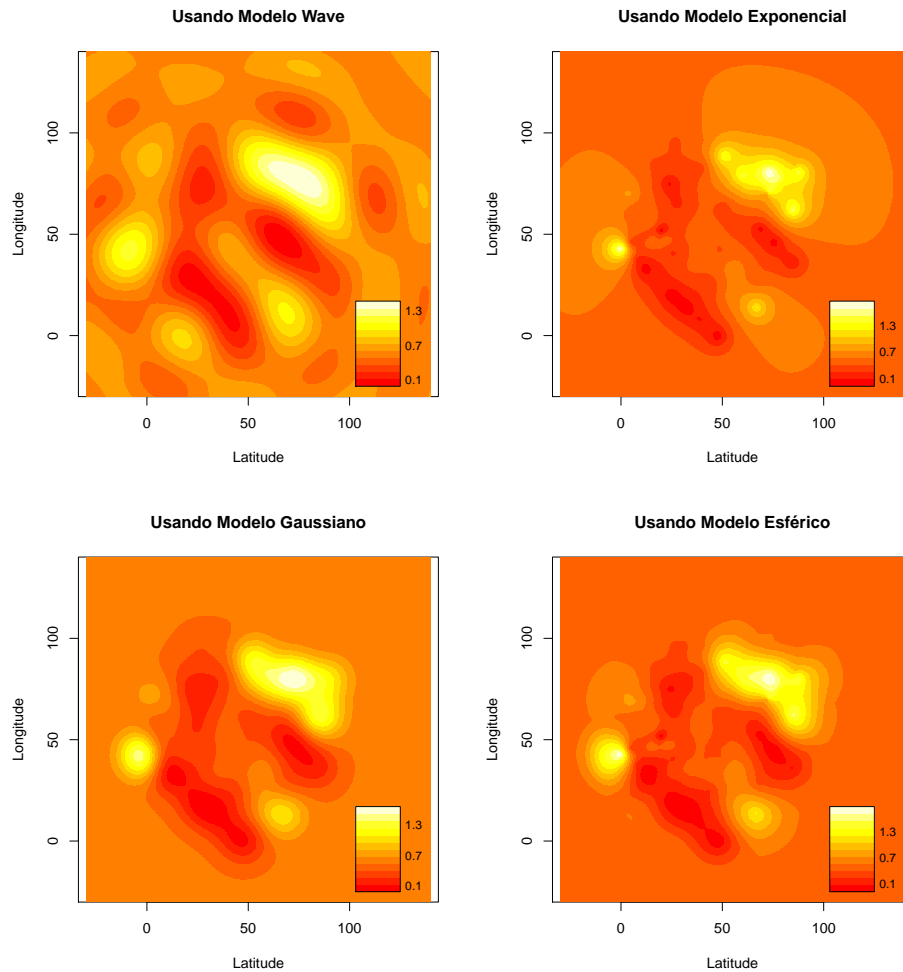


Figura 2 Mapas de krigagem utilizando diferentes modelos de semivariograma

Observando a Figura 2, o uso de diferentes modelos ajustados ao mesmo semivariograma resultam em diferentes mapas de krigagem. A geração de diferen-

tes mapas ilustra a importância da seleção do modelo que mais bem se ajusta ao semivariograma, para que o mapa selecionado seja o mais coerente com a realidade da variável desejada na área em estudo.

No caso desta ilustração, que são dados de cafeicultura, com o mapa mais coerente, pode-se realizar uma correção do solo em relação ao teor de zinco, deixando toda área com nível contido no intervalo que proporciona maior produtividade. Percebe-se que se o analista utilizar um dos outros mapas que não seja gerado pelo modelo que mais bem se ajusta, iria realizar uma correção equivocada do solo, prejudicando uma maior produtividade.

Até os dias de hoje, a inexistência de um método recomendado para seleção é um problema a ser solucionado.

O presente trabalho enfoca nesse problema de pesquisa, buscando responder a seguinte questão:

Utilizando a técnica de autovalidação, qual deveria ser o método recomendado para proceder a seleção de um modelo que mais bem se ajusta ao semivariograma?

Essa questão de pesquisa pode ser desdobrada nas seguintes subquestões:

- a)** Quais métodos usuais de seleção?
- b)** Aplicando os métodos usuais de seleção e os propostos, qual é o comportamento deles? Eles selecionam o mesmo modelo?
- c)** Caso não selecione, quais métodos seriam recomendados como os melhores?

## 1.1 Objetivos

O objetivo geral do presente trabalho é:

Recomendar um método de seleção de um modelo que mais bem se ajusta ao semivariograma.

Os objetivos específicos do presente trabalho são:

- a) Revisar a literatura em busca dos métodos usuais de seleção para modelo de semivariograma;
- b) Programar os métodos usuais de seleção e avaliá-los;
- c) Realizar a proposição de métodos de seleção para modelo de semivariograma;
- d) Programar os métodos propostos de seleção e avaliá-los.

## 2 REFERENCIAL TEÓRICO

Neste tópico, são revisados os assuntos fundamentais da Geoestatística para execução deste trabalho, tais como: semivariograma, modelo de semivariograma, importância do modelo, avaliação da sua adequabilidade, autovalidação, métodos usuais de seleção do modelo, métodos de seleção utilizados em outras áreas e simulação.

A Estatística Espacial é classificada por Cressie (1993) em três categorias em relação à natureza estocástica da observação: dados de superfícies contínuas, dados de processos pontuais e dados de área. A Geoestatística, que trabalha com dados de superfícies contínuas, é o foco do presente trabalho.

### 2.1 Geoestatística

Na África do Sul, no século XX - mais especificamente no início dos anos 50 - o engenheiro de minas Daniel Gerhardus Krige e o estatístico Hebert Simon Sichel desenvolveram empiricamente uma técnica própria de predição para o cálculo de reservas minerais, baseada em conceito de regressão, a qual posteriormente recebeu tratamento formal pelo geólogo e matemático Georges Matheron, na França, com o nome Geoestatística. De acordo com Matheron (1962), a base da Geoestatística vem da teoria das variáveis regionalizadas, sendo variáveis contínuas com condicionamento espacial. Matheron (1962) destacava que a diferença nos valores de uma variável tomados em dois pontos no campo pode depender da distância entre eles.

Inicialmente a Geoestatística era aplicada apenas para situações em geologia de minas e, posteriormente, se estendeu para outros campos, inclusive para

as Ciências Agrárias. De acordo com Chiles e Delfiner (2012), para descrever as variações espaciais da variável desejada, tais como grau, espessura, acúmulo, entre outras, a análise estrutural da variável regionalizada é fundamental.

A Geoestatística engloba um conjunto de ferramentas que possibilita mapear, com mais precisão, as propriedades dos dados geográficos, permitindo o estudo da dependência espacial. Um requisito básico para utilizar Geoestatística é que as amostras sejam georreferenciadas, não necessariamente por coordenadas geográficas, mas algum tipo de referência que informe a localização de cada ponto.

De acordo com Burrough (1987), a teoria das variáveis regionalizadas pressupõe que uma variável pode ser expressa por

$$Z(s) = \mu(s) + \varepsilon'(s) + \varepsilon''(s),$$

em que  $Z(s)$  é o valor do atributo  $Z$  no ponto  $s$ ,  $s = (x_1, \dots, x_d)$  é formado pelas coordenadas espaciais  $x_i$ , com  $i = 1, \dots, d$  e dimensão  $d$ ,  $\mu(s)$  é uma componente estrutural associada a uma tendência constante,  $\varepsilon'(s)$  é uma componente aleatória espacialmente correlacionada e  $\varepsilon''(s)$  um ruído aleatório.

Na Geoestatística, o processo envolvido é um processo estocástico, sendo um conjunto de variáveis aleatórias que são indexadas (localizadas) num mesmo espaço, representado por

$$\{Z(s) : s \in D \subset \mathbb{R}^d\}.$$

O processo estocástico é referido como um campo aleatório quando a dimensão  $d > 1$ . O campo aleatório é chamado de Gaussiano quando a distribuição para qualquer conjunto de  $n$  variáveis aleatórias é uma Gaussiana  $n$ -variada. Por

propriedades da distribuição Gaussiana multivariada, cada uma de suas variáveis aleatórias tem distribuição marginal Gaussiana univariada.

Uma observação em relação ao conjunto de  $n$  observações georreferenciadas numa área de estudo é que, ao contrário da Estatística Clássica em que se trabalha com  $n$  observações de uma variável aleatória, na Geoestatística quase sempre têm-se  $n$  variáveis aleatórias com apenas uma observação cada, não existindo a repetição de uma mesma variável. Essa restrição ao número de repetições limita o estudo da distribuição dessas variáveis. A solução para essa limitação é suposição de que o processo satisfaça certas propriedades de estacionariedade.

A existência de estacionariedade permite a criação de repetições no experimento, mesmo que as amostras sejam coletadas em pontos diferentes. Esse fato é justificado em função de que todas as amostras pertencem a populações com os mesmos momentos estatísticos.

Entre os tipos de estacionariedade num campo aleatório, destacam-se:

### **Campo aleatório estacionário estrito**

Essa estacionariedade ocorre quando a distribuição espacial é invariante sob a translação das coordenadas, isto é,

$$\begin{aligned} &P[Z(s_1) < z_1, Z(s_2) < z_2, \dots, Z(s_n) < z_n] \\ &= P[Z(s_1 + h) < z_1, Z(s_2 + h) < z_2, \dots, Z(s_n + h) < z_n], \forall h, n, s_i. \end{aligned}$$

A distribuição conjunta de  $(Z(s_1), Z(s_2), \dots, Z(s_n))$  é a mesma de  $(Z(s_1 + h), Z(s_2 + h), \dots, Z(s_n + h))$ , para qualquer translação da distância  $h$ , para todo conjunto de pontos, para todo número de pontos.

### **Campo aleatório estacionário de primeira ordem**

Essa estacionariedade ocorre quando a média é constante em toda área, isto é,

$$E[Z(s)] = \mu, \forall s,$$

em que  $\mu$  é a média populacional.

### **Campo aleatório estacionário de segunda ordem**

Essa estacionariedade ocorre quando a média é constante e a covariância entre os atributos em diferentes locais é uma função apenas de  $h$ , isto é,

$$\begin{aligned} E[Z(s)] &= \mu, \\ C[Z(s), Z(s+h)] &= C(h), \forall s. \end{aligned}$$

Dado que  $C(h)$  não depende das coordenadas absolutas e que

$$C[Z(s), Z(s+0)] = \text{Var}[Z(s)] = C(0),$$

então a variabilidade desse campo é a mesma em todos os lugares.

Portanto, esse tipo de campo aleatório tem média e variância constantes, e covariâncias que dependem apenas da distância entre dois diferentes pontos.

### **Campo aleatório estacionário intrínseco**

Ocorre quando, além da média constante, a variância dos incrementos  $Z(s) - Z(s+h)$  existe e é independente da localização na região, sendo em função

apenas da distância  $h$ , isto é,

$$E[Z(s)] = \mu,$$

$$\frac{Var[Z(s) - Z(s+h)]}{2} = \gamma(h),$$

em que a função  $\gamma(h)$  recebe o nome de semivariância do processo espacial.

A função  $\gamma(h)$  mede a variabilidade das diferenças entre as realizações da variável aleatória de interesse, de tal maneira que quanto menores, menor a variação dessas diferenças.

A relação entre  $\gamma(h)$  e  $C[Z(s), Z(s+h)]$  é encontrada da seguinte forma:

$$\begin{aligned} \gamma(h) &= \frac{Var[Z(s) - Z(s+h)]}{2} \\ &= \frac{E\{[Z(s) - Z(s+h)]^2\} - \{E[Z(s) - Z(s+h)]\}^2}{2} \\ &= \frac{E\{[Z(s) - Z(s+h)]^2\} - \left\{ \underbrace{E[Z(s)]}_{\mu} - \underbrace{E[Z(s+h)]}_{\mu} \right\}^2}{2} \\ &= \frac{E\{[Z(s) - Z(s+h)]^2\} - \underbrace{[\mu - \mu]^2}_0}{2} \\ &= \frac{E\{[Z(s) - Z(s+h)]^2\}}{2} \\ &= \frac{E\{[Z(s)]^2 - 2Z(s)Z(s+h) + [Z(s+h)]^2\}}{2} \\ &= \frac{\underbrace{E\{[Z(s)]^2\}}_{Var[Z(s)]+\mu^2} - 2\underbrace{E\{Z(s)Z(s+h)\}}_{2C[Z(s),Z(s+h)]+2\mu^2} + \underbrace{E\{[Z(s+h)]^2\}}_{Var[Z(s+h)]+\mu^2}}{2} \\ &= \frac{Var[Z(s)] + \mu^2 - 2C[Z(s), Z(s+h)] - 2\mu^2 + Var[Z(s+h)] + \mu^2}{2}. \end{aligned}$$

Portanto,

$$\gamma(h) = \frac{\text{Var}[Z(s)]}{2} + \frac{\text{Var}[Z(s+h)]}{2} - C[Z(s), Z(s+h)].$$

Observe que não foi exigido que o campo aleatório fosse estacionário de segunda ordem. Caso fosse, teria-se que:

$$\begin{aligned} \gamma(h) &= \frac{\text{Var}[Z(s)]}{2} + \frac{\text{Var}[Z(s+h)]}{2} - \underbrace{C[Z(s), Z(s+h)]}_{=C(h)} \\ &= \underbrace{\frac{\text{Var}[Z(s)]}{2} + \frac{\text{Var}[Z(s+h)]}{2}}_{\text{Var}[Z(s)]=\text{Var}[Z(s+h)]} - C(h) \\ &= \text{Var}[Z(s)] - C(h) \\ &= C[Z(s), Z(s+0)] - C(h) \\ &= C(0) - C(h). \end{aligned}$$

Destaca-se que o campo aleatório estacionário intrínseco é menos restritivo que os demais, sendo o mais utilizado pelos pesquisadores. Trabalhar com  $\gamma(h)$  ao invés de  $C(h)$  tem uma grande vantagem: se o processo é intrínseco mas não estacionário de segunda de ordem, então  $C(h) \neq C[Z(s), Z(s+h)]$ , logo deve-se trabalhar com  $\gamma(h)$ . Veja que  $C(h)$  expressa a similaridade dos valores e  $\gamma(h)$  a dissimilaridade desses.

O gráfico das semivariâncias em relação as distâncias  $h$  recebe o nome de semivariograma.

### 2.1.1 Semivariograma

O conhecimento do semivariograma é fundamental para desenvolver as predições de valores para locais não amostrados. Para que os valores preditos sejam mais próximos dos valores reais, espera-se que o semivariograma represente

o comportamento real da dependência espacial envolvida.

A função de semivariância é definida por

$$\gamma(h) = \frac{\text{Var} [Z(s) - Z(s+h)]}{2} = \frac{E \{ [Z(s) - Z(s+h)]^2 \}}{2}, \quad (1)$$

em que a segunda igualdade vale se a média for constante.

Na expressão (1), a variância está sendo dividida por dois, então se utiliza o prefixo “semi” para distinguir da variância e daí vem o nome semivariância para  $\gamma(h)$  e semivariograma para o gráfico de  $\gamma(h)$  em função de  $h$ . Analisando a expressão (2.1), quanto mais próximos estiverem os pontos, maior a semelhança entre eles e, portanto, menor a semivariância; e quanto mais distantes, menor a semelhança e, conseqüentemente, maior a semivariância.

Na Figura 3 ilustra-se um semivariograma com suas características.

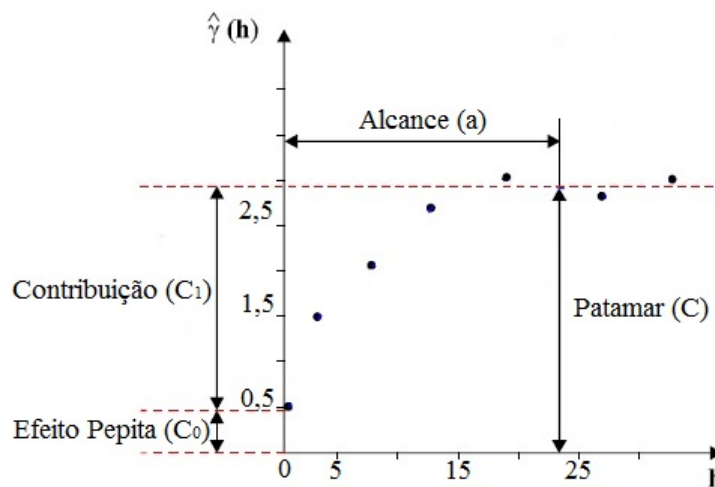


Figura 3 Exemplo de semivariograma com suas características

Observando a Figura 3, tem-se que:

1. *Alcance ( $a$ )*: é a distância  $h$  até a semivariância  $\gamma(h)$  atingir um valor aproximadamente constante, sendo que dentro desse intervalo  $[0, h]$  implica que as amostras apresentam dependência espacial;
2. *Efeito Pepita ( $C_0$ )*: na teoria para a distância  $h = 0$  tem-se que a semivariância  $\gamma(0) = 0$ . O uso de dados amostrais na estimativa da semivariância e na construção do semivariograma, revela que, frequentemente, a semivariância  $\gamma(0)$  difere de zero. A impossibilidade de se fazer reamostragem exatamente sobre um ponto já amostrado, podendo ocorrer variações a distâncias menores do que a menor distância entre as amostras e erros de medição, são justificativas dessa descontinuidade na origem. Quando  $\gamma(0)$  difere de 0 surge um novo termo chamado de efeito pepita ( $C_0$ );
3. *Contribuição ( $C_1$ )*: é a diferença entre o patamar ( $C$ ) e o Efeito Pepita ( $C_0$ );
4. *Patamar ( $C$ )*: as medições realizadas a distâncias maiores que  $a$  não apresentam dependência espacial entre as amostras, com variações que não são justificadas pela semelhança de um ponto com outro, sendo independentes entre si. O valor de  $\gamma(h)$  é constante sendo chamado de patamar  $C = C_0 + C_1$ . Para essas amostras, a Estatística Clássica pode ser aplicada sem restrições.

Destaca-se que  $a$  é a divisão para aplicação de Geoestatística ou Estatística Clássica e por isso o cálculo do semivariograma deveria ser feito rotineiramente, garantindo as hipóteses estatísticas sob as quais são analisados.

De acordo com Yamamoto e Landim (2013), o efeito pepita reflete a incerteza em pequenas distâncias, sendo que, quanto maior esse efeito, maior a variabilidade. Para diminuir os valores do efeito pepita é necessário que a amostragem

seja realizada a distâncias menores que a utilizada para que assim se possa detectar a estrutura da variância, ou seja, a escala de variabilidade natural do fenômeno (JOURNEL; HUIJBREGTS, 1978).

Se o semivariograma for constante e igual ao patamar para qualquer valor de  $h$ , recebe o nome de efeito pepita puro. Nesse caso, tem-se a ausência total de dependência espacial e, se existir, manifesta-se a distâncias menores do que o menor espaçamento entre amostras. Um outro tipo de semivariograma é aquele que apresenta a semivariância com flutuações, chamado de semivariograma periódico ou cíclico. Também pode-se ter um semivariograma sem patamar definido, em que as semivariâncias crescem, sem limites, para todos os valores de  $h$ , indicando que a hipótese de estacionaridade de segunda ordem não foi atendida e que a máxima distância  $h$  entre as amostras não foi capaz de exibir toda a variabilidade dos dados, provavelmente existindo tendência dos dados para determinada direção.

Se for verificada tendência, deve-se realizar a modelagem e trabalhar com os resíduos. Logo após, constrói-se novamente o semivariograma e verifica-se se esse apresenta patamar (estacionaridade de segunda ordem).

A distância  $h$  é encontrada por meio de um vetor, sendo assim, o semivariograma depende de sua magnitude e direção. Quando o semivariograma é idêntico para qualquer direção de  $h$ , recebe-se o nome de isotrópico e quando o semivariograma é diferente dependendo da direção de  $h$ , recebe-se o nome de anisotrópico.

Para realizar a construção do semivariograma, utiliza-se um estimador para semivariância. Dentre os mais utilizados, destacam-se: estimador clássico de Matheron (1962) e de Cressie e Hawkins (1980).

O estimador clássico de Matheron (1962) é dado por:

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{N(h)} [z(s_i) - z(s_j)]^2,$$

em que  $\hat{\gamma}(h)$  é o estimador da semivariância para distância  $h$ ,  $N(h)$  são os pares separados por  $h$ , sendo  $N(h) = \{(s_i, s_j), \|s_i - s_j\| = h; i, j = 1, 2, \dots, n, i \neq j\}$  e  $|N(h)|$  a cardinalidade de  $N(h)$ .

O estimador clássico de Matheron é pouco robusto sendo afetado por observações atípicas (*outliers*), sobretudo aquelas que produzem distribuições de caudas pesadas (assimétricas, leptocúrticas). Nesses casos, é adequado o uso de um estimador que não seja afetado por pontos discrepantes, que não possam ser eliminados ou substituídos (RIBEIRO JR., 1995). Visando dados que não tem distribuição normal de probabilidades e com presença de *outliers*, Cressie e Hawkins (1980) propuseram um estimador robusto, dado por:

$$\hat{\gamma}(h) = \frac{\left[ \frac{1}{2|N(h)|} \sum_{N(h)} |z(s_i) - z(s_j)|^{\frac{1}{2}} \right]^4}{\left( 0,457 + \frac{0,494}{|N(h)|} \right)}.$$

Segundo Oliveira (1991), o cálculo de semivariâncias não é algo simples e direto, pois é bastante sensível à distribuição dos pontos amostrais, bem como ao tipo de distribuição estatística associada. Com relação à distribuição espacial dos pontos, ela pode ser regular ou irregular.

Quando a distribuição é regular, o semivariograma pode ser calculado diretamente com base nos pontos amostrais. Neste caso a malha pode ser quadrada ou retangular. Quando é quadrada têm-se as direções correspondentes aos ângulos  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ,  $180^\circ$ ,  $225^\circ$ ,  $270^\circ$  e  $315^\circ$ . Quando é retangular têm-se as direções correspondentes aos ângulos  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$  e os ângulos das diagonais que precisam ser calculados com base nos lados do retângulo.

Quando a distribuição é irregular, há necessidade de se definir parâmetros adicionais, para que a malha de pontos seja regularizada. Para cada ponto, define-se uma janela, dentro da qual pode haver nenhum, um ou mais pontos. Essa janela é definida pela direção, tolerância angular e largura máxima, bem como pelo tamanho do passo (distância) e tolerância do passo. O parâmetro largura máxima limita a abertura indefinida da janela de pesquisa dada pela tolerância angular.

O gráfico do semivariograma é formado por uma série de estimativas de  $\gamma(h)$ , conforme ilustra a Figura 2.1. Nesse gráfico, objetiva-se ajustar um modelo que seja representativo para que as predições obtidas a partir da krigagem sejam mais confiáveis, fornecendo informações corretas do fenômeno em estudo. Como regra, quanto mais simples puder ser o modelo ajustado, melhor, e não se deve dar importância excessiva a pequenas flutuações que podem ser causadas por uma pequena cardinalidade do número de pares de uma determinada distância.

### 2.1.2 Modelos de semivariograma

Dentre vários modelos de semivariograma existentes, adotam-se no presente trabalho os modelos: Exponencial, Esférico, Gaussiano e Wave.

#### Modelo Exponencial

Sua equação é dada por:

$$\gamma(h) = \begin{cases} 0, & \text{se } h = 0, \\ C_0 + C_1 \left( 1 - e^{-\frac{h}{a}} \right), & \text{se } 0 < h < \infty. \end{cases}$$

### Modelo Esférico

Sua equação é dada por:

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

### Modelo Gaussiano

Sua equação é dada por:

$$\gamma(h) = \begin{cases} 0, & \text{se } h = 0, \\ C_0 + C_1 \left[ 1 - e^{-\left( \frac{h}{a} \right)^2} \right], & \text{se } 0 < h < \infty. \end{cases}$$

### Modelo Wave

Sua equação é dada por:

$$\gamma(h) = \begin{cases} 0, & \text{se } h = 0, \\ C_0 + C_1 \frac{a}{h} \text{sen} \left( \frac{h}{a} \right), & 0 < h < \infty. \end{cases}$$

Observe que, com exceção do modelo esférico, os demais atingem o patamar apenas assintoticamente.

Para realizar o ajuste de cada modelo, diversos métodos são apresentados na literatura. Dentre os mais utilizados, destacam-se os de quadrados mínimos

(ordinários, ponderados e generalizados). Esses métodos têm como base o estimador de semivariância, os quais buscam parâmetros para o modelo que minimizem a soma dos quadrados das diferenças entre os valores observados e estimados de semivariância. Maiores detalhes sobre os métodos destacados podem ser encontrados em Cressie (1993). De acordo com Ribeiro Jr. (1995), espera-se que o ajuste seja eficiente, principalmente, para pequenas distâncias. Além disso, é frequente que as semivariâncias para as maiores distâncias sejam estimadas com menor número de pares, tendo uma menor precisão. Segundo Isaaks e Srivastava (1989), esse fato justifica a especificação de uma distância máxima para qual a semivariância é estimada.

### 2.1.3 Importância do modelo de semivariograma

Na análise geoestatística, objetiva-se gerar um mapa da área em estudo que apresente valores de uma variável para qualquer posição. Para realizar a predição dos valores em locais não amostrados, pode-se utilizar o método de krigagem. Esse método utiliza a dependência espacial expressa no modelo de semivariograma para realizar a predição. Para que as predições sejam as mais próximas possíveis dos valores verdadeiros, a seleção de um modelo adequado é fundamental.

Dentre vários tipos de krigagem, a ordinária é a mais utilizada, sendo um método local de predição onde um ponto não amostrado resulta da combinação linear dos valores encontrados na vizinhança próxima. Seu preditor é dado por:

$$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i),$$

em que  $s_0$  é o ponto não amostrado,  $\lambda_i$  é o peso do  $i$ -ésimo dado e  $Z(s_i)$  é o valor da variável estudada no  $i$ -ésimo ponto.

Os pesos ótimos são calculados sob as condições de que o preditor seja não viesado e de variância do erro de predição mínima.

De acordo com Journel e Huijbregts (1978), a condição de não viés da estimativa significa que, em média, a diferença entre os valores observados e preditos para o mesmo ponto deve ser igual a zero,

$$E \left[ \hat{Z}(s_0) - Z(s_0) \right] = 0. \quad (2)$$

Desenvolvendo a expressão (2), chega-se à condição de não viés:

$$\begin{aligned} E \left[ \hat{Z}(s_0) - Z(s_0) \right] &= E \left[ \hat{Z}(s_0) \right] - E \left[ Z(s_0) \right] \\ \Rightarrow E \left[ \hat{Z}(s_0) \right] - E \left[ Z(s_0) \right] &= 0 \\ \Rightarrow E \left[ \hat{Z}(s_0) \right] &= E \left[ Z(s_0) \right] \\ \Rightarrow E \left[ \sum_{i=1}^n \lambda_i Z(s_i) \right] &= E \left[ Z(s_0) \right] \\ \Rightarrow \sum_{i=1}^n \lambda_i \underbrace{E \left[ Z(s_i) \right]}_{\mu} &= \underbrace{E \left[ Z(s_0) \right]}_{\mu} \\ \Rightarrow \sum_{i=1}^n \lambda_i \mu &= \mu \\ \Leftrightarrow \sum_{i=1}^n \lambda_i &= 1. \end{aligned}$$

Logo, com a condição de não viés, tem-se que os pesos ótimos devem satisfazer:

$$\sum_{i=1}^n \lambda_i = 1.$$

A condição de variância do erro de predição seja mínima significa que, em-

bora possam existir diferenças ponto por ponto entre o valor predito e o observado, essas diferenças devem ser mínimas.

A variância do erro de predição  $\sigma_E^2(s_0)$  é calculada como:

$$\sigma_E^2(s_0) = Var \left[ Z(s_0) - \hat{Z}(s_0) \right].$$

Quando utiliza-se o preditor de krigagem  $\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i)$ , a variância do erro de predição recebe o nome de variância de krigagem, sendo dada por:

$$\sigma_k^2(s_0) = \sum_{i=1}^n \lambda_i \gamma(s_i, s_0) + \mu.$$

A minimização de  $\sigma_k^2$  leva em conta a restrição  $\sum_{i=1}^n \lambda_i = 1$  e a técnica utilizada é o Multiplicador de Lagrange. Após a minimização chega-se ao resultado que possibilita encontrar os valores dos pesos ótimos, dado por

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ b \end{bmatrix} = \begin{bmatrix} \gamma(s_1 - s_1) & \gamma(s_1 - s_2) & \dots & \gamma(s_1 - s_n) & 1 \\ \gamma(s_2 - s_1) & \gamma(s_2 - s_2) & \dots & \gamma(s_2 - s_n) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(s_n - s_1) & \gamma(s_n - s_2) & \dots & \gamma(s_n - s_n) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma(s_0 - s_1) \\ \gamma(s_0 - s_2) \\ \vdots \\ \gamma(s_0 - s_n) \\ 1 \end{bmatrix}.$$

De acordo com Gorsich e Genton (2000), a escolha dos pesos ótimos é completamente determinada pelo modelo de semivariograma selecionado e de acordo com Genton (1998), esse modelo também afeta a eficácia dos mapas de krigagem obtidos pela predição espacial linear. Ambas afirmações corroboram a importância da boa adequabilidade do modelo de semivariograma selecionado.

### 2.1.4 Avaliação da adequabilidade do modelo de semivariograma

Uma vez estimada a nuvem de pontos do semivariograma e ajustado vários modelos a esta nuvem, de acordo com Guerra (1988), o objetivo fundamental é decidir qual o modelo que mais bem se ajusta ao semivariograma, de tal modo que a partir desse modelo possam ser feitas inferências em relação ao semivariograma verdadeiro. Portanto, após o ajuste do modelo é necessário verificar se ele é adequado em termos estatísticos.

A avaliação da adequabilidade do modelo de semivariograma é um passo essencial no processo de modelagem pois indica o nível de precisão e acurácia das predições do modelo. Essa é uma fase importante para aumentar a confiança nesse modelo ou para permitir seleção de modelos alternativos.

De acordo com Tedeschi (2006), na decisão da seleção de um modelo baseando-se em um método, pode-se levar a dois tipos de erro: erro tipo I e erro tipo II. O erro tipo I é quando rejeita-se um modelo apropriado e o erro tipo II é quando não rejeita-se um modelo inadequado. A ocorrência do erro tipo II pode ser tão mal quanto ou pior do que o erro tipo I.

Na Tabela 1 apresenta-se as decisões possíveis na seleção de um modelo.

Tabela 1 Decisões possíveis na seleção de um modelo

Decisão tomada	Modelo selecionado	Modelo selecionado
	Apropriado	Inadequado
Não rejeita-se	Decisão correta	Decisão errada
	Probabilidade: $1-\alpha$ (Coeficiente de confiança)	Probabilidade: $\beta$ (Erro tipo II)
Rejeita-se	Decisão errada	Decisão correta
	Probabilidade: $\alpha$ (Erro tipo I)	Probabilidade: $1-\beta$ (Poder do teste)

Ambos os tipos de erros podem ser minimizados com uma análise estatística para detectar a adequação do modelo e o uso de observações não viesadas.

### Conceitos de acurácia e precisão

A acurácia mede a capacidade do modelo predizer os valores verdadeiros e a precisão mede a capacidade do modelo predizer valores semelhantes de forma consistente. A Figura 4 ilustra a diferença entre acurácia e precisão usando a analogia da prática do alvo, com a seguinte interpretação: o caso 1 é sem acurácia e sem precisão, o caso 2 é sem acurácia e com precisão, o caso 3 é com acurácia e sem precisão e o caso 4 é com acurácia e com precisão.

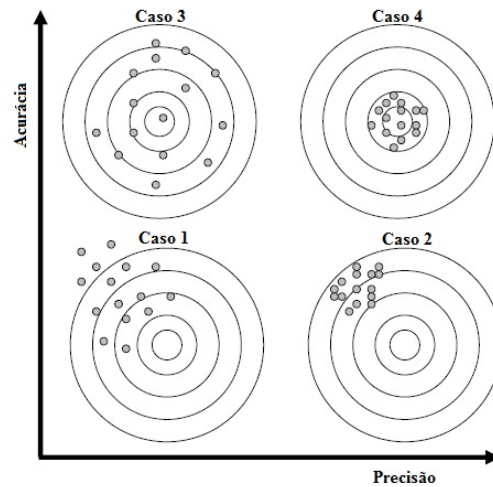


Figura 4 Esquematização da acurácia versus precisão

Em relação à precisão, a estimativa do coeficiente de determinação ( $R^2$ ), apresentado na seção 2.1.6, é um bom indicador, sendo que, quanto maior o  $R^2$  maior a precisão. Em relação a acurácia, as estimativas dos coeficientes linear

$(\beta_0)$  e angular  $(\beta_1)$ , ambos do modelo de regressão linear simples, apresentados na seção 2.1.6, são bons indicadores, sendo que, quanto mais perto de 0 e 1, respectivamente, maior a acurácia. Segundo Tedeschi (2006), o julgamento da adequação do modelo tem que ser feito em comparações relativas da acurácia e precisão.

Para realizar a avaliação da adequabilidade do modelo, uma técnica consagrada na Geoestatística é a autovalidação.

### 2.1.5 Autovalidação

O processo de autovalidação, também referido como *leave-one-out*, tem como objetivo calcular os erros de predição, possibilitando a avaliação da adequabilidade do modelo para o semivariograma. De acordo com Ribeiro Jr. (1995), além de avaliar a adequação do modelo, toda a modelagem do processo estocástico está também sendo avaliada, incluindo-se a decisão da hipótese estacionária, os estimadores utilizados, o tratamento dos valores discrepantes e decisão relativa à anisotropia.

De acordo com Davis (1987), o processo de autovalidação usa apenas informações disponíveis na amostra, sendo da seguinte forma:

1. Selecione um dos modelos de semivariograma que deseja comparar;
2. Seja uma amostra de  $n$  pontos, retire um ponto e realize a sua predição pelo método de krigagem (seção 2.1.4), usando o modelo selecionado e os valores dos outros pontos;
3. Refaça esse procedimento para todos os pontos amostrais;

4. Portanto, para cada  $i$ -ésimo ponto,  $i = 1, 2, \dots, n$ , haverá dois valores:

$Z(s_i)$  (valor observado)

$\hat{Z}(s_i)$  (valor predito)

Com os valores de  $Z(s_i)$  e  $\hat{Z}(s_i)$ , vários métodos de seleção do modelo de semivariograma foram e podem ser propostos. Segundo Gorsich e Genton (2000), a seleção do modelo de semivariograma utilizando os métodos usuais é frequentemente realizada de forma bastante subjetiva.

### 2.1.6 Métodos usuais de seleção do modelo de semivariograma

Após revisar na literatura, foram encontrados os seguintes métodos de seleção:

#### Método A

Teo et al. (1992) e Retta et al. (1996) utilizaram como método de seleção o coeficiente linear  $\beta_0$ , angular  $\beta_1$  e a correlação  $r$ , calculados em relação ao gráfico de  $Z(s_i)$  e  $\hat{Z}(s_i)$ .

Por meio de um gráfico dos valores emparelhados de  $Z(s_i)$  e  $\hat{Z}(s_i)$ , pode-se ajustar uma reta (modelo de regressão linear simples) entre eles

$$\widehat{Z}(s_i) = \hat{\beta}_0 + \hat{\beta}_1 \hat{Z}(s_i),$$

em que  $\hat{\beta}_0$  é a estimativa do coeficiente linear e  $\hat{\beta}_1$  é a estimativa do coeficiente angular da reta ajustada.

O estimador do coeficiente angular  $\beta_1$  é dado por

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n \hat{Z}(s_i)Z(s_i) - \frac{\left[\sum_{i=1}^n \hat{Z}(s_i)\right] \left[\sum_{i=1}^n Z(s_i)\right]}{n}}{\sum_{i=1}^n \left[\hat{Z}(s_i)\right]^2 - \frac{\left[\sum_{i=1}^n \hat{Z}(s_i)\right]^2}{n}}.$$

O estimador do coeficiente linear  $\beta_0$  é dado por

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n Z(s_i)}{n} - \hat{\beta}_1 \frac{\sum_{i=1}^n \hat{Z}(s_i)}{n}.$$

Esses coeficientes podem ser estimados utilizando a função  $lm()$  do *software* R (R CORE TEAM, 2017).

O coeficiente de correlação linear de Pearson ( $r$ ), utilizado para verificar a relação entre os valores de  $Z(s_i)$  e  $\hat{Z}(s_i)$ , é calculado da seguinte forma:

$$r = \frac{\sum_{i=1}^n \hat{Z}(s_i)Z(s_i) - \frac{\left[\sum_{i=1}^n \hat{Z}(s_i)\right] \left[\sum_{i=1}^n Z(s_i)\right]}{n}}{\sqrt{\left\{ \sum_{i=1}^n \left[\hat{Z}(s_i)\right]^2 - \frac{\left[\sum_{i=1}^n \hat{Z}(s_i)\right]^2}{n} \right\} \left\{ \sum_{i=1}^n \left[Z(s_i)\right]^2 - \frac{\left[\sum_{i=1}^n Z(s_i)\right]^2}{n} \right\}}}$$

O  $r$  pode ser calculado usando a função  $cor()$  do *software* R (R CORE TEAM, 2017).

Segundo Dent e Blackie (1979), no ajuste da reta para comparar valores que esperadamente deveriam ser iguais, o ideal é que a reta passe pela origem ( $\beta_0 = 0$ ) e tenha uma inclinação com ângulo de  $45^\circ$  ( $\beta_1 = 1$ ). Assim, se  $\hat{Z}(s_i)$  for idêntico a  $Z(s_i)$ , então  $\hat{\beta}_0 = 0$ ,  $\hat{\beta}_1 = 1$  e  $r = 1$ .

O método de seleção do modelo é aquele que apresentar a estimativa de  $\beta_0$  mais próxima de 0, a de  $\beta_1$  mais próxima de 1 e a de  $r$  mais próxima de 1.

### Método B

Journel e Huijbregts (1978) e Alves et al. (2014) utilizaram como método de seleção o erro médio de predição (EM) e variância do erro de predição (VE).

O estimador do erro médio de predição ( $\widehat{EM}$ ) é dado por

$$\widehat{EM} = \frac{\sum_{i=1}^n [\hat{Z}(s_i) - Z(s_i)]}{n},$$

e o estimador da variância do erro de predição ( $\widehat{VE}$ ) é dado por

$$\widehat{VE} = \frac{\sum_{i=1}^n \left\{ [\hat{Z}(s_i) - Z(s_i)] - \widehat{EM} \right\}^2}{n - 1}.$$

Aplicando-se as condições de não tendência e de variância mínima nos erros de predição, tem-se que

$$EM = E [\hat{Z}(s_i) - Z(s_i)] = 0,$$

$$VE = \text{Var} [\hat{Z}(s_i) - Z(s_i)] = \text{mínimo}.$$

O modelo selecionado é aquele que apresentar as estimativas de  $EM$  e

$VE$  mais próximas de 0.

### Método C

Esse método foi o mais utilizado nos trabalhos revisados. Como exemplo, vejam Castrignanò et al. (2009), Journel e Huijbregts (1978), McBratney e Webster (1986), Cressie (1993) e Vieira (2000).

O estimador do erro médio padronizado ( $\widehat{EMP}$ ) é dado por

$$\widehat{EMP} = \frac{1}{n} \sum_{i=1}^n \left[ \frac{\hat{Z}(s_i) - Z(s_i)}{\hat{\sigma}_k(s_i)} \right],$$

em que  $\hat{\sigma}_k(s_i)$  é a  $i$ -ésima variância de krigagem apresentada na seção 2.1.3.

O estimador da variância do erro de predição padronizada ( $\widehat{VEP}$ ) é dado por

$$\widehat{VEP} = \frac{\sum_{i=1}^n \left\{ \left[ \frac{\hat{Z}(s_i) - Z(s_i)}{\hat{\sigma}_k(s_i)} \right] - \widehat{EMP} \right\}^2}{n - 1}.$$

O método de seleção do modelo é aquele que apresentar a estimativa do  $EMP$  mais próxima de 0 e a estimativa de  $VEP$  mais próxima de 1.

### Método D

Souza Filho (2013) utilizou como método de seleção o erro quadrático médio ( $EQM$ ).

O estimador do erro quadrático médio ( $\widehat{EQM}$ ) é dado por

$$\widehat{EQM} = \frac{\sum_{i=1}^n [\hat{Z}(s_i) - Z(s_i)]^2}{n}.$$

O método de seleção do modelo é aquele que resultar a menor estimativa do  $EQM$ .

### **Método E**

O coeficiente de determinação  $R^2$ , em relação ao modelo de regressão linear simples, é calculado da seguinte forma

$$R^2 = r^2.$$

em que  $r$  foi apresentado no método 1.

O método de seleção do modelo é aquele que apresentar o valor mais próximo de 1.

### **Método F**

Gutiérrez e Miguel (2011), Mc Bratney e Webster (1986) utilizaram como método de seleção o AIC.

O  $AIC$  é estimado por

$$\widehat{AIC} = n \ln \left\{ \sum_{i=1}^n [Z(s_i) - \widehat{Z}(s_i)]^2 \right\} + 2p,$$

em que  $p$  é o número de parâmetros no modelo.

O método de seleção do modelo é aquele que apresentar a menor estimativa de  $AIC$ .

### **Método G**

Mohammadi e Safa (2016) utilizaram como método de seleção o erro combinado ( $EC$ ).

O  $EC$  é definido por eles como

$$EC = 1 - |r| + EQM + |EM|.$$

O método de seleção do modelo é aquele que apresentar a menor estimativa do  $EC$  deve ser o selecionado.

Dado que na Geoestatística diferentes métodos de seleção são utilizados, não havendo uma recomendação, uma pesquisa foi realizada em outras áreas e alguns métodos de seleção foram encontrados.

### **Método H**

A avaliação da adequabilidade de modelo uma questão importante a ser abordada é se os valores preditos pelo modelo são precisos e com acurácia, simultaneamente. Proposto por Lin (1989), o coeficiente de correlação de concordância ( $CCC$ ) é utilizado como método de seleção, avaliando simultaneamente a precisão e acurácia.

O *CCC*, é dado por

$$\hat{\rho}_c = r \frac{2}{\underbrace{\frac{DP_{Z(s_i)}}{DP_{\hat{Z}(s_i)}} + \frac{DP_{\hat{Z}(s_i)}}{DP_{Z(s_i)}} + \left[ \frac{\bar{Z} - \bar{\hat{Z}}}{\sqrt{DP_{Z(s_i)} DP_{\hat{Z}(s_i)}}} \right]^2}_{=C_b}},$$

em que  $DP_{Z(s_i)} = \sqrt{\frac{\sum_{i=1}^n Z(s_i) - \bar{Z}}{n-1}}$ ,  $DP_{\hat{Z}(s_i)} = \sqrt{\frac{\sum_{i=1}^n \hat{Z}(s_i) - \bar{\hat{Z}}}{n-1}}$  e  $r$  é dado no método A na seção 2.1.6.

O primeiro componente,  $r$ , é a estimativa do coeficiente de correlação que mede a precisão e o segundo componente,  $C_b$ , mede a acurácia, sendo o fator de correção de viés que indica até que ponto a reta se desvia do  $\beta_1$  igual a 1.

As propriedades de  $\hat{\rho}_c$  são apresentadas a seguir:

1.  $-1 \leq \hat{\rho}_c \leq 1$ ;
2.  $\hat{\rho}_c = 0 \Leftrightarrow r = 0$ ;
3.  $\hat{\rho}_c = r \Leftrightarrow DP_Y = DP_X$  e  $\bar{X} = \bar{Y}$ ;
4.  $\hat{\rho}_c = \pm 1 \Leftrightarrow r = \pm 1, DP_Y = DP_X$  e  $\bar{X} = \bar{Y}$ .

Como método de seleção, compare as estimativas de  $\rho_c$  referentes aos modelos adotados e selecione o modelo que apresentar a maior estimativa.

O *CCC* tem sido também utilizado na avaliação de dispositivos de medição tais como comparações de métodos químicos e autovalidação de instrumentos ou ensaios (DHANOA ET AL., 1999).

### 2.1.7 Métodos de seleção exógenos à Geoestatística

Alguns métodos utilizados em outras áreas poderão ser aplicados neste trabalho. Os métodos encontrados foram: gráficos de Bland-Altman, desvio quadrático médio particionado em três componentes e análise gráfica via postos.

#### Gráfico de Bland-Altman

Proposto por Bland-Altman (1983), o gráfico é utilizado para comparar vários métodos, com objetivo de avaliar as suas similaridades na Medicina.

Na verificação da concordância entre dois métodos ( $M1$  e  $M2$ ), vários pesquisadores utilizam a seguinte técnica gráfica: consideram os valores do método  $M1$  no eixo X e os valores do método  $M2$  no eixo Y, estabelecem uma reta de igualdade (considerando a mesma escala na horizontal e vertical) tendo um ângulo de  $45^\circ$  para ambos os eixos. Um exemplo é ilustrado na Figura 5:

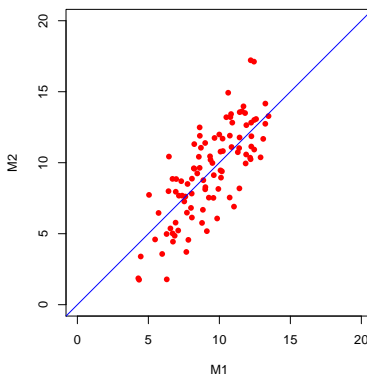


Figura 5 Pontos formados pelos valores de  $M2$  versus  $M1$

Bland e Altman (1983) cita que essa reta é essencial para obter uma avaliação visual correta da relação, mas quando o intervalo de variação das medições é grande em comparação com as diferenças entre os métodos, esse gráfico pode obscurecer informações úteis.

Uma maneira melhor de avaliar a concordância entre dois métodos, considerando variáveis de natureza quantitativa, foi apresentada por Bland e Altman (1983) sendo o primeiro artigo com uma metodologia alternativa ao cálculo do coeficiente de correlação. De acordo com Ryan e Woodall (2005), o segundo artigo, Bland e Altman (1986), destacava-se ocupando o 6º lugar da lista dos artigos da área da estatística mais citados em todos os tempos, considerando dados analisados até o ano de 2003.

Para examinar os dados, a metodologia proposta por Bland e Altman (1983) recomenda, ao invés de plotar o gráfico de um método  $M2$  em relação a um método  $M1$ , plotar o gráfico de dispersão da diferença entre os dois métodos ( $M2 - M1$ ) para cada sujeito ( $d_i = M2_i - M1_i$ ) no eixo Y, em relação a média dos dois métodos  $[(M1 + M2)/2]$  no eixo X. Um exemplo é ilustrado na Figura 6:

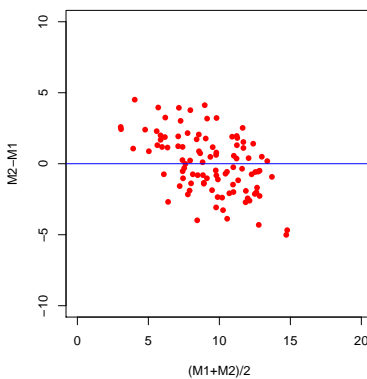


Figura 6 Pontos formados pelos valores de  $M2 - M1$  versus  $(M1 + M2)/2$

Neste gráfico é possível:

1. Testar a hipótese do viés (o quanto as diferenças se afastam do valor zero) ser ou não igual a 0, por meio da média das diferenças  $\bar{d} = \left( \sum_{i=1}^n d_i \right) / n$ , em que  $n$  é o tamanho amostral;
2. Testar a hipótese do coeficiente de correlação ( $r$ ) entre as diferenças absolutas  $|M2 - M1|$  e a média  $(M1 + M2)/2$ , ser ou não igual a 0;
3. Examinar os limites de concordância,  $\bar{d} \pm 1,96S_d$ .

### **Desvio quadrático médio particionado em três componentes**

Proposto por Kobayashi e Salam (2000), utiliza o desvio quadrático médio ( $DQM$ ), também chamado de erro quadrático médio ( $EQM$ ), particionado em três componentes, para comparar valores medidos e simulados ( $M$  e  $S$ ) na computação. O  $DQM$  é dado por

$$DQM = QV + DQDP + FCPDP,$$

em que  $QV$  é o quadrado do viés,  $DQDP$  é a diferença quadrática entre os desvios padrões e  $FCPDP$  é a falta de correlação ponderada pelo desvio padrão.

De acordo com Kobayashi e Salam (2000), um critério mais relevante que testar os coeficientes do modelo de regressão linear simples é trabalhar com o desvio ( $d$ ) entre o valor simulado ( $s$ ) e o valor medido ( $m$ ).

Quando a comparação é feita para um tamanho amostral igual a  $n$ , o  $i$ -ésimo  $d$  é calculado da seguinte forma

$$d_i = m_i - s_i, \quad \text{para } i=1,2,\dots,n,$$

em que  $s_i$  e  $m_i$  são os valores simulados e medidos, respectivamente, para o  $i$ -ésimo dado.

A estatística mais comumente utilizada em relação ao desvio é a raiz do desvio quadrático médio ( $RDQM$ ), dada por

$$RDQM = \sqrt{\frac{1}{n} \sum_{i=1}^n (s_i - m_i)^2}.$$

Elevando ao quadrado o  $RDQM$ , tem-se o desvio quadrático médio ( $DQM$ ), dado por

$$DQM = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2.$$

O  $DQM$  pode ser particionado em dois componentes

$$DQM = \underbrace{(\bar{s} - \bar{m})^2}_{=QV} + \underbrace{\frac{1}{n} \sum_{i=1}^n [(s_i - \bar{s}) - (m_i - \bar{m})]^2}_{=VQM}.$$

em que  $\bar{s}$  e  $\bar{m}$  são as médias dos  $s_i$  e  $m_i$  ( $i = 1, 2, \dots, n$ ), respectivamente, e  $VQM$  é o valor médio da variação centralizada na média ao quadrado.

O coeficiente de correlação entre os valores simulados e observados é dado por

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (s_i - \bar{s})(m_i - \bar{m})}{DP_S DP_M}.$$

Com esse coeficiente, o  $VQM$  pode ser particionado em dois componentes, ou seja

$$\begin{aligned}
 VQM &= \frac{1}{n} \sum_{i=1}^n [(s_i - \bar{s}) - (m_i - \bar{m})]^2 \\
 &= \frac{1}{n} \sum_{i=1}^n [(s_i - \bar{s})^2 - 2(s_i - \bar{s})(m_i - \bar{m}) + (m_i - \bar{m})^2] \\
 &= \frac{1}{n} \sum_{i=1}^n (s_i - \bar{s})^2 - \frac{2}{n} \sum_{i=1}^n (s_i - \bar{s})(m_i - \bar{m}) + \frac{1}{n} \sum_{i=1}^n (m_i - \bar{m})^2 \\
 &= DP_S^2 - 2rDP_SDP_M + DP_M^2 \\
 &= DP_S^2 - 2DP_SDP_M + DP_M^2 + 2DP_SDP_M - 2rDP_SDP_M \\
 &= \underbrace{(DP_S - DP_M)^2}_{DQDP} + \underbrace{2DP_SDP_M(1-r)}_{FCPDP},
 \end{aligned}$$

em que  $DQDP$  é a diferença quadrática entre os desvios padrões e  $FCPDP$  é a falta de correlação ponderada pelo desvio padrão.

Logo, tem-se

$$\begin{aligned}
 VQM &= DQDP + FCPDP \\
 \Rightarrow DQM &= QV + DQDP + FCPDP.
 \end{aligned}$$

A comparação dos valores simulados e observados utilizando o  $DQM$  é direta, com seus componentes representando diferentes aspectos do desvio.

### **Método de análise gráfica via postos**

Proposto por Laslett et al. (1987), utilizado para avaliar a similaridade de

vários métodos, por meio de um método de análise gráfica via postos.

O procedimento executado por Laslett et al. (1987) foi:

1. Seleciona-se  $n$  métodos, com objetivo de avaliar suas similaridades, amostrando  $m$  locais;
2. Considera-se postos de 1 a  $n$  e classifica-se as estimativas de cada local de acordo com a ordem crescente, conforme a Tabela 2:

Tabela 2 Estimativas classificadas dos métodos selecionados, com as médias e os desvios padrões dos postos calculados

Local	Método			
	1	2	...	n
1	3,69	3,41	...	2,92
Posto	11	10	...	4
2	8,42	4,00	...	2,22
Posto	$n$	5	...	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$m$	3,98	4,06	...	7,30
Posto	3	4	...	8

3. Calculou a média ( $\overline{\text{Posto}}$ ) e o desvio padrão  $DP_{\text{Posto}}$  dos postos para cada modelo,

$$\overline{\text{Posto}}_{\text{Método } i} = \sum_{j=1}^m \frac{\text{Posto}_{ij}}{m},$$

$$DP_{\text{PostoMétodo } i} = \sqrt{\sum_{j=1}^m \frac{(\text{Posto}_{ij} - \overline{\text{Posto}}_{\text{Método } i})^2}{m-1}}.$$

4. Plotou um gráfico dos pontos formados pelos desvios padrões em relação as médias dos postos conforme a Figura 7:

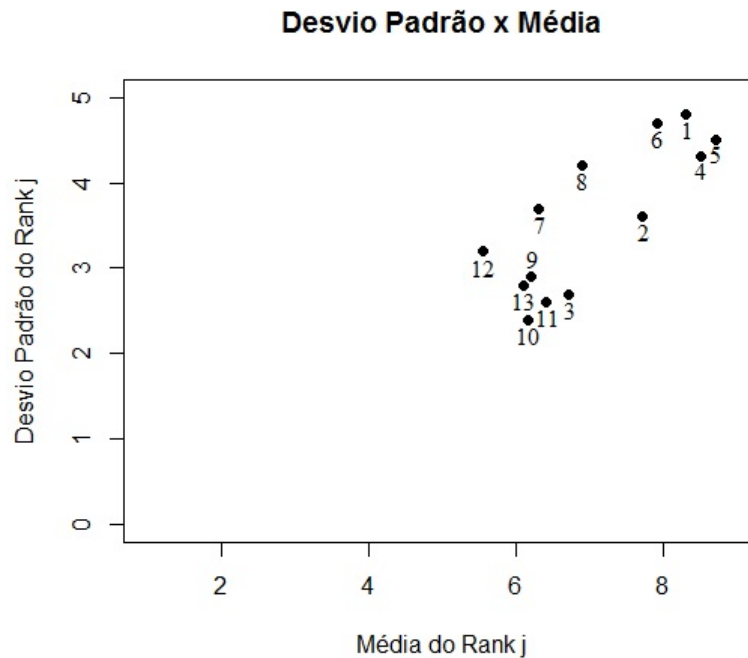


Figura 7 Pontos formados pelos desvios padrões versus médias dos postos

5. Por meio da Figura 7 realiza-se uma análise informando as similaridades dos métodos, ou seja, aqueles que apresentam comportamento parecidos.

O uso de ferramentas computacionais para auxiliar a aplicação de técnicas estatísticas possibilita que análises gráficas e métodos numéricos que necessitariam de maior esforço sejam realizados, como no caso de uma grande quantidade de dados e de cálculos laboriosos.

A estatística computacional ocupa lugar de destaque na solução de problemas que envolvem grande complexidade, por exemplo, situações em que a distri-

buição dos dados é desconhecida, na avaliação de testes e de métodos de estimação, na simulação de dados de modelos que possuem componentes estocásticos e em outras diversas circunstâncias.

Uma técnica que vêm desempenhando um papel cada vez mais importante para resolver problemas em diversas áreas é a simulação.

## **2.2 Simulação**

Segundo Dachs (1988), a simulação tem como objetivo imitar o funcionamento de um sistema real, averiguando o que aconteceria se fossem feitas alterações no seu funcionamento. Esse tipo de procedimento pode ter um custo baixo, evitar prejuízos por não utilizar procedimentos inadequados e otimizar a decisão e o funcionamento do sistema real.

Segundo Ferreira (2013), quando é realizado a simulação devem ser tomadas precauções contra erros, tais como: escolha inadequada das distribuições, simplificação inadequada da realidade e erros de implementação. Para que a simulação seja realizada corretamente, deve-se fazer o sistema simulado operar nas condições do sistema real, verificando por meio de testes se os resultados que são observados estão coerentes com os que ocorrem no sistema real. Para o uso da simulação ser justificável, a solução alcançada deve ser obtida mais rapidamente, com menor custo e de fácil interpretação em relação a outro método qualquer.

### **Simulação de campos aleatórios**

A simulação de dados espaciais é importante por várias razões. Dentre elas, pode-se destacar que o valor de um método estatístico pode ser estabelecido de modo convincente apenas com o método exibindo comportamento satisfatório.

De acordo com Cressie (1993), validar métodos sobre dados dos quais se conhecem os verdadeiros parâmetros pode ser uma parte essencial para reconhecer a sua importância. Um requisito para realizar simulação é o de que todos os parâmetros do processo sejam especificados sendo que a fonte de valores aleatórios é, usualmente, um gerador de números pseudoaleatórios.

### **Simulação de campos aleatórios gaussianos**

O campo aleatório Gaussiano ocupa uma posição central na teoria da análise de dados espaciais, seguindo a mesma linha de que a distribuição Gaussiana é a principal nas abordagens clássicas de inferência estatística. As propriedades estatísticas de estimadores derivados da forma Gaussiana são fáceis de analisar e as estatísticas de teste geralmente têm conhecida uma distribuição simples. Assim, a criação de realizações desse tipo de campo é importante.

O *software* R (R CORE TEAM, 2017) é apropriado para realizar as simulações computacionalmente intensivas, pois é muito fácil e eficiente obter as implementações de funções. Esse *software* é um dos mais acessíveis de análise estatística do mundo, por ser livre (gratuito e com código fonte aberto), além de ser multiplataforma (VENABLES; RIPLEY, 2002). O *download* desse *software* pode ser realizado gratuitamente a partir do endereço <http://www.cran.r-project.org/> sendo possível também obter o manual e orientações de auto-ajuda.

A geração de números aleatórios que seguem uma distribuição normal multivariada com a inclusão da dependência espacial que o pesquisador desejar pode ser realizada utilizando a função *grf()* (*Gaussian Random Fields*) do pacote *geoR* (RIBEIRO JR.; DIGGLE, 2001) do *software* R (R CORE TEAM, 2017). As análises geoestatísticas podem ser realizadas também utilizando o pacote *geoR*.

### 3 MATERIAL E MÉTODOS

#### 3.1 Material

Por meio do *software* R (R CORE TEAM, 2017), o presente trabalho propõe métodos de seleção, aplica os métodos usuais e propostos em dados simulados, e realiza simulações com objetivo de recomendar um método de seleção.

##### 3.1.1 Dados reais com estrutura de dependência espacial

Os dados reais são referentes ao teor de zinco ( $mg/dm^3$ ) no solo, gentilmente cedidos por Bruno Manoel Rezende de Melo do IFSULDEMINAS, Campus Inconfidentes-MG. A distribuição espacial dos pontos é dada na Figura 8:

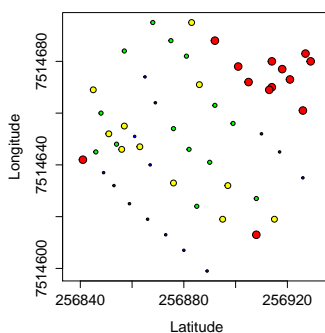


Figura 8 Malha dos dados reais de teor de zinco ( $mg/dm^3$ ) no solo

Os dados foram amostrados na região de Inconfidentes-MG em uma lavoura de café da cultivar Rubi, implantada no espaçamento  $2,2m \times 1,3m$ , em dezembro de 2011. As coletas de solos ocorreram há uma profundidade de 20cm e

foram enviadas ao Laboratório de Análise de Solos do IFSULDEMINAS, Campus Inconfidentes-MG. Observa-se na Figura 8 que foram amostrados 52 pontos, identificados espacialmente por meio da latitude e longitude. A diferença dos níveis do teor de zinco nessa área são visualizadas por meio dos diferentes tamanhos e cores dos círculos, sendo que quanto maior, maior o nível.

### 3.1.2 Dados simulados com estrutura de dependência espacial

Para realizar o estudo dos métodos de seleção, os dados analisados neste trabalho foram gerados por meio da função *grff()* do pacote *geoR* (RIBEIRO JR.; DIGGLE, 2001), seguindo uma distribuição normal multivariada com a inclusão de quatro tipos de dependência espacial: esférica, exponencial, gaussiana e wave. Os valores paramétricos utilizados na inclusão da dependência espacial esférica, exponencial e gaussiana foram  $C_0=0$ ,  $C_1=10$ ,  $a=30$  e  $C=10$ , e da dependência espacial wave foram  $C_0=0$ ,  $C_1=10$ ,  $a=5$  e  $C=10$ . Para uma maior compreensão, na Figura 9 tem-se o modelo esférico com os valores paramétricos adotados.

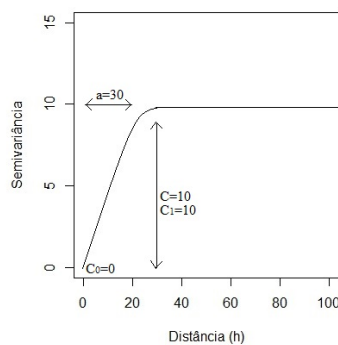


Figura 9 Modelo esférico considerando  $C_0=0$ ,  $C_1=10$ ,  $a=30$  e  $C=10$

A inclusão das dependências espaciais podem ser apresentadas algebricamente, da seguinte forma:

$$\vec{Z}(s) = [Z(s_1), Z(s_2), \dots, Z(s_n)]' \sim N_n(\vec{\mu}, \Sigma),$$

em que  $\vec{Z}(s)$  é o vetor contendo  $n$  variáveis  $Z(s_i)$ ,  $\vec{\mu}$  é o vetor de médias e  $\Sigma$  a matriz de covariâncias, ambos dados por:

$$\vec{\mu} = [\mu_1, \mu_2, \dots, \mu_n]' = [10, 10, \dots, 10]',$$

e

$$\Sigma = \begin{bmatrix} \text{Var}[Z(s_1)] = 10 & C[Z(s_1), Z(s_2)] & \dots & C[Z(s_1), Z(s_n)] \\ C[Z(s_2), Z(s_1)] & \text{Var}[Z(s_2)] = 10 & \dots & C[Z(s_2), Z(s_n)] \\ \vdots & \vdots & \ddots & \vdots \\ C[Z(s_n), Z(s_1)] & C[Z(s_n), Z(s_2)] & \dots & \text{Var}[Z(s_n)] = 10 \end{bmatrix}$$

em que:

$$\begin{aligned} C[Z(s_i), Z(s_j)] &= C(s_i - s_j) \\ &= \text{Var}[Z(s)] - \gamma_{\text{Modelo}}(s_i - s_j) \\ &= 10 - \gamma_{\text{Modelo}}(s_i - s_j), \end{aligned}$$

### 1. Considerando modelo esférico

$$\gamma_{\text{Esférico}}(s_i - s_j) = C_0 + C_1 \left[ \frac{3}{2} \left( \frac{s_i - s_j}{a} \right) - \frac{1}{2} \left( \frac{s_i - s_j}{a} \right)^3 \right],$$

$$0 \leq |s_i - s_j| \leq a,$$

$$\gamma_{\text{Esférico}}(s_i - s_j) = C_0 + C_1, |s_i - s_j| > a.$$

Tomando  $C_0 = 0$ ,  $C_1 = 10$  e  $a = 30$ , tem-se que:

$$\gamma_{\text{Esférico}}(s_i - s_j) = \frac{s_i - s_j}{2} - 5 \left( \frac{s_i - s_j}{30} \right)^3, 0 \leq |s_i - s_j| \leq 30,$$

$$\gamma_{\text{Esférico}}(s_i - s_j) = 10, |s_i - s_j| > 30.$$

## 2. Considerando modelo exponencial

$$\gamma_{\text{Exponencial}}(s_i - s_j) = C_0 + C_1 \left( 1 - e^{-\frac{s_i - s_j}{a}} \right), 0 \leq |s_i - s_j| < \infty.$$

Tomando  $C_0 = 0$ ,  $C_1 = 10$  e  $a = 30$ , tem-se que:

$$\gamma_{\text{Exponencial}}(s_i - s_j) = 10 \left( 1 - e^{-\frac{s_i - s_j}{30}} \right), 0 \leq |s_i - s_j| < \infty.$$

## 3. Considerando modelo gaussiano

$$\gamma_{\text{Gaussiano}}(s_i - s_j) = C_0 + C_1 \left[ 1 - e^{-\left( \frac{s_i - s_j}{a} \right)^2} \right], 0 \leq |s_i - s_j| < \infty.$$

Tomando  $C_0 = 0$ ,  $C_1 = 10$  e  $a = 30$ , tem-se que:

$$\gamma_{\text{Gaussiano}}(s_i - s_j) = 10 \left[ 1 - e^{-\left( \frac{s_i - s_j}{30} \right)^2} \right], 0 \leq |s_i - s_j| < \infty.$$

#### 4. Considerando modelo wave

$$\gamma_{\text{Wave}}(s_i - s_j) = C_0 + C_1 \frac{a}{s_i - s_j} \text{sen} \left( \frac{s_i - s_j}{a} \right), 0 \leq |s_i - s_j| < \infty.$$

Tomando  $C_0 = 0$ ,  $C_1 = 10$  e  $a = 5$ , tem-se que:

$$\gamma_{\text{Wave}}(s_i - s_j) = \frac{50}{s_i - s_j} \text{sen} \left( \frac{s_i - s_j}{5} \right), 0 \leq |s_i - s_j| < \infty.$$

Nas simulações foram utilizadas malhas quadradas regulares de 100 pontos, com espaçamento entre os pontos de 10 unidades, tanto no sentido do eixo  $x$  quanto do  $y$ .

Na Figura 10, apresenta-se um conjunto de dados simulados considerando o modelo esférico para a dependência espacial.

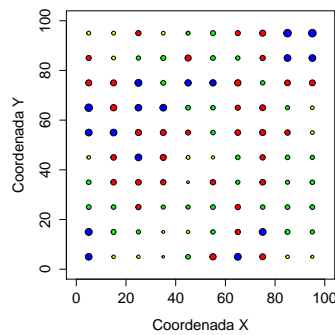


Figura 10 Malha de um exemplo de dados simulados com dependência espacial esférica

## 3.2 Métodos

### 3.2.1 Métodos utilizados na análise geoestatística dos dados de teor de zinco

Nos dados de teor de zinco foi modelada a tendência de grau 2. Para construção do semivariograma foi utilizado o estimador clássico de Matheron (1962). Na obtenção do semivariograma, foi estabelecida a distância máxima de 60% por meio do comando `max.dist` da função `variog()` do pacote `geoR` (RIBEIRO JR.; DIGGLE, 2001) e foi especificado o número de pares mínimos de uma determinada distância igual a 5 utilizando o comando `pairs.min` também dessa função. Em relação aos modelos ajustados ao semivariograma, foram escolhidos: Wave, Exponencial, Gaussiano e Esférico.

### 3.2.2 Métodos utilizados na análise geoestatística dos dados simulados

Para construir os semivariogramas referentes as dependências espaciais inclusas, as semivariâncias foram estimadas por meio do estimador clássico de Matheron (1962).

Com objetivo de estudar a interferência da variabilidade entre as semivariâncias nas maiores distâncias dos semivariogramas na seleção do modelo que mais bem se ajusta, foram consideradas variabilidade alta, média e baixa. Para encontrar as especificações da distância máxima que resultem os tipos de variabilidades desejadas, foram construídos os semivariogramas em relação as especificações de 100%, 95%, 90%, 80%, 75% e 70% da distância máxima.

Na Figura 11 tem-se os semivariogramas em relação a essas especificações, considerando dados simulados com dependência espacial esférica.

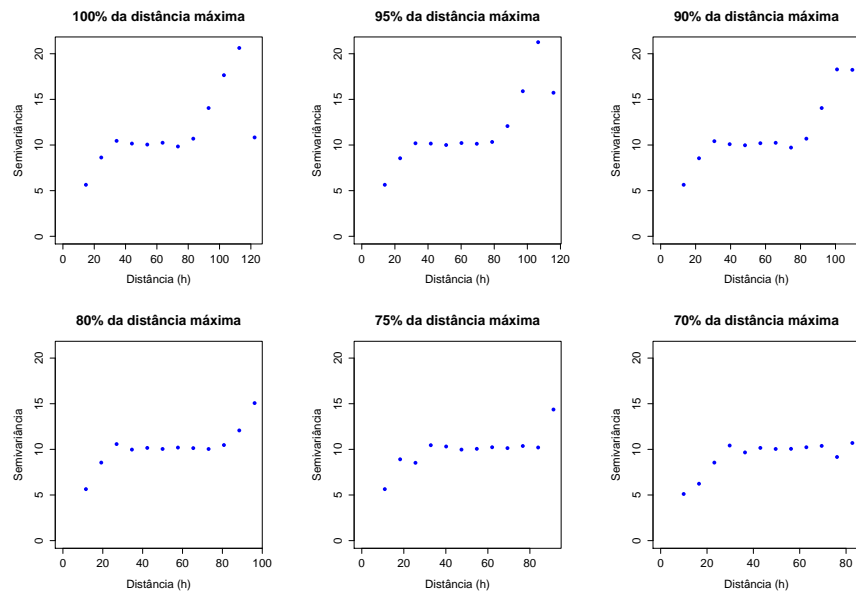


Figura 11 Semivariogramas referentes aos dados simulados com dependência espacial esférica, com 100%, 95%, 90%, 80%, 75% e 70% da distância máxima

Por meio desses semivariogramas torna-se possível especificar as três distâncias máximas que atendem as variabilidades desejadas. De acordo com a Figura 11, as especificações de 100%, 80% e 70% possibilitam realizar o estudo da variabilidade alta, média e baixa nas maiores distâncias do semivariograma, respectivamente. Essas especificações foram adotadas para os quatro tipos de dependências espaciais inclusas no presente trabalho.

Com os quatro modelos de semivariograma escolhidos, com os quatro tipos de dependência espacial inclusa na simulação e com as três especificações da distância máxima do semivariograma, possibilita-se o estudo de diferentes cenários para verificar o comportamento dos métodos usuais e propostos de seleção.

Para realizar o ajuste dos modelos aos semivariogramas referentes a cada tipo de dependência espacial inclusa, foi utilizado o método de mínimo quadrados ordinários.

Para aplicar os métodos usuais e propostos, para cada  $i$ -ésima coordenada do valor observado  $Z(s_i)$ , foi encontrado seu  $i$ -ésimo valor predito  $\hat{Z}(s_i)$  por meio da autovalidação. As predições dos  $i$ -ésimos valores  $\hat{Z}(s_i)$  realizadas na autovalidação foram utilizando a krigagem ordinária com informações do modelo ajustado.

Após realizar a autovalidação, para cada coordenada do  $i$ -ésimo valor observado  $Z(s_i)$  tem-se o  $i$ -ésimo valor predito  $\hat{Z}(s_i)$ . Esses valores podem ser apresentados graficamente por pontos, sendo que cada  $i$ -ésimo ponto  $p_i$  é dado por:

$$p_i = \left( \hat{Z}(s_i), Z(s_i) \right).$$

Nesses pontos podem-se ajustar infinitas retas, dentre elas:

$$Z(s_i) = \beta_0 + \beta_1 \hat{Z}(s_i) + \varepsilon_i \left\{ \begin{array}{l} \widehat{Z}(s_i) = \hat{\beta}_0 + \hat{\beta}_1 \hat{Z}(s_i) + \varepsilon'_i \text{(Reta ajustada)} \\ \widetilde{Z}(s_i) = \beta_0 + \beta_1 \hat{Z}(s_i) + \varepsilon''_i \text{(Reta paramétrica)} \\ \underbrace{Z(s_i)} = 0 + 1\hat{Z}(s_i) + \varepsilon'''_i \text{(Reta ideal)}. \end{array} \right.$$

em que  $\varepsilon_i, \varepsilon''_i$  e  $\varepsilon'''_i$  são os  $i$ -ésimos erros.

A seguir, foram propostos métodos para seleção do modelo que mais bem se ajusta ao semivariograma.

### 3.2.3 Proposição de métodos de seleção do modelo de semivariograma

Perante a ausência de um método recomendado para proceder a seleção de um modelo de semivariograma, foram propostos 7 métodos utilizando diferentes técnicas e unindo aos métodos usuais, foram aplicados em dados simulados perante diferentes cenários, com objetivo de identificar os que irão apresentar melhores resultados.

#### Proposição do método 1

O método 1 consiste em utilizar uma ferramenta eficiente na avaliação da variabilidade nos Gráficos de Controle Estatístico de Processo, apresentada por Montgomery (2009), acrescentando um intervalo de confiança em relação a  $Z(s)$ .

Mitchell e Sheenhy (1997) propuseram um método de avaliação baseado na análise de desvios traçados em função dos valores observados e verificado a porcentagem de pontos dentro de um intervalo (envelope), o que é utilizado como critério de adequação do modelo. A faixa aceitável é baseada na finalidade do modelo, geralmente são especificados os limites do intervalo com 95% de confiança.

O uso de intervalos com limites especificados por  $1,96\sigma$  garante uma confiança de 95% considerando a distribuição Normal. Dado que o campo aleatório simulado é Normal, cada variável segue uma distribuição Normal. Logo, o intervalo com 95% de confiança para  $Z(s_i)$ , também chamado de intervalo com 95% de confiança para medida individual, é dado por

$$IC (Z(s_i))_{95\%} : z(s_i) \pm 1,96\hat{\sigma},$$

em que  $\hat{\sigma}$  é o estimador da medida de variabilidade adotada.

A medida de variabilidade adotada foi a amplitude móvel ( $AM$ ).  
Sendo assim,

$$\begin{aligned}
 IC(Z(s_i))_{95\%} &: z(s_i) \pm 1,96\hat{\sigma} \\
 &: z(s_i) \pm 1,96 \frac{\overline{AM}_{\text{Min}}}{d_2(2)} \\
 &: z(s_i) \pm 1,96 \frac{\sum_{j=1}^{n-1} AM_j}{(n-1)d_2(2)} \\
 &: z(s_i) \pm 1,96 \frac{\sum_{j=1}^{n-1} |z(s_{j+1}) - z(s_j)|}{(n-1)d_2(2)},
 \end{aligned}$$

em que  $\overline{AM}_{\text{Min}}$  é a menor amplitude móvel média em relação aos modelos adotados,  $d_2(2) = 1,128$  é o fator de correção de viés e  $z(s_j)$  é o valor observado referente a classificação do valor predito ordenado de forma crescente.

A Figura 12 apresenta um exemplo desse intervalo de confiança.

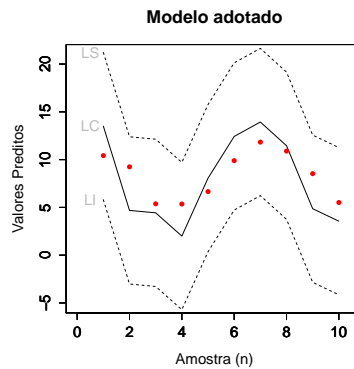


Figura 12 Exemplo de intervalo de confiança para  $Z(s_i)$

em que

$\bullet = \hat{z}(s_i)$  é o  $i$ -ésimo valor predito;

$LC_i = z(s_i)$  é a Linha Central;

$LI_i = z(s_i) - 1,96 \frac{\sum_{j=1}^{n-1} |z(s_{j+1}) - z(s_j)|}{(n-1)d_2(2)}$  é o Limite Inferior;

$LS_i = z(s_i) + 1,96 \frac{\sum_{j=1}^{n-1} |z(s_{j+1}) - z(s_j)|}{(n-1)d_2(2)}$  é o Limite Superior.

O próximo passo é calcular a proporção de pontos referentes aos  $\hat{Z}(s_i)$ 's dentro dos intervalos com 95% de confiança ( $IC(Z(s_i))$ ), sendo dada por

$$\hat{p}_{\text{Modelo}} = \frac{\sum_{i=1}^n I(\hat{Z}(s_i) \in IC(Z(s_i)))}{n},$$

em que  $\hat{p}_{\text{Modelo}}$  é a proporção estimada do modelo,  $I(\hat{Z}(s_i) \in IC(Z(s_i)))$  é a função indicadora que retorna 1 se  $\hat{Z}(s_i)$  pertence ao  $IC(Z(s_i))$  ou retorna 0 se não pertencer.

O modelo selecionado é aquele que resultar a maior proporção, com 95% de confiança.

### Proposição do método 2

O método consiste em acrescentar um intervalo de confiança para o erro de predição  $Z(s) - \hat{Z}(s)$ . A medida de variabilidade adotada foi a variância de krigagem.

Sendo assim,

$$IC(Z(s_i) - \hat{Z}(s_i))_{95\%} : z(s_i) - \hat{z}(s_i) \pm 1,96\hat{\sigma}_k(s_i).$$

Mas, observe na Figura 13 que as variâncias de krigagem variam de um modelo para outro, gerando intervalos com amplitudes diferentes, tornando a comparação inadequada.

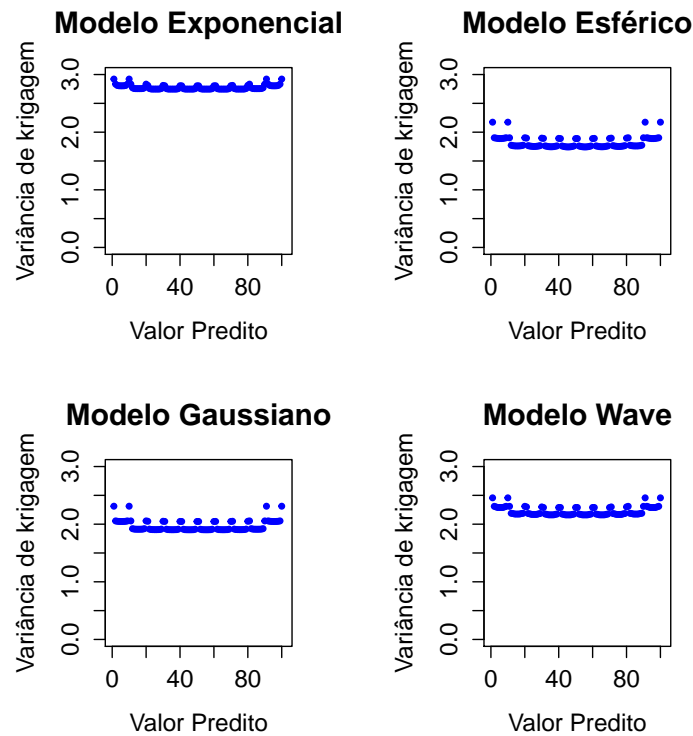


Figura 13 Variâncias de krigagem em relação aos valores preditos dos modelos adotados

Uma solução para isso é impor um critério tal que todos intervalos tenham amplitudes iguais para todos pontos.

O critério imposto é selecionar a menor variância de krigagem entre os modelos para cada  $i$ -ésimo  $s_i$ , isto é

$$\overrightarrow{\hat{\sigma}_k^2 M(s_i)} = \{ \min [\hat{\sigma}_k^2 \text{Modelo } j(s_1)], \min [\hat{\sigma}_k^2 \text{Modelo } j(s_2)], \dots, \min [\hat{\sigma}_k^2 \text{Modelo } j(s_n)] \},$$

em que  $j = 1, \dots, m$ , sendo  $m$  o número de modelos adotados.

Para uma maior compreensão do procedimento, observe a Tabela 3:

Tabela 3 Processo de seleção das  $i$ -ésima variâncias de krigagem entre os modelos

<b>i</b>	$\hat{\sigma}_k^2 \text{Modelo } 1(s_i)$	$\hat{\sigma}_k^2 \text{Modelo } 2(s_i)$	$\hat{\sigma}_k^2 \text{Modelo } 3(s_i)$	$\hat{\sigma}_k^2 \text{Modelo } 4(s_i)$
1	6,2235	<b>5,1337</b>	5,6989	6,0006
2	6,2369	5,2264	<b>5,1322</b>	6,0012
⋮	⋮	⋮	⋮	⋮
n	5,2626	5,1379	5,6506	<b>5,0004</b>

Na Tabela 4 são apresentadas as  $i$ -ésima  $\hat{\sigma}_M^2(p_i)$  selecionadas:

Tabela 4 Variâncias de krigagem selecionadas

<b>i</b>	$\hat{\sigma}_k^2(s_i)$
1	5,1337
2	5,1322
⋮	⋮
n	5,0004

Com essas variâncias, tem-se o seguinte intervalo de confiança para o erro predição

$$IC(Z(s_i) - \hat{Z}(s_i))_{1-\alpha_i} : z(s_i) - \hat{z}(s_i) \pm 1,96\hat{\sigma}_k M(s_i), i = 1, \dots, n.$$

*Justificativa:* a imposição do mínimo tem como objetivo dificultar que o modelo tenha um bom desempenho e aquele que tiver, mostra que realmente é um bom modelo.

Na Figura 14 apresenta-se um exemplo desse intervalo de confiança.

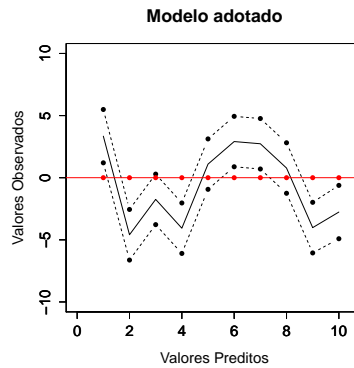


Figura 14 Exemplo de intervalo de confiança para  $Z(s_i) - \hat{Z}(s_i)$

O próximo passo é calcular a proporção de vezes que o valor 0 esteve contido dentro do intervalo de confiança ( $IC(Z(s_i) - \hat{Z}(s_i))$ ). Essa proporção é dada por

$$\hat{p}_{\text{Modelo}} = \frac{\sum_{i=1}^n I(0 \in IC(Z(s_i) - \hat{Z}(s_i)))}{n},$$

em que  $\hat{p}_{\text{Modelo}}$  é a proporção estimada do modelo,  $I(0 \in IC(Z(s_i) - \hat{Z}(s_i)))$  é a função indicadora que retorna 1 se 0 pertence ao  $IC(Z(s_i) - \hat{Z}(s_i))$  ou retorna 0 se não pertencer.

O modelo selecionado é aquele que resultar a maior proporção.

### Proposição do método 3

Ao ajustar um modelo ao semivariograma, suponha que o ajuste seja perfeito. Com isso, ao ajustar uma reta aos  $p_i$ 's ocorre também um ajuste perfeito e a reta será a ideal.

Na Figura 15 ilustra-se esse caso:

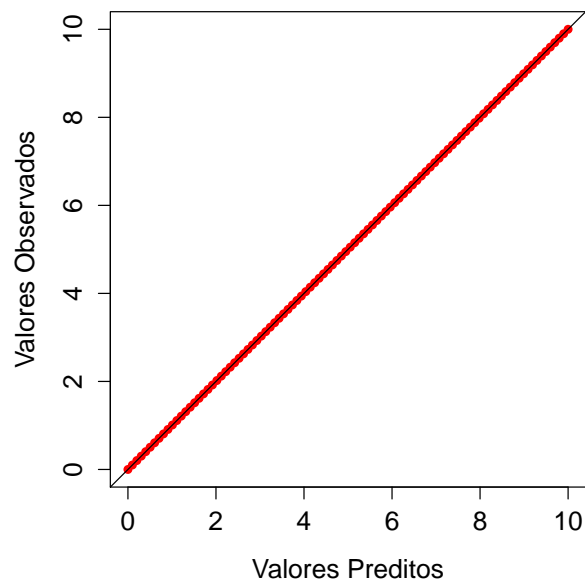


Figura 15 Ajuste perfeito da reta aos pontos  $p_i$ 's,  $i = 1, 2, \dots, n$

Nos estudos dos fenômenos, o ajuste perfeito do modelo,  $z(s_i) = \beta_0 + \beta_1 \hat{z}(s_i)$  com  $\beta_0 = 0$  e  $\beta_1 = 1$ , ao semivariograma raramente acontece, pois os  $p_i$ 's apresentarão uma variabilidade em relação aos respectivos pontos  $pm_i = (\hat{Z}(s_i), \hat{Z}(s_i))$ , pertencentes a reta ideal. Neste caso, a reta ajustada apresenta estimativas de  $\beta_0$  e  $\beta_1$  diferentes de 0 ou 1, respectivamente.

Na Figura 16 ilustram-se  $p_i$ 's com a reta ajustada e a reta ideal acrescentada:

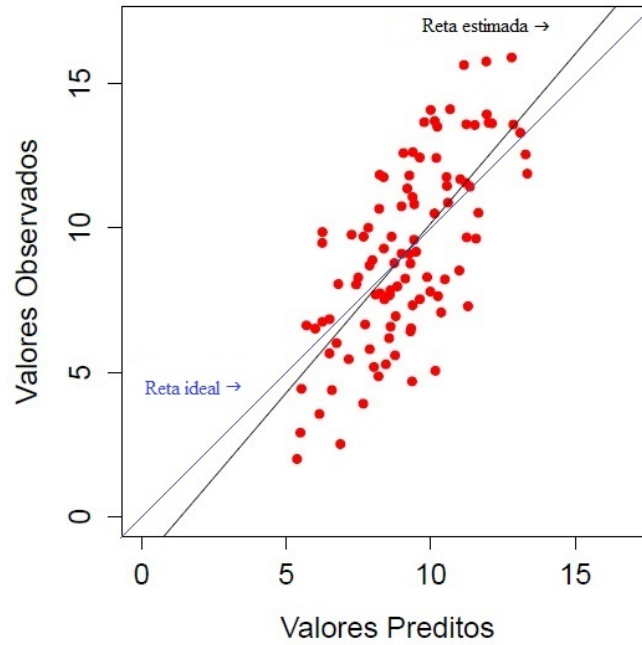


Figura 16 Reta ajustada aos  $p_i$ 's e a reta ideal acrescentada

Observando a Figura 16, realizar apenas a comparação entre essas retas, estaria avaliando apenas a acurácia. Portanto, após realizar a comparação deve-se avaliar a precisão por meio da variabilidade dos  $p_i$ 's em relação aos  $pm_i$ 's.

Na Figura 17 ilustra-se os  $i$ -ésimo  $p_i$ 's e  $pm_i$ 's,  $i = 1, 2, 3, 4$ .

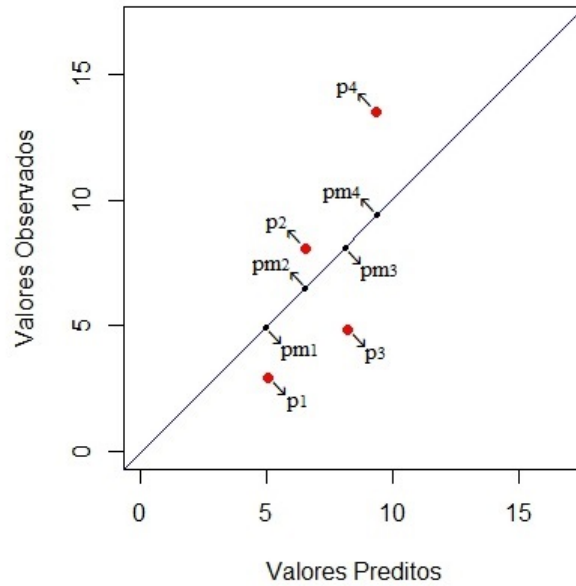


Figura 17 Especificação dos  $i$ -ésimo  $p_i$ 's e  $pm_i$ 's,  $i = 1, 2, 3, 4$

O método 3 tem como objetivo melhorar o método A apresentado na seção 2.1.6, sendo realizado em dois passos:

### Passo 1

Consiste em realizar dois testes de hipóteses, um para verificar se  $\beta_0 = 0$  e outro para  $\beta_1 = 1$ , com objetivo de verificar se a reta ajustada é igual a reta ideal.

As estatísticas de teste são:

$$t_{c\beta_0} = \frac{\hat{\beta}_0}{\sqrt{\frac{\sum_{i=1}^n [Z(s_i) - \widehat{Z}(s_i)]^2}{n-2} \left\{ \frac{1}{n} + \frac{\left[ \frac{\sum_{i=1}^n \hat{Z}(s_i)}{n} \right]^2}{\sum_{i=1}^n [\hat{Z}(s_i)]^2 - \frac{\left[ \sum_{i=1}^n \hat{Z}(s_i) \right]^2}{n}} \right\}}},$$

e

$$t_{c\beta_1} = \frac{\hat{\beta}_1 - 1}{\sqrt{\frac{\sum_{i=1}^n [Z(s_i) - \widehat{Z}(s_i)]^2}{(n-2) \left\{ \sum_{i=1}^n [\hat{Z}(s_i)]^2 - \frac{\left[ \sum_{i=1}^n \hat{Z}(s_i) \right]^2}{n} \right\}}}}.$$

A estatística de cada um desses testes seguem uma distribuição  $t$  com  $n-2$  graus de liberdade.

Após realizar os dois testes para os modelos adotados, ocorrerá um dos casos abaixo:

1. Se todos modelos rejeitaram  $H_0$ , nenhum deles é recomendado e não realiza-se o passo 2;
2. Se apenas um dos modelos não rejeitar  $H_0$  tanto para  $\beta_0$  quanto  $\beta_1$ , é recomendado utilizá-lo e não realiza-se o passo 2;
3. Se mais de um dos modelos não rejeitar  $H_0$  tanto para  $\beta_0$  quanto  $\beta_1$ , é preciso realizar o passo 2 para selecionar o melhor desses modelos.

## Passo 2

Para selecionar o melhor modelo dentre os referentes ao teste que não rejeitaram  $H_0$ , o passo 2 consiste em analisar a variabilidade entre os seus  $i$ -ésimo  $p_i$  e  $pm_i$ , utilizando a  $i$ -ésima distância ( $D_i$ ) definida como:

$$D_i(p_i, pm_i) = |Z(s_i) - \hat{Z}(s_i)|.$$

Observe que o ideal seria todas distâncias iguais a 0, informando que o ajuste do modelo ao semivariograma foi perfeito.

O passo 2 tem o seguinte procedimento:

1. Para cada modelo, acrescente a reta ideal ao gráfico dos  $i$ -ésimo  $p_i$ 's;
2. Para analisar a variabilidade entre a reta ideal e os  $p_i$ 's, calcule a distância média referente ao modelo adotado, da seguinte forma:

$$\bar{D}_{\text{Modelo}} = \frac{\sum_{i=1}^n |Z(s_i) - \hat{Z}(s_i)|}{n};$$

3. Como método de seleção, compare as estimativas das distâncias médias referentes aos modelos adotados e escolha a menor. O modelo da qual ela foi gerada é o selecionado.

## Proposição do método 4

Dent e Blackie (1979) sugerem uma hipótese nula mais relevante na qual são testados se o coeficiente linear e angular simultaneamente não são diferentes

de 0 e 1, respectivamente.

Portanto, o método 4 consiste em aplicar o teste F recomendado por Dent e Blackie (1979), ao invés de utilizar o teste t conforme o método 3. Esse método também tem como objetivo melhorar o método A apresentado na seção 2.1.6, sendo realizado em dois passos:

### **Passo 1**

Consiste em realizar um teste de hipóteses, para verificar se  $\beta_0 = 0$  e  $\beta_1 = 1$  simultaneamente, com objetivo de verificar se a reta ajustada é igual a reta ideal.

A estatística de teste recomendada por Dent e Blackie (1979) é dada por

$$F_c = \frac{n\hat{\beta}_0^2 + 2\hat{\beta}_0(\hat{\beta}_1 - 1) \sum_{i=1}^n \hat{Z}(s_i) + (\hat{\beta}_1 - 1)^2 \sum_{i=1}^n \hat{Z}(s_i)^2}{2 \frac{\sum_{i=1}^n [Z(s_i) - \hat{Z}(s_i)]^2}{n - 2}},$$

A estatística de cada um desses testes seguem uma distribuição F com 2 e  $n - 2$  graus de liberdade.

Após realizar esse teste para os modelos adotados, ocorrerá um dos casos abaixo:

1. Se todos modelos rejeitaram  $H_0$ , nenhum deles é recomendado e não realiza-se o passo 2, que é o mesmo do método 3 proposto;
2. Se apenas um dos modelos não rejeitar  $H_0$ , é recomendado utilizá-lo e não realiza-se o passo 2;

3. Se mais de um dos modelos não rejeitar  $H_0$ , é preciso realizar o passo 2 para selecionar o melhor desses modelos.

### **Proposição do método 5**

O método 5 consiste em aplicar a metodologia proposta por Bland e Altman (1983), comparando os valores observados e preditos pela autovalidação. Esse método tem o seguinte procedimento:

1. Realize o teste para correlação e selecione os modelos que não rejeitarem  $H_0$ ;
2. Considerando os modelos selecionados, examine o viés realizando o teste t para amostras emparelhadas e selecione os modelos que não rejeitarem a hipótese nula;
3. Considerando os novos modelos selecionados e  $S_d = \sqrt{\frac{\left[ \sum_{i=1}^n (d_i - \bar{d})^2 \right]}{(n-1)}}$ , calcule o  $1,96S_d$  e selecione o modelo que resultar a menor estimativa.

### **Proposição do método 6**

O método 6 consiste em aplicar o desvio quadrático médio (*DQM*), avaliando os seus componentes *QV*, *DQDP* e *FCPDP*, como método de seleção, conforme recomendado por Kobayashi e Salam (2000), sendo mais informativo do que o método D.

O  $DQM$  é dado por

$$DQM = \underbrace{(\bar{Z} - \hat{Z})^2}_{QV} + \underbrace{(DP_Z - DP_{\hat{Z}})^2}_{DQDP} + \underbrace{2DP_Z DP_{\hat{Z}}(1-r)}_{FCPDP}.$$

Como método de seleção, compare as estimativas dos componentes do  $DQM$  referentes aos modelos adotados, dando ênfase à variabilidade entre as estimativas de cada componente, e selecione o modelo que apresentar as menores.

Após aplicar os métodos usuais e propostos do 1 ao 6, é possível verificar quais estão selecionando corretamente o modelo em relação à dependência espacial inclusa nos dados.

Escolher os métodos que apresentarem um bom comportamento na verificação e propor um novo método realizando a junção desses métodos escolhidos em apenas um, espera-se que seja um método mais eficiente.

### **Proposição do método 7**

O método 7 consiste em aprimorar a ideia de Laslett et al. (1987), realizando a junção de vários métodos em apenas um. Para que este método apresente um melhor comportamento, os métodos juntados serão escolhidos após a verificação do comportamento dos métodos usuais e propostos do 1 ao 6. Portanto, será proposta sua ideia e nos resultados serão apresentados os métodos escolhidos para a junção.

O procedimento deste método é da seguinte forma:

1. Selecione  $n$  métodos de seleção;
2. Ao invés de utilizar os postos conforme Laslett et al. (1987), será utilizado a seguinte forma de ranqueamento:

$$\text{Ranqueamento}_{\text{Método } j} \left| \frac{EstMM_j - EstM_{ji}}{EstMM_j} \right|,$$

em que  $EstMM_j$  é a estimativa do melhor modelo em relação ao método  $j$  e  $EstM_{ji}$  é a estimativa do  $i$ -ésimo modelo em relação ao método  $j$ , sendo  $i$ =Exponencial, Esférico, Gaussiano e Wave.

Destaca-se que essa forma de ranqueamento considera a variabilidade existente entre as estimativas, tornando o método mais eficiente.

3. Classifique as estimativas conforme o ranqueamento proposto de acordo com a ordem que o método informa do melhor para o pior e calcule as médias ( $\overline{\text{Rank}}$ ) e os desvios padrão ( $DP_{\text{Rank}}$ ) dos ranks em relação a cada modelo.

Um exemplo é dado na Tabela 5:

Tabela 5 Estimativas classificadas dos  $n$  métodos selecionados, com as médias e os desvios padrões dos ranks calculados

Método	Modelo			
	Exponencial	Esférico	Gaussiano	Wave
1	4,614	3,760	4,015	4,146
Rank	0,227	0,000	0,068	0,103
2	5,472	5,600	4,013	5,122
Rank	0,364	0,395	0,000	0,276
n=3	5,098	4,060	4,004	7,030
Rank	0,273	0,014	0,000	0,756
$\overline{\text{Rank}}_i$	0,288	0,136	0,023	0,378
$DP_{\text{Rank}_i}$	0,070	0,224	0,039	0,338

4. Plote um gráfico sendo um diagrama de dispersão dos pontos formados pelos desvios padrões versus as médias dos ranks. Ao invés de analisar a similaridade, conforme Laslett et al. (1987), acrescente ao gráfico o ponto mínimo ( $\overline{\text{Rank}} = 0, DP = 0$ ) e calcule a distância dele em relação a cada ponto formado pelos desvio padrão em relação as médias dos ranks. Veja a Figura 18:

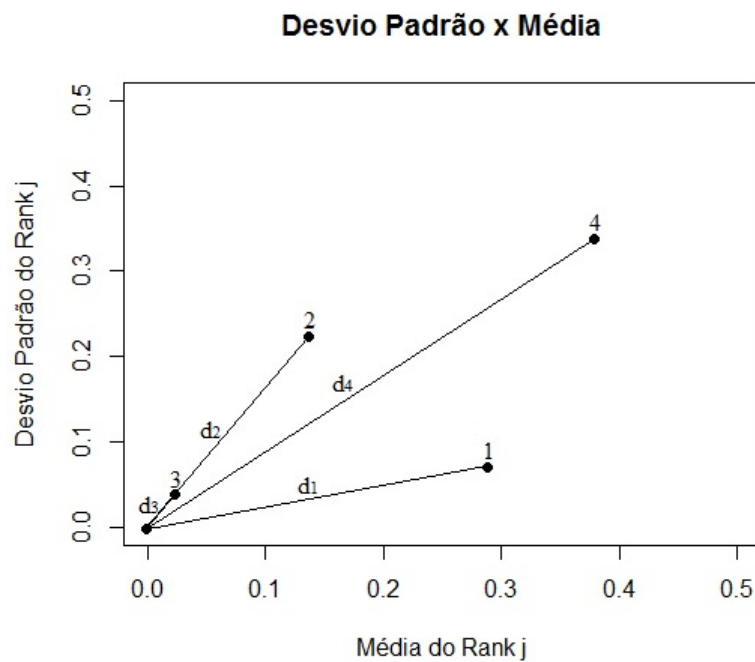


Figura 18 Pontos formados pelos desvios padrões em relação as médias dos postos, com as distâncias desses pontos em relação ao ponto (0,0)

5. Com as distâncias calculadas, selecione o modelo que resultar a menor distância, por exemplo, na Figura 18 seria selecionado o modelo 3.

Utilizando o *software* R (R CORE TEAM, 2017) e a função *geoR()* (RIBEIRO JR.; DIGGLE, 2001) desse pacote, foram programadas as funções que permitem realizar os métodos de seleção usuais e propostos neste trabalho. Os nomes das funções referentes aos métodos são apresentados na Tabela 6:

Tabela 6 Nomes das funções dos métodos usuais e propostos de seleção

Método de Seleção	Função
A	MA (vo, vpexp, vpesf, vpgau, vpwav)
B	MB (vo, vpexp, vpesf, vpgau, vpwav)
C	MC (vo, vpexp, vpesf, vpgau, vpwav, vkexp, vkesf, vkgau, vkwav)
D	MD (vo, vpexp, vpesf, vpgau, vpwav)
E	ME (vo, vpexp, vpesf, vpgau, vpwav)
F	MF (vo, vpexp, vpesf, vpgau, vpwav)
G	MG (vo, vpexp, vpesf, vpgau, vpwav)
H	MH (vo, vpexp, vpesf, vpgau, vpwav)
1	M1 (vo, vpexp, vpesf, vpgau, vpwav)
2	M2 (vo, vpexp, vpesf, vpgau, vpwav, vkexp, vkesf, vkgau, vkwav)
3	M3 (vo, vpexp, vpesf, vpgau, vpwav)
4	M4 (vo, vpexp, vpesf, vpgau, vpwav)
5	M5 (vo, vpexp, vpesf, vpgau, vpwav)
6	M6 (vo, vpexp, vpesf, vpgau, vpwav)
7	M7 (vo, vpexp, vpesf, vpgau, vpwav)

em que

- vo é o vetor dos valores observados;
- vpexp é o vetor dos valores preditos em relação ao modelo exponencial;
- vpesf é o vetor dos valores preditos em relação ao modelo esférico;
- vpgau é o vetor dos valores preditos em relação ao modelo gaussiano;
- vpwav é o vetor dos valores preditos em relação ao modelo wave.

Com as funções programadas, facilitam-se as aplicações dos métodos usuais e propostos.

### 3.2.4 Estratégia utilizada para recomendação de um ou mais métodos de seleção

A realização de milhares de simulações com dependência espacial utilizando a função  $grf()$  do pacote *geoR* (RIBEIRO JR.; DIGGLE, 2001) apresenta uma limitação. Essa limitação consiste que, em algumas das milhares de simulações, os semivariogramas estimados apresentarão comportamentos bem diferentes em relação ao desejado semivariograma paramétrico.

Na Figura 19, tem-se nove semivariogramas referentes a nove simulações de dados com dependência espacial esférica.

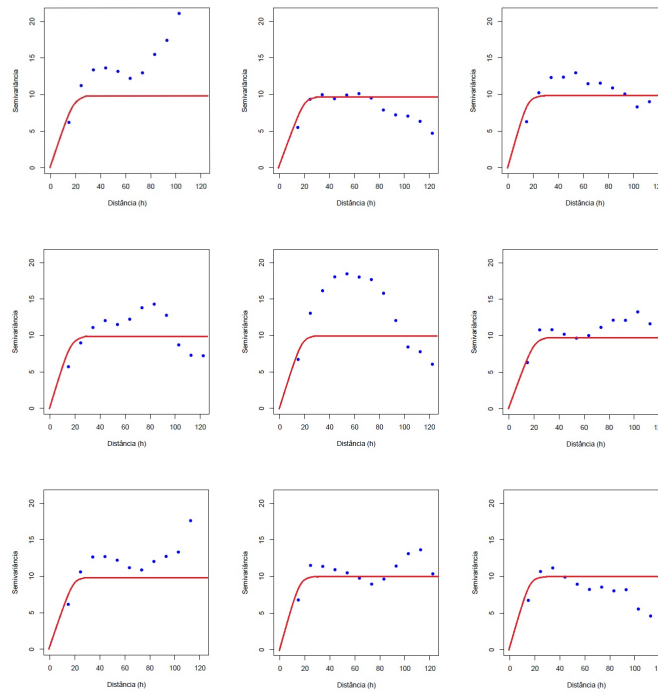


Figura 19 Semivariogramas referentes a dados simulados com dependência espacial esférica de acordo com a seção 3.1.2

Na Figura 19 observa-se que os semivariogramas estimados, referentes aos dados simulados, oscilam muito em relação ao semivariograma paramétrico.

Para uma maior ilustração dessa oscilação, na Figura 20 têm-se, em cada um dos 12 gráficos, 1000 semivariogramas referentes a 1000 simulações de dados com dependência espacial exponencial, esférica, gaussiana e wave, de acordo com a seção 3.1.2, considerando as três especificações da distância máxima.

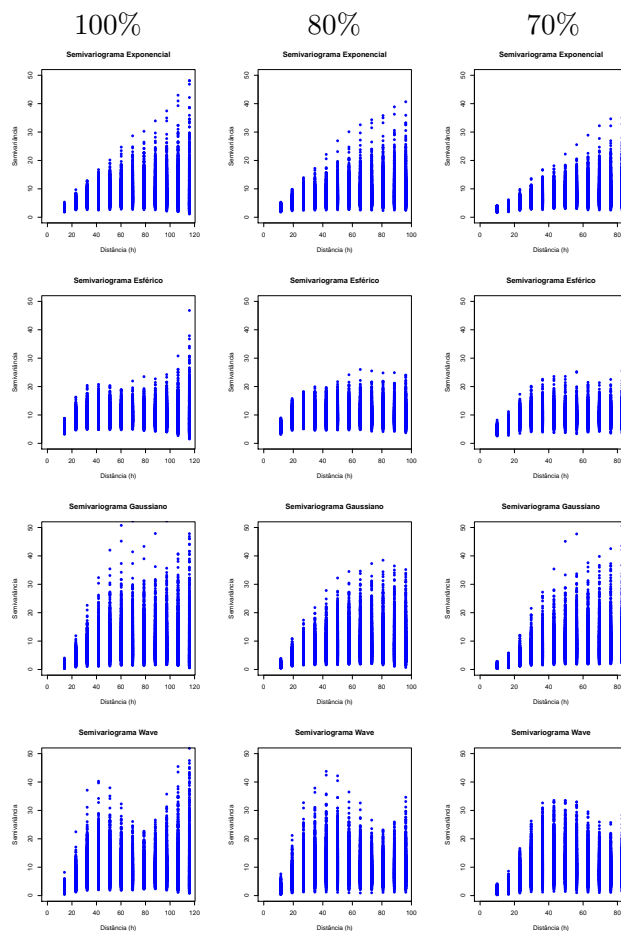


Figura 20 Semivariogramas estimados referentes a 1000 simulações de dados com dependência espacial exponencial, esférico, gaussiano e wave, considerando 100%, 80% e 70% da distância máxima

Com objetivo de tornar-se a função  $grf()$  um gerador de números aleatórios eficiente diante da imposição de dependência espacial e mais próximo do que realmente acontece na natureza dos fenômenos, foi condicionado na simulação gerar números aleatórios que estimam semivariograma com coeficiente de determinação  $R^2$  acima de 0,7 em relação ao semivariograma paramétrico desejado.

Para ilustrar a eficiência da função  $grf()$  condicionada, vejam a Figura 21:

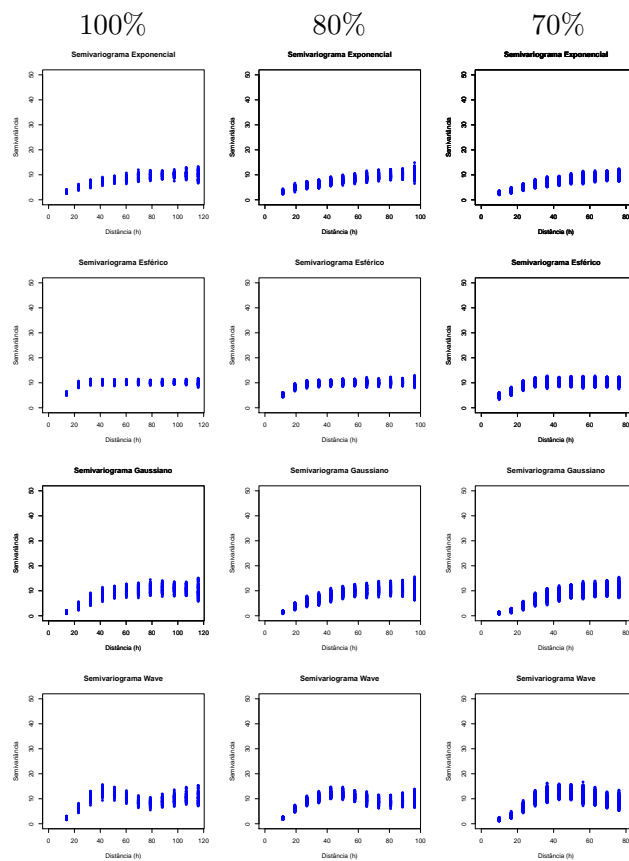


Figura 21 Semivariogramas estimados referentes a 1000 simulações de dados com dependência espacial exponencial, esférico, gaussiano e wave, condicionados ao  $R^2$  acima de 0,7 em relação ao semivariograma paramétrico desejado, considerando 100%, 80% e 70% da distância máxima

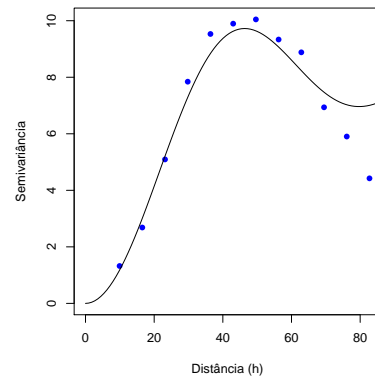
O valor do  $R^2$  acima de 0,7 permite que exista uma variabilidade entre os semivariogramas estimados em relação aos dados simulados, mas que não seja uma variabilidade tão grande, diferente das situações típicas ocorridas na natureza dos fenômenos. O valor do  $R^2$  acima de 0,7 foi convencionado neste trabalho, mas o pesquisador pode alterá-lo segundo a sua conveniência. Destaca-se que o condicionamento do  $R^2$  não é uma declaração de que a função  $grf()$  tenha defeitos intrínsecos e sim de que existe uma dificuldade notória em simulações de dados com dependência espacial.

Mesmo automatizando o ajuste do modelo e tornando eficiente a imposição da dependência espacial nos dados em milhares de simulações, um outro problema, frequentemente, ocorre.

Quando o semivariograma estimado é muito eficiente em relação à imposição, o comportamento das semivariâncias é muito próximo do modelo ajustado. Veja, por exemplo, a Figura 22 (a). O ajuste muito bom do modelo ao semivariograma ocasiona um erro no processo de autovalidação, conforme pode ser visto na Figura 22 (b).

O erro descrito na Figura 22 (b) significa que a matriz é singular, não admitindo uma matriz inversa. Uma matriz é singular somente se seu determinante é nulo. Por exemplo: se uma matriz quadrada tiver pelo menos uma linha ou coluna nula, terá determinante nulo. No processo de predição da autovalidação, a matriz tem que admitir inversa.

Com objetivo de não permitir que este tipo de erro interrompa o processo em milhares de simulações, foi utilizada a função  $try()$  que executa uma expressão que possa falhar.



(a)

```

Console
> pepita4<-va4[[1]]$nugget
> var.ols4<-variofit(semivarM,cov.model = "wave",ini=c(sillp4,alcance4),
+ nugget=pepita4)
variofit: covariance model used is wave
variofit: weights used: npairs
variofit: minimisation function used: optim
> plot(semivarM, xlab="Distância (h)", ylab="Semivariância",col='blue',pch=16)
> lines(var.ols4,lwd=1,lty=1)
> xv.va4<-xvalid(dados,model=var.ols4)
xvalid: number of data locations = 100
xvalid: number of validation locations = 100
xvalid: performing cross-validation at location ... 1,
Error in solve.default(vcov, trend.d) :
system is computationally singular: reciprocal condition number = 9.30755e-19
>

```

(b)

Figura 22 (a) Semivariograma referente a dados simulados com dependência espacial wave com valores paramétricos  $C_0=0$ ,  $C_1=10$ ,  $a=5$  e  $C=10$ , e (b) o erro ocasionado no processo de autovalidação

Com o uso crescente de métodos de otimização em estudos geoestatísticos, segundo Mohammadi e Safa (2016), nos últimos anos muito esforço tem sido feito com relação à montagem automática eficiente do ajuste do modelo ao semivariograma.

Para a recomendação de um ou mais métodos de seleção, a estratégia utilizada no presente trabalho foi realizar a aplicação dos métodos em vários cenários, conforme a Tabela 7.

Tabela 7 Cenários adotados na verificação do comportamento dos métodos

Semivariograma Simulado	Distância Máxima	Métodos de Seleção Aplicados
Semivariograma Esférico	100%	Usuais (A, B, C, D, E, F, G, H)
		Propostos (1, 2, 3, 4, 5, 6, 7)
	80%	Usuais (A, B, C, D, E, F, G, H)
		Propostos (1, 2, 3, 4, 5, 6, 7)
	70%	Usuais (A, B, C, D, E, F, G, H)
		Propostos (1, 2, 3, 4, 5, 6, 7)
Semivariograma Exponencial	100%	Usuais (A, B, C, D, E, F, G, H)
		Propostos (1, 2, 3, 4, 5, 6, 7)
	80%	Usuais (A, B, C, D, E, F, G, H)
		Propostos (1, 2, 3, 4, 5, 6, 7)
	70%	Usuais (A, B, C, D, E, F, G, H)
		Propostos (1, 2, 3, 4, 5, 6, 7)
Semivariograma Gaussiano	100%	Usuais (A, B, C, D, E, F, G, H)
		Propostos (1, 2, 3, 4, 5, 6, 7)
	80%	Usuais (A, B, C, D, E, F, G, H)
		Propostos (1, 2, 3, 4, 5, 6, 7)
	70%	Usuais (A, B, C, D, E, F, G, H)
		Propostos (1, 2, 3, 4, 5, 6, 7)
Semivariograma Wave	100%	Usuais (A, B, C, D, E, F, G, H)
		Propostos (1, 2, 3, 4, 5, 6, 7)
	80%	Usuais (A, B, C, D, E, F, G, H)
		Propostos (1, 2, 3, 4, 5, 6, 7)
	70%	Usuais (A, B, C, D, E, F, G, H)
		Propostos (1, 2, 3, 4, 5, 6, 7)

Foram realizadas 3000 simulações de dados referentes a cada cenário da Tabela 7 e para aqueles que apresentaram semivariogramas respeitando o condicionamento do  $R^2$  adotado, com processo de autovalidação não apresentando erro, foram aplicados os métodos usuais e propostos. Destaca-se que foi fixado o mesmo valor da semente para todas simulações. Por meio das proporções de acerto em cada cenário, é possível recomendar um método de seleção.

## 4 RESULTADOS E DISCUSSÃO

Nesse tópico, são apresentadas as aplicações dos métodos usuais e propostos em dados simulados com dependência espacial esférica, um resumo da aplicação dos métodos usuais e propostos em dados simulados com dependência espacial exponencial, gaussiana e wave, a recomendação de um método de seleção por meio das simulações e a sua aplicação em dados reais de teor de zinco do solo.

### 4.1 Aplicações dos métodos usuais e propostos em dados simulados com dependência espacial esférica

Exemplos de semivariogramas estimados, utilizando dados simulados com dependência espacial esférica e considerando as três especificações da distância máxima são apresentados na Figura 23.

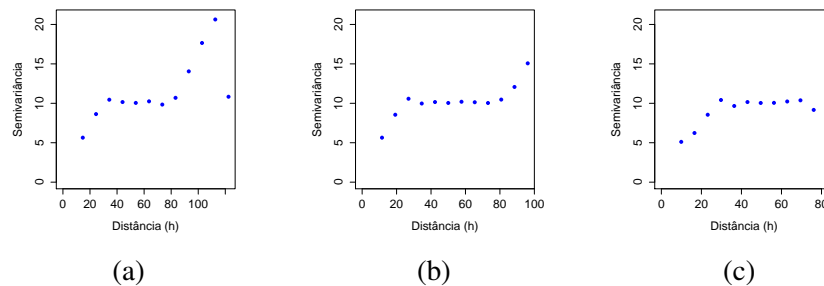


Figura 23 Semivariograma esférico com (a) 100%, (b) 80% e (c) 70% da distância máxima

Vale ressaltar que as três especificações da distância máxima do semivariograma teve como objetivo estudar os seguintes casos:

1. 100% da distância máxima (variabilidade alta nas maiores distâncias);

2. 80% da distância máxima (variabilidade média nas maiores distâncias);
3. 70% da distância máxima (variabilidade baixa nas maiores distâncias).

Na Figura 24 são apresentados os semivariogramas da Figura 23, porém com os modelos exponencial, esférico, gaussiano e wave ajustados em cada caso.

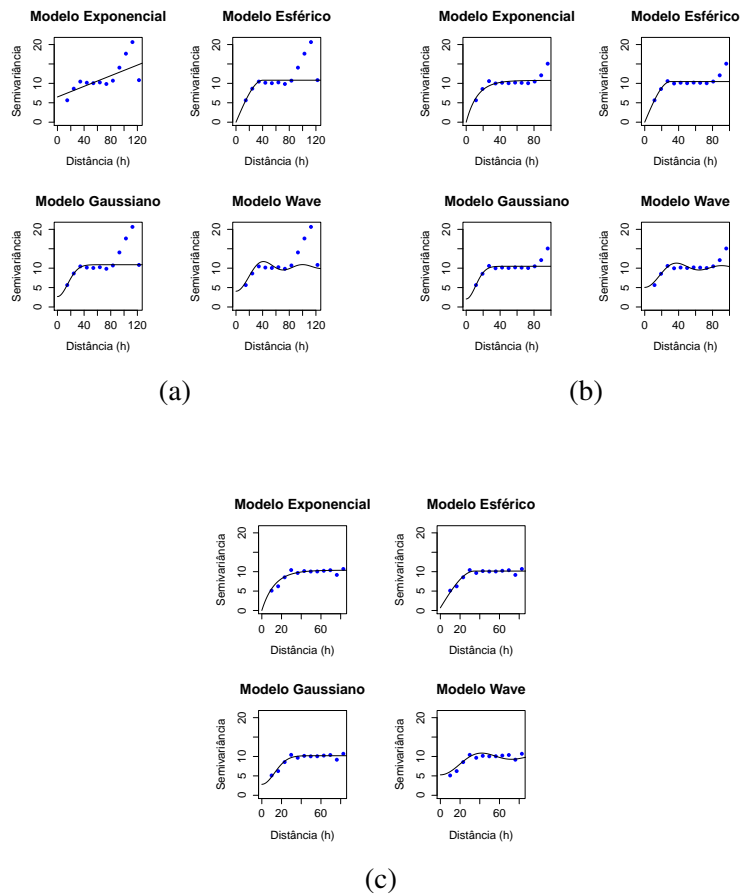


Figura 24 Modelo Exponencial, Esférico, Gaussiano e Wave ajustados ao semi-variograma esférico com (a) 100%, (b) 80% e (c) 70% da distância máxima

Na Tabela 8 tem-se as estimativas dos parâmetros desses modelos:

Tabela 8 Estimativas dos parâmetros dos modelos ajustados, considerando 100%, 80% e 70% da distância máxima do semivariograma esférico

100% da distância máxima do semivariograma				
Modelo	$a$	$C_0$	$C_1$	$C$
Exponencial	<b>Indefinido</b>	6,4218	<b>Indefinido</b>	<b>Indefinido</b>
Esférico	40,3755	0,0000	10,8143	10,8143
Gaussiano	21,7028	2,6815	8,2081	10,8896
Wave	9,1699	4,0622	6,2788	10,3410
80% da distância máxima do semivariograma				
Modelo	$a$	$C_0$	$C_1$	$C$
Exponencial	13,2301	0,0000	10,7629	10,7629
Esférico	30,3365	0,0000	10,4621	10,4621
Gaussiano	15,5862	2,0704	8,4144	10,4848
Wave	8,3192	5,0499	5,1345	10,1844
70% da distância máxima do semivariograma				
Modelo	$a$	$C_0$	$C_1$	$C$
Exponencial	12,8290	0,0000	10,3752	10,3752
Esférico	34,5473	0,6893	9,4638	10,1531
Gaussiano	18,1042	2,8298	7,3658	10,1956
Wave	9,3333	5,2686	4,6011	9,8697

Observando a Tabela 8, considerando dados simulados com dependência espacial esférica com  $a = 30$ ,  $C_0 = 0$ ,  $C_1 = 10$  e  $C = 10$ , retiram-se as seguintes informações:

1. Considerando 100% e 70% da distância máxima, realizar a comparação do  $a$  não é plausível, pois o único modelo que expressa realmente o alcance é o esférico. Em relação a  $C_0$  e  $C_1$ , as estimativas do modelo esférico foram as mais próximas de 0 e 10, respectivamente. Em relação a  $C$ , a estimativa do wave foi a mais próxima de 10;

2. Considerando 80% da distância máxima, em relação a  $a$  não é plausível realizar a comparação. Em relação a  $C_0$ , as estimativas dos modelos exponencial e esférico foram iguais a 0. Em relação a  $C_1$ , a estimativa do modelo esférico foi a mais próxima de 10. Em relação a  $C$ , a estimativa do modelo wave foi a mais próxima de 10.

Após realizar o ajuste dos modelos, deve-se selecionar o que mais bem se ajusta ao semivariograma. Pela análise das estimativas da Tabela 8, como os dados são simulados, os valores paramétricos do modelo são conhecidos, portanto, sugere-se que o modelo esférico seja selecionado. Mas, numa análise de dados reais, a informação dos valores paramétricos é desconhecida, tornando-se impossível selecionar o melhor modelo por meio de suas estimativas.

Nesse caso, uma maneira de selecionar o modelo é a realizar a autovalidação, apresentada na seção 2.1.5, gerando um valor predito para cada coordenada onde há também um valor observado, e aplicar os métodos da seção 2.1.6 e os da seção 3.2.3.

### **Resultados das aplicações dos métodos usuais em dados simulados com dependência espacial esférica**

A Tabela 9 apresenta os resultados das aplicações dos métodos usuais em relação ao semivariograma com variabilidade alta nas maiores distâncias. Observando-a, os métodos D, E, F, G e H selecionaram corretamente o modelo esférico, os métodos A e B apresentaram inconsistência, e o método C selecionou errado o modelo exponencial. Portanto, os pesquisadores que utilizam os métodos A e B iriam ficar indecisos e os que utilizam o método C iriam selecionar errado.

Tabela 9 Resultados das aplicações dos métodos usuais em dados simulados com dependência espacial esférica, considerando 100% da distância máxima

Método	Estatística	Modelo				Modelo Selecionado
		Exponencial	Esférico	Gaussiano	Wave	
A	$\beta_0$	-4,430	0,277	<b>-0,167</b>	-0,753	<b>Inconsistente</b>
	$\beta_1$	1,483	0,975	<b>1,019</b>	1,083	
	$r$	0,531	<b>0,731</b>	0,726	0,685	
B	$EM$	<b>0,001</b>	0,006	0,006	0,007	<b>Inconsistente</b>
	$VE$	7,601	<b>4,733</b>	4,806	5,426	
C	$EMP$	<b>-0,001</b>	0,002	0,002	0,002	<b>Exponencial</b> <b>(Seleção Errada)</b>
	$VEP$	<b>0,978</b>	1,449	0,806	1,094	
D	$EQM$	7,638	<b>4,733</b>	4,806	5,426	<b>Esférico</b> <b>(Seleção Correta)</b>
E	$R^2$	0,282	<b>0,534</b>	0,527	0,469	<b>Esférico</b> <b>(Seleção Correta)</b>
F	$AIC$	487,537	<b>444,160</b>	445,734	457,380	<b>Esférico</b> <b>(Seleção Correta)</b>
G	$EC$	8,108	<b>5,008</b>	5,086	5,748	<b>Esférico</b> <b>(Seleção Correta)</b>
H	$\hat{\rho}_c$	0,332	<b>0,702</b>	0,686	0,618	<b>Esférico</b> <b>(Seleção Correta)</b>

A Tabela 10 apresenta os resultados das aplicações dos métodos usuais em relação ao semivariograma com variabilidade média nas maiores distâncias. Observando-a, os métodos A, D, E, F, G e H selecionaram corretamente o modelo esférico, o método B apresentou inconsistência e o método C selecionou errado o modelo exponencial. Portanto, os resultados sugerem que os pesquisadores que utilizam o método B iriam ficar indecisos e os que utilizam o método C iriam selecionar errado. A diferença dos resultados dessa análise com a análise considerando variabilidade alta nas maiores distâncias é que o método A, que havia apresentado inconsistência, selecionou corretamente o modelo esférico.

Tabela 10 Resultados das aplicações dos métodos usuais em dados simulados com dependência espacial esférica, considerando 80% da distância máxima

Método	Estatística	Modelo				Modelo Selecionado
		Exponencial	Esférico	Gaussiano	Wave	
A	$\beta_0$	-1,141	<b>0,372</b>	0,781	-1,588	<b>Esférico</b> (Seleção Correta)
	$\beta_1$	1,125	<b>0,960</b>	0,916	1,174	
	$r$	0,699	<b>0,709</b>	0,691	0,695	
B	$EM$	<b>0,003</b>	0,007	0,008	0,007	<b>Inconsistente</b>
	$VE$	5,249	<b>5,052</b>	5,354	5,363	
C	$EMP$	<b>0,001</b>	0,002	0,002	0,001	<b>Exponencial</b> (Seleção Errada)
	$VEP$	<b>0,872</b>	0,188	1,311	0,866	
D	$EQM$	5,249	<b>5,052</b>	5,354	5,363	<b>Esférico</b> (Seleção Correta)
E	$R^2$	0,489	<b>0,503</b>	0,477	0,482	<b>Esférico</b> (Seleção Correta)
F	$AIC$	453,426	<b>450,592</b>	455,789	454,713	<b>Esférico</b> (Seleção Correta)
G	$EC$	5,553	<b>5,350</b>	5,671	5,675	<b>Esférico</b> (Seleção Correta)
H	$\hat{\rho}_c$	0,627	<b>0,678</b>	0,664	0,609	<b>Esférico</b> (Seleção Correta)

A Tabela 11 apresenta os resultados das aplicações dos métodos usuais em relação ao semivariograma com variabilidade baixa nas maiores distâncias. Observando-a, os métodos D, E, F, G e H selecionaram corretamente, os métodos A e B apresentaram inconsistência, e o método C selecionou errado. Portanto, os resultados sugerem que os pesquisadores que utilizam os métodos A e B iriam ficar indecisos, e os que utilizam o método C iriam selecionar errado. A diferença dos resultados dessa análise com a análise considerando variabilidade média é que o método A que havia selecionado corretamente voltou a apresentar inconsistência.

Tabela 11 Resultados das aplicações dos métodos usuais em dados simulados com dependência espacial esférica, considerando 70% da distância máxima

Método	Estatística	Modelo				Modelo Selecionado
		Exponencial	Esférico	Gaussiano	Wave	
A	$\beta_0$	-1,223	-0,237	<b>-0,062</b>	-1,263	<b>Inconsistente</b>
	$\beta_1$	1,134	1,026	<b>1,007</b>	1,138	
	$r$	0,699	<b>0,732</b>	0,711	0,651	
B	$EM$	0,003	0,005	0,006	<b>0,002</b>	<b>Inconsistente</b>
	$VE$	5,265	<b>4,715</b>	5,023	5,920	
C	$EMP$	0,001	0,001	0,001	<b>0,000</b>	<b>Wave</b> (Seleção Errada)
	$VEP$	0,885	1,104	1,125	<b>0,944</b>	
D	$EQM$	5,265	<b>4,715</b>	5,023	5,920	<b>Esférico</b> (Seleção Correta)
E	$R^2$	0,488	<b>0,536</b>	0,505	0,423	<b>Esférico</b> (Seleção Correta)
F	$AIC$	453,563	<b>443,772</b>	450,179	465,544	<b>Esférico</b> (Seleção Correta)
G	$EC$	5,569	<b>4,988</b>	5,318	6,269	<b>Esférico</b> (Seleção Correta)
H	$\hat{\rho}_c$	0,624	<b>0,692</b>	0,670	0,561	<b>Esférico</b> (Seleção Correta)

Diante do comportamento dos métodos usuais, justifica-se a importância da proposição de métodos de seleção e recomendação de qual deve ser utilizado. Destaca-se que o método C, um dos mais utilizados pelos pesquisadores, se não for o mais, selecionou errado nos três casos. Por outro lado, os métodos D, E, F, G e H selecionaram corretamente nos três casos. Em relação as especificações da distância máxima, as estimativas dos métodos D, E, F, G e H referentes ao modelo esférico que mais bem se ajustou ao semivariograma, foram parecidas. Interpreta-se, então, que as especificações não interferiram nas estimativas, po-

dendo ser explicada pela teoria de predição via método de krigagem (veja seção 2.1.3). O método de krigagem utiliza a dependência espacial para prever que é definido pelo alcance do semivariograma encontrado nas menores distâncias.

### **Resultados das aplicações dos métodos propostos em dados simulados com dependência espacial esférica**

A seguir, são apresentados em passos detalhados os 7 métodos de seleção propostos aplicados em dados simulados com dependência espacial esférica.

#### **Aplicação do método 1**

Na Figura 25 apresenta-se os valores preditos pela validação do modelo esférico plotados nos intervalos com 95% de confiança para medida individual em relação a amplitude móvel, considerando 100% da distância máxima.

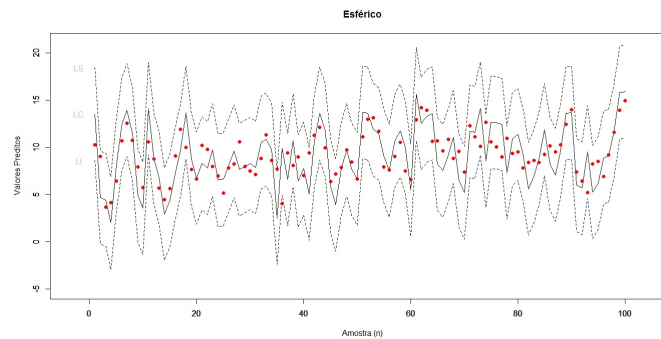


Figura 25 Valores preditos pela validação do modelo esférico plotados nos intervalos com 95% de confiança para medida individual, considerando 100% da distância máxima

Na Figura 26 apresenta-se os valores preditos pela validação do modelo esférico plotados nos intervalos com 95% de confiança para medida individual em relação a amplitude móvel, considerando 80% da distância máxima.

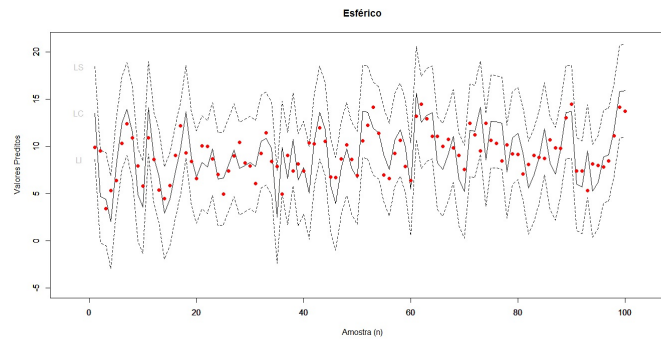


Figura 26 Valores preditos pela validação do modelo esférico plotados nos intervalos com 95% de confiança para medida individual, considerando 80% da distância máxima

Na Figura 27 apresenta-se os valores preditos pela validação do modelo esférico plotados nos intervalos com 95% de confiança para medida individual em relação a amplitude móvel, considerando 70% da distância máxima.

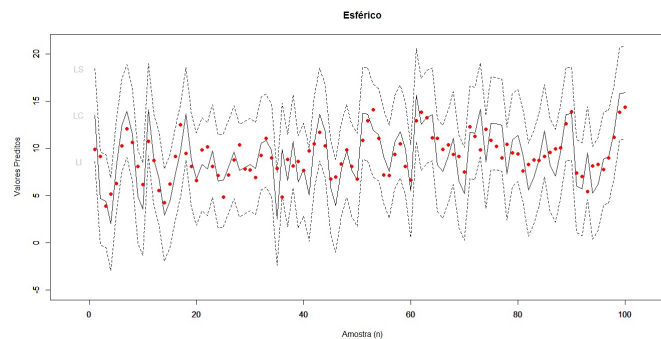


Figura 27 Valores preditos pela validação do modelo esférico plotados nos intervalos com 95% de confiança para medida individual, considerando 70% da distância máxima

Calculando a proporção de valores preditos dentro dos intervalos com 95% de confiança para medida individual em relação a amplitude móvel, tem-se as proporções estimadas, que estão apresentadas na Tabela 12.

Tabela 12 Estimativas das proporções do Método 1

$\hat{p}_{\text{Modelo}}$	Distância Máxima		
	100%	80%	70%
$\hat{p}_{\text{Exponencial}}$	93%	98%	98%
$\hat{p}_{\text{Esférico}}$	98%	98%	<b>99%</b>
$\hat{p}_{\text{Gaussiano}}$	<b>99%</b>	96%	98%
$\hat{p}_{\text{Wave}}$	<b>99%</b>	<b>100%</b>	<b>99%</b>

Observando as proporções na Tabela 12, considerando intervalos com 95% de confiança, tem-se que: Considerando 100% da distância máxima, as maiores proporções foram dos modelos gaussiano e wave, apresentando inconsistência na seleção. Considerando 80%, a maior proporção foi do modelo wave, sendo selecionado erroneamente. Considerando 70%, as maiores proporções foram dos modelos esférico e wave, apresentando inconsistência na seleção.

### Aplicação do método 2

Na Figura 28 apresenta-se o intervalo de confiança para o erro de predição em relação a validação do modelo esférico, considerando 100% da distância máxima.

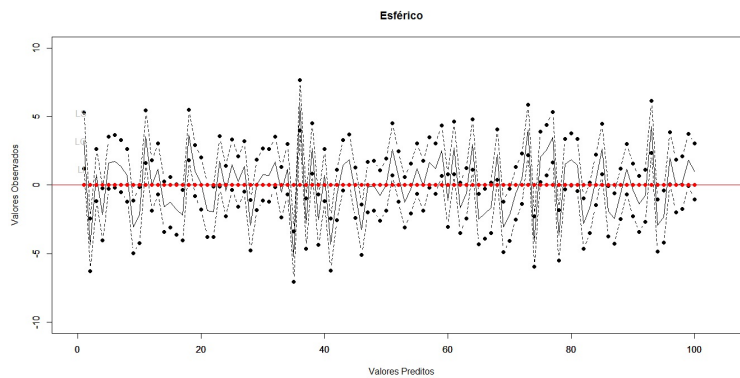


Figura 28 Intervalo de confiança para o erro de predição em relação a validação do modelo esférico, considerando 100% da distância máxima

Na Figura 29 apresenta-se o intervalo de confiança para o erro de predição em relação a validação do modelo esférico, considerando 80% da distância máxima.

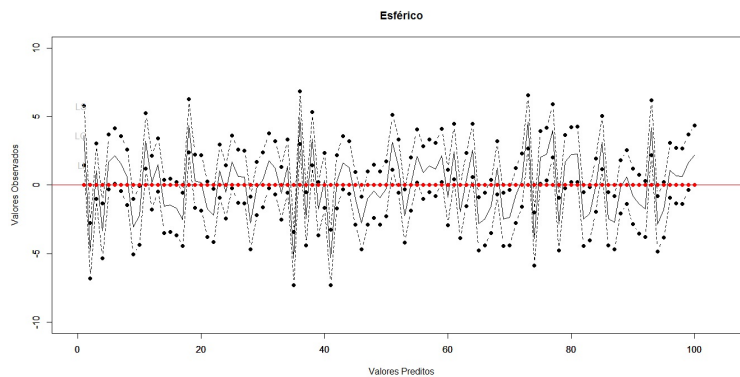


Figura 29 Intervalo de confiança para o erro de predição em relação a validação do modelo esférico, considerando 80% da distância máxima

Na Figura 30 apresenta-se o intervalo de confiança para o erro de predição em relação a validação do modelo esférico, considerando 70% da distância

máxima.

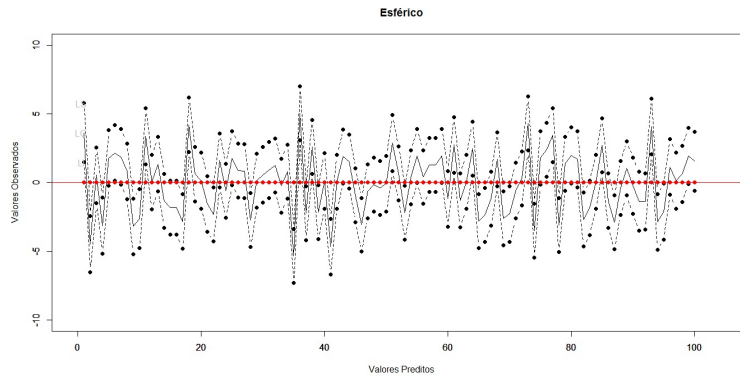


Figura 30 Intervalo de confiança para o erro de predição em relação a validação do modelo esférico, considerando 70% da distância máxima

Calculando a proporção de vezes que o valor 0 esteve contido nos intervalos de confiança para o erro de predição em relação a variância de krigagem, tem-se as seguintes proporções na Tabela 13:

Tabela 13 Estimativas das proporções do Método 2

$\hat{p}_{\text{Modelo}}$	Distância Máxima		
	100%	80%	70%
$\hat{p}_{\text{Exponencial}}$	52%	63%	57%
$\hat{p}_{\text{Esférico}}$	<b>68%</b>	<b>64%</b>	<b>69%</b>
$\hat{p}_{\text{Gaussiano}}$	62%	59%	65%
$\hat{p}_{\text{Wave}}$	53%	52%	54%

Observando as proporções na Tabela 13, considerando 100%, 80% e 70% da distância máxima, a maior foi do modelo esférico para os três casos, sendo selecionado corretamente.

### Aplicação do método 3

Na Tabela 14 tem-se os valores- $p$  do teste  $t$  e as distâncias médias:

Tabela 14 Valores- $p$  do teste  $t$  e as distâncias médias entre os valores observados e preditos, considerando 100%, 80% e 70% da distância máxima

100% da distância máxima do semivariograma			
Teste t		Distância média	
Valor- $p_{\beta_0}$ Exponencial	0,9757	$\bar{D}$ Exponencial	2,2745
Valor- $p_{\beta_1}$ Exponencial	0,0236		
Valor- $p_{\beta_0}$ Esférico	<b>0,3929</b>	$\bar{D}$ Esférico	<b>1,7792</b>
Valor- $p_{\beta_1}$ Esférico	<b>0,6031</b>		
Valor- $p_{\beta_0}$ Gaussiano	<b>0,5718</b>	$\bar{D}$ Gaussiano	1,8046
Valor- $p_{\beta_1}$ Gaussiano	<b>0,4233</b>		
Valor- $p_{\beta_0}$ Wave	<b>0,7538</b>	$\bar{D}$ Wave	1,9281
Valor- $p_{\beta_1}$ Wave	<b>0,2391</b>		
80% da distância máxima do semivariograma			
Teste t		Distância média	
Valor- $p_{\beta_0}$ Exponencial	<b>0,8512</b>	$\bar{D}$ Exponencial	1,8677
Valor- $p_{\beta_1}$ Exponencial	<b>0,1427</b>		
Valor- $p_{\beta_0}$ Esférico	<b>0,3422</b>	$\bar{D}$ Esférico	<b>1,8590</b>
Valor- $p_{\beta_1}$ Esférico	<b>0,6595</b>		
Valor- $p_{\beta_0}$ Gaussiano	<b>0,1982</b>	$\bar{D}$ Gaussiano	1,9148
Valor- $p_{\beta_1}$ Gaussiano	<b>0,8073</b>		
Valor- $p_{\beta_0}$ Wave	<b>0,7026</b>	$\bar{D}$ Wave	1,9349
Valor- $p_{\beta_1}$ Wave	<b>0,0798</b>		
70% da distância máxima do semivariograma			
Teste t		Distância média	
Valor- $p_{\beta_0}$ Exponencial	<b>0,8491</b>	$\bar{D}$ Exponencial	1,8711
Valor- $p_{\beta_1}$ Exponencial	<b>0,1283</b>		
Valor- $p_{\beta_0}$ Esférico	<b>0,3424</b>	$\bar{D}$ Esférico	<b>1,7725</b>
Valor- $p_{\beta_1}$ Esférico	<b>0,3925</b>		
Valor- $p_{\beta_0}$ Gaussiano	<b>0,1982</b>	$\bar{D}$ Gaussiano	1,8544
Valor- $p_{\beta_1}$ Gaussiano	<b>0,4708</b>		
Valor- $p_{\beta_0}$ Wave	<b>0,6665</b>	$\bar{D}$ Wave	2,0103
Valor- $p_{\beta_1}$ Wave	<b>0,1533</b>		

Observando a Tabela 14, retiram-se as seguintes informações: Considerando 100% da distância máxima, ambos os valores- $p$  referentes aos modelos esférico, gaussiano e wave indicaram a não rejeição de  $H_0$ , a 5% de significância, informando que suas retas ajustadas são iguais à ideal. Realizando o passo 2, a menor estimativa foi do modelo esférico, sendo selecionado corretamente. Consi-

derando 80% e 70% da distância máxima, ambos valores- $p$  referentes aos modelos também indicaram a não rejeição de  $H_0$ , a 5% de significância. Realizando o passo 2, a menor estimativa foi do modelo esférico, sendo selecionado corretamente.

#### Aplicação do método 4

Na Tabela 15 tem-se os valores- $p$  do teste  $F$  e as distâncias médias:

Tabela 15 Valores- $p$  do teste  $F$  e as distâncias médias entre os valores observados e preditos, considerando 100%, 80% e 70% da distância máxima

100% da distância máxima do semivariograma			
Teste F		Distância média	
Valor- $p$ Exponencial	<b>0,7772</b>	$\bar{D}_{\text{Exponencial}}$	2,2745
Valor- $p$ Esférico	<b>0,9659</b>	$\bar{D}_{\text{Esférico}}$	<b>1,7792</b>
Valor- $p$ Gaussiano	<b>0,9810</b>	$\bar{D}_{\text{Gaussiano}}$	1,8046
Valor- $p$ Wave	<b>0,7772</b>	$\bar{D}_{\text{Wave}}$	1,9281
80% da distância máxima do semivariograma			
Teste F		Distância média	
Valor- $p$ Exponencial	<b>0,5672</b>	$\bar{D}_{\text{Exponencial}}$	1,8677
Valor- $p$ Esférico	<b>0,9182</b>	$\bar{D}_{\text{Esférico}}$	<b>1,8590</b>
Valor- $p$ Gaussiano	<b>0,6865</b>	$\bar{D}_{\text{Gaussiano}}$	1,9148
Valor- $p$ Wave	<b>0,3773</b>	$\bar{D}_{\text{Wave}}$	1,9349
70% da distância máxima do semivariograma			
Teste F		Distância média	
Valor- $p$ Exponencial	<b>0,5291</b>	$\bar{D}_{\text{Exponencial}}$	1,8711
Valor- $p$ Esférico	<b>0,9630</b>	$\bar{D}_{\text{Esférico}}$	<b>1,7725</b>
Valor- $p$ Gaussiano	<b>0,9970</b>	$\bar{D}_{\text{Gaussiano}}$	1,8544
Valor- $p$ Wave	<b>0,5948</b>	$\bar{D}_{\text{Wave}}$	2,0103

Observando as estimativas na Tabela 15, considerando 100%, 80% e 70% da distância máxima do semivariograma esférico, os valores- $p$  em relação aos modelos adotados indicaram a não rejeição da hipótese nula, a 5% de significância, informando que suas retas ajustadas são iguais a reta ideal. Realizando o passo 2, a menor estimativa foi referente ao modelo esférico, sendo selecionado corretamente.

### Aplicação do método 5

Na Tabela 16 são apresentados os valores- $p$  referentes ao teste da correlação entre as diferenças absolutas e a média dos valores observados e preditos, os valores- $p$  referente ao teste do viés em relação aos valores observados e preditos, e os limites de concordância.

Tabela 16 Valores- $p$  referentes ao teste da correlação e teste do viés, e estimativas dos limites de concordância, considerando 100%, 80% e 70% da distância máxima

100% da distância máxima do semivariograma				
Teste da Correlação		Teste do Viés		Limites
Valor- $p$ Exponencial	<b>0,4761</b>	Valor- $p$ Exponencial	<b>0,5009</b>	$\widehat{LC}$ Exponencial 5,4167
Valor- $p$ Esférico	<b>0,7456</b>	Valor- $p$ Esférico	<b>0,4895</b>	$\widehat{LC}$ Esférico <b>4,2642</b>
Valor- $p$ Gaussiano	<b>0,7613</b>	Valor- $p$ Gaussiano	<b>0,4883</b>	$\widehat{LC}$ Gaussiano 4,2969
Valor- $p$ Wave	<b>0,5540</b>	Valor- $p$ Wave	<b>0,4875</b>	$\widehat{LC}$ Wave 4,5654
80% da distância máxima do semivariograma				
Teste da Correlação		Teste do Viés		Limites
Valor- $p$ Exponencial	<b>0,6461</b>	Valor- $p$ Exponencial	<b>0,4946</b>	$\widehat{LC}$ Exponencial 4,4907
Valor- $p$ Esférico	<b>0,7276</b>	Valor- $p$ Esférico	<b>0,4867</b>	$\widehat{LC}$ Esférico <b>4,4055</b>
Valor- $p$ Gaussiano	<b>0,7292</b>	Valor- $p$ Gaussiano	<b>0,4866</b>	$\widehat{LC}$ Gaussiano 4,5351
Valor- $p$ Wave	<b>0,4738</b>	Valor- $p$ Wave	<b>0,4879</b>	$\widehat{LC}$ Wave 4,5391
70% da distância máxima do semivariograma				
Teste da Correlação		Teste do Viés		Limites
Valor- $p$ Exponencial	<b>0,6418</b>	Valor- $p$ Exponencial	<b>0,4948</b>	$\widehat{LC}$ Exponencial 4,4972
Valor- $p$ Esférico	<b>0,7649</b>	Valor- $p$ Esférico	<b>0,4906</b>	$\widehat{LC}$ Esférico <b>4,2558</b>
Valor- $p$ Gaussiano	<b>0,7350</b>	Valor- $p$ Gaussiano	<b>0,4892</b>	$\widehat{LC}$ Gaussiano 4,3928
Valor- $p$ Wave	<b>0,5615</b>	Valor- $p$ Wave	<b>0,4971</b>	$\widehat{LC}$ Wave 4,7689

Observando as estimativas na Tabela 16, considerando 100%, 80% e 70% da distância máxima do semivariograma esférico, tanto os valores- $p$  do teste da correlação quanto aos do teste do viés indicaram a não rejeição da hipótese nula, a 5% de significância, concluindo-se que tanto o coeficiente de correlação populacional quanto o viés populacional são iguais a 0. Para selecionar o modelo que mais bem se ajustou ao semivariograma, foram observadas as estimativas dos limites de concordância e a menor foi do modelo esférico, sendo selecionado corretamente.

### Aplicação do método 6

Na Tabela 17 são apresentadas as estimativas do  $QV$ ,  $DQDP$ ,  $FCPDP$  e  $DQM$ , para os modelos adotados.

Tabela 17  $DQM$ ,  $QV$ ,  $DQDP$  e  $FCPDP$ , em relação aos modelos adotados, considerando 100%, 80% e 70% da distância máxima

100% da distância máxima do semivariograma				
Modelo	$QV$	$DQDP$	$FCPDP$	$DQM$
Exponencial	4e-07	4,2323	3,4055	7,6377
Esférico	3e-05	0,6398	4,0934	<b>4,7332</b>
Gaussiano	4e-05	0,8399	3,9663	4,8062
Wave	5e-05	1,3751	4,0505	5,4257
80% da distância máxima do semivariograma				
Modelo	$QV$	$DQDP$	$FCPDP$	$DQM$
Exponencial	9e-06	1,4526	3,7970	5,2495
Esférico	6e-05	0,6926	4,3596	<b>5,0523</b>
Gaussiano	6e-05	0,6133	4,7404	5,3538
Wave	5e-05	1,6936	3,6697	5,3634
70% da distância máxima do semivariograma				
Modelo	$QV$	$DQDP$	$FCPDP$	$DQM$
Exponencial	9e-06	1,4943	3,7703	5,2646
Esférico	3e-05	0,8343	3,8802	<b>4,7146</b>
Gaussiano	4e-05	0,8795	4,1435	5,0230
Wave	3e-06	1,8623	4,0577	5,9200

Observando a Tabela 17, as estimativas do  $QV$  e  $FCPDP$  foram parecidas, sendo irrelevante considerá-las na seleção, já as estimativas do  $DQDP$  apresentaram uma maior variabilidade, sendo que as menores foram dos modelos esférico e gaussiano. Dentre as estimativas do  $DQM$ , o modelo esférico resultou a menor. As estimativas dos componentes do modelo esférico apresentaram um bom comportamento, sendo selecionado corretamente.

### Aplicação do método 7

Na Tabela 18 tem-se os seguintes resultados:

Tabela 18 Estimativas ranqueadas, com as médias e os desvios padrões dos ranks, considerando 100%, 80% e 70% da distância máxima

<b>100% da distância máxima do semivariograma</b>				
Modelo				
Método	Exponencial	Esférico	Gaussiano	Wave
D	7,638	4,733	4,806	5,426
Rank	0,614	0,000	0,015	0,146
E	0,282	0,534	0,527	0,469
Rank	0,472	0,000	0,013	0,122
H	0,332	0,702	0,686	0,618
Rank	0,527	0,000	0,023	0,120
$\bar{D}$	2,274	1,779	1,805	1,928
Rank	0,278	0,000	0,015	0,084
$\overline{\text{Rank}}_j$	0,473	0,000	0,016	0,118
$DP_{\text{Rank}_j}$	0,142	0,000	0,004	0,026
<b>80% da distância máxima do semivariograma</b>				
Modelo				
Método	Exponencial	Esférico	Gaussiano	Wave
D	5,249	5,052	5,354	5,363
Rank	0,039	0,000	0,060	0,062
E	0,489	0,503	0,477	0,482
Rank	0,028	0,000	0,052	0,042
H	0,627	0,678	0,664	0,609
Rank	0,075	0,000	0,021	0,102
$\bar{D}$	1,868	1,860	1,915	1,935
Rank	0,004	0,000	0,030	0,040
$\overline{\text{Rank}}_j$	0,036	0,000	0,041	0,061
$DP_{\text{Rank}_j}$	0,029	0,000	0,018	0,029
<b>70% da distância máxima do semivariograma</b>				
Modelo				
Método	Exponencial	Esférico	Gaussiano	Wave
D	5,265	4,715	5,023	5,920
Rank	0,117	0,000	0,065	0,256
E	0,488	0,536	0,505	0,423
Rank	0,089	0,000	0,058	0,211
H	0,624	0,692	0,670	0,560
Rank	0,098	0,000	0,032	0,191
$\bar{D}$	1,871	1,772	1,854	2,010
Rank	0,056	0,000	0,046	0,134
$\overline{\text{Rank}}_j$	0,090	0,000	0,050	0,198
$DP_{\text{Rank}_j}$	0,025	0,000	0,014	0,051

Observando a Tabela 18, os métodos escolhidos para serem compostos foram: D, E, H e  $\bar{D}$  dado nos métodos 3 e 4. Suas estimativas foram ranqueadas. Após o ranqueamento, foram calculadas as médias e os desvios padrões dos ranks,

em relação a cada modelo. Na Figura 31 tem-se o gráfico dos pontos formados por  $\overline{\text{Rank}}$  e  $DP_{\text{Rank}}$ , com as distâncias em relação ao ponto mínimo.

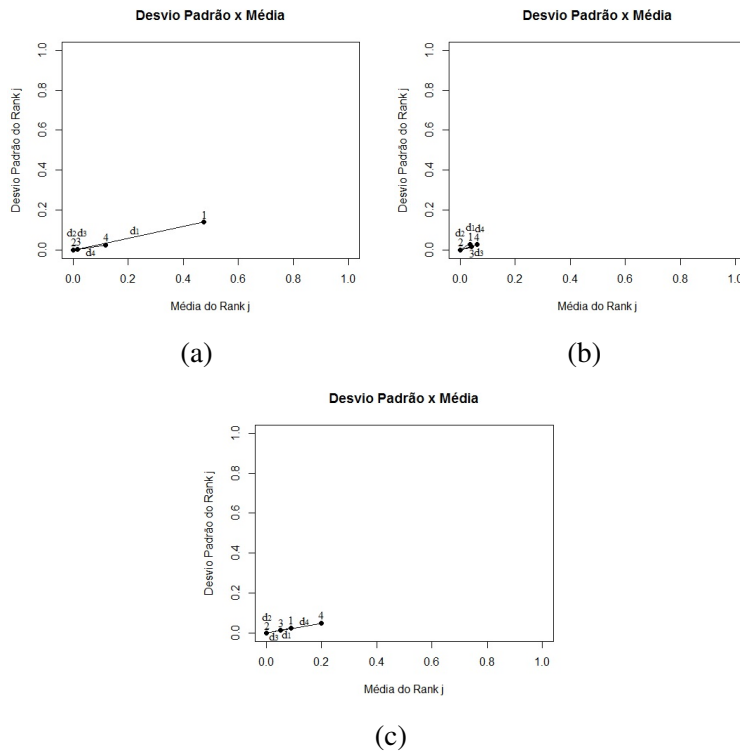


Figura 31 Pontos formados pelos desvios padrões em relação as médias dos ranks, com as distâncias desses pontos em relação ao ponto (0,0), considerando (a) 100% , (b) 80% e (c) 70% da distância máxima

Essas distâncias são apresentadas na Tabela 19:

Tabela 19 Distâncias referentes a Figura 31

$d_{\text{Modelo}}$	Distância Máxima		
	100%	80%	70%
$d_{\text{Exponencial}} (d_1)$	0,494	0,046	0,093
$d_{\text{Esférico}} (d_2)$	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>
$d_{\text{Gaussiana}} (d_3)$	0,016	0,002	0,052
$d_{\text{Wave}} (d_4)$	0,121	0,067	0,204

Observando a Tabela 19, para as três especificações da distância máxima, a menor distância foi do modelo esférico, sendo selecionado corretamente.

### Resultados das aplicações dos métodos propostos em dados simulados com dependência espacial esférica

A Tabela 20 apresenta os resultados das aplicações dos métodos propostos em relação ao semivariograma com variabilidade alta nas maiores distâncias.

Tabela 20 Resultados das aplicações dos métodos propostos em dados simulados com dependência espacial esférica, considerando 100% da distância máxima

Método	Estatística	Modelo				Modelo Selecionado
		Exponencial	Esférico	Gaussiano	Wave	
1	$\hat{p}$	93%	98%	<b>99%</b>	<b>99%</b>	Inconsistente
2	$\hat{p}$	52%	<b>68%</b>	62%	53%	<b>Esférico (Seleção Correta)</b>
3	Valor- $p_{\beta_0}$	0,9757	<b>0,3927</b>	<b>0,5718</b>	<b>0,7538</b>	<b>Esférico (Seleção Correta)</b>
	Valor- $p_{\beta_1}$	0,0236	<b>0,6031</b>	<b>0,4233</b>	<b>0,2391</b>	
	$\bar{D}$	2,2745	<b>1,7792</b>	1,8046	1,9281	
4	Valor-p	<b>0,7772</b>	<b>0,9659</b>	<b>0,9810</b>	<b>0,7772</b>	<b>Esférico (Seleção Correta)</b>
	$\bar{D}$	2,2745	<b>1,7792</b>	1,8046	1,9281	
5	Valor-p	0,4761	0,7456	0,7613	0,5540	<b>Esférico (Seleção Correta)</b>
	Valor-p	0,5009	0,4895	0,4883	0,4875	
	Limite	5,4167	<b>4,2642</b>	4,2969	4,5654	
6	$QV$	4e-07	3e-05	4e-05	5e-05	<b>Esférico (Seleção Correta)</b>
	$DQDP$	4,2323	0,6398	0,8399	1,3751	
	$FCDP$	3,4055	4,0934	3,9663	4,0505	
	$DQM$	7,6377	<b>4,7332</b>	4,8062	5,4257	
7	$d_{Modelo}$	0,4940	<b>0,0000</b>	0,0160	0,1210	<b>Esférico (Seleção Correta)</b>

Observando a Tabela 20, o método 1 apresentou inconsistência na seleção e o restante dos métodos propostos selecionaram corretamente o modelo esférico.

A Tabela 21 apresenta os resultados das aplicações dos métodos propostos em relação ao semivariograma com variabilidade média nas maiores distâncias.

Tabela 21 Resultados das aplicações dos métodos propostos em dados simulados com dependência espacial esférica, considerando 80% da distância máxima

Método	Estatística	Modelo				Modelo Selecionado
		Exponencial	Esférico	Gaussiano	Wave	
1	$\hat{p}$	98%	98%	96%	<b>100%</b>	<b>Wave</b> (Seleção Errada)
2	$\hat{p}$	63%	<b>64%</b>	59%	52%	<b>Esférico</b> (Seleção Correta)
3	Valor-p $\beta_0$	<b>0,8512</b>	<b>0,3422</b>	<b>0,1982</b>	<b>0,7026</b>	<b>Esférico</b> (Seleção Correta)
	Valor-p $\beta_1$	<b>0,1427</b>	<b>0,6595</b>	<b>0,8073</b>	<b>0,0798</b>	
	$\bar{D}$	1,8677	<b>1,8590</b>	1,9148	1,9349	
4	Valor-p	<b>0,4946</b>	<b>0,4867</b>	<b>0,4866</b>	<b>0,4971</b>	<b>Esférico</b> (Seleção Correta)
	$\bar{D}$	1,8677	<b>1,8590</b>	1,9148	1,9349	
5	Valor-p	0,6461	0,7276	0,7292	0,4738	<b>Esférico</b> (Seleção Correta)
	Valor-p	0,4946	0,4867	0,4866	0,4879	
	Limite	4,4907	<b>4,4055</b>	4,5351	4,5391	
6	$QV$	9e-06	6e-05	6e-05	5e-05	<b>Esférico</b> (Seleção Correta)
	$DQDP$	1,4526	0,6926	0,6133	1,6936	
	$FCPDP$	3,7970	4,3596	4,7404	3,6697	
	$DQM$	5,2495	<b>5,0523</b>	5,3538	5,3634	
7	$d_{Modelo}$	0,0460	<b>0,0000</b>	0,0020	0,0670	<b>Esférico</b> (Seleção Correta)

Observando a Tabela 21, o método 1 apresentou inconsistência na seleção e o restante selecionaram corretamente o modelo esférico. Comparando os resultados dessa análise com a análise considerando variabilidade alta nas maiores distâncias do semivariograma, as informações retiradas são iguais.

A Tabela 22 apresenta os resultados das aplicações dos métodos propostos em relação ao semivariograma com variabilidade baixa nas maiores distâncias.

Tabela 22 Resultados das aplicações dos métodos propostos em dados simulados com dependência espacial esférica, considerando 70% da distância máxima

Método	Estatística	Modelo				Modelo Selecionado
		Exponencial	Esférico	Gaussiano	Wave	
1	$\hat{p}$	98%	<b>99%</b>	98%	<b>99%</b>	<b>Inconsistente</b>
2	$\hat{p}$	57%	<b>69%</b>	65%	54%	<b>Esférico (Seleção Correta)</b>
3	Valor-p $\beta_0$	<b>0,8491</b>	<b>0,3426</b>	<b>0,1984</b>	<b>0,6665</b>	<b>Esférico (Seleção Correta)</b>
	Valor-p $\beta_1$	<b>0,1283</b>	<b>0,3925</b>	<b>0,4708</b>	<b>0,1533</b>	
	$\bar{D}$	1,8711	<b>1,7725</b>	1,8544	2,0103	
4	Valor-p	<b>0,5291</b>	<b>0,9630</b>	<b>0,9970</b>	<b>0,5948</b>	<b>Esférico (Seleção Correta)</b>
	$\bar{D}$	1,8711	<b>1,7725</b>	1,8544	2,0103	
5	Valor-p	0,6418	0,7649	0,7350	0,5615	<b>Esférico (Seleção Correta)</b>
	Valor-p	0,4948	0,4906	0,4892	0,4971	
	Limite	4,4972	<b>4,2558</b>	4,3928	4,7689	
6	$QV$	9e-06	3e-05	4e-05	3e-06	<b>Esférico (Seleção Correta)</b>
	$DQDP$	1,4943	0,8343	0,8795	1,8623	
	$FCDP$	3,7703	3,8802	4,1435	4,0577	
	$DQM$	5,2646	<b>4,7146</b>	5,0230	5,9200	
7	$d_{Modelo}$	0,4940	<b>0,0000</b>	0,0160	0,1210	<b>Esférico (Seleção Correta)</b>

Observando a Tabela 22, o método 1 apresentou inconsistência na seleção e o restante selecionaram corretamente o modelo esférico. Comparando os resultados dessa análise com as análises considerando variabilidade alta e média nas maiores distâncias do semivariograma, as informações retiradas são iguais.

Diante as aplicações dos métodos propostos, os resultados sugerem que o método 1 não se comportou bem. Se os pesquisadores utilizarem os métodos

propostos do 2 ao 7, os resultados sugerem que irão selecionar corretamente o modelo esférico. Em relação as especificações da distância máxima, as estimativas dos métodos 2 ao 7 referentes ao modelo esférico, que mais bem se ajustou, foram parecidas. As especificações não interferiram nas estimativas desses métodos.

Em seguida, apresenta-se apenas os resultados dos métodos usuais e propostos aplicados em dados simulados com dependência espacial exponencial, gaussiana e wave, para evitar cansar o leitor com a repetição de todos os detalhes.

#### 4.2 Aplicações dos métodos usuais e propostos em dados simulados com dependência espacial exponencial, gaussiana e wave

Nesta seção serão apresentados apenas as tabelas que resumem os resultados dos casos exponencial, gaussiano e wave. Ressalta-se que para a construção dessas tabelas foram realizados os mesmos passos detalhados do caso esférico.

##### Dados simulados com dependência espacial exponencial

Na Figura 32 apresenta-se o semivariograma em relação aos dados simulados com dependência espacial exponencial, considerando as três especificações.

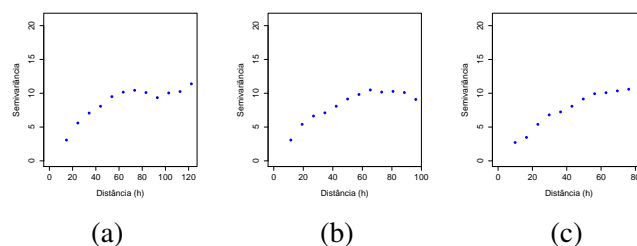


Figura 32 Semivariograma exponencial com (a) 100%, (b) 80% e (c) 70% da distância máxima

A Tabela 23 apresenta os resultados das aplicações dos métodos usuais e propostos, ao semivariograma com variabilidade alta nas maiores distâncias.

Tabela 23 Resultados das aplicações dos métodos usuais e propostos em dados simulados com dependência espacial exponencial, considerando 100% da distância máxima

Método	Estatística	Modelo				Modelo Selecionado
		Exponencial	Esférico	Gaussiano	Wave	
A	$\beta_0$	-0,058	<b>0,017</b>	-0,026	-0,057	<b>Inconsistente</b>
	$\beta_1$	1,006	0,996	<b>1,002</b>	1,007	
	$r$	<b>0,836</b>	0,832	0,802	0,756	
B	$EM$	-0,010	-0,010	-0,007	<b>-0,001</b>	<b>Inconsistente</b>
	$VE$	<b>2,517</b>	2,573	2,987	3,589	
C	$EMP$	-0,003	-0,003	-0,002	<b>-0,001</b>	<b>Inconsistente</b>
	$VEP$	<b>0,922</b>	1,164	1,151	1,025	
D	$EQM$	<b>2,492</b>	2,548	2,958	3,553	<b>Exponencial (Seleção Correta)</b>
E	$R^2$	<b>0,699</b>	0,692	0,643	0,571	<b>Exponencial (Seleção Correta)</b>
F	$AIC$	<b>381,099</b>	383,302	398,224	416,572	<b>Exponencial (Seleção Correta)</b>
G	$EC$	<b>2,666</b>	2,726	3,163	3,798	<b>Exponencial (Seleção Correta)</b>
H	$\hat{\rho}_c$	<b>0,8221</b>	0,8189	0,7824	0,7257	<b>Exponencial (Seleção Correta)</b>
1	$\hat{p}$	<b>98%</b>	<b>98%</b>	97%	95%	<b>Inconsistente</b>
2	$\hat{p}$	<b>79%</b>	76%	67%	67%	<b>Exponencial (Seleção Correta)</b>
3	Valor-p $\beta_0$	<b>0,5434</b>	<b>0,4871</b>	<b>0,5172</b>	<b>0,5328</b>	<b>Exponencial (Seleção Correta)</b>
	Valor-p $\beta_1$	<b>0,4623</b>	<b>0,5215</b>	<b>0,4868</b>	<b>0,4664</b>	
	$\bar{D}$	<b>1,3021</b>	1,3214	1,3864	1,5358	
4	Valor-p $\bar{D}$	<b>0,9935</b>	<b>0,9965</b>	<b>0,9986</b>	<b>0,9964</b>	<b>Exponencial (Seleção Correta)</b>
	$\bar{D}$	<b>1,3021</b>	1,3214	1,3864	1,5358	
5	Valor-p	0,4872	0,4918	0,4514	0,9166	<b>Exponencial (Seleção Correta)</b>
	Valor-p	0,5248	0,5252	0,5156	0,5014	
	Limite	<b>3,1098</b>	3,1442	3,3877	3,7133	
6	$QV$	9e-05	0,0001	4e-05	4e-07	<b>Exponencial (Seleção Correta)</b>
	$DQDP$	0,2391	0,2272	0,3350	0,5222	
	$FCPDP$	2,2783	2,3462	2,6525	3,0672	
	$DQM$	<b>2,5175</b>	2,5735	2,9875	3,5893	
7	$d_{Modelo}$	<b>0,0000</b>	0,0149	0,1138	0,2642	<b>Exponencial (Seleção Correta)</b>

A Tabela 24 apresenta os resultados das aplicações dos métodos usuais e propostos, ao semivariograma com variabilidade média nas maiores distâncias.

Tabela 24 Resultados das aplicações dos métodos usuais e propostos em dados simulados com dependência espacial exponencial, considerando 80% da distância máxima

Método	Estatística	Modelo				Modelo Selecionado
		Exponencial	Esférico	Gaussiano	Wave	
A	$\beta_0$	<b>-0,099</b>	-0,373	-0,265	-0,134	<b>Exponencial</b> (Seleção Correta)
	$\beta_1$	<b>1,012</b>	1,048	1,034	1,017	
	$r$	<b>0,836</b>	0,827	0,794	0,755	
B	$EM$	-0,009	-0,006	-0,003	<b>-0,001</b>	<b>Inconsistente</b>
	$VE$	<b>2,516</b>	2,654	3,098	3,601	
C	$EMP$	-0,002	-0,002	-0,001	<b>0,000</b>	<b>Wave</b> (Seleção Errada)
	$VEP$	0,842	0,785	0,859	<b>0,905</b>	
D	$EQM$	<b>2,491</b>	2,627	3,067	3,565	<b>Exponencial</b> (Seleção Correta)
E	$R^2$	<b>0,699</b>	0,684	0,630	0,570	<b>Exponencial</b> (Seleção Correta)
F	$AIC$	<b>381,033</b>	385,922	401,670	416,857	<b>Exponencial</b> (Seleção Correta)
G	$EC$	<b>2,664</b>	2,806	3,276	3,811	<b>Exponencial</b> (Seleção Correta)
H	$\hat{\rho}_c$	<b>0,8214</b>	0,8047	0,7672	0,7226	<b>Exponencial</b> (Seleção Correta)
1	$\hat{p}$	<b>98%</b>	<b>98%</b>	95%	91%	<b>Inconsistente</b>
2	$\hat{p}$	<b>80%</b>	75%	73%	69%	<b>Exponencial</b> (Seleção Correta)
3	Valor-p $\beta_0$	0,5736	0,7427	0,6615	0,5750	<b>Exponencial</b> (Seleção Correta)
	Valor-p $\beta_1$	0,4300	0,2518	0,3341	0,4223	
	$\bar{D}$	<b>1,3018</b>	1,3365	1,4280	1,5370	
4	Valor-p $\bar{D}$	<b>0,9829</b>	<b>0,7990</b>	<b>0,9119</b>	<b>0,9809</b>	<b>Exponencial</b> (Seleção Correta)
	$\bar{D}$	<b>1,3018</b>	1,3365	1,4280	1,5370	
5	Valor-p	0,5020	0,6183	0,6145	0,9216	<b>Exponencial</b> (Seleção Correta)
	Valor-p	0,5231	0,5152	0,5077	0,5007	
	Limite	<b>3,1092</b>	3,1929	3,4498	3,7192	
6	$QV$	8e-05	3e-05	1e-05	1e-07	<b>Exponencial</b> (Seleção Correta)
	$DQDP$	0,2518	0,3719	0,4515	0,5572	
	$FCPDP$	2,2646	2,2819	2,6464	3,0435	
	$DQM$	<b>2,5164</b>	2,6538	3,0980	3,6007	
7	$d_{Modelo}$	<b>0,0000</b>	0,0346	0,1435	0,2674	<b>Exponencial</b> (Seleção Correta)

A Tabela 25 apresenta os resultados das aplicações dos métodos usuais e propostos, ao semivariograma com variabilidade baixa nas maiores distâncias.

Tabela 25 Resultados das aplicações dos métodos usuais e propostos em dados simulados com dependência espacial exponencial, considerando 70% da distância máxima

Método	Estatística	Modelo				Modelo Selecionado
		Exponencial	Esférico	Gaussiano	Wave	
A	$\beta_0$	<b>-0,034</b>	-0,161	-0,108	-0,043	<b>Exponencial</b> (Seleção Correta)
	$\beta_1$	<b>1,003</b>	1,020	1,014	1,006	
	$r$	<b>0,836</b>	0,831	0,799	0,758	
B	$EM$	-0,010	-0,008	-0,005	<b>-0,001</b>	<b>Inconsistente</b>
	$VE$	<b>2,518</b>	2,591	3,032	3,557	
C	$EMP$	-0,003	-0,002	-0,001	<b>0,000</b>	<b>Inconsistente</b>
	$VEP$	<b>0,945</b>	0,957	1,026	1,024	
D	$EQM$	<b>2,493</b>	2,565	3,002	3,521	<b>Exponencial</b> (Seleção Correta)
E	$R^2$	<b>0,699</b>	0,691	0,638	0,575	<b>Exponencial</b> (Seleção Correta)
F	$AIC$	<b>381,144</b>	383,905	399,672	415,662	<b>Exponencial</b> (Seleção Correta)
G	$EC$	<b>2,667</b>	2,742	3,208	3,764	<b>Exponencial</b> (Seleção Correta)
H	$\hat{\rho}_c$	<b>0,8225</b>	0,8140	0,7766	0,7292	<b>Exponencial</b> (Seleção Correta)
1	$\hat{p}$	<b>98%</b>	<b>98%</b>	96%	92%	<b>Inconsistente</b>
2	$\hat{p}$	<b>79%</b>	76%	68%	68%	<b>Exponencial</b> (Seleção Correta)
3	Valor-p $\beta_0$	<b>0,5252</b>	<b>0,6148</b>	<b>0,5702</b>	<b>0,5246</b>	<b>Exponencial</b> (Seleção Correta)
	Valor-p $\beta_1$	<b>0,4817</b>	<b>0,3862</b>	<b>0,4303</b>	<b>0,4746</b>	
	$\bar{D}$	<b>1,3028</b>	1,3242	1,4049	1,5264	
4	Valor-p $\bar{D}$	<b>0,9968</b>	<b>0,9575</b>	<b>0,9842</b>	<b>0,9980</b>	<b>Exponencial</b> (Seleção Correta)
	Limite	<b>3,1104</b>	3,1550	3,4128	3,6964	
5	Valor-p	0,4814	0,5466	0,5144	0,9002	<b>Exponencial</b> (Seleção Correta)
	Valor-p	0,5258	0,5212	0,5122	0,5009	
	Limite	<b>3,1104</b>	3,1550	3,4128	3,6964	
	Limite	<b>3,1104</b>	3,1550	3,4128	3,6964	
6	$QV$	0,0001	7e-05	2e-05	2e-07	<b>Exponencial</b> (Seleção Correta)
	$DQDP$	0,2317	0,2870	0,3762	0,5058	
	$FCPDP$	2,2867	2,3040	2,6558	3,0509	
	$DQM$	<b>2,5185</b>	2,5912	3,0320	3,5567	
7	$d_{Modelo}$	<b>0,0000</b>	0,0186	0,1248	0,2556	<b>Exponencial</b> (Seleção Correta)

Diante das estimativas dos métodos usuais apresentadas nas Tabelas 23, 24 e 25, pode-se dizer que o problema de pesquisa foi comprovado. Em relação às três especificações da distância máxima, os resultados sugerem que os métodos D, E, F, G e H selecionaram corretamente e o método C novamente não apresentou um bom comportamento. As estimativas dos métodos D, E, F, G e H referentes ao modelo exponencial, que mais bem se ajustou, foram parecidas. As especificações não interferiram nas estimativas desses métodos.

Diante das estimativas dos métodos propostos apresentadas nas Tabelas 23, 24 e 25, em relação às três especificações da distância máxima, apenas o método 1 não se comportou bem. Se os pesquisadores utilizarem os métodos 2 ao 7, os resultados sugerem que irão selecionar corretamente. As estimativas dos métodos 2 ao 7 referentes ao modelo exponencial, que mais bem se ajustou, foram parecidas. As especificações não interferiram nas estimativas desses métodos.

### Dados simulados com dependência espacial gaussiana

Na Figura 33 apresenta-se o semivariograma em relação aos dados simulados com dependência espacial gaussiana, considerando três especificações.

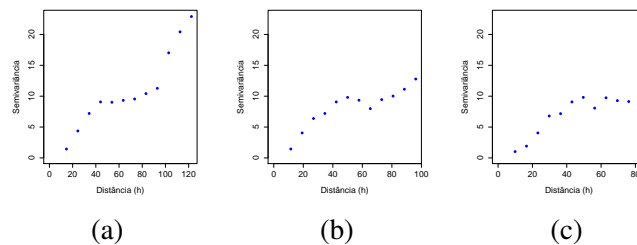


Figura 33 Semivariograma gaussiano com (a) 100%, (b) 80% e (c) 70% da distância máxima

A Tabela 26 apresenta os resultados das aplicações dos métodos usuais e propostos, ao semivariograma com variabilidade alta nas maiores distâncias.

Tabela 26 Resultados das aplicações dos métodos usuais e propostos em dados simulados com dependência espacial gaussiana, considerando 100% da distância máxima

Método	Estatística	Modelo				Modelo Selecionado
		Exponencial	Esférico	Gaussiano	Wave	
A	$\beta_0$	-0,473	-1,111	<b>-0,149</b>	-1,207	<b>Inconsistente</b>
	$\beta_1$	1,061	1,143	<b>1,020</b>	1,155	
	$r$	<b>0,993</b>	0,980	0,988	0,750	
B	$EM$	0,001	<b>0,000</b>	0,003	-0,002	<b>Inconsistente</b>
	$VE$	<b>0,138</b>	0,464	0,203	3,746	
C	$EMP$	<b>0,000</b>	<b>0,000</b>	0,001	<b>0,000</b>	<b>Exponencial</b> (Seleção Errada)
	$VEP$	<b>0,068</b>	0,140	0,234	0,811	
D	$EQM$	<b>0,136</b>	0,460	0,201	3,709	<b>Exponencial</b> (Seleção Errada)
E	$R^2$	<b>0,986</b>	0,960	0,976	0,562	<b>Exponencial</b> (Seleção Errada)
F	$AIC$	<b>68,266</b>	180,435	127,748	418,561	<b>Exponencial</b> (Seleção Errada)
G	$EC$	<b>0,144</b>	0,480	0,216	3,961	<b>Exponencial</b> (Seleção Errada)
H	$\hat{\rho}_c$	<b>0,9912</b>	0,9680	0,9875	0,6847	<b>Exponencial</b> (Seleção Errada)
1	$\hat{p}$	<b>100%</b>	<b>100%</b>	<b>100%</b>	69%	<b>Inconsistente</b>
2	$\hat{p}$	<b>100%</b>	96%	98%	48%	<b>Exponencial</b> (Seleção Errada)
3	Valor-p $\beta_0$	0,9999	1,0000	<b>0,8676</b>	<b>0,9275</b>	<b>Gaussiano</b> (Seleção Correta)
	Valor-p $\beta_1$	0,0001	0,0001	<b>0,1136</b>	<b>0,0673</b>	
	$\bar{D}$	2,1888	2,2986	<b>0,3498</b>	1,6371	
4	Valor-p $\bar{D}$	0,0001	0,0001	<b>0,4847</b>	<b>0,3329</b>	<b>Gaussiano</b> (Seleção Correta)
	$\bar{D}$	2,1888	2,2986	<b>0,3498</b>	1,6371	
5	Valor-p	0,8872	0,8699	0,9353	0,5714	<b>Exponencial</b> (Seleção Errada)
	Valor-p	0,4891	0,4975	0,4741	0,5042	
	Limite	<b>0,7273</b>	1,3357	0,8827	3,7936	
6	$QV$	1e-06	1e-07	8e-06	4e-06	<b>Exponencial</b> (Seleção Errada)
	$DQDP$	0,0341	0,1715	0,0080	1,0307	
	$FCPDP$	0,1035	0,2929	0,1948	2,7154	
	$DQM$	<b>0,1377</b>	0,4645	0,2028	3,7462	
7	$d_{Modelo}$	2,939	3,302	<b>0,266</b>	14,664	<b>Gaussiano</b> (Seleção Correta)

A Tabela 27 apresenta os resultados das aplicações dos métodos usuais e propostos, ao semivariograma com variabilidade média nas maiores distâncias.

Tabela 27 Resultados das aplicações dos métodos usuais e propostos em dados simulados com dependência espacial gaussiana, considerando 80% da distância máxima

Método	Estatística	Modelo				Modelo Selecionado
		Exponencial	Esférico	Gaussiano	Wave	
A	$\beta_0$	-0,589	-0,390	<b>-0,067</b>	-0,543	<b>Gaussiano</b> (Seleção Correta)
	$\beta_1$	1,076	1,051	<b>1,009</b>	1,070	
	$r$	0,993	0,993	<b>0,998</b>	0,933	
B	$EM$	<b>0,002</b>	0,003	0,005	<b>-0,002</b>	<b>Inconsistente</b>
	$VE$	0,149	0,142	<b>0,036</b>	1,106	
C	$EMP$	<b>0,001</b>	<b>0,001</b>	0,003	<b>-0,001</b>	<b>Exponencial</b> (Seleção Errada)
	$VEP$	<b>0,054</b>	0,066	0,086	0,374	
D	$EQM$	0,147	0,141	<b>0,036</b>	1,095	<b>Gaussiano</b> (Seleção Correta)
E	$R^2$	0,986	0,986	<b>0,996</b>	0,870	<b>Gaussiano</b> (Seleção Correta)
F	$AIC$	65,535	79,274	<b>-44,536</b>	295,999	<b>Gaussiano</b> (Seleção Correta)
G	$EC$	0,156	0,151	<b>0,043</b>	1,164	<b>Gaussiano</b> (Seleção Correta)
H	$\hat{\rho}_c$	0,9904	0,9910	<b>0,9978</b>	0,9249	<b>Gaussiano</b> (Seleção Correta)
1	$\hat{p}$	<b>100%</b>	<b>100%</b>	<b>100%</b>	98%	<b>Inconsistente</b>
2	$\hat{p}$	99%	98%	<b>100%</b>	65%	<b>Gaussiano</b> (Seleção Correta)
3	Valor-p $\beta_0$	1,0000	0,9997	<b>0,8862</b>	<b>0,9442</b>	<b>Gaussiano</b> (Seleção Correta)
	Valor-p $\beta_1$	0,0001	0,0001	<b>0,0849</b>	<b>0,0480</b>	
	$\bar{D}$	1,2872	1,1514	<b>0,1145</b>	0,8509	
4	Valor-p	0,0001	0,0019	<b>0,3842</b>	<b>0,2584</b>	<b>Gaussiano</b> (Seleção Correta)
	$\bar{D}$	1,2872	1,1514	<b>0,1145</b>	0,8509	
5	Valor-p	0,7959	0,9343	0,0483	0,9457	<b>Gaussiano</b> (Seleção Correta)
	Valor-p	0,4784	0,4654	0,4047	0,5085	
	Limite	0,7560	0,7394	<b>0,3738</b>	2,0610	
6	$QV$	4e-06	1e-05	2e-05	5e-06	<b>Gaussiano</b> (Seleção Correta)
	$DQDP$	0,0494	0,0256	0,0011	0,1356	
	$FCPDP$	0,0993	0,1167	0,0353	0,9702	
	$DQM$	0,1488	0,1423	<b>0,0363</b>	1,1058	
7	$d_{Modelo}$	5,867	5,213	<b>0,000</b>	16,588	<b>Gaussiano</b> (Seleção Correta)

A Tabela 28 apresenta os resultados das aplicações dos métodos usuais e propostos, ao semivariograma com variabilidade baixa nas maiores distâncias.

Tabela 28 Resultados das aplicações dos métodos usuais e propostos em dados simulados com dependência espacial gaussiana, considerando 70% da distância máxima

Método	Estatística	Modelo				Modelo Selecionado
		Exponencial	Esférico	Gaussiano	Wave	
A	$\beta_0$	-0,588	-0,389	<b>0,000</b>	-0,104	<b>Gaussiano</b> (Seleção Correta)
	$\beta_1$	1,076	1,051	<b>0,999</b>	1,014	
	$r$	0,993	0,993	<b>0,999</b>	0,955	
B	$EM$	0,002	0,003	<b>0,000</b>	0,003	<b>Gaussiano</b> (Seleção Correta)
	$VE$	0,149	0,141	<b>0,000</b>	0,735	
C	$EMP$	<b>0,001</b>	<b>0,001</b>	-0,003	<b>0,001</b>	<b>Exponencial</b> (Seleção Errada)
	$VEP$	<b>0,055</b>	0,068	0,247	0,528	
D	$EQM$	0,147	0,140	<b>0,001</b>	0,728	<b>Gaussiano</b> (Seleção Correta)
E	$R^2$	0,986	0,986	<b>0,998</b>	0,912	<b>Gaussiano</b> (Seleção Correta)
F	$AIC$	65,539	78,470	<b>-835,878</b>	257,871	<b>Gaussiano</b> (Seleção Correta)
G	$EC$	0,156	0,150	<b>0,002</b>	0,776	<b>Gaussiano</b> (Seleção Correta)
H	$\hat{\rho}_c$	0,9904	0,9911	<b>0,9999</b>	0,9534	<b>Gaussiano</b> (Seleção Correta)
1	$\hat{p}$	<b>100%</b>	<b>100%</b>	<b>100%</b>	99%	<b>Inconsistente</b>
2	$\hat{p}$	20%	29%	<b>100%</b>	8%	<b>Gaussiano</b> (Seleção Correta)
3	Valor-p $\beta_0$	1,0000	0,9998	<b>0,3855</b>	<b>0,6545</b>	<b>Gaussiano</b> (Seleção Correta)
	Valor-p $\beta_1$	0,0000	0,0000	<b>0,7137</b>	<b>0,3320</b>	
	$\overline{D}$	1,0113	0,9878	<b>0,0012</b>	0,6729	
4	Valor-p	0,0000	0,0019	<b>0,6761</b>	<b>0,9090</b>	<b>Gaussiano</b> (Seleção Correta)
	$\overline{D}$	1,0113	0,9878	<b>0,0012</b>	0,6729	
5	Valor-p	0,7969	0,9411	0,0758	0,9963	<b>Gaussiano</b> (Seleção Correta)
	Valor-p	0,4785	0,4655	0,7538	0,4845	
	Limite	0,7555	0,7366	<b>0,0071</b>	1,6809	
6	$QV$	4e-06	2e-05	6e-08	1e-05	<b>Gaussiano</b> (Seleção Correta)
	$DQDP$	0,0492	0,0255	4e-08	0,0280	
	$FCPDP$	0,0993	0,1158	1e-05	0,7074	
	$DQM$	0,1486	0,1412	<b>1e-05</b>	0,7355	
7	$d_{Modelo}$	472,180	460,991	<b>0,000</b>	496,099	<b>Gaussiano</b> (Seleção Correta)

Diante das estimativas dos métodos usuais das Tabelas 26, 27 e 28, os re-

sultados sugerem que, com 100% da distância máxima, todos os métodos apresentaram inconsistência ou selecionaram erroneamente. Com 80% e 70%, os métodos A, D, E, F, G e H selecionaram corretamente. O método C selecionou errado nos três casos. As estimativas dos métodos A, D, E, F, G e H referentes ao modelo gaussiano, que mais bem se ajustou, não foram parecidas. As especificações interferiram nas estimativas e a explicação está no comportamento das semivariâncias nas menores distâncias, afetando os valores dos pesos ótimos na predição.

Diante das estimativas dos métodos propostos das Tabelas 26, 27 e 28, com 100% da distância máxima, os resultados sugerem que os métodos 3, 4 e 7 apresentaram bom comportamento, pois selecionaram corretamente. Com 80% e 70%, os resultados sugerem que os métodos 2 ao 7 apresentaram bom comportamento, pois selecionaram corretamente. As estimativas dos métodos propostos referentes ao modelo gaussiano, que mais bem se ajustou, não foram parecidas. As especificações interferiram nas estimativas pelo mesmo motivo já mencionado.

### Dados simulados com dependência espacial wave

Na Figura 34 apresenta-se o semivariograma em relação aos dados simulados com dependência espacial wave, considerando as três especificações.

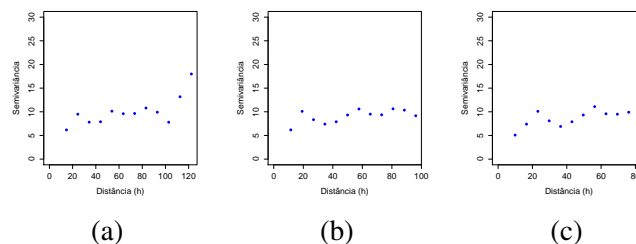


Figura 34 Semivariograma wave com (a) 100%, (b) 80% e (c) 70% da distância máxima

A Tabela 29 apresenta os resultados das aplicações dos métodos usuais e propostos, ao semivariograma com variabilidade alta nas maiores distâncias.

Tabela 29 Resultados das aplicações dos métodos usuais e propostos em dados simulados com dependência espacial wave, considerando 100% da distância máxima

Método	Estatística	Modelo				Modelo Selecionado
		Exponencial	Esférico	Gaussiano	Wave	
A	$\beta_0$	-1,094	<b>0,145</b>	2,115	-1,246	<b>Inconsistente</b>
	$\beta_1$	1,115	<b>0,984</b>	0,775	1,131	
	$r$	0,445	0,373	0,286	<b>0,982</b>	
B	$EM$	-0,010	-0,006	<b>-0,003</b>	-0,011	<b>Inconsistente</b>
	$VE$	7,323	7,845	8,425	<b>0,438</b>	
C	$EMP$	-0,002	<b>-0,001</b>	<b>-0,001</b>	-0,003	<b>Inconsistente</b>
	$VEP$	<b>1,001</b>	1,035	1,089	0,170	
D	$EQM$	7,250	7,767	8,341	<b>0,434</b>	<b>Wave</b> (Seleção Correta)
E	$R^2$	0,198	0,139	0,082	<b>0,964</b>	<b>Wave</b> (Seleção Correta)
F	$AIC$	487,624	494,771	501,153	<b>174,940</b>	<b>Wave</b> (Seleção Correta)
G	$EC$	7,815	8,400	9,058	<b>0,463</b>	<b>Wave</b> (Seleção Correta)
H	$\hat{\rho}_c$	0,3067	0,2468	0,1863	<b>0,9726</b>	<b>Wave</b> (Seleção Correta)
1	$\hat{p}$	89%	88%	88%	<b>100%</b>	<b>Wave</b> (Seleção Correta)
2	$\hat{p}$	49%	44%	50%	<b>98%</b>	<b>Wave</b> (Seleção Correta)
3	Valor-p $\beta_0$	<b>0,6937</b>	<b>0,4754</b>	<b>0,1985</b>	<b>0,9201</b>	<b>Wave</b> (Seleção Correta)
	Valor-p $\beta_1$	<b>0,3065</b>	<b>0,5257</b>	<b>0,8035</b>	<b>0,0902</b>	
	$\bar{D}$	2,1451	2,2103	2,2732	<b>0,4671</b>	
4	Valor-p	<b>0,8791</b>	<b>0,9977</b>	<b>0,6948</b>	<b>0,0883</b>	<b>Wave</b> (Seleção Correta)
	$\bar{D}$	2,1451	2,2103	2,2732	<b>0,4671</b>	
5	Valor-p	0,8908	0,8939	0,8914	0,4595	<b>Wave</b> (Seleção Correta)
	Valor-p	0,5145	0,5086	0,5046	0,5670	
	Limite	5,3040	5,4899	5,6890	<b>1,2971</b>	
6	$QV$	9e-05	3e-05	1e-05	0,0001	<b>Wave</b> (Seleção Correta)
	$DQDP$	3,2867	3,5165	3,6207	0,1572	
	$FCPDP$	4,0365	4,3289	4,8041	0,2807	
	$DQM$	7,3233	7,8454	8,4249	<b>0,4381</b>	
7	$d_{Modelo}$	8,825	9,483	10,213	<b>0,000</b>	<b>Wave</b> (Seleção Correta)

A Tabela 30 apresenta os resultados das aplicações dos métodos usuais e propostos, ao semivariograma com variabilidade média nas maiores distâncias.

Tabela 30 Resultados das aplicações dos métodos usuais e propostos em dados simulados com dependência espacial wave, considerando 80% da distância máxima

Método	Estatística	Modelo				Modelo Selecionado
		Exponencial	Esférico	Gaussiano	Wave	
A	$\beta_0$	<b>-0,126</b>	-5,492	1,685	-0,708	<b>Inconsistente</b>
	$\beta_1$	<b>1,013</b>	1,580	0,8211	1,073	
	$r$	0,358	0,930	0,283	<b>0,990</b>	
B	$EM$	-0,004	-0,014	<b>-0,001</b>	-0,014	<b>Inconsistente</b>
	$VE$	7,942	2,294	8,415	<b>0,227</b>	
C	$EMP$	-0,001	-0,013	<b>0,000</b>	0,004	<b>Inconsistente</b>
	$VEP$	<b>0,991</b>	0,382	1,042	0,136	
D	$EQM$	7,863	2,271	8,331	<b>0,225</b>	<b>Wave</b> (Seleção Correta)
E	$R^2$	0,128	0,865	0,080	<b>0,980</b>	<b>Wave</b> (Seleção Correta)
F	$AIC$	495,998	309,578	501,376	<b>120,368</b>	<b>Wave</b> (Seleção Correta)
G	$EC$	8,509	2,355	9,049	<b>0,249</b>	<b>Wave</b> (Seleção Correta)
H	$\hat{\rho}_c$	0,2250	0,8129	0,1742	<b>0,9865</b>	<b>Wave</b> (Seleção Correta)
1	$\hat{p}$	89%	99%	87%	<b>100%</b>	<b>Wave</b> (Seleção Correta)
2	$\hat{p}$	44%	70%	39%	<b>99%</b>	<b>Wave</b> (Seleção Correta)
3	Valor-p $\beta_0$	<b>0,5199</b>	1,0000	<b>0,2645</b>	1,0000	<b>Exponencial</b> (Seleção Errada)
	Valor-p $\beta_1$	<b>0,4806</b>	0,0000	<b>0,7369</b>	0,0000	
	$\bar{D}$	<b>2,216</b>	1,152	2,265	0,317	
4	Valor-p	<b>0,9987</b>	0,0001	<b>0,8178</b>	0,0002	<b>Exponencial</b> (Seleção Errada)
	$\bar{D}$	<b>2,216</b>	1,152	2,265	0,317	
5	Valor-p	0,8978	0,8273	0,8996	0,4166	<b>Wave</b> (Seleção Correta)
	Valor-p	0,5057	0,5369	0,5016	0,6119	
	Limite	5,5236	2,9684	5,6858	<b>0,9349</b>	
6	$QV$	1e-05	0,0002	1e-06	0,0002	<b>Wave</b> (Seleção Correta)
	$DQDP$	3,8077	1,5429	3,9135	0,0555	
	$FCPDP$	4,1343	0,7508	4,5017	0,1720	
	$DQM$	7,9420	2,2939	8,4153	<b>0,2277</b>	
7	$d_{Modelo}$	18,987	5,184	20,139	<b>0,000</b>	<b>Wave</b> (Seleção Correta)

A Tabela 31 apresenta os resultados das aplicações dos métodos usuais e propostos, ao semivariograma com variabilidade baixa nas maiores distâncias.

Tabela 31 Resultados das aplicações dos métodos usuais e propostos em dados simulados com dependência espacial wave, considerando 70% da distância máxima

Método	Estatística	Modelo				Modelo Selecionado
		Exponencial	Esférico	Gaussiano	Wave	
A	$\beta_0$	<b>-1,151</b>	-4,028	1,957	-1,412	<b>Inconsistente</b>
	$\beta_1$	1,159	1,424	0,792	<b>1,149</b>	
	$r$	0,466	0,928	0,286	<b>0,982</b>	
B	$EM$	-0,010	-0,020	<b>0,003</b>	-0,008	<b>Inconsistente</b>
	$VE$	7,170	1,957	8,415	<b>0,479</b>	
C	$EMP$	-0,002	-0,004	<b>0,000</b>	-0,002	<b>Inconsistente</b>
	$VEP$	<b>0,978</b>	0,385	1,070	0,165	
D	$EQM$	7,098	1,937	8,331	<b>0,474</b>	<b>Wave</b> (Seleção Correta)
E	$R^2$	0,217	0,861	0,082	<b>0,964</b>	<b>Wave</b> (Seleção Correta)
F	$AIC$	485,252	311,927	501,175	<b>178,286</b>	<b>Wave</b> (Seleção Correta)
G	$EC$	7,642	2,029	9,048	<b>0,500</b>	<b>Wave</b> (Seleção Correta)
H	$\hat{\rho}_c$	0,3224	0,8492	0,1829	<b>0,9696</b>	<b>Wave</b> (Seleção Correta)
1	$\hat{p}$	90%	99%	84%	<b>100%</b>	<b>Wave</b> (Seleção Correta)
2	$\hat{p}$	46%	82%	50%	<b>98%</b>	<b>Wave</b> (Seleção Correta)
3	Valor-p $\beta_0$	<b>0,7615</b>	1,0000	<b>0,2216</b>	1,0000	<b>Exponencial</b> (Seleção Errada)
	Valor-p $\beta_1$	<b>0,2380</b>	0,0000	<b>0,7802</b>	0,0000	
	$\bar{D}$	<b>2,1244</b>	1,0627	2,2685	0,4801	
4	Valor-p	<b>0,7755</b>	0,0001	<b>0,7420</b>	0,0000	<b>Exponencial</b> (Seleção Errada)
	$\bar{D}$	<b>2,1244</b>	1,0627	2,2685	0,4801	
5	Valor-p	0,8912	0,7566	0,8943	0,5024	<b>Wave</b> (Seleção Correta)
	Valor-p	0,5150	0,5572	0,5035	0,5459	
	Limite	5,2482	2,7418	5,6858	<b>1,3560</b>	
6	QV	0,0001	0,0004	6e-06	6e-05	<b>Wave</b> (Seleção Correta)
	DQDP	3,2580	1,1049	3,7164	0,1929	
	FCPDP	3,9119	0,8520	4,6989	0,2856	
	DQM	7,1700	1,9573	8,4153	<b>0,4787</b>	
7	$d_{Modelo}$	7,872	1,802	9,306	<b>0,000</b>	<b>Wave</b> (Seleção Correta)

Diante das estimativas dos métodos usuais das Tabelas 29, 30 e 31, em relação as especificações da distância máxima, os resultados sugerem que os métodos D, E, F, G e H selecionaram corretamente e suas estimativas não foram parecidas. As especificações interferiram nas estimativas.

Diante das estimativas dos métodos propostos das Tabelas 29, 30 e 31, com 100% da distância máxima, os resultados sugerem que todos os métodos selecionaram corretamente. Com 80% e 70%, os resultados sugerem que os métodos 3 e 4 não selecionaram corretamente. As estimativas dos métodos propostos referentes ao modelo wave, que mais bem se ajustou, não foram parecidas. As especificações interferiram nas estimativas.

Diante os cenários adotados, um resumo do comportamento dos métodos usuais e propostos são apresentados na Tabela 32:

Tabela 32 Comportamento dos métodos usuais e propostos diante os cenários

Semivariograma simulado	Distância máxima	Métodos que selecionaram corretamente
Esférico	100%	D, E, F, G, H, 2 ao 7
	80%	A, D, E, F, G, H, 2 ao 7
	70%	D, E, F, G, H, 2 ao 7
Exponencial	100%	D, E, F, G, H, 2 ao 7
	80%	A, D, E, F, G, H, 2 ao 7
	70%	A, D, E, F, G, H, 2 ao 7
Gaussiano	100%	3, 4 e 7
	80%	A, D, E, F, G, H, 2 ao 7
	70%	A, B, D, E, F, G, H, 2 ao 7
Wave	100%	D, E, F, G, H, 1 ao 7
	80%	D, E, F, G, H, 1, 2, 5 ao 7
	70%	D, E, F, G, H, 1, 2, 5 ao 7

### 4.3 Discussão em relação aos métodos usuais e propostos

Tanto os métodos usuais quanto aos propostos apresentam limitações, dentre elas:

1. Os métodos que utilizam mais de uma estimativa pontual para realizar a seleção são subjetivos, pois pode ocorrer de um modelo ser melhor de acordo com uma das estimativas, mas outro ser melhor de acordo com outra dessas estimativas;
2. Olea (2006) menciona que as conclusões derivadas dos métodos que utilizam os erros de predição devem ser tomadas com cautela pois eles não são independentes;
3. Kobayashi e Salam (2000) mencionam que quando os valores observados são comparados aos preditos, é comum realizar o ajuste do modelo de regressão linear simples e calcular o  $r$ . No entanto, há problemas nesta abordagem, pois a suposição da regressão de que  $\hat{Z}(s_i)$  está linearmente relacionado ao  $Z(s_i)$  não é garantida;
4. Hirakata e Camey (2009) mencionam que o  $r$  não avalia a concordância e sim a associação. Por exemplo, se um aparelho fornece os valores 80, 95, 96 e 100 e outro fornece os valores 90, 105, 106 e 110, o  $r$  seria igual a 1, uma associação perfeita e a discordância entre os valores é óbvia. Neter et al. (1996) menciona que mesmo se o  $r$  for alto não indica que as predições podem ser feitas por um determinado modelo, uma vez que mede a precisão e não acurácia;
5. Vieira et al. (1983) mencionam que o conceito de  $VE = \text{mínimo}$  torna-se um método subjetivo quando não se tem uma referência;

6. O  $\widehat{EQM}$  e o  $R^2$  são estimadores sensíveis a grandes variabilidades entre os valores observados e preditos, por serem muito afetados devido as suas fórmulas terem o expoente igual a 2, ocasionando decisões errôneas. Mitchell (1997) menciona também que o  $R^2$  é irrelevante uma vez que não se pretende fazer previsões a partir da reta ajustada e essa reta não é, necessariamente, a reta ideal mencionada por Dent e Blackie (1979);
7. O  $AIC$  seleciona o melhor modelo dentre os comparados mas a estimativa final não apresenta uma noção da variabilidade. Pode ocorrer do modelo selecionado ser o melhor dentre os comparados mas que não esteja ajustando bem devido a grande variabilidade;
8. Harrison (1990) menciona que a violação da suposição do teste  $t$  de independência entre as amostras, resultam valores de  $\beta_1$  e  $\beta_0$  menor que 1 e maior que 0, respectivamente;
9. Analla (1998) menciona que o teste F simultâneo não é informativo quando vários modelos estão disponíveis, por não ser capaz de selecionar o melhor.

Na próxima seção, foram realizadas 3000 simulações de dados referentes a cada cenário. Para os dados que apresentarem semivariogramas estimados respeitando  $R^2 > 0,7$  em relação ao semivariograma paramétrico desejado, com a autovalidação não apresentando erro, foram aplicados os métodos usuais e propostos. Destaca-se que, após realizar a aplicação dos métodos nas simulações, é possível verificar a proporção de acerto de cada método, sendo possível realizar a recomendação do método de seleção que apresentar a maior proporção.

#### 4.4 Resultados das simulações para identificação dos método recomendado

Na Tabela 33, são apresentadas as proporções de acerto dos modelos selecionados pelos métodos, aplicados em dados com dependência espacial esférica.

Tabela 33 Proporções de acerto dos modelos selecionados pelos métodos usuais e propostos, aplicados em dados com dependência espacial esférica

100% da distância máxima do semivariograma esférico				
Método	Modelo			
	Exponencial	Esférico	Gaussiano	Wave
A	26,3%	<b>27,1%</b>	21,6%	25,0%
B	20,1%	<b>27,6%</b>	25,2%	27,1%
C	6,4%	23,3%	25,2%	<b>45,1%</b>
D	25,0%	<b>71,4%</b>	3,6%	0,0%
E	28,6%	<b>67,8%</b>	3,6%	0,0%
F	28,6%	<b>67,8%</b>	3,6%	0,0%
G	25,0%	<b>71,4%</b>	3,6%	0,0%
H	7,1%	<b>78,6%</b>	14,3%	0,0%
1	<b>45,1%</b>	41,3%	10,3%	3,3%
2	34,8%	<b>35,2%</b>	15,2%	14,8%
3	33,9%	<b>55,5%</b>	10,6%	0,0%
4	32,1%	<b>60,7%</b>	3,6%	3,6%
5	31,8%	<b>51,2%</b>	14,9%	2,1%
6	34,2%	<b>54,2%</b>	11,6%	0,0%
7	17,9%	<b>75,0%</b>	7,1%	0,0%
80% da distância máxima do semivariograma esférico				
Método	Modelo			
	Exponencial	Esférico	Gaussiano	Wave
A	<b>28,4%</b>	26,6%	23,9%	21,1%
B	<b>40,0%</b>	20,5%	29,2%	10,3%
C	19,6%	<b>36,6%</b>	21,9%	21,9%
D	16,4%	<b>72,7%</b>	9,1%	1,8%
E	18,8%	<b>68,6%</b>	10,7%	1,8%
F	18,8%	<b>68,6%</b>	10,7%	1,8%
G	16,4%	<b>72,7%</b>	9,1%	1,8%
H	9,1%	<b>73,9%</b>	17,0%	0,0%
1	29,9%	<b>50,1%</b>	17,1%	2,9%
2	23,1%	<b>51,9%</b>	17,3%	7,7%
3	28,5%	<b>58,7%</b>	10,1%	2,7%
4	16,4%	<b>74,5%</b>	9,1%	0,0%
5	25,9%	<b>57,4%</b>	13,0%	3,7%
6	26,7%	<b>58,7%</b>	11,9%	2,7%
7	9,1%	<b>78,2%</b>	10,9%	1,8%
70% da distância máxima do semivariograma esférico				
Método	Modelo			
	Exponencial	Esférico	Gaussiano	Wave
A	32,1%	28,6%	17,9%	21,4%
B	<b>35,3%</b>	2,3%	32,0%	30,4%
C	<b>37,5%</b>	13,5%	26,0%	23,0%
D	10,1%	<b>71,6%</b>	16,3%	2,0%
E	13,1%	<b>71,1%</b>	14,3%	1,5%
F	13,1%	<b>71,1%</b>	14,3%	1,5%
G	10,1%	<b>71,6%</b>	16,3%	2,0%
H	2,1%	<b>79,0%</b>	16,8%	2,1%
1	14,3%	<b>54,3%</b>	31,4%	0,0%
2	17,2%	<b>61,6%</b>	17,7%	3,5%
3	14,4%	<b>49,8%</b>	32,4%	0,4%
4	12,6%	<b>83,5%</b>	1,9%	2,0%
5	18,6%	<b>48,1%</b>	33,3%	0,0%
6	14,3%	<b>52,0%</b>	35,7%	2,0%
7	8,0%	<b>85,4%</b>	4,5%	2,1%

Na Tabela 34, são apresentadas as proporções de acerto dos modelos selecionados pelos métodos escolhidos, aplicados em dados com dependência espacial exponencial.

Tabela 34 Proporções de acerto dos modelos selecionados pelos métodos usuais e propostos, aplicados em dados com dependência espacial exponencial

100% da distância máxima do semivariograma exponencial				
Método	Modelo			
	Exponencial	Esférico	Gaussiano	Wave
A	23,4%	<b>29,2%</b>	22,6%	24,8%
B	24,9%	<b>30,1%</b>	24,8%	20,2%
C	26,9%	21,1%	<b>28,8%</b>	23,2%
D	<b>72,7%</b>	22,7%	4,6%	0,0%
E	<b>71,8%</b>	23,6%	4,6%	0,0%
F	<b>71,8%</b>	23,6%	4,6%	0,0%
G	<b>72,7%</b>	22,7%	4,6%	0,0%
H	<b>74,5%</b>	20,0%	5,5%	0,0%
1	<b>58,5%</b>	27,8%	8,2%	5,5%
2	<b>55,7%</b>	29,1%	11,4%	3,8%
3	<b>63,8%</b>	25,4%	8,7%	2,1%
4	<b>71,8%</b>	18,2%	7,3%	2,7%
5	<b>67,9%</b>	22,6%	6,6%	2,9%
6	<b>69,6%</b>	23,9%	5,8%	0,7%
7	<b>73,6%</b>	22,7%	3,7%	0,0%
80% da distância máxima do semivariograma exponencial				
Método	Modelo			
	Exponencial	Esférico	Gaussiano	Wave
A	24,4%	<b>25,4%</b>	25,3%	24,9%
B	23,5%	<b>25,9%</b>	23,5%	27,1%
C	26,2%	<b>29,8%</b>	23,8%	20,2%
D	<b>78,5%</b>	16,3%	4,7%	0,5%
E	<b>77,8%</b>	17,7%	4,0%	0,5%
F	<b>77,8%</b>	17,7%	4,0%	0,5%
G	<b>78,5%</b>	16,3%	4,7%	0,5%
H	<b>83,8%</b>	12,7%	3,0%	0,5%
1	<b>68,3%</b>	21,6%	6,0%	4,1%
2	<b>67,0%</b>	21,4%	9,3%	2,3%
3	<b>77,5%</b>	16,3%	5,7%	4,5%
4	<b>75,8%</b>	17,7%	5,0%	1,5%
5	<b>76,3%</b>	14,7%	7,3%	1,7%
6	<b>79,4%</b>	14,4%	5,7%	0,5%
7	<b>79,8%</b>	14,7%	4,8%	0,7%
70% da distância máxima do semivariograma exponencial				
Método	Modelo			
	Exponencial	Esférico	Gaussiano	Wave
A	24,5%	<b>26,5%</b>	24,1%	24,9%
B	<b>29,6%</b>	24,1%	20,4%	25,9%
C	<b>33,3%</b>	15,2%	30,3%	21,2%
D	<b>73,0%</b>	20,5%	6,2%	0,3%
E	<b>73,3%</b>	20,6%	6,1%	0,0%
F	<b>73,3%</b>	20,6%	6,1%	0,0%
G	<b>73,0%</b>	20,5%	6,2%	0,3%
H	<b>73,6%</b>	19,6%	6,4%	0,4%
1	<b>61,1%</b>	18,9%	15,8%	4,2%
2	<b>57,0%</b>	25,5%	12,1%	5,5%
3	<b>72,6%</b>	19,8%	6,4%	1,2%
4	<b>73,4%</b>	19,4%	6,0%	0,2%
5	<b>70,5%</b>	20,9%	7,8%	0,8%
6	<b>73,2%</b>	19,6%	6,8%	0,4%
7	<b>73,7%</b>	19,6%	6,7%	0,0%

Na Tabela 35, são apresentadas as proporções de acerto dos modelos selecionados pelos métodos escolhidos, aplicados em dados com dependência espacial gaussiana.

Tabela 35 Proporções de acerto dos modelos selecionados pelos métodos usuais e propostos, aplicados em dados com dependência espacial gaussiana

100% da distância máxima do semivariograma gaussiano				
Modelo				
Método	Exponencial	Esférico	Gaussiano	Wave
A	<b>39,3%</b>	18,1%	11,8%	30,8%
B	21,7%	13,3%	<b>40,4%</b>	24,6%
C	20,2%	13,6%	23,2%	<b>43,0%</b>
D	14,4%	18,9%	<b>66,4%</b>	0,0%
E	18,4%	14,9%	<b>66,4%</b>	0,0%
F	18,4%	14,9%	<b>66,4%</b>	0,0%
G	14,4%	18,9%	<b>66,4%</b>	0,0%
H	3,1%	20,5%	<b>76,4%</b>	0,0%
1	13,3%	12,1%	<b>74,6%</b>	0,0%
2	8,3%	0,0%	<b>91,7%</b>	0,0%
3	7,7%	17,7%	<b>74,6%</b>	0,0%
4	0,0%	0,0%	<b>95,0%</b>	5,0%
5	16,9%	32,5%	<b>49,9%</b>	0,7%
6	16,3%	25,8%	<b>57,9%</b>	0,0%
7	3,1%	20,5%	<b>76,4%</b>	0,0%
80% da distância máxima do semivariograma gaussiano				
Modelo				
Método	Exponencial	Esférico	Gaussiano	Wave
A	<b>29,6%</b>	17,3%	18,5%	34,6%
B	21,9%	<b>29,3%</b>	24,4%	24,4%
C	13,3%	13,3%	<b>40,0%</b>	33,3%
D	13,4%	23,8%	<b>63,8%</b>	0,0%
E	18,0%	19,2%	<b>62,8%</b>	0,0%
F	18,0%	19,2%	<b>62,8%</b>	0,0%
G	13,4%	23,8%	<b>63,8%</b>	0,0%
H	10,1%	19,4%	<b>70,5%</b>	0,0%
1	10,1%	24,9%	<b>65,0%</b>	0,0%
2	8,5%	7,6%	<b>83,9%</b>	0,0%
3	0,0%	0,0%	<b>57,0%</b>	43,0%
4	0,8%	2,0%	<b>89,7%</b>	7,5%
5	8,6%	28,4%	<b>56,8%</b>	6,2%
6	9,6%	26,5%	<b>63,9%</b>	0,0%
7	7,1%	18,2%	<b>74,8%</b>	0,0%
70% da distância máxima do semivariograma gaussiano				
Modelo				
Método	Exponencial	Esférico	Gaussiano	Wave
A	8,3%	16,7%	<b>51,0%</b>	24%
B	24,9%	12,6%	<b>62,5%</b>	0,0%
C	25,2%	9,8%	<b>65%</b>	0,0%
D	6,7%	15,7%	<b>77,6%</b>	0,0%
E	12,5%	12,7%	<b>74,8%</b>	0,0%
F	12,5%	12,7%	<b>74,8%</b>	0,0%
G	6,7%	15,7%	<b>77,6%</b>	0,0%
H	3,7%	10,5%	<b>85,8%</b>	0,0%
1	<b>90,2%</b>	0,0%	9,8%	0,0%
2	2,0%	0,0%	<b>92,5%</b>	5,5%
3	0,0%	0,0%	<b>90,1%</b>	9,9%
4	0,0%	0,0%	<b>96,9%</b>	3,1%
5	18,0%	22,0%	<b>60,0%</b>	0,0%
6	14,7%	18,7%	<b>66,6%</b>	0,0%
7	2,7%	11,5%	<b>85,8%</b>	0,0%

Na Tabela 36, são apresentadas as proporções de acerto dos modelos selecionados pelos métodos escolhidos, aplicados em dados com dependência espacial wave.

Tabela 36 Proporções de acerto dos modelos selecionados pelos métodos usuais e propostos, aplicados em dados com dependência espacial wave

100% da distância máxima do semivariograma wave				
Modelo				
Método	Exponencial	Estérico	Gaussiano	Wave
A	24,5%	24,8%	<b>26,2%</b>	24,5%
B	24,3%	24,9%	<b>26,0%</b>	24,8%
C	21,0%	28,5%	<b>29,5%</b>	21,0%
D	0,2%	0,0%	13,5%	<b>86,3%</b>
E	0,0%	<b>100,0%</b>	0,0%	0,0%
F	0,0%	<b>100,0%</b>	0,0%	0,0%
G	0,2%	0,0%	13,5%	<b>86,3%</b>
H	0,1%	<b>99,7%</b>	0,1%	0,1%
1	0,0%	0,0%	<b>59,1%</b>	40,9%
2	0,0%	0,0%	22,1%	<b>77,9%</b>
3	16,6%	0,0%	0,5%	<b>82,9%</b>
4	3,6%	0,0%	0,5%	<b>95,9%</b>
5	0,9%	0,4%	24,2%	<b>74,5%</b>
6	0,0%	0,0%	23,1%	<b>76,9%</b>
7	0,2%	0,0%	13,5%	<b>86,3%</b>
80% da distância máxima do semivariograma wave				
Modelo				
Método	Exponencial	Estérico	Gaussiano	Wave
A	24,1%	24,3%	<b>27,2%</b>	24,4%
B	23,8%	24,1%	<b>27,6%</b>	24,5%
C	0,0%	39,4%	<b>60,6%</b>	0,0%
D	0,0%	0,0%	8,1%	<b>91,9%</b>
E	0,0%	<b>100,0%</b>	0,0%	0,0%
F	0,0%	<b>100,0%</b>	0,0%	0,0%
G	0,0%	0,0%	8,1%	<b>91,9%</b>
H	0,0%	<b>100,0%</b>	0,0%	0,0%
1	0,0%	0,0%	17,5%	<b>82,5%</b>
2	0,0%	0,0%	7,8%	<b>92,2%</b>
3	8,7%	0,0%	0,0%	<b>91,3%</b>
4	2,3%	0,0%	0,0%	<b>97,7%</b>
5	0,0%	0,0%	8,3%	<b>91,7%</b>
6	1,3%	0,0%	7,3%	<b>91,4%</b>
7	0,0%	0,0%	8,1%	<b>91,9%</b>
70% da distância máxima do semivariograma wave				
Modelo				
Método	Exponencial	Estérico	Gaussiano	Wave
A	23,5%	24,6%	<b>27,4%</b>	24,5%
B	22,7%	24,7%	<b>27,6%</b>	25,0%
C	0,0%	7,9%	<b>92,1%</b>	0,0%
D	0,0%	0,0%	5,9%	<b>94,1%</b>
E	0,0%	<b>100,0%</b>	0,0%	0,0%
F	0,0%	<b>100,0%</b>	0,0%	0,0%
G	0,0%	0,0%	5,9%	<b>94,1%</b>
H	0,0%	<b>100,0%</b>	0,0%	0,0%
1	0,0%	0,0%	9,1%	<b>90,9%</b>
2	0,0%	0,0%	10,4%	<b>89,6%</b>
3	0,0%	0,0%	0,0%	<b>100,0%</b>
4	0,0%	0,0%	0,0%	<b>100,0%</b>
5	0,7%	0,4%	12,5%	<b>86,4%</b>
6	0,0%	0,0%	6,0%	<b>94,0%</b>
7	0,0%	0,0%	7,1%	<b>92,9%</b>

Analisando as proporções de acerto apresentadas nas Tabelas 33, 34, 35 e 36, apresenta-se na Tabela 37 uma classificação dos métodos usuais e propostos em relação as maiores proporções de acerto para cada cenário adotado:

Tabela 37 Classificação dos métodos usuais e propostos em relação as maiores proporções de acerto diante os cenários adotados

Semivariograma Simulado	Distância Máxima	Classificação dos Métodos			
		1º	2º	3º	4º
Esférico	100%	H	7	D ou G	E ou F
	80%	7	4	H	D ou G
	70%	7	4	H	D ou G
Exponencial	100%	H	7	D ou G	E, F ou 4
	80%	H	7	6	D ou G
	70%	7	H	4	E ou F
Gaussiano	100%	4	2	H ou 7	1 ou 3
	80%	4	2	7	H
	70%	4	2	3	H ou 7
Wave	100%	4	D, G ou 7	3	2
	80%	4	2	D, G ou 7	5
	70%	3 ou 4	D ou G	6	7

Analisando a Tabela 37, pode-se dizer que os métodos H, 4 e 7 apresentaram um bom comportamento na classificação. Destaca-se que o método proposto 7 esteve entre as primeiras posições em todos os cenários adotados e, também, é composto pelo método H e  $\bar{D}$  utilizado no método 4.

Portanto, o método proposto 7 é o recomendado para seleção de um modelo que mais bem se ajusta ao semivariograma.

#### 4.5 Aplicação do método recomendado em dados reais

Com intuito de recomendar ao pesquisador qual mapa de krigagem utilizar na análise geoestatística aplicada nos dados reais do teor de zinco ( $mg/dm^3$ ) do solo, dentre os apresentados na Figura 1.2, foi aplicado o método recomendado no presente trabalho.

Ranqueando as estimativas dos métodos D, E, H e  $\bar{D}$  de acordo com a ordem que informa do melhor para o pior e calculando, em relação a cada modelo, as médias dos ranks ( $\overline{\text{Rank}}$ ) e os desvios padrões dos ranks ( $DP_{\text{Rank}}$ ), tem-se as seguintes estimativas na Tabela 38:

Tabela 38 Estimativas ranqueadas, com as médias e os desvios padrões dos ranks, considerando dados reais de teor de zinco

Método	Modelo			
	Exponencial	Esférico	Gaussiano	Wave
D	0,146	0,137	0,142	0,111
Rank	0,316	0,235	0,274	0,000
E	0,404	0,441	0,426	0,544
Rank	0,258	0,190	0,216	0,000
H	0,605	0,639	0,630	0,720
Rank	0,159	0,112	0,124	0,000
$\bar{D}$	0,290	0,281	0,285	0,252
Rank	0,150	0,116	0,132	0,000
$\overline{\text{Rank}}_j$	0,221	0,163	0,187	0,000
$DP_{\text{Rank}_j}$	0,080	0,060	0,072	0,000
$d_{\text{Modelo}}$	0,235	0,174	0,200	0,000

Portanto, o modelo de semivariograma selecionado foi o wave.

## 5 CONCLUSÕES

As conclusões retiradas no presente trabalho foram baseadas no estudo de 8 métodos de seleção já usuais encontrados na revisão literária e de 7 métodos propostos, sendo todos programados no software R.

Perante a classificação dos métodos usuais e propostos em relação às proporções de acerto referentes às simulações realizadas, destaca-se que o método proposto 7 esteve entre as primeiras posições em todos cenários adotados.

O método proposto 7 possui uma grande vantagem, pois é composto pelos métodos  $D$ ,  $E$ ,  $H$  e  $\bar{D}$  (utilizado no método 4) que também estiveram entre as primeiras posições.

Portanto, recomenda-se utilizá-lo na seleção do modelo que mais bem se ajusta ao semivariograma.

Como os objetivos foram atingidos, espera-se que este trabalho colabore para uma análise geoestatística mais eficiente.

## REFERÊNCIAS

ALVES, K. C. C. de. L. F. et al. Precipitação provável na Bacia Hidrográfica do Rio Formoso, Tocantins. **Water Resources and Irrigation Management**, v. 3, p. 65-78, 2014.

ANALLA, M. Model Validation Through the Linear Regression Fit to Actual Versus Predicted Values. **Agricultural Systems**, v. 57, n.1, p. 115-119, 1998.

BLAND, J. M.; ALTMAN, D. G. Measurement in medicine: the analysis of method comparison studies. **Statistician**, p. 307-317, 1983.

BLAND, J. M.; ALTMAN, D. G. Statistical methods for assessing agreement between two methods of clinical measurement. **Lancet**, p. 307-310, 1986.

BLAND, J. M.; ALTMAN, D. G. Measuring agreement in method comparison studies. **Statistical Methods in Medical Research**, v. 8, p. 135-160, 1999.

BURROUGH, P. A. **Principles of geographical information systems for land resources assessment**. Oxford: Clarendon Press, 1987. 193 p.

CASTRIGNANÒ, A. et al. Uncertainty assessment of a soil quality index using geostatistics. **Environmetrics**, v. 20, p. 298-311, 2009.

CHILES, J. P.; DELFINER, P. **Geostatistics: modeling spatial uncertainty**. 2<sup>a</sup> ed. Wiley, 2012. 734 p.

CRESSIE, N. **Statistics for spatial data**. Wiley, 1993. 900 p.

CRESSIE, N. A. C.; HAWKINS, D. M. Robust estimator of the variogram. **Mathematical Geology**, v. 12, n. 2, p. 115-125, 1980.

DACHS, J. N. W. **Estatística computacional**: Uma introdução ao Turbo Pascal. Rio de Janeiro: Livros Técnicos e Científicos, 1988. 236 p.

DAVIS, B. M. Uses and abuses of cross-validation in Geostatistics. **Mathematical Geology**, New York, v. 19, n. 13, p. 1-8, 1987.

DENT, J. B.; BLACKIE, M. J. **Systems Simulation in Agriculture**. Applied Science, London, 1979. 180 p.

DHANOVA, M. S. et al. Use of mean square prediction error analysis and reproducibility measures to study near infrared calibration equation performance. **Journal of Near Infrared Spectroscopy**, v. 7, p. 133-143, 1999.

FERREIRA, D. F. **Estatística Computacional em Java**. 1ª ed. Lavras: Editora UFLA, 2013. 695 p.

GENTON, M. Highly Robust Variogram Estimation. **Mathematical Geology**, v. 30, n. 2, p. 213-221, 1998.

GORSICH, D. J.; GENTON, M. G. Variogram Model Selection via Nonparametric

tric Derivative Estimation. **Mathematical Geology**, v. 32, p. 249-270, 2000.

GUERRA, P. A. G. **Geoestatística operacional**. Brasília: Ministério das Minas e Energia, Departamento Nacional da Produção Mineral, 1988. 145 p.

GUTIÉRREZ, J. S.; MIGUEL, S. M. de. A Variogram Model Comparison for Predicting Forest Changes. **Procedia Environmental Sciences**, v.7, p.383-388, 2011.

HARRISON, S. R. Regression of a model on real-system output: an invalid test of model validity. **Agricultural Systems**, v. 34, p. 183-190, 1990.

HIRAKATA, V. N.; CAMEY, S. A. Análise de concordância entre métodos de Bland-Altman. **HCPA**, v. 29, n. 3, p. 261-268, 2009.

ISAAKS, E. H.; SRIVASTAVA, R. M. **An introduction to applied geostatistics**. New York: Oxford University Press, 1989. 561 p.

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

KOBAYASHI, K.; SALAM, M. U. Comparing Simulated and Measured Values Using Mean Squared Deviation and its Components. **Agron.** v. 92, p. 345-352, 2000.

LASLETT, G. M. et al. Comparison of several spatial prediction methods for soil pH. **Journal of Soil Science**, v. 38, p. 325-341, 1987.

LIN, L. I. K. A concordance correlation coefficient to evaluate reproducibility. **Biometrics**, v. 45, p. 255-268, 1989.

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

MC BRATNEY, A. G.; WEBSTER, A. G. Choosing functions for semi-variograms and fitting them to sampling estimates. **Journal of Soil Science**, v. 37, p. 617-639, 1986.

MITCHELL, P. L. Misuse of regression for empirical validation of models. **Agricultural Systems**, v. 54, p. 313-326, 1997.

MITCHELL, P. L.; SHEENY, J. E. Comparison of predictions and observations to assess model performance: a method of empirical validation. In: Kropff, M. J.; Teng, P. S.; Aggarwal, P. K.; Bouma, J.; Bouman, B. A. M.; Jones, J. W.; Van Laar, H. H. (Eds.) **Applications of Systems Approaches at the Field Level**. Kluwer Academic Publishers, Boston, p. 437-451, 1997.

MOHAMMADI, S. S.; SAFA, M. A simulated annealing based optimization algorithm for automatic variogram model fitting. **Archives of Mining Sciences**, v. 61, n. 3, p. 635-649, 2016.

MONTGOMERY, D. C. **Introdução ao controle estatístico da qualidade**. 4<sup>a</sup>ed. LTC, 2009. 513 p.

NETER, J. et al. **Applied Linear Statistical Model**, 4<sup>a</sup>ed. McGraw-Hill, 1996. 1408 p.

OLEA, R. A. A six-step practical approach to semivariogram modeling. **Stoch Environ Res Risk Assess**, v. 20, p. 307-318, 2006.

OLIVEIRA, M. S. de. **Planos amostrais para variáveis espaciais utilizando Geoestatística**. 1991. 110 p. Dissertação (Mestrado em Estatística) - Universidade Estadual de Campinas, Campinas, 1991.

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

RETTA, A. et al. Application of SORKAM to simulate shattercane growth using forage sorghum. **Agron**. v. 88, p. 596-601, 1996.

RIBEIRO JR., P. J. **Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo**. 1995. 99 p. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1995.

RIBEIRO JR, P. J.; DIGGLE, P. J. geoR: a package for geostatistical analysis. **R NEWS**, v. 1, n. 2, p. 14-18, June 2001. ISSN 1609-3631. Disponível em: <http://CRAN.R-project.org/doc/Rnews/>.

RYAN, T. P.; WOODALL, W. H. The Most-Cited Statistical Papers. **Journal of Applied Statistics**. v. 32, n. 5, p. 461-474, 2005.

SOUZA FILHO, L. A. de. **Análise espacial em modelos de geração de viagens**. 2013. 90 p. Dissertação (Mestrado em Engenharia de Transportes) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2013.

TEDESCHI, L. O. Assessment of the adequacy of mathematical models. **Agricultural systems**, v. 89, p. 225-247, 2006.

TEO, Y. H.; BEYROUTY, C. A.; GBUR, E. E. Evaluating a model for predicting nutrient uptake by rice during vegetative growth. **Agron**. v. 84, p. 1064-1070, 1992.

VENABLES, W. N.; RIPLEY, B. D. **Modern applied statistics with S**. 4ed. New York: Springer, 2002. 498 p.

VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAES, R. S.; ALVAREZ, V. H.; SCHAESER, C. E. G. R. (Ed.) **Tópicos em Ciências do Solo**, Viçosa: Sociedade Brasileira de Ciência do Solo, p. 1-54, 2000.

VIEIRA, S. R. et al. Geostatistical theory and application to variability of some agronomical properties. **Hilgardia**, v. 51, p. 65-75, 1983.

YAMAMOTO, J. K.; LANDIM, P. M. B. **Geoestatística**: conceitos e aplicações. São Paulo: Editora Oficina de textos, 2013. 216 p.