



CHARLES SHALIMAR FELIPPE DA SILVA

**PROPOSTA DE CORREÇÃO DO VIÉS NA ESTIMAÇÃO DA
SEMIVARIÂNCIA DO RESÍDUO NA PRESENÇA DE
TENDÊNCIA**

**LAVRAS – MG
2018**

CHARLES SHALIMAR FELIPPE DA SILVA

**PROPOSTA DE CORREÇÃO DO VIÉS NA ESTIMAÇÃO DA SEMIVARIÂNCIA DO
RESÍDUO NA PRESENÇA DE TENDÊNCIA**

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

Prof. Dr. Marcelo Silva de Oliveira
Orientador

**LAVRAS – MG
2018**

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

Silva, Charles Shalimar Felipe da.

Proposta de correção do viés na estimação da semivariância do
resíduo na presença de tendência / Charles Shalimar Felipe da
Silva. - 2018.

90 p. : il.

Orientador(a): Marcelo Silva de Oliveira.

Tese (doutorado) - Universidade Federal de Lavras, 2018.
Bibliografia.

1. Semivariância dos erros. 2. Viés corrigido. 3. Tendência
espacial. I. Oliveira, Marcelo Silva de. II. Título.

CHARLES SHALIMAR FELIPPE DA SILVA

**PROPOSTA DE CORREÇÃO DO VIÉS NA ESTIMAÇÃO DA SEMIVARIÂNCIA DO
RESÍDUO NA PRESENÇA DE TENDÊNCIA**

**PROPOSAL FOR CORRECTION OF BIAS IN ESTIMATION OF SEMIVARIANCE
OF RESIDUAL IN PRESENCE OF TREND**

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

APROVADA em 06 de fevereiro de 2018.

Dr. João Domingos Scalon UFLA

Dr. Marcelo de Carvalho Alves UFLA

Dr. Eric Batista Ferreira UNIFAL

Dr. Renato Ribeiro de Lima UFLA

Prof. Dr. Marcelo Silva de Oliveira
Orientador

**LAVRAS – MG
2018**

DEDICATÓRIA

À pessoa em quem tenho encontrado, a cada dia, iluminação, inspiração, força e refrigério.

A ele que, sendo Deus, por amor se fez homem para doar sua vida. Um homem perfeitamente humilde, puro e belo, no esplendor da sua santidade. Sua graça é melhor do que a vida.

Ao Senhor Jesus Cristo, o Filho de Deus, Salvador de todos aqueles que nele crêem, referência maior para o meu viver, dedico.

Ao meu amado filho Christian, herança de Deus, o grande Eu Sou, dedico.

AGRADECIMENTOS

A Deus Pai, Filho e Espírito Santo.

Aos meus pais, Joaquim (*in memoriam*) e Lila, pelo incentivo e apoio nos estudos, desde criança.

À minha irmã Fabiana e meu cunhado Wagner, pelo carinho e apoio.

Ao amigo Paulo César Moraes, pela solidariedade e pelos momentos de reflexão conjunta.

Ao governo da Republica Federativa do Brasil, pelo investimento na universidade pública.

Ao Brigadeiro Izmailov, pela liberação concedida.

À Ten-Cel Denise, pelo apoio.

Ao colega Marcelo Coelho, pelo suporte nas aulas.

Aos colegas e amigos da EPCAR, pela cooperação e pelos incentivos.

À professora Maria do Carmo, pela receptividade e pelo exemplo de conduta pedagógica.

Ao professor Daniel, pelas primeiras aulas de Estatística Básica na UFLA e pela generosidade no partilhar do saber.

Ao professor Paulo “Bola”, pelas primeiras aulas de Estatística Experimental.

À colega Marcília Bruna, pela amizade e prestatividade, e pelo compartilhar do conhecimento.

Às secretarias, Josie “P” e Josie “G”, Nádia, Maria, Magali e demais funcionários do DES e DEX, pelo auxílio e pela disponibilidade.

Aos professores do DEX e DES, pela oportunidade do aprendizado.

Aos amigos e irmãos da Comunidade Evangélica Universitária (CEU), Peter, Marcelo, Ismael, Rogério, Paulinho, Tatiane, Leila, Rafael Agostinho, Sérgio e Vítor, pelos momentos preciosos de reflexão bíblica, oração e troca de experiências.

Aos colegas Guido, Gilberto e Cristina Nogueira, pela conversa proveitosa.

Às colegas Kelly e Allana, pela amizade, pelos momentos de descontração e pelo trabalho conjunto.

Ao colega Ricardo Vítor, pela confiança e pelo trabalho conjunto.

Aos funcionários do RU, pelas refeições diárias, inclusive nos finais de semana.

Aos colegas e amigos do Grupo de Pesquisa em Estatística Espacial (GPS), do DEX, DES e demais departamentos da UFLA, pela amizade, pelos conselhos, pelos momentos de descontração e pelo compartilhar do conhecimento.

Ao professor Marcelo Oliveira, meu orientador, pela oportunidade do trabalho conjunto, pela generosidade nos ensinamentos e conselhos, pelo exemplo de redobrada disposição em função dos compromissos assumidos e pela palavra “boa ao seu tempo”, nos inúmeros momentos de ponderações concernentes a esta tese.

Aos professores da banca de defesa, pela participação e pelos conselhos.

A todos que, de alguma forma, contribuíram para a concretização deste trabalho.

RESUMO

O objetivo da análise geoestatística é a predição de dados em localizações não amostradas na região de um fenômeno com dependência espacial. A Geostatística possibilita a implementação de um mapa de predição ou krigagem que caracterize a variabilidade espacial desse fenômeno. A qualidade do mapa, assim como de todos os produtos da análise geoestatística, depende da qualidade da estimação das semivariâncias. Para situações em que a média da variável regionalizada não pode ser considerada constante, pode-se recorrer à predição com a krigagem universal ou com regressão. No entanto, a semivariância dos resíduos ou erros preditos, utilizada por ambos os tipos de krigagem, apresenta viés, subestimando os valores de semivariância dos erros, com prejuízo significativo na modelagem da variação espacial. Por isso, esforços têm sido feitos para corrigir o viés na semivariância dos resíduos. O presente trabalho foi realizado, então, com o objetivo de desenvolver uma metodologia para corrigir tal viés, a qual foi denominada critério IRWGLS (do inglês *Iteratively Re-weighted Generalized Least Squares* ou Mínimos Quadrados Generalizados Iterativamente Reponderados) com correção de viés, e implementá-la no programa R. Os modelos de semivariância ajustados de acordo com o semivariograma foram usados para compor a matriz de covariância dos erros. Na estimação dos parâmetros dos modelos foram usados os critérios dos mínimos quadrados ordinários e generalizados, e uma correção de viés, segundo a metodologia proposta. Para avaliar o desempenho da metodologia proposta foi feita uma análise com três conjuntos de dados, que foram submetidos a teste de validação. Com os resultados obtidos, observou-se melhoria efetiva na qualidade da estimação dos parâmetros do modelo de semivariância. Além disso, o uso da metodologia proposta potencializou a correção de viés no modelo de semivariância considerado como o mais adequado para descrever a variabilidade espacial do fenômeno, fato percebido pelo maior aumento percentual nos valores das médias das semivariâncias estimadas, que chegou a 9,96% numa das amostras. A metodologia proposta tem a vantagem de ser de validade geral e, aliada à experiência de especialistas com conhecimento do aspecto físico do fenômeno e de estatísticos, se constitui num avanço para que o objetivo da análise geoestatística seja alcançado com mais acurácia e precisão.

Palavras-chave: Semivariância dos erros. Viés corrigido. Tendência espacial. Análise geoestatística.

ABSTRACT

The aim of geostatistical analysis is the prediction of data in locations that were not sampled in the region of a phenomenon with spatial dependence. Geostatistics allows implementation of a prediction or kriging map that characterizes the spatial variability of this phenomenon. The quality of this map, as well as all products of geostatistical analysis depends on quality of semivariances estimation. For situations that mean of regionalized variable is not constant, one can resort to prediction with universal or regression kriging. However, the semivariance of residuals or predicted errors, used by both types of kriging presents bias, underestimating values of semivariance of errors, with significant loss in modeling of spatial variation. Therefore, efforts have been made to correct the bias in semivariance of residuals. Thus, the present work was done with objective of developing a methodology to correct such bias, termed criterion IRWGLS (*Iteratively Re-weighted Generalized Least Squares*) with bias correction, and implement it in software R. The semivariance models adjusted according to semivariogram were used to compose error covariance matrix. To estimation of parameters of these models, the criteria of ordinary and generalized least squares were used and a bias correction according to the proposed methodology. To evaluate the performance of the proposed methodology, an analysis was performed with three data sets that were submitted to a validation test. The results obtained showed an effective improvement in quality of estimation of parameters in the semivariance model. In addition, use of the proposed methodology potentiated bias correction in semivariance model considered as most adequate to describe the spatial variability of phenomenon, fact perceived by highest percentage increase in mean values of estimated semivariances, which reached 9.96 % in one sample set. The proposed methodology has advantage of being of general validity and, together with experience of specialists with knowledge of physical aspect of the phenomenon and statisticians, it constitutes an advance so that the objective of geostatistical analysis is reached with more accuracy and precision.

Keywords: Semivariance of errors. Corrected bias. Spatial trend. Geostatistical analysis.

LISTA DE FIGURAS

Figura 1	– Componentes da variação espacial.....	11
Figura 2	– Semivariograma experimental típico (sequência de pontos no gráfico) e curva crescente ajustada à sequência de pontos.....	16
Figura 3	– Amostragem regular em malha quadrada e amostragem irregular	18
Figura 4	– Representação gráfica dos modelos de semivariâncias exponencial, esférico e gaussiano	19
Figura 5	– Representação gráfica do modelo de semivariância potência com diferentes valores para o expoente B	20
Figura 6	– Semivariância (viesada) dos resíduos $\gamma_{Res}(h)$ e semivariância real $\gamma_Z(h)$	34
Figura 7	– Pontos formados pelos desvios padrões em relação às médias dos postos, com as distâncias desses pontos em relação ao ponto (0,0), conforme o método de validação apresentado por Bastos (2017).....	42
Figura 8	– Gráficos de tendência para o conjunto de dados Parana	47
Figura 9	– Gráfico de semivariâncias para o modelo esférico ajustado com MQNP e MQLO (com linha tracejada), e com MQNP e IRWGLS com correção de viés (com linha contínua), usando o conjunto de dados Parana	50
Figura 10	– Gráfico de semivariâncias para o modelo exponencial ajustado com MQNP e MQLO (com linha tracejada) e com MQNP e IRWGLS com correção de viés (com linha contínua), usando o conjunto de dados Meuse	54
Figura 11	– Gráfico de semivariâncias para o modelo esférico ajustado com MQNP e MQLO (com linha tracejada) e com MQNP e IRWGLS com correção de viés (com linha contínua), usando o conjunto de dados Castanha.....	58
Figura 12	– Mapas de krigagem do conjunto de dados Parana, com modelo de semivariância ajustado (a) com MQLO e (b) com IRWGLS com correção de viés.....	61
Figura 13	– Mapas de erro de krigagem do conjunto de dados Parana, com modelo de semivariância ajustado (a) com MQLO e (b) com IRWGLS com correção de viés	62
Figura 14	– Mapas de krigagem do conjunto de dados Meuse, com modelo de semivariância ajustado (a) com MQLO e (b) com IRWGLS com correção de viés.....	63
Figura 15	– Mapas de erro de krigagem do conjunto de dados Meuse, com modelo de semivariância ajustado (a) com MQLO e (b) com IRWGLS com correção de viés	64

Figura 16 – Mapas de krigagem do conjunto de dados Castanha, com modelo de semivariância ajustado (a) com MQLO e (b) com IRWGLS com correção de viés.	64
Figura 17 – Mapas de erro de krigagem do conjunto de dados Castanha, com modelo de semivariância ajustado (a) com MQLO e (b) com IRWGLS com correção de viés	65

LISTA DE TABELAS

Tabela 1 – Estimativas classificadas de $r = 3$ métodos selecionados, conforme o método de validação apresentado por Bastos (2017), com as médias e os desvios padrão dos postos calculados.....	41
Tabela 2 – Estimativas dos parâmetros de pepita, contribuição e alcance, com MQLO, IRWGLS e IRWGLS com correção de viés, para os modelos exponencial, gaussiano e esférico, usando o conjunto de dados Parana	49
Tabela 3 – Médias das semivariâncias estimadas com MQLO, IRWGLS e IRWGLS com correção de viés, para os modelos ajustados com MQNP, e aumento percentual envolvido, para o conjunto de dados Parana	51
Tabela 4 – Resultados [em unidades de distância estabelecida pelo método de Bastos (2017)] do teste de validação para os modelos ajustados com MQLO, IRWGLS e IRWGLS com correção de viés (utilizando MQNP e MQNO), usando o conjunto de dados Parana.....	52
Tabela 5 – Estimativas dos parâmetros de pepita, contribuição e alcance, com MQLO, IRWGLS e IRWGLS com correção de viés, para os modelos exponencial, gaussiano e esférico, usando o conjunto de dados Meuse.....	53
Tabela 6 – Médias das semivariâncias estimadas com MQLO, IRWGLS e IRWGLS com correção de viés, para os modelos ajustados com MQNP, e aumento percentual envolvido, para o conjunto de dados Meuse	55
Tabela 7 – Resultados [em unidades de distância estabelecida pelo método de Bastos (2017)] do teste de validação para os modelos ajustados com MQLO, IRWGLS e IRWGLS com correção de viés, usando o conjunto de dados Meuse.....	56
Tabela 8 – Estimativas dos parâmetros de pepita, contribuição e alcance, com MQLO, IRWGLS e IRWGLS com correção de viés, para os modelos gaussiano e esférico, usando o conjunto de dados Castanha	57
Tabela 9 – Médias das semivariâncias estimadas com MQLO, IRWGLS e IRWGLS com correção de viés, para os modelos ajustados com MQNP, e aumento percentual envolvido, para o conjunto de dados Castanha	59
Tabela 10 – Resultados [em unidades de distância estabelecida pelo método de Bastos (2017)] do teste de validação para os modelos ajustados com MQLO, IRWGLS e IRWGLS com correção de viés, usando o conjunto de dados Castanha.....	60

LISTA DE SIGLAS

BLUE	<i>Best Linear Unbiased Estimator</i>
BLUP	<i>Best Linear Unbiased Predictor</i>
IRWGLS	<i>Iteratively Re-weighted Generalized Least Squares</i>
MQLO	Mínimos Quadrados Lineares Ordinários
MLG	Mínimos Quadrados Lineares Generalizados
MQNG	Mínimos Quadrados Não-lineares Generalizados
MQNO	Mínimos Quadrados Não-lineares Ordinários
MQNP	Mínimos Quadrados Não-lineares Ponderados

SUMÁRIO

1	INTRODUÇÃO	10
2	REFERENCIAL TEÓRICO	13
2.1	Teoria das variáveis regionalizadas	13
2.2	Estacionariedade	14
2.3	Semivariograma experimental	15
2.4	Modelo de semivariância	18
2.5	Krigagem	21
2.5.1	Krigagem simples	23
2.5.2	Krigagem universal	25
2.5.3	Krigagem ordinária	28
2.5.4	Krigagem com regressão	29
2.6	Estimação da semivariância na presença de tendência	30
3	MATERIAL E MÉTODOS	39
3.1	Dados	39
3.2	Validação dos resultados	40
4	RESULTADOS E DISCUSSÃO	43
4.1	Método proposto para a correção de viés na semivariância dos resíduos	43
4.2	Aplicações do método proposto para correção de viés na semivariância dos resíduos	46
4.2.1	Aplicação do método proposto usando o conjunto de dados Parana	48
4.2.2	Aplicação do método proposto usando o conjunto de dados Meuse	52
4.2.3	Aplicação do método proposto usando o conjunto de dados Castanha	56
4.2.4	Mapas de krigagem e de erro de krigagem dos conjuntos de dados	60
4.2.5	Ponderações sobre as aplicações do método proposto	65
4.2.6	Considerações finais	67
5	CONCLUSÃO	69
	REFERÊNCIAS	71
	APÊNDICE	74

1 INTRODUÇÃO

Existem áreas da Ciência e Tecnologia que demandam o conhecimento de valores distribuídos em espaços estruturados, tais como a superfície terrestre, entre outros. Alguns exemplos incluem a investigação de jazidas mineiras, estudos para prospecção petrolífera, exame de distribuição de chuvas e concentração de poluentes no ar. Em comum, esses problemas apresentam fenômenos ou populações em que deve ser levada em consideração uma continuidade ou dependência espacial entre os valores do atributo¹ de interesse e a necessidade de predições feitas com base nessa dependência e nas posições dos dados (pontos) amostrais.

A Geoestatística é o ramo da Estatística Espacial voltado ao estudo e resolução dos problemas citados, os quais envolvem variáveis aleatórias atreladas a uma localização espacial e com dependência espacial entre si. Matheron (1963), com o fim de representar e modelar essas variáveis, desenvolveu o conceito de variável regionalizada, no qual ele fundamentou a teoria geoestatística.

Burrough (1987) disse que a variação espacial de uma variável regionalizada pode ser expressa pela soma de três componentes. De acordo com o autor, o valor da variável regionalizada Z , na posição espacial u , é dado por:

$$z(u) = \mu(u) + \varepsilon'(u) + \varepsilon''(u), \quad (1)$$

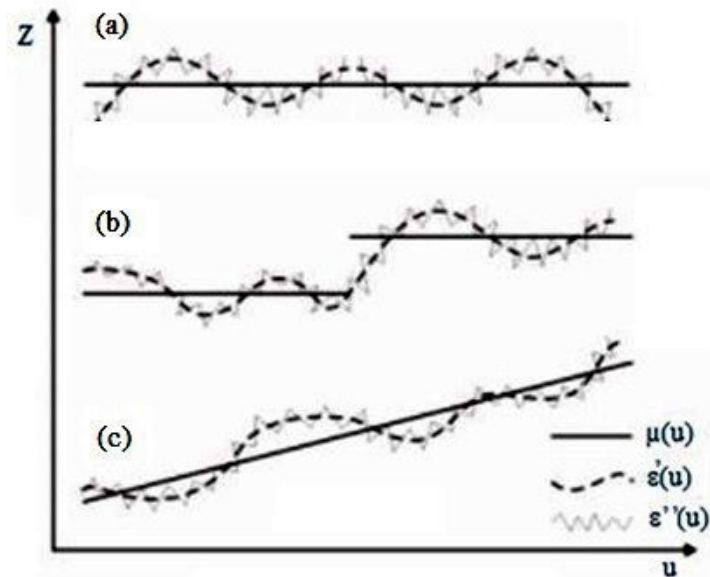
em que:

- $\mu(u)$ é um componente determinístico, uma tendência associada a um valor médio;
- $\varepsilon'(u)$ é uma variável aleatória com dependência espacial e média zero, e variância φ^2 ;
- $\varepsilon''(u)$ é um ruído, isto é, uma variável aleatória com média zero, não correlacionada e variância τ^2 ;
- a posição u é um vetor de coordenadas.

Na Figura 1.1a, podem ser vistos os três componentes da variação espacial, com destaque para a média constante. Na Figura 1.1b, são exibidos os mesmos três componentes, mas com uma média em dois níveis distintos. Na Figura 1.1c, novamente, são exibidos os mesmos três componentes, mas com uma média crescente.

¹ Atributo é uma propriedade física medida num indivíduo (amostra física) de uma população (GOOVAERTS, 1997).

Figura 1 – Componentes da variação espacial, (a) com média $\mu(u)$ constante, (b) com média $\mu(u)$ em dois níveis distintos e (c) com média $\mu(u)$ crescente.



Fonte: Adaptado de Burrough (1987).

O objetivo da análise geoestatística é a predição de dados, em localizações não amostradas, na região em estudo de um fenômeno com dependência espacial. Com isso, ela possibilita a implementação de um mapa que caracterize a variabilidade espacial inerente a esse fenômeno. A qualidade desse mapa de predição ou de krigagem, assim como de todos os produtos da análise geoestatística, tais como simulações, entre outros, depende da qualidade da estimação das semivariâncias.

Para situações em que a média da variável regionalizada [componente $\mu(u)$ da Equação (1)] não pode ser considerada constante, pode-se recorrer à predição com a krigagem universal ou com a krigagem com regressão (esta última usa variáveis auxiliares para realizar a predição). No entanto, a semivariância dos erros preditos ou resíduos, a qual é utilizada por ambos os tipos de krigagem apresenta viés, subestimando os valores de semivariâncias dos erros (MATHERON, 1971). Esse viés pode causar prejuízo significativo na modelagem da variação espacial da variável regionalizada e, por isso, esforços têm sido feitos para corrigi-lo (NEUMAN E JACOBSON, 1984; CRESSIE, 1985; BECKERS E BOGAERT, 1998; KIM E BOOS, 2004; SCHABENBERGER E GOTWAY, 2005; BLINYUK, 2012), porém com resultados não definitivos.

Assim sendo, o problema de pesquisa abordado nesta tese é: como corrigir o viés existente na estimação da semivariância na presença de tendência espacial? Para atender esse problema de pesquisa, o presente trabalho foi realizado com o objetivo de desenvolver uma

metodologia para corrigir o viés na semivariância dos resíduos e implementar um algoritmo na linguagem de programação do ambiente do software R com esta metodologia, a qual foi denominada critério IRWGLS (do inglês *Iteratively Re-weighted Generalized Least Squares* ou Mínimos Quadrados Generalizados Iterativamente Reponderados) com correção de viés, para que a predição de dados seja realizada com mais acurácia e precisão, na presença de média não constante da variável regionalizada.

O gráfico de semivariância ou semivariograma é usado neste trabalho como a principal ferramenta para a detecção e descrição da estrutura de variabilidade espacial. Os modelos de semivariância, ajustados de acordo com o gráfico, foram usados para compor a matriz de covariância dos erros com dependência espacial. Na estimação dos parâmetros dos modelos foram usados os critérios dos mínimos quadrados ordinários e generalizados. Os modelos ajustados com o uso do critério de mínimos quadrados generalizados receberam uma correção de viés na semivariância dos resíduos, conforme a metodologia proposta nesta tese.

Para avaliar o desempenho da metodologia proposta, foram utilizados três conjuntos de dados. Os modelos de semivariância ajustados aos conjuntos de dados foram, então, submetidos ao método de validação proposto por Bastos (2017), de modo que o modelo que resultou no menor valor para a variável distância no método de validação foi considerado o mais adequado para descrever a estrutura da variabilidade espacial.

2 REFERENCIAL TEÓRICO

A Geoestatística começou a se desenvolver a partir do trabalho de Krige (1951), considerando jazidas de ouro na África do Sul. Em tal trabalho o autor argumenta sobre a importância da localização espacial das amostras na interpretação da variabilidade do fenômeno. Tendo isso em mente, Matheron (1963), com o objetivo de representar e modelar variáveis atreladas a localizações espaciais, desenvolveu o conceito de variável regionalizada, no qual ele fundamentou a teoria geoestatística. Este conceito engloba uma função numérica com distribuição espacial que varia de um ponto a outro com continuidade aparente e cuja variação, segundo Matheron, não pode ser representada por uma função matemática simples.

2.1 Teoria das variáveis regionalizadas

Como um primeiro passo, com vistas à compreensão da incerteza espacial que reveste os dados espaciais, pode-se entendê-los como procedentes de uma realização de um processo estocástico, que é um conjunto de variáveis aleatórias reais $\{Z(u): u \in D \subset \mathcal{R}^d\}$, definidas num mesmo espaço de probabilidade, em que $Z(u)$ representa a variável Z na localização $u=[u_1, u_2, \dots, u_d]$, sendo u_c , com $c = 1, \dots, d$, as coordenadas espaciais com dimensão d e D é o domínio ou região do fenômeno em estudo, subconjunto do espaço d -dimensional \mathcal{R}^d . Para este trabalho, sem perda de generalidade, será considerado o espaço bidimensional ($d = 2$), de modo que o vetor de localização é definido por $u=[u_1, u_2]$. É usado, no entanto, $u=[x, y]$, ou seja, as coordenadas x e y correspondendo a u_1 e u_2 , respectivamente, a fim de facilitar a identificação costumeira dos sentidos horizontal e vertical dos gráficos com os eixos x e y . O conceito de processo estocástico, também conhecido como campo aleatório quando a dimensão $d > 1$ (SCHABENBERGER E GOTWAY, 2005), se encontra incluso na teoria das variáveis regionalizadas que consiste em que:

- a) cada ponto no espaço não apresenta um único valor, mas uma distribuição de probabilidade de ocorrência de valores;
- b) num ponto u , o atributo de interesse $Z(u)$ é uma variável aleatória com média μ , variância σ^2 e uma distribuição referentes a esse ponto;
- c) no espaço existem infinitos pontos u , em cada um dos quais $Z(u)$ tem sua própria distribuição;
- d) os valores observados de $Z(u)$ constituem suas realizações $z(u)$;
- e) o conjunto das variáveis aleatórias $\{Z(u): u \in D \subset \mathcal{R}^d\}$ constitui a função aleatória, processo aleatório, processo estocástico ou campo aleatório;

- f) o conjunto dos valores reais de $z(u)$, os quais compreendem a realização da função aleatória é o que se conhece como variável regionalizada (WEBSTER E OLIVER, 2007).

2.2 Estacionariedade

Schabenberger e Gotway (2005) lembram que, pela teoria das variáveis regionalizadas, o conjunto coletado de n observações não representa uma variável com n realizações, mas n variáveis, cada qual, com uma única realização. Assim, para que seja possível avaliar a variabilidade de um processo estocástico e realizar inferência estatística, a fim de construir o mapa de predição, é imprescindível a restrição de que o campo aleatório seja estacionário. A estacionariedade permite a repetição no espaço, usando pares de dados amostrais (CHILÈS E DELFINER, 2012). Para o estudo geoestatístico é suficiente o uso das hipóteses de estacionariedade de segunda ordem ou intrínseca.

Estacionariedade de segunda ordem

Um campo aleatório é estacionário de segunda ordem (JOURNEL, 1989) se:

- i) $E[Z(u)] = \mu, \forall u \in D$, sendo μ uma constante;
- ii) $Cov[Z(u), Z(u+h)] = C(h), \forall u \in D$.

Assim, a média é constante e a covariância existe e depende apenas do vetor distância h entre os pontos. Esta dependência do vetor distância h pode ser apenas de seu tamanho (isto é, da distância h): neste caso, o fenômeno é dito ser isotrópico. Caso contrário, se depender também da direção do vetor h , é dito ser anisotrópico. O desenvolvimento desta tese não se preocupará com isotropia ou anisotropia, pois a metodologia vale para ambas as situações. Portanto, sem perda de generalidade, o texto seguirá tratando do vetor h como distância h . Quando for necessária a menção ao vetor h , esta será feita explicitamente. Observe também que

$$Cov[Z(u), Z(u+0)] = Var[Z(u)] = C(0),$$

isto é, a variância no campo aleatório é a mesma em todos os lugares.

Estacionariedade intrínseca

A hipótese de estacionariedade de segunda ordem implica na existência de variância finita dos dados observados, ou seja, $Var[Z(u)] = C(0)$. Alguns fenômenos, no entanto, possuem uma capacidade de dispersão infinita. Para tais situações, uma hipótese menos restritiva pode ser aplicada. Trata-se da hipótese de estacionariedade intrínseca.

Um campo aleatório é intrinsecamente estacionário (CRESSIE, 1993) se:

- i) $E[Z(u)] = \mu, \forall u \in D$;

$$\text{ii) } \frac{1}{2} \text{Var}[Z(u) - Z(u+h)] = \frac{1}{2} \text{E}[Z(u) - Z(u+h)]^2 = \gamma(h), \forall u \in D;$$

em que $\gamma(h)$ é uma função real não-negativa, chamada semivariância do processo espacial. Essa função mede a variabilidade da diferença entre as realizações da variável aleatória, separadas por uma distância h , se constituindo numa medida que descreve estatisticamente como os valores em dois pontos se tornam diferentes, conforme o aumento da distância entre esses pontos (CHILÈS E DELFINER, 2012).

2.3 Semivariograma experimental

A hipótese de estacionariedade de segunda ordem permite que a dependência espacial de uma variável regionalizada possa ser estimada, de forma equivalente, pela função covariância $C(h)$ e pela semivariância $\gamma(h)$, dada a relação existente entre ambas (VIEIRA, 2000):

$$\gamma(h) = C(0) - C(h). \quad (2)$$

A hipótese intrínseca, no entanto, é mais geral, possibilitando que um maior número de fenômenos possa ser abrangido por $\gamma(h)$ do que por $C(h)$. Outra vantagem da semivariância sobre a covariância é que não requer o conhecimento da média. Assim, neste trabalho será enfatizado o uso de $\gamma(h)$ para descrever a estrutura de dependência espacial.

A função semivariância é estimada pela semivariância experimental ou amostral. Um estimador de semivariância bastante conhecido é o estimador clássico de Matheron, definido como:

$$\hat{\gamma}(h_{ij}) = \frac{1}{2N(h_{ij})} \sum_{\substack{\forall (u_i, u_j) \\ \text{tais que} \\ |u_i - u_j| = h_{ij}}} [Z(u_i) - Z(u_j)]^2,$$

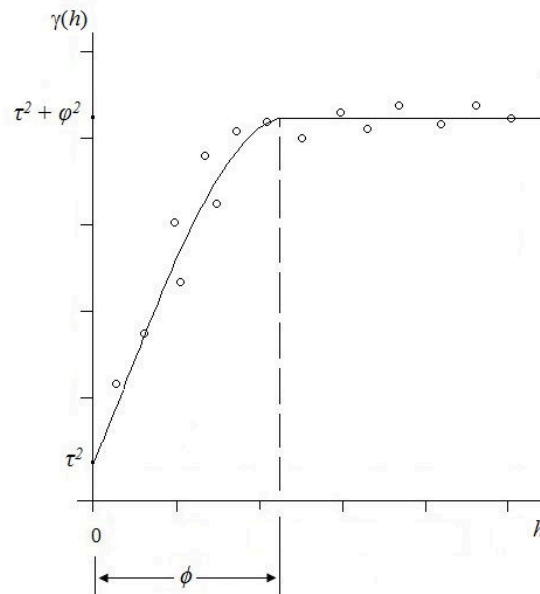
em que $N(h_{ij})$ é o número de pares de valores a serem medidos $Z(u_i)$ e $Z(u_j)$, separados pelo vetor h_{ij} .

O semivariograma é o gráfico no qual são plotados, no eixo das ordenadas, os valores da semivariância em função das distâncias, que são representadas no eixo das abscissas.

Um semivariograma experimental com comportamento próximo do ideal é apresentado na Figura 2. As semivariâncias estimadas para cada distância h_{ij} , representadas pelos pontos no gráfico, mostram que as diferenças $[z(u_i) - z(u_j)]$ aumentam conforme aumenta a distância h_{ij} que as separa. Isto está de acordo com o fato de que valores próximos no espaço tendem a assumir valores mais parecidos entre si, ou seja, são mais correlacionados, do que aqueles separados a maiores distâncias (TOBLER, 1970). Assim, a

série de pontos pode ser ajustada por um modelo de curva crescente. Podem ser vistos na Figura 2 os componentes de um modelo de semivariância: pepita (τ^2), patamar ($\tau^2 + \phi^2$) e alcance (ϕ).

Figura 2 – Semivariograma experimental típico (sequência de pontos no gráfico) e curva crescente ajustada à sequência de pontos.



Fonte: Do autor (2016).

Patamar, alcance, contribuição e pepita

A ilustração apresentada na Figura 2 mostra que, à medida que h_{ij} aumenta, $\gamma(h_{ij})$ também aumenta até um valor máximo no qual se estabiliza. Este valor, ao redor do qual $\gamma(h_{ij})$ passa a flutuar, se chama patamar, sendo representado por $\tau^2 + \phi^2$, em que τ^2 é o efeito pepita e ϕ^2 é conhecido como contribuição. O patamar é aproximadamente igual à variância amostral dos dados (VIEIRA, 2000).

A distância na qual $\gamma(h_{ij})$ atinge o patamar é chamada alcance (representado por ϕ). A partir do alcance os dados passam a ser não-correlacionados. Isto é, em distâncias maiores do que o alcance não há dependência espacial entre as observações, restando somente variação aleatória.

A contribuição (ϕ^2) é a diferença entre o patamar e o efeito pepita. Esse valor está associado à variância estrutural, ou seja, à variabilidade regionalizada, que pode ser explicada por um modelo com uma certa estrutura de dependência espacial.

Por definição $\gamma(0)=0$, porém, na prática, quando h_{ij} se aproxima de zero, o semivariograma pode tender para um valor positivo. Tal descontinuidade na origem do semivariograma é chamada efeito pepita (τ^2). Para entendê-lo, pode-se assumir um outro modelo para a variação espacial de uma variável regionalizada, diferente daquele visto no

início deste trabalho [EQUAÇÃO (1)]. Tal modelo é definido por Schabenberger e Gotway (2005) como:

$$z(u) = \mu(u) + \omega(u) + \eta(u) + \varepsilon(u), \quad (3)$$

em que $\mu(u)$ é a função tendência em larga escala do campo aleatório, equivalente a $\mu(u)$ da Equação (1); $\omega(u)$ é a variação de pequena escala, equivalente a $\varepsilon'(u)$ da Equação (1); $\eta(u)$ é a variação de microescala, e $\varepsilon(u)$ é um ruído aleatório não correlacionado, com variância σ_ε^2 .

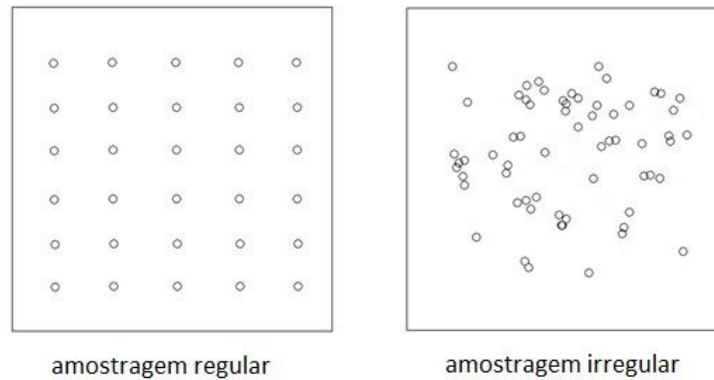
Segundo Goovaerts (1997), o efeito pepita pode ser causado por erro de medição, referente ao componente $\varepsilon(u)$ do modelo (este também inclui a variação aleatória do processo espacial que não pode ser removida). Também pode ser ocasionado por fontes de variabilidade espacial numa escala menor do que a amostrada, as quais estão relacionadas com o componente $\eta(u)$. O erro de medição pode ser diminuído mediante uso de melhor dispositivo tecnológico e treinamento mais apurado da pessoa que mede, enquanto que a detecção da variabilidade espacial numa escala menor é conseguida por meio de amostragem mais fina.

Se o semivariograma experimental não for crescente e dependente de h_{ij} , mas for constante e flutuar ao redor do patamar para qualquer valor de h_{ij} , está caracterizado o efeito pepita puro. Ele implica na total ausência de dependência espacial entre os dados na escala considerada, visto que o componente alcance é menor do que o menor espaçamento entre as observações. Nesse caso, em que somente variação aleatória está presente na distribuição dos dados, não se aplica a geoestatística, com vistas à predição de valores.

Classes de distância (*lags*)

A qualidade do ajuste do modelo de semivariância depende da adequada detecção da variabilidade espacial, a qual, por sua vez, depende da malha de amostragem adotada. Esta envolve tanto a escala de medição, como o posicionamento dos dados amostrais no espaço. Uma distribuição espacial dos dados amostrais, como na malha quadrada da amostragem regular ou sistemática (FIGURA 3), é preferível à da amostragem irregular, porque facilita o cálculo da semivariância experimental. Esta leva em conta pares separados por um vetor h_{ij} e, em outras palavras, não somente a distância, mas também a direção deve ser considerada no cálculo dos pares. A amostragem sistemática permite uma cobertura amostral mais homogênea do campo em estudo, contribuindo para a otimização da precisão da predição (OLIVEIRA, 1991). Em muitas situações, no entanto, por motivos de ordem financeira, falta de precisão de instrumentos de posicionamento ou dificuldades de acesso geográfico, o estudo é feito com dados distribuídos de forma irregular.

Figura 3 – Amostragem regular em malha quadrada e amostragem irregular.



Fonte: Do autor (2016).

Para regularizar a malha de dados, o cálculo da semivariância experimental é feito considerando classes de distância ou *lags* que se constituem dos espaçamentos entre as observações, com acréscimo de uma tolerância para a distância (tolerância de passo) e uma tolerância para a direção (tolerância angular) (JOURNEL, 1987).

Journel e Huijbregts (1978) recomendam que cada *lag* deve abranger, no mínimo, trinta pares de dados para uma melhor estruturação do semivariograma. Além disso, recomendam que a distância limite entre dados para o cálculo da semivariância seja de cerca da metade da maior distância entre os dados, na direção considerada no momento.

2.4 Modelo de semivariância

O semivariograma experimental apresenta uma série de pontos discretos de $\gamma(h_{ij})$ correspondentes a cada valor de h_{ij} . Modelos de funções contínuas devem ser ajustados aos pontos discretos (FIGURA 2), de maneira a se obter valores estimados da semivariância para quaisquer valores de h_{ij} requeridos na etapa de predição.

A fim de garantir que a variância da combinação linear das variáveis aleatórias do campo aleatório intrinsecamente estacionário seja não-negativa, o modelo de semivariância deve ser condicionalmente negativo definido (CRESSIE, 1993; GOOVAERTS, 1997). Para que tal condição seja satisfeita, é preciso que

$$\sum_{i=1}^k \sum_{j=1}^k c_i c_j \gamma(h_{ij}) \leq 0, \quad h_{ij} = |u_i - u_j|,$$

para quaisquer posições espaciais $u_1, \dots, u_k \in D$, e quaisquer $c_1, \dots, c_k \in \mathcal{R}$, de tal modo que

$\sum_{i=1}^k c_i = 0, \quad \forall k \in \mathbb{Z}^+$. Além disso, pela definição da semivariância, o modelo deve ter as

seguintes propriedades:

- i) $\gamma(0) = 0$;

ii) $\gamma(h_{ij}) = \gamma(h_{ji}), \forall u_i \text{ e } u_j \in D$.

Diante de tais restrições, surge uma classe de modelos que são definidos como modelos autorizados. Alguns desses modelos, comumente usados na literatura (GOOVAERTS, 1997), são os modelos esférico, gaussiano e exponencial, cujas semivariâncias são dadas por:

Modelo esférico:

$$\gamma(h) = \begin{cases} \tau^2 + \varphi^2 \left[\frac{3h}{2\phi} - \frac{1}{2} \left(\frac{h}{\phi} \right)^3 \right], & 0 < h < \phi; \\ \tau^2 + \varphi^2, & h \geq \phi \end{cases}$$

Modelo gaussiano:

$$\gamma(h) = \tau^2 + \varphi^2 \left[1 - e^{-3 \left(\frac{h}{\phi} \right)} \right], \quad 0 < h < d,$$

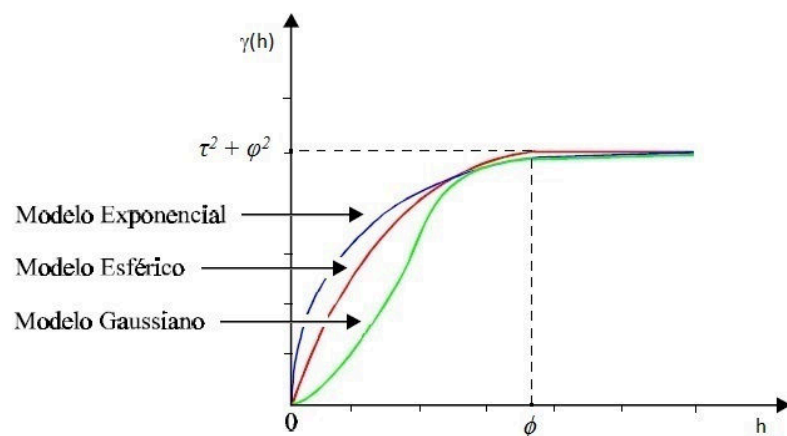
em que d é a distância máxima na qual o modelo é definido;

Modelo exponencial:

$$\gamma(h) = \tau^2 + \varphi^2 \left[1 - e^{-3 \left(\frac{h}{\phi} \right)^2} \right], \quad 0 < h < d.$$

Na Figura 4 são apresentados exemplos desses modelos.

Figura 4 – Representação gráfica dos modelos de semivariâncias exponencial, esférico e gaussiano.



Fonte: Do autor (2016).

Os modelos exponencial e gaussiano se caracterizam por alcançar o patamar assintoticamente. Para esses modelos, o alcance (ϕ) é determinado como sendo o valor da

distância na qual se alcança 95% da contribuição (ϕ^2). Esse alcance também é conhecido como alcance prático.

Na Figura 4 pode ser visto que, próximo à origem, o modelo gaussiano exibe um comportamento parabólico, enquanto que os modelos exponencial e esférico mostram um comportamento linear. O modelo gaussiano, então, é mais apropriado para descrever fenômenos altamente regulares, por exemplo, elevações topográficas em relevos suavemente ondulados (GOOVAERTS, 1997). Deve ser notado, ainda, que para o mesmo alcance prático, o modelo exponencial cresce mais rapidamente do que o esférico.

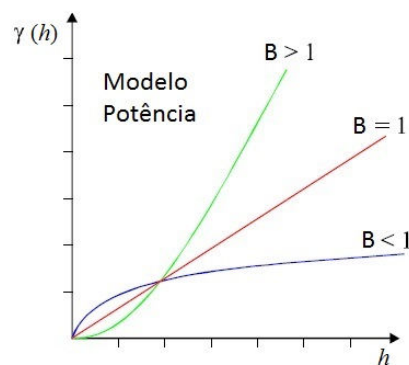
Existem fenômenos que apresentam uma capacidade de dispersão infinita, atendendo somente à hipótese intrínseca. Nesses fenômenos o crescimento de $\gamma(h)$ é contínuo com o incremento de h . Para tal situação pode-se recorrer ao ajuste com o modelo potência, que se caracteriza por não ter patamar definido. A função de semivariância desse modelo é dada por:

$$\gamma(h) = \begin{cases} 0, & \text{se } h = 0; \\ \tau^2 + Ah^B, & \text{se } h \neq 0, \end{cases}$$

em que $\tau^2 > 0$, $A > 0$ e $0 < B < 2$, sendo que τ^2 se refere ao efeito pepita, A e B representam, respectivamente, parâmetros de declividade e intensidade de variação.

Na Figura 5 podem ser vistos exemplos de modelos potência para diferentes valores do expoente B .

Figura 5 – Representação gráfica do modelo de semivariância potência com diferentes valores para o expoente B .



Fonte: Do autor (2016).

Existem fenômenos que exibem semivariogramas experimentais com estruturas aninhadas ou entrelaçadas, com mais de um patamar e mais de um alcance. Tais semivariogramas aninhados ocorrem por causa da mistura de diferentes escalas de variação nas observações, podendo ser ajustados pela combinação dos modelos teóricos vistos. Por exemplo, um modelo gaussiano até determinado alcance, seguido por um esférico. Esse ajuste

é possível porque, como salientam Issaks e Srivastava (1989), qualquer combinação linear dos modelos autorizados é também um modelo autorizado.

Entre os critérios que são frequentemente usados para o ajuste do modelo de semivariância estão os de mínimos quadrados não-lineares ordinários, ponderados e generalizados, que têm como base o estimador de semivariância (CRESSIE, 1993; CHILÈS E DELFINER, 2012). Além desses, também são usados os critérios de máxima verossimilhança e máxima verossimilhança restrita, os quais são aplicados diretamente no conjunto de dados.

Os critérios de mínimos quadrados não-lineares ordinários, ponderados e generalizados não exigem distribuição gaussiana da variável regionalizada e não sofrem de viés nas estimativas (CRESSIE, 1993). Estes critérios são definidos por:

Mínimos quadrados não-lineares ordinários (MQNO):

$$\sum_{k=1}^K [\hat{\gamma}(h_k) - \gamma(h_k)]^2. \quad (4)$$

Mínimos quadrados não-lineares ponderados (MQNP):

$$\sum_{k=1}^K \omega_k [\hat{\gamma}(h_k) - \gamma(h_k)]^2,$$

em que o peso ω_k , referente a cada *lag*, é proporcional ao número $N(h_k)$ de pares de dados que contribuem para a estimativa $\hat{\gamma}(h_k)$.

Mínimos quadrados não-lineares generalizados (MQNG):

$$\sum_{k=1}^K [\hat{\gamma}(h_k) - \gamma(h_k)]^T \Sigma^{-1} [\hat{\gamma}(h_k) - \gamma(h_k)],$$

em que Σ é a matriz de covariâncias de $\hat{\gamma}(h_k)$.

O critério MQNO tem um apelo geométrico (distância mínima entre os valores experimentais e os estimados por meio do critério), enquanto que o critério MQNG leva em consideração também a correlação entre os diferentes valores na semivariância experimental. O critério MQNP leva em conta o número de pares de dados em cada *lag* da semivariância experimental (GOOVAERTS, 1997).

2.5 Krigagem

O objetivo da análise geoestatística é a predição ou interpolação de dados, em pontos não amostrados, na região em estudo de um fenômeno com dependência espacial. Com isso, ela possibilita a implementação de um mapa que caracterize a variabilidade espacial inerente a esse fenômeno.

A krigagem, nome dado por Matheron (1963) em homenagem ao engenheiro sul-africano Krige, um dos pioneiros no estudo relacionado à geoestatística, se refere a um conjunto de métodos de predição de valores de variáveis aleatórias distribuídas no espaço e/ou tempo. A predição, expressa como um modelo linear, é feita com base em valores amostrais vizinhos (vizinhança), que são considerados interdependentes pelo modelo de semivariância.

Predição ótima

O sistema de krigagem, o qual determina o peso associado a cada dado amostral vizinho, pondera com base na distância entre os dados amostrais e o ponto a ser predito. Além disso, também pondera com base nas separações dos dados amostrais entre si (interposicionamento). Os pesos que daí advêm são tais que o resultado da predição é considerado ótimo. Para que o preditor linear seja ótimo é necessário que ele seja não-viesado ou exato e tenha variância mínima, o que dá origem ao termo em inglês – *Best Linear Unbiased Predictor* (BLUP) ou Melhor Preditor Linear Não-viesado. Matematicamente, a condição de não-viés implica que:

$$E[\hat{Z}(u_0) - Z(u_0)] = 0$$

e a condição de variância mínima pode ser expressa por:

$$Var[\hat{Z}(u_0) - Z(u_0)] = \left\{ E[\hat{Z}(u_0) - Z(u_0)]^2 \right\} \text{ mínima,}$$

em que $\hat{Z}(u_0)$ e $Z(u_0)$ são, respectivamente, as variáveis aleatórias associadas aos valores predito e observado do atributo Z no ponto $u_0 \in D$, D sendo o domínio do fenômeno sob exame.

Goovaerts (1997) lembra que os métodos de krigagem são algoritmos de regressão linear, obtidos por meio do critério dos mínimos quadrados generalizados, de modo que os preditores dos diversos tipos de krigagem são variantes do preditor de regressão linear básico $\hat{Z}(u_0)$, definido como

$$\hat{Z}(u_0) - \mu(u_0) = \sum_{i=1}^n \lambda_i [Z(u_i) - \mu(u_i)],$$

em que λ_i é o peso atribuído ao dado $z(u_i)$, interpretado como uma realização da variável regionalizada $Z(u_i)$; $\mu(u_0)$ e $\mu(u_i)$ são os valores esperados (médias) das variáveis regionalizadas $Z(u_0)$ e $Z(u_i)$; n é o número de dados vizinhos envolvidos na predição de $Z(u_0)$.

Alguns métodos de krigagem que podem ser destacados são a krigagem simples, krigagem universal, krigagem ordinária e krigagem com regressão.

2.5.1 Krigagem simples

Neste tipo de krigagem se considera que a média $\mu(u)$ da variável regionalizada é conhecida e constante em todo domínio D do fenômeno em estudo, segundo Goovaerts (1997). Porém, segundo Journel e Huijbregts (1978), a média deve ser conhecida, mas, não necessariamente, constante. Seguindo Goovaerts (1997), tem-se, então, que:

$$\mu(u) = \mu_0, \text{ conhecida } \forall u \in D.$$

O sistema de krigagem simples inclui $n+1$ equações:

$$\begin{cases} \lambda_0 + \sum_{i=0}^n \lambda_i^{ks} \mu_i = \mu_0. \\ \sum_{i=0}^n \lambda_i^{ks} \sigma_{ij} = \sigma_{i0}, \quad j = 1, \dots, n, \end{cases}$$

em que $\sigma_{ij} = \text{Cov}[Z(u_i), Z(u_j)]$ e $\sigma_{i0} = \text{Cov}[Z(u_i), Z(u_0)]$, em notação simplificada.

Usando notação matricial, o sistema de krigagem simples pode ser escrito como:

$$\mathbf{A}_{ks} \boldsymbol{\lambda}_{ks} = \mathbf{B}_{ks},$$

em que \mathbf{A}_{ks} é a matriz particionada de dimensão $(n+1) \times (n+1)$:

$$\mathbf{A}_{ks} = \begin{bmatrix} 1 & | & \boldsymbol{\Psi}' \\ \text{---} & | & \text{---} \\ \mathbf{O}_{ks} & | & \boldsymbol{\Sigma} \end{bmatrix},$$

na qual, por sua vez:

$\boldsymbol{\Sigma}$ é uma submatriz de dimensão $n \times n$, com as covariâncias entre os dados amostrais:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{bmatrix};$$

$\boldsymbol{\Psi}$ é um subvetor de dimensão $n \times 1$, com as médias nas posições dos dados amostrais, e $\boldsymbol{\Psi}'$ é a sua transposta:

$$\boldsymbol{\Psi} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix};$$

\mathbf{O}_{ks} é um subvetor de dimensão $n \times 1$, com valores iguais a zero;

$\boldsymbol{\lambda}_{ks}$ é um vetor de dimensão $(n+1) \times 1$, com os pesos da krigagem simples:

$$\boldsymbol{\lambda}_{ks} = \begin{bmatrix} \lambda_0 \\ \text{---} \\ \boldsymbol{\Lambda}_{ks} \end{bmatrix},$$

em que o subvetor Λ_{ks} , de dimensão $n \times 1$, é definido por:

$$\Lambda_{ks} = \begin{bmatrix} \lambda_1^{ks} \\ \vdots \\ \lambda_n^{ks} \end{bmatrix};$$

\mathbf{B}_{ks} é um vetor de dimensão $(n+1) \times 1$, com a média no ponto que se quer interpolar e com as covariâncias entre os dados amostrais vizinhos e o ponto que se quer interpolar:

$$\mathbf{B}_{ks} = \begin{bmatrix} \mu_0 \\ - \\ \mathbf{K} \end{bmatrix},$$

em que o subvetor \mathbf{K} , de dimensão $n \times 1$, com as covariâncias entre os dados amostrais vizinhos e o ponto que se quer interpolar, é definido por:

$$\mathbf{K} = \begin{bmatrix} \sigma_{10} \\ \vdots \\ \sigma_{n0} \end{bmatrix}.$$

Os pesos atribuídos aos dados amostrais vizinhos do ponto que se quer interpolar são calculados a partir de:

$$\Lambda_{ks} = \mathbf{A}_{ks}^{-1} \mathbf{B}_{ks}.$$

Assim, o peso λ_0 e o subvetor de pesos Λ_{ks} , resultantes do sistema de krigagem simples, são definidos por:

$$\lambda_0 = \mu_0 - \Psi' \Sigma^{-1} \mathbf{K}.$$

$$\Lambda_{ks} = \Sigma^{-1} \mathbf{K}.$$

O preditor da krigagem simples $\hat{Z}_{ks}(u_0)$ pode ser definido (CRESSIE, 1993) como:

$$\hat{Z}_{ks}(u_0) = \lambda_0 + \Lambda'_{ks} \mathbf{Z},$$

em que \mathbf{Z} é o vetor de dimensão $n \times 1$ com as observações, os dados amostrais vizinhos do ponto que se quer interpolar:

$$\mathbf{Z} = \begin{bmatrix} z(u_1) \\ \vdots \\ z(u_n) \end{bmatrix}.$$

Assim, o preditor da krigagem simples $\hat{Z}_{ks}(u_0)$ pode ser escrito como:

$$\hat{Z}_{ks}(u_0) = \mu_0 - \mathbf{K}' \Sigma^{-1} \Psi + \mathbf{K}' \Sigma^{-1} \mathbf{Z}.$$

A variância do erro de predição minimizada ou variância de krigagem simples $\sigma_{ks}^2(u_0)$ pode ser escrita (CRESSIE, 1993) como:

$$\begin{aligned}\sigma_{ks}^2(u_0) &= Var[Z(u_0) - \hat{Z}_{ks}(u_0)], \\ \sigma_{ks}^2(u_0) &= Var[Z(u_0)] + Var\left[\sum_{i=1}^n \lambda_i^{ks} Z(u_i)\right] - 2Cov\left[Z(u_0), \sum_{i=1}^n \lambda_i^{ks} Z(u_i)\right], \\ \sigma_{ks}^2(u_0) &= \sigma_{00} - \mathbf{K}'\Sigma^{-1}\mathbf{K}.\end{aligned}$$

A variância de krigagem simples é importante por ser uma medida da precisão associada à predição. Tanto aqui como nos demais tipos de krigagem é recomendável que, juntamente com o mapa de valores preditos, seja visualizado também o mapa de erro ou desvio padrão da krigagem, baseado na variância de krigagem. Por esse mapa pode-se averiguar os locais em que a confiança na predição é limitada ou suficiente (VIEIRA, 2000). Quanto mais próximo dos pontos amostrados, tanto menor é a variância visualizada no mapa e maior a precisão, e conseqüentemente, a confiança envolvida no valor predito.

O preditor de krigagem simples é um interpolador exato, pois ao se tentar prever um valor num local que já foi amostrado, será devolvido o valor amostral como resultado da predição e a variância de krigagem simples será igual a zero. Assim, a solução única do sistema determinado de krigagem, assegurada por um modelo de covariância positivo definido, implica em que a predição num ponto deve coincidir com o valor da amostra daquele ponto.

2.5.2 Krigagem universal

Para situações em que a média da variável regionalizada é desconhecida e não pode ser considerada constante, pode ser utilizada a krigagem universal. Nesse tipo de krigagem a média ou tendência, a qual se constitui na componente estrutural de larga escala da variável regionalizada, como foi visto na Equação (3), na seção 2.3, é modelada na suposição de variar de modo suave em função das coordenadas, conforme a expressão (CRESSIE, 1993; GOOVAERTS, 1997):

$$\mu(u) = \sum_{l=0}^p \beta_l f_l(u),$$

em que $f_l(u)$ são funções conhecidas e β_l são coeficientes desconhecidos.

As características conhecidas do fenômeno podem sugerir o tipo de função a ser usada na modelagem. Por exemplo, funções seno e coseno costumam ser usadas na modelagem de tendência periódica de um atributo (GOOVAERTS, 1997). Polinômios de baixa ordem

(menor ou igual a dois) têm uso corriqueiro na descrição de tendências suaves, lineares ou quadráticas.

O sistema de krigagem universal inclui $n+p+1$ equações:

$$\begin{cases} \sum_{j=1}^n \lambda_j^{ku} \sigma_{ij} - \sum_{l=0}^p \xi_l^{ku} f_l(u_i) = \sigma_{i0}, & i = 1, \dots, n; \\ \sum_{j=1}^n f_l(u_j) = f_l(u_0), & l = 0, \dots, p, \end{cases}$$

em que ξ_l^{ku} são multiplicadores de Lagrange, usados na minimização da variância do erro de predição, devido à restrição de não-viés do preditor $\left(\sum_{j=1}^n f_l(u_j) = f_l(u_0) \right)$.

Usando notação matricial, o sistema de krigagem universal pode ser escrito como:

$$\mathbf{A}_{ku} \boldsymbol{\Lambda}_{ku} = \mathbf{B}_{ku},$$

em que \mathbf{A}_{ku} é a matriz particionada de dimensão $(n+p+1) \times (n+p+1)$:

$$\mathbf{A}_{ku} = \begin{bmatrix} \boldsymbol{\Sigma} & | & \mathbf{X} \\ \hline & & \\ \mathbf{X}' & | & \mathbf{O}_{ku} \end{bmatrix},$$

na qual, por sua vez, $\boldsymbol{\Sigma}$ é uma submatriz de dimensão $n \times n$, com as covariâncias entre os dados amostrais, definida na seção 2.5.1; \mathbf{X} é uma submatriz de dimensão $n \times (p+1)$, com os monômios da função polinomial de tendência nas posições dos dados amostrais, e \mathbf{X}' é a sua transposta:

$$\mathbf{X} = \begin{bmatrix} f_0(u_1) & \cdots & f_p(u_1) \\ \vdots & \ddots & \vdots \\ f_0(u_n) & \cdots & f_p(u_n) \end{bmatrix};$$

\mathbf{O}_{ku} é uma submatriz de dimensão $(p+1) \times (p+1)$, com valores iguais a zero;

$\boldsymbol{\Lambda}_{ku}$ é um vetor de dimensão $(n+p+1) \times 1$, com os pesos da krigagem universal e parâmetros de Lagrange:

$$\boldsymbol{\Lambda}_{ku} = \begin{bmatrix} \boldsymbol{\Lambda}_{ku} \\ \hline -\boldsymbol{\Gamma}_{ku} \end{bmatrix},$$

em que o subvetor $\boldsymbol{\Lambda}_{ku}$, de dimensão $n \times 1$, com os pesos da krigagem universal, é definido por:

$$\boldsymbol{\Lambda}_{ku} = \begin{bmatrix} \lambda_1^{ku} \\ \vdots \\ \lambda_n^{ku} \end{bmatrix}.$$

O subvetor $\mathbf{\Gamma}_{ku}$, de dimensão $(p+1) \times 1$, com os multiplicadores de Lagrange, é definido por:

$$\mathbf{\Gamma}_{ku} = \begin{bmatrix} \mathfrak{L}_1^{ku} \\ \vdots \\ \mathfrak{L}_n^{ku} \end{bmatrix};$$

\mathbf{B}_{ku} é um vetor de dimensão $(n+p+1) \times 1$, com as covariâncias entre os dados amostrais vizinhos e o ponto que se quer interpolar, e com os monômios da função de tendência:

$$\mathbf{B}_{ku} = \begin{bmatrix} \mathbf{K} \\ - \\ \mathbf{M} \end{bmatrix},$$

em que \mathbf{K} é um subvetor de dimensão $n \times 1$, com as covariâncias entre os dados amostrais vizinhos e o ponto que se quer interpolar, definido na seção 2.5.1. O subvetor \mathbf{M} , de dimensão $(p+1) \times 1$, com os monômios da função de tendência na posição do ponto que se quer interpolar, é definido por:

$$\mathbf{M} = \begin{bmatrix} f_0(u_0) \\ \vdots \\ f_p(u_0) \end{bmatrix}.$$

Os pesos atribuídos aos dados amostrais vizinhos ao ponto que se quer interpolar são calculados a partir de:

$$\mathbf{\Lambda}_{ku} = \mathbf{A}_{ku}^{-1} \mathbf{B}_{ku}.$$

Assim, o subvetor de pesos $\mathbf{\Lambda}_{ku}$, resultante do sistema de krigagem universal, é definido por:

$$\mathbf{\Lambda}'_{ku} = \mathbf{K}'\mathbf{\Sigma}^{-1} - \mathbf{K}'\mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}^{-1} + \mathbf{M}'(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})\mathbf{X}'\mathbf{\Sigma}^{-1}.$$

O preditor da krigagem universal $\hat{Z}_{ku}(u_0)$ pode ser definido (CRESSIE, 1993) como:

$$\hat{Z}_{ku}(u_0) = \mathbf{\Lambda}'_{ku} \mathbf{Z}, \text{ com } \mathbf{\Lambda}'_{ku} \mathbf{X} = \mathbf{M}',$$

em que \mathbf{Z} é o vetor de dimensão $n \times 1$ com os dados amostrais vizinhos do ponto que se quer interpolar, definido na seção 2.5.1. $\mathbf{\Lambda}'_{ku} \mathbf{X} = \mathbf{M}'$ é a condição necessária e suficiente para assegurar que o preditor seja não-viesado.

A variância do erro de predição minimizada ou variância de krigagem universal, $\sigma_{ku}^2(u_0)$, pode ser escrita (CRESSIE, 1993) como:

$$\sigma_{ku}^2(u_0) = \text{Var}[Z(u_0) - \hat{Z}_{ku}(u_0)],$$

$$\sigma_{ku}^2(u_0) = Var[Z(u_0)] + Var\left[\sum_{i=1}^n \lambda_i^{ku} Z(u_i)\right] - 2Cov\left[Z(u_0), \sum_{i=1}^n \lambda_i^{ku} Z(u_i)\right],$$

$$\sigma_{ku}^2(u_0) = \sigma_{00} + \mathbf{M}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{M} - 2\mathbf{M}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{K} - \mathbf{K}'\boldsymbol{\Sigma}^{-1}\mathbf{K} + \mathbf{K}'\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{K}.$$

O preditor de krigagem universal, como o dos demais tipos de krigagem, é um interpolador exato. A solução única é assegurada não somente por um modelo de covariância positivo definido, mas também pelas $p+1$ funções $f_i(u)$ serem linearmente independentes (GOOVAERTS, 1997).

2.5.3 Krigagem ordinária

Em situações em que a média da variável regionalizada é desconhecida e suposta ser constante, pode ser utilizada a krigagem ordinária (JOURNAL, 1989; CHILÈS E DELFINER, 2012). A krigagem ordinária é um caso particular da krigagem universal. O sistema de krigagem ordinária tem apenas a primeira função de tendência que, por convenção, é a constante um ($f_0(u)=1$) (GOOVAERTS, 1997). Assim, esse sistema apresenta $n+1$ equações:

$$\begin{cases} \sum_{j=1}^n \lambda_j^{ko} \sigma_{ij} - \mathfrak{L}^{ko} = \sigma_{i0}. \\ \sum_{j=1}^n \lambda_j^{ko} = 1, \end{cases}$$

em que \mathfrak{L}^{ko} é um multiplicador de Lagrange, usado na minimização da variância do erro de predição, devido à restrição de não-viés do preditor $\left(\sum_{j=1}^n \lambda_j^{ko} = 1\right)$.

Os pesos Λ_{ko} , resultantes do sistema de krigagem ordinária, são dados por:

$$\Lambda'_{ko} = \mathbf{K}'\boldsymbol{\Sigma}^{-1} + \mathbf{K}'\boldsymbol{\Sigma}^{-1}\mathbf{X}_0(\mathbf{X}'_0\boldsymbol{\Sigma}^{-1}\mathbf{X}_0)^{-1}\mathbf{X}'_0\boldsymbol{\Sigma}^{-1} + (\mathbf{X}'_0\boldsymbol{\Sigma}^{-1}\mathbf{X}_0)\mathbf{X}'_0\boldsymbol{\Sigma}^{-1},$$

$$\Lambda'_{ko} = \mathbf{K}'\boldsymbol{\Sigma}^{-1} + \frac{(1 - \mathbf{K}'\boldsymbol{\Sigma}^{-1}\mathbf{X}_0)\mathbf{X}'_0\boldsymbol{\Sigma}^{-1}}{(\mathbf{X}'_0\boldsymbol{\Sigma}^{-1}\mathbf{X}_0)},$$

em que \mathbf{K} e $\boldsymbol{\Sigma}$ são, respectivamente, vetor e matriz de covariâncias, definidos na seção 2.5.1. \mathbf{X}_0 é uma matriz de dimensão $n \times 1$, com valores iguais a um, correspondentes ao caso de média constante nas posições dos dados amostrais:

$$\mathbf{X}_0 = \begin{bmatrix} f_0(u_1) \\ \vdots \\ f_0(u_n) \end{bmatrix}.$$

O preditor da krigagem ordinária $\hat{Z}_{ko}(u_0)$ pode ser definido (CRESSIE, 1993) como:

$$\hat{Z}_{ko}(u_0) = \Lambda'_{ko} \mathbf{Z}, \text{ com } \Lambda'_{ko} \mathbf{X}_0 = 1,$$

em que \mathbf{Z} é o vetor de dimensão $n \times 1$, com os dados amostrais vizinhos do ponto que se quer interpolar, definido na seção 2.5.1. $\mathbf{\Lambda}'_{ko} \mathbf{X}_0 = 1$ é a condição necessária e suficiente para assegurar que o preditor seja não-viesado.

A variância do erro de predição minimizada ou variância de krigagem ordinária $\sigma_{ko}^2(u_0)$ pode ser escrita como:

$$\begin{aligned}\sigma_{ko}^2(u_0) &= \text{Var}[Z(u_0) - \hat{Z}_{ko}(u_0)], \\ \sigma_{ko}^2(u_0) &= \text{Var}[Z(u_0)] + \text{Var}\left[\sum_{i=1}^n \lambda_i^{ko} Z(u_i)\right] - 2\text{Cov}\left[Z(u_0), \sum_{i=1}^n \lambda_i^{ko} Z(u_i)\right], \\ \sigma_{ko}^2(u_0) &= \sigma_{00} - \mathbf{K}' \mathbf{\Sigma}^{-1} \mathbf{K} - \frac{1}{(\mathbf{X}_0' \mathbf{\Sigma}^{-1} \mathbf{X}_0)} - \frac{2\mathbf{X}_0' \mathbf{\Sigma}^{-1} \mathbf{K}}{(\mathbf{X}_0' \mathbf{\Sigma}^{-1} \mathbf{X}_0)} + \frac{\mathbf{K}' \mathbf{\Sigma}^{-1} \mathbf{X}_0 \mathbf{X}_0' \mathbf{\Sigma}^{-1} \mathbf{K}}{(\mathbf{X}_0' \mathbf{\Sigma}^{-1} \mathbf{X}_0)}.\end{aligned}$$

O preditor de krigagem ordinária, como o dos demais tipos de krigagem, é um interpolador exato. A solução única é assegurada por um modelo de covariância positivo definido.

2.5.4 Krigagem com regressão

Assim como a krigagem universal, a krigagem com regressão pode ser utilizada em situações nas quais a média da variável regionalizada é desconhecida e não pode ser considerada constante. A diferença entre ambas reside em que, na krigagem com regressão, a tendência é modelada com a suposição de variar segundo variáveis secundárias ou covariáveis $w(u)$ diferentes das coordenadas, conforme a expressão:

$$\mu(u) = \sum_{k=0}^K a_k f_k(w(u)),$$

em que $f_k(w(u))$ são funções conhecidas e a_k são coeficientes desconhecidos.

O sistema de krigagem com regressão inclui, então, $n+p+1$ equações e, em notação matricial, pode ser escrito de modo similar ao sistema de krigagem universal. Deve-se observar que, enquanto na krigagem universal se usa as coordenadas como covariáveis, na krigagem com regressão se usa variáveis auxiliares como covariáveis. Apesar dessa diferença de uso de covariáveis, a formulação matricial e matemática é a mesma para ambas.

Goovaerts (1997), com base em resultados experimentais, comenta que o uso de informação adicional a partir de variáveis secundárias corrige parcialmente o efeito indesejado de suavização da krigagem ordinária, o qual se constitui de falta de detalhes sobre a variabilidade (menos precisão) do fenômeno, principalmente em locais com amostras mais esparsas. O efeito de suavização ocorre porque a krigagem é um método de ponderação linear, retornando um valor que está ao redor da média dos dados considerados, de modo que as

predições de krigagem são menos dispersas do que os dados reais. Seguindo nessa linha, o autor verifica também que a correlação entre valores observados e preditos aumenta significativamente com o uso da informação da covariável. Isto é, o uso da krigagem com regressão é importante para reduzir (melhorar) a variância do erro de predição.

2.6 Estimação da semivariância na presença de tendência

Seja um campo aleatório $\{Z(u): u \in D \subset \mathcal{R}^d\}$, intrinsecamente estacionário, com função semivariância dada por $\gamma(\cdot)$ e com média espacial dada por $\mu(\cdot)$ não-estacionária, isto é:

$$Z(u) = \mu(u) + \varepsilon(u).$$

A proposição seguinte não está formalizada na literatura, mas sua ideia geral se encontra dissolvida na mesma. Então, para melhor compreensão do assunto envolvendo a estimação da semivariância na presença de tendência, a proposição é apresentada aqui, explicitamente.

Proposição 1: se $\mu(\cdot)$ for conhecida, pode-se estimar $\gamma(\cdot)$ sem viés, utilizando-se o estimador de semivariância de Matheron sobre os erros.

Demonstração: dada uma amostra espacial $\{Z(u_1), \dots, Z(u_n)\}$, calculam-se os erros $\{\varepsilon(u_1), \dots, \varepsilon(u_n)\}$ pela equação $\varepsilon(u) = Z(u) - \mu(u)$. A partir destes erros, calcula-se o estimador de semivariância de Matheron. Deve-se demonstrar que a esperança do estimador de semivariância de Matheron sobre os erros é igual à semivariância dos dados.

Seja o estimador de semivariância de Matheron sobre os erros

$$\hat{\gamma}_\varepsilon(h_{ij}) = \frac{1}{2N(h_{ij})} \sum_{\substack{\forall (u_i, u_j) \\ \text{tais que} \\ |u_i - u_j| = h_{ij}}} [\varepsilon(u_i) - \varepsilon(u_j)]^2, \text{ em que } N(h_{ij}) \text{ é o número de pares de valores a serem}$$

medidos $\varepsilon(u_i)$ e $\varepsilon(u_j)$, separados pelo vetor h_{ij} . Veja que $\hat{\gamma}_\varepsilon(h_{ij})$ é não-viesado para $\gamma(h_{ij})$:

$$E[\hat{\gamma}_\varepsilon(h_{ij})] = E\left\{ \frac{1}{2N(h_{ij})} \sum [\varepsilon(u_i) - \varepsilon(u_j)]^2 \right\} = \frac{1}{2N(h_{ij})} \sum E[\varepsilon(u_i) - \varepsilon(u_j)]^2,$$

$$E[\hat{\gamma}_\varepsilon(h_{ij})] = \frac{1}{2N(h_{ij})} \sum E[Z(u_i) - \mu(u_i) - Z(u_j) + \mu(u_j)]^2,$$

$$E[\hat{\gamma}_\varepsilon(h_{ij})] = \frac{1}{2N(h_{ij})} \sum E\{[Z(u_i) - Z(u_j)] + [\mu(u_j) - \mu(u_i)]\}^2,$$

$$E[\hat{\gamma}_\varepsilon(h_{ij})] = \frac{1}{2N(h_{ij})} \sum E\{[Z(u_i) - Z(u_j)]^2 + 2[Z(u_i) - Z(u_j)][\mu(u_j) - \mu(u_i)] + [\mu(u_j) - \mu(u_i)]^2\},$$

$$E[\hat{\gamma}_\varepsilon(h_{ij})] = \frac{1}{2N(h_{ij})} \sum E[Z(u_i) - Z(u_j)]^2 + \frac{1}{N(h_{ij})} \sum E[Z(u_i) - Z(u_j)][\mu(u_j) - \mu(u_i)] +$$

$$\frac{1}{2N(h_{ij})} \sum E[\mu(u_j) - \mu(u_i)]^2.$$

Observe que:

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

$$2\gamma(h) = E[Z(u) - Z(u+h)]^2 - [\mu(u) - \mu(u+h)]^2,$$

portanto,

$$E[\hat{\gamma}_\varepsilon(h_{ij})] = \frac{1}{2N(h_{ij})} \sum \{2\gamma(h_{ij}) + [\mu(u_j) - \mu(u_i)]^2\} - \frac{1}{N(h_{ij})} \sum [\mu(u_i) - \mu(u_j)][-\mu(u_j) + \mu(u_i)] +$$

$$\frac{1}{2N(h_{ij})} \sum [\mu(u_i) - \mu(u_j)]^2,$$

$$E[\hat{\gamma}_\varepsilon(h_{ij})] = \frac{1}{2N(h_{ij})} \sum 2\gamma(h_{ij}) + \frac{1}{2N(h_{ij})} \sum [\mu(u_i) - \mu(u_j)]^2 - \frac{1}{N(h_{ij})} \sum [\mu(u_i) - \mu(u_j)]^2 +$$

$$\frac{1}{2N(h_{ij})} \sum [\mu(u_i) - \mu(u_j)]^2,$$

$$E[\hat{\gamma}_\varepsilon(h_{ij})] = \frac{1}{2N(h_{ij})} N(h_{ij}) 2\gamma(h_{ij}) + \frac{1}{N(h_{ij})} \sum_{\substack{\forall (u_i, u_j) \\ \text{tais que} \\ |u_i - u_j| = h_{ij}}} [\mu(u_i) - \mu(u_j)]^2 - \frac{1}{N(h_{ij})} \sum_{\substack{\forall (u_i, u_j) \\ \text{tais que} \\ |u_i - u_j| = h_{ij}}} [\mu(u_i) - \mu(u_j)]^2,$$

$$E[\hat{\gamma}_\varepsilon(h_{ij})] = \gamma(h_{ij}).$$

Assim, provou-se que $E[\hat{\gamma}_\varepsilon(h_{ij})] = \gamma(h_{ij})$, isto é, $\hat{\gamma}_\varepsilon(h_{ij})$ é um estimador não-viesado para $\gamma(h_{ij})$.

Na prática, a média $\mu(\cdot)$ frequentemente não é conhecida, de modo que não é possível o cálculo dos erros. Assim, deve-se atentar para a proposição seguinte, a qual se encontra presente na literatura (CRESSIE, 1993; CHILÈS E DELFINER, 2012), porém sem demonstração.

Proposição 2: se, em vez de se usar o estimador da semivariância dos erros $\hat{\gamma}_\varepsilon(h)$, usa-se o estimador da semivariância dos dados observados $\hat{\gamma}(h)$, então este estimador $\hat{\gamma}(h)$ é viesado.

Demonstração: deve-se demonstrar que a esperança do estimador de semivariância de Matheron sobre os dados é diferente da semivariância dos dados.

Seja o estimador de semivariância de Matheron sobre os dados

$$\hat{\gamma}(h_{ij}) = \frac{1}{2N(h_{ij})} \sum_{\substack{\forall (u_i, u_j) \\ \text{tais que} \\ |u_i - u_j| = h_{ij}}} [Z(u_i) - Z(u_j)]^2, \text{ em que } N(h_{ij}) \text{ é o número de pares de valores a serem}$$

medidos $Z(u_i)$ e $Z(u_j)$, separados pelo vetor h_{ij} . Veja que $\hat{\gamma}(h_{ij})$ é viesado para $\gamma(h_{ij})$:

$$E[\hat{\gamma}(h_{ij})] = E \left\{ \frac{1}{2N(h_{ij})} \sum_{\substack{\forall (u_i, u_j) \\ \text{tais que} \\ |u_i - u_j| = h_{ij}}} [Z(u_i) - Z(u_j)]^2 \right\} = \frac{1}{2N(h_{ij})} \sum E \{ [Z(u_i) - Z(u_j)]^2 \},$$

pela demonstração da proposição 1, tem-se que:

$$\begin{aligned} E[\hat{\gamma}(h_{ij})] &= \frac{1}{2N(h_{ij})} \sum \{ 2\gamma(h_{ij}) + [\mu(u_i) - \mu(u_j)]^2 \}, \\ E[\hat{\gamma}(h_{ij})] &= \frac{1}{2N(h_{ij})} \sum 2\gamma(h_{ij}) + \frac{1}{2N(h_{ij})} \sum [\mu(u_i) - \mu(u_j)]^2, \\ E[\hat{\gamma}(h_{ij})] &= \frac{1}{2N(h_{ij})} N(h_{ij}) 2\gamma(h_{ij}) + \frac{1}{2N(h_{ij})} \sum [\mu(u_i) - \mu(u_j)]^2, \\ E[\hat{\gamma}(h_{ij})] &= \gamma(h_{ij}) + \frac{1}{2N(h_{ij})} \sum_{\substack{\forall (u_i, u_j) \\ \text{tais que} \\ |u_i - u_j| = h_{ij}}} [\mu(u_i) - \mu(u_j)]^2. \end{aligned} \quad (5)$$

Assim, provou-se que $E[\hat{\gamma}(h_{ij})] \neq \gamma(h_{ij})$, isto é, $\hat{\gamma}(h_{ij})$ é um estimador viesado para $\gamma(h_{ij})$.

Observe que $\hat{\gamma}(h_{ij})$ produz um viés que superestima o valor de $\gamma(h_{ij})$, como pode ser visto pela Equação (5).

Como a média $\mu(\cdot)$ não é conhecida, o estimador da semivariância dos erros não pode ser calculado. Além disso, como o estimador da semivariância dos dados observados apresenta viés para o caso de média $\mu(\cdot)$ não-estacionária, deve-se utilizar um outro estimador de semivariância em seu lugar. O estimador da semivariância dos resíduos ou erros preditos pode ser usado no lugar do estimador da semivariância dos erros. No entanto, o estimador da semivariância dos resíduos apresenta um outro tipo de viés, conforme será visto na proposição seguinte, a qual está presente na literatura (MATHERON, 1971; CRESSIE, 1993; CHILÈS E DELFINER, 2012), porém sem demonstração.

Proposição 3: se $\mu(\cdot)$ não for conhecida, sendo, portanto, estimada por $\hat{\mu}(\cdot)$, o cálculo dos resíduos $R(u_i) = Z(u_i) - \hat{\mu}(u_i)$, $\forall i = 1, \dots, n$ da amostra, não produzirá um estimador de Matheron não-viesado.

Demonstração: deve-se demonstrar que a semivariância dos resíduos é diferente da semivariância dos dados.

Seja o resíduo $R(u_i) = Z(u_i) - \hat{\mu}(u_i)$. A semivariância dos resíduos é dada por:

$$\gamma_{\text{Res}}(h) = \frac{1}{2} \text{Var}[R(u) - R(u+h)],$$

$$\gamma_{\text{Res}}(h) = \frac{1}{2} E[R(u) - R(u+h)]^2 = \frac{1}{2} E[Z(u) - \hat{\mu}(u) - Z(u+h) + \hat{\mu}(u+h)]^2,$$

$$\gamma_{\text{Res}}(h) = \frac{1}{2} E[(Z(u) - Z(u+h)) - (\hat{\mu}(u) - \hat{\mu}(u+h))]^2,$$

$$\gamma_{\text{Res}}(h) = \frac{1}{2} E[(Z(u) - Z(u+h))^2 - 2(Z(u) - Z(u+h))(\hat{\mu}(u) - \hat{\mu}(u+h)) + (\hat{\mu}(u) - \hat{\mu}(u+h))^2],$$

$$\gamma_{\text{Res}}(h) = \frac{1}{2} E[(Z(u) - Z(u+h))^2 - E[(Z(u) - Z(u+h))(\hat{\mu}(u) - \hat{\mu}(u+h))] + \frac{1}{2} E[(\hat{\mu}(u) - \hat{\mu}(u+h))^2],$$

$$\gamma_{\text{Res}}(h) = \frac{1}{2} E[Z(u) - Z(u+h)]^2 - \frac{1}{2} [\mu(u) - \mu(u+h)]^2 - E[(Z(u) - Z(u+h))(\hat{\mu}(u) - \hat{\mu}(u+h))] + \frac{1}{2} E[(\hat{\mu}(u) - \hat{\mu}(u+h))^2] - \frac{1}{2} [\mu(u) - \mu(u+h)]^2 + [\mu(u) - \mu(u+h)]^2,$$

observe que:

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

$$-E[(Z(u) - Z(u+h))(\hat{\mu}(u) - \hat{\mu}(u+h))] = -\text{Cov}[Z(u) - Z(u+h), \hat{\mu}(u) - \hat{\mu}(u+h)] - [\mu(u) - \mu(u+h)][\mu(u) - \mu(u+h)],$$

$$\frac{1}{2} \text{Var}[(\hat{\mu}(u) - \hat{\mu}(u+h))] = \frac{1}{2} E[(\hat{\mu}(u) - \hat{\mu}(u+h))]^2 - \frac{1}{2} \{E[\hat{\mu}(u) - \hat{\mu}(u+h)]\}^2,$$

$$\frac{1}{2} \text{Var}[(\hat{\mu}(u) - \hat{\mu}(u+h))] = \frac{1}{2} E[(\hat{\mu}(u) - \hat{\mu}(u+h))]^2 - \frac{1}{2} [\mu(u) - \mu(u+h)]^2,$$

portanto,

$$\gamma_{\text{Res}}(h) = \frac{1}{2} \text{Var}[(Z(u) - Z(u+h))] - \text{Cov}[Z(u) - Z(u+h), \hat{\mu}(u) - \hat{\mu}(u+h)] -$$

$$-[\mu(u) - \mu(u+h)][\mu(u) - \mu(u+h)] + \frac{1}{2} \text{Var}[(\hat{\mu}(u) - \hat{\mu}(u+h))] + [\mu(u) - \mu(u+h)]^2,$$

$$\gamma_{\text{Res}}(h) = \gamma(h) - \text{Cov}[Z(u) - Z(u+h), \hat{\mu}(u) - \hat{\mu}(u+h)] + \frac{1}{2} \text{Var}[\hat{\mu}(u) - \hat{\mu}(u+h)]. \quad (6)$$

Assim, provou-se que $\gamma_{Res}(h) \neq \gamma(h)$, isto é, ao se utilizar um estimador da semivariância dos resíduos não se obtém um estimador não-viesado da semivariância dos dados, visto que se tratam de semivariâncias distintas.

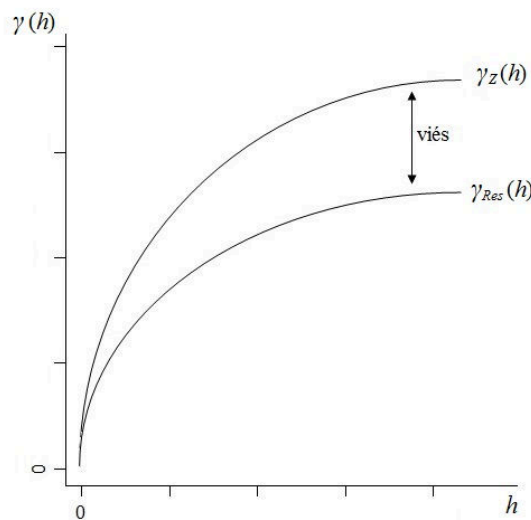
A proposição seguinte é dada por Chilès e Delfiner (2012), sem demonstração. Nesta tese também não será demonstrada.

Proposição 4: se $\mu(\cdot)$ desconhecida for estimada por $\hat{\mu}(\cdot) = \tilde{\mu}(\cdot)$ que é um BLUE (*Best Linear Unbiased Estimator*), isto é, um estimador que é não-viesado e que tem variância mínima, então os resíduos $R(u_i) = Z(u_i) - \tilde{\mu}(u_i)$, $\forall i = 1, \dots, n$ da amostra, produzirão um viés que pode ser visto na Equação (7):

$$\gamma_{Res}(h) = \gamma(h) - \frac{1}{2} \text{Var}[\tilde{\mu}(u) - \tilde{\mu}(u+h)]. \quad (7)$$

De acordo com a Equação (7), os valores das semivariâncias dos resíduos são viesados “para baixo”, isto é, os valores são sistematicamente subestimados e o patamar do gráfico de semivariância é menor do que deveria ser, conforme ilustrado na Figura 6.

Figura 6 – Semivariância (viesada) dos resíduos $\gamma_{Res}(h)$ e semivariância real $\gamma_Z(h)$.



Fonte: Do autor (2017).

Cressie (1993) observou que o viés, visto na Equação (7) para a semivariância dos resíduos, é pequeno em *lags* próximos à origem, mas mais substancial em *lags* mais distantes, nos quais há menos pares de dados envolvidos. Esse autor e Kim e Boos (2004) afirmam que tal viés afeta a variância estimada de krigagem, de modo que essa variância será menor do que deveria.

Beckers e Bogaert (1998), por meio de estudos de caso, nos quais usaram um estimador para a média $\mu(\cdot)$ que não era um BLUE, mostram que o viés na Equação (6) pode

ser negligenciável em alguns casos, mas em outros pode mesmo ser maior do que 20% do patamar, ou seja, pode não ser negligenciável. Os autores comentam que a magnitude do viés em questão depende de fatores, tais como: configuração da amostragem, número de parâmetros usados para modelar a média, quantidade de dados amostrais e força da dependência espacial entre as observações, a qual diz respeito ao modelo de semivariância com seus parâmetros. Eles reconhecem ainda a dificuldade de formular regras gerais acerca do impacto de cada fator na constituição do viés, em virtude das múltiplas combinações que se pode ter com esses fatores em conjunto.

Pelletier et al. (2009), em análise via simulação, identificam dois componentes no viés do estimador da semivariância dos resíduos: um viés devido à substituição da tendência desconhecida por estimativas e um viés devido a qualquer má especificação do modelo de tendência. Este segundo viés ocorre, por exemplo, quando se usa um modelo polinomial de 1º grau quando o correto seria usar um modelo de 2º grau. Em seu trabalho, os autores percebem que o viés aumenta para maiores *lags* de distância, o que está de acordo com Cressie (1993) e é diretamente proporcional à força da dependência espacial envolvida. Segundo Pelletier et al. (2009), o viés também cresce com o aumento da ordem do polinômio usado para estimar a tendência, o que está em consonância com os resultados obtidos por Beckers e Bogaert (1998). Ademais, relatam que o viés apresenta diminuição para um crescente tamanho de amostra.

Cressie (1993) salienta que modelos de semivariância ajustados pelos critérios de mínimos quadrados não-lineares ponderados e generalizados corrigem, em parte, o efeito do viés na estimação da semivariância dos resíduos porque, no cálculo do estimador, dão mais peso aos pequenos *lags*, nos quais o efeito do viés é menor. Além disso, a krigagem realizada com vizinhança local também contribui para diminuir o impacto negativo do viés, pois o modelo de semivariância usado na predição é calculado para *lags* não distantes da origem.

Beckers e Bogaert (1998) desenvolveram uma expressão para o viés envolvido na estimação da semivariância dos resíduos, para o caso em que essa estimação é realizada com o uso de coeficientes de tendência estimados por meio do critério de mínimos quadrados lineares ordinários. Os autores usam a expressão para compor um estimador de semivariância com viés corrigido, o qual é utilizado em modelos de semivariância ajustados com o auxílio de um procedimento iterativo que atualiza os parâmetros *pepita*, *alcance* e *contribuição*, a cada iteração, até que haja convergência desses parâmetros, segundo algum critério definido pelo usuário. Esse método de correção de viés deixa a cargo do usuário a escolha do modelo de semivariância mais adequado dentre os candidatos. Os autores dizem que seu método é

uma extensão do critério de mínimos quadrados não-lineares ponderados proposto por Cressie (1985) e que não faz pressuposição acerca da distribuição dos dados. Além disso, com base num estudo de caso, comentam que a correção de viés pelo método pode ser não significativa para o caso em que se tem um número de parâmetros, usado para modelar a média, pequeno em relação à quantidade de observações amostradas.

Kim e Boos (2004), assumindo que correlações são monótonas decrescentes com a distância, propõe um método que usa um fator de correção do viés na semivariância dos resíduos, o qual é expresso pela razão entre a semivariância dos erros e a média da semivariância dos resíduos, preditos com o uso de coeficientes de tendência estimados por meio do critério dos mínimos quadrados lineares ordinários. O fator é calculado com o uso da matriz de covariâncias estimadas, obtida a partir de uma versão monotonizada do estimador da semivariância dos resíduos. Essa versão monotonizada, a qual resulta num gráfico de semivariância monótono crescente, se baseia num procedimento de substituição de valores de semivariâncias que pode ser descrito da seguinte maneira: seja $\hat{\gamma}(h_i)$ o valor da semivariância estimada para o i -ésimo *lag* de distância. Começando com $\hat{\gamma}(h_1)$ no gráfico de semivariância, deve-se mover para a direita e parar se $\hat{\gamma}(h_i) > \hat{\gamma}(h_{i+1})$; neste caso, deve-se substituir $\hat{\gamma}(h_i)$ e $\hat{\gamma}(h_{i+1})$ por sua média ponderada $\bar{\gamma}(h_i)$; então, deve-se mover para a esquerda para se assegurar de que $\hat{\gamma}(h_{i-1}) \leq \bar{\gamma}(h_i)$ usando a média ponderada. Deve-se continuar a mover para a esquerda e checar, da mesma maneira, se $\hat{\gamma}(h_{i-1}) \leq \bar{\gamma}(h_i)$ para todos os valores de semivariâncias estimadas até chegar ao primeiro *lag*. Então, procede-se novamente para a direita. Esse processo de substituição de valores de semivariâncias continua até que o último *lag* à direita seja alcançado.

Após o término do procedimento de substituição de valores de semivariâncias, são considerados, para cálculos posteriores, somente os valores das semivariâncias estimadas até a distância limite de cerca da metade da maior distância entre os dados na direção considerada, de acordo com a recomendação de Journel e Huijbregts (1978). Finalmente, o estimador da semivariância dos resíduos, que é submetido à correção de viés, deve ser, também, monotonizado (KIM E BOOS, 2004). A partir de análise via simulação e resultados medidos em logaritmos dos valores estimados de semivariâncias, os autores comentam que seu método produz um estimador de semivariância dos resíduos que apresenta menor (melhor) erro quadrático médio, quando comparado ao estimador de Matheron. Além disso, afirmam que a sua abordagem é invariante à presença ou ausência de efeito pepita.

Bliznyuk et al. (2012) propõem um método para correção do viés na estimação da semivariância dos resíduos, o qual envolve a minimização de uma soma de quadrados dos resíduos que leva em consideração a tendência, além dos parâmetros de semivariância. Essa minimização é realizada por meio de um critério de mínimos quadrados não-lineares ordinários no ajuste do modelo de semivariância. O critério pode ser expresso por:

$$\sum_{k=1}^K [\gamma(h_k) + \beta h_k^2 - \hat{\gamma}(h_k) - \hat{\beta} h_k^2]^2,$$

em que βh_k^2 é referente ao componente de tendência, presente em cada *lag*, e $\hat{\beta}$ é o estimador de β .

O método de Bliznyuk et al. (2012) se baseia na teoria assintótica e pressupõe que a tendência é uma função de regressão que varia de modo suave. Os autores argumentam que o estimador da semivariância dos resíduos, resultante do método, é consistente sob condições de regularidade padrão da teoria assintótica de domínio crescente, além de ser computacionalmente mais eficiente do que o estimador de máxima verossimilhança. Porém, advertem que a eficiência do estimador, que é baseado no critério de mínimos quadrados não-lineares ordinários, pode ser melhorada ao se utilizar o critério de mínimos quadrados não-lineares ponderados ou generalizados.

Desassis e Renard (2014) desenvolveram uma expressão para o viés, presente em cada *lag* de distância, no estimador da semivariância dos resíduos, preditos com o uso de coeficientes de tendência estimados por meio do critério dos mínimos quadrados lineares ordinários. Seu método inclui um algoritmo iterativo que envolve sucessivos ajustes do modelo de semivariância que são corrigidos com o uso da expressão desenvolvida para o viés, a qual, por sua vez, é recalculada a cada iteração com os valores atualizados de semivariâncias estimadas. No entanto, os autores não apresentam resultados práticos do desempenho do método.

Cressie (1993) comenta que estimadores de covariância baseados na mediana (como o estimador *median-polish*) são menos viesados do que os baseados na média. Porém, segundo o mesmo autor, tais estimadores exigem gaussianidade da distribuição dos dados, pressuposição esta que pode não ser atendida pelos dados sob exame.

Ainda em relação ao viés no estimador da semivariância dos resíduos, parece soar atual e apropriada a afirmação de Cressie (1993) e de Kim e Boos (2004), quanto a considerar ainda em aberto o desafio de se encontrar um estimador adequado para a expressão do viés, com vistas à redução bem sucedida de sua influência indesejada.

Em vista dessa revisão bibliográfica, pode-se dizer que o problema do viés na estimação da semivariância na presença de tendência, não está esgotado, o que justifica o desenvolvimento desta tese.

3 MATERIAL E MÉTODOS

A implementação da proposta metodológica foi feita na linguagem do ambiente do software estatístico R (R CORE TEAM, 2017), cujo *script* pode ser visto no Apêndice. Essa implementação envolveu:

- a) a construção do algoritmo com o esquema de estimação IRWGLS (SCHABENBERGER E GOTWAY, 2005), descrito na seção de resultados e discussão, o qual é usado na estimação dos coeficientes de tendência da variável regionalizada e, concomitantemente, na estimação da semivariância dos resíduos, que é submetida ao ajuste de modelo;
- b) a implementação de um termo de correção para o viés presente na estimação da semivariância dos resíduos na Equação (7), vista na seção 2.6, termo este a ser usado no modelo de semivariância dos resíduos ajustado com o auxílio do algoritmo com o esquema de estimação IRWGLS.

3.1 Dados

Os dados utilizados nas análises se referem a três conjuntos de observações. O primeiro conjunto, de nome Parana, apresenta 143 medições de índice pluviométrico feitas em estações meteorológicas localizadas em distintos pontos no estado do Paraná, região sul do Brasil. As medições foram realizadas durante os meses de maio e junho, em diferentes anos. O conjunto de dados Parana está disponível no pacote *geoR* (RIBEIRO JÚNIOR E DIGGLE, 2001), do software R (R CORE TEAM, 2017).

O segundo conjunto de dados, de nome Meuse, apresenta 164 medições de concentrações de metais pesados no solo, em partes por milhão (ppm), juntamente com outras variáveis, feitas na região dentro do raio de inundação do rio Meuse, próximo à vila de Stein, na Holanda. Nesta tese, foram usadas as medições referentes ao metal cobre. O conjunto de dados Meuse está disponível no pacote *gstat* (PEBESMA, 2004), do software R (R CORE TEAM, 2017).

O terceiro conjunto de dados, aqui denominado Castanha, é referente a ocorrência de árvores de castanha-do-pará numa área da Amazônia brasileira e apresenta 60 medições de densidade de castanha-do-pará. Essa densidade é calculada pela razão entre o número de árvores e a área de um hectare (árvores/ha). O conjunto de dados Castanha foi gentilmente cedido por Batista (2017), do Programa de Pós-Graduação em Engenharia Florestal, área de concentração em Ciências Florestais, da UFLA.

Os dois primeiros conjuntos de dados são bem conhecidos na literatura de Geoestatística. O terceiro conjunto de dados foi usado com o fim de se testar o método proposto nesta tese com uma amostra de tamanho pequeno.

Eles foram analisados, de acordo com os objetivos estabelecidos neste trabalho, no software R (R CORE TEAM, 2017), com o auxílio do pacote geoR (RIBEIRO JÚNIOR E DIGGLE, 2001).

3.2 Validação dos resultados

Para avaliar e comparar o desempenho do método proposto com o método em que a estimação da semivariância na presença de tendência é realizada sem que o viés presente seja corrigido, os resultados obtidos nas análises foram submetidos a teste de autovalidação (*leave-one-out*). Foi utilizado o método de validação proposto por Bastos (2017) o qual, com base numa medida de distância, permite a seleção do melhor modelo de semivariância dentre os concorrentes. Mais especificamente, quanto menor for a referida distância resultante para um modelo, mais adequado é considerado este modelo para descrever a estrutura da variabilidade espacial do fenômeno sob exame.

O método de validação é uma combinação ponderada, com base em postos, de outros quatro métodos. Dentre esses métodos, o primeiro utiliza o erro quadrático médio (EQM) como medida de validação e o segundo usa o coeficiente de determinação (R^2). O terceiro método utiliza um teste que verifica se os coeficientes linear b_0 e angular b_1 de uma reta de regressão linear no gráfico de dispersão dos valores observados *versus* os preditos pela validação são, respectivamente, diferentes de 0 e 1. Isto é, o terceiro método avalia, para cada modelo de semivariância testado, se os coeficientes linear e angular diferem, significativamente, da reta de regressão ideal, na qual os valores observados são iguais aos preditos. Além disso, esse método também avalia a variabilidade dos pontos em torno da reta de regressão, cujos coeficientes não foram diferentes de 0 e 1, de modo que o modelo de semivariância escolhido é o que apresenta pontos menos dispersos em relação a essa reta. Finalmente, o quarto método usa o coeficiente de correlação de concordância (CCC), proposto por Lin (1989), para avaliar o grau de concordância entre os valores observados e preditos pela validação.

O funcionamento do método de validação de Bastos (2017) pode ser descrito nos seguintes passos:

1. selecione r métodos de seleção (nesta tese utilizou-se $r = 4$ métodos);
2. os postos são impostos da seguinte forma:

$$\left| \frac{EstMM - EstM_i}{EstMM} \right|,$$

em que $EstMM$ é a estimativa do melhor modelo e $EstM_i$ é a estimativa do modelo i , $i =$ exponencial, gaussiano e esférico, nesta tese;

3. classifique as estimativas com esses postos, de acordo com a ordem que o método informa, do melhor modelo (posto de menor valor) para o pior, e calcule as médias ($Média_{Posto}$) e os desvios padrão (DP_{Posto}) dos postos em relação a cada modelo.

Um exemplo é dado na Tabela 1:

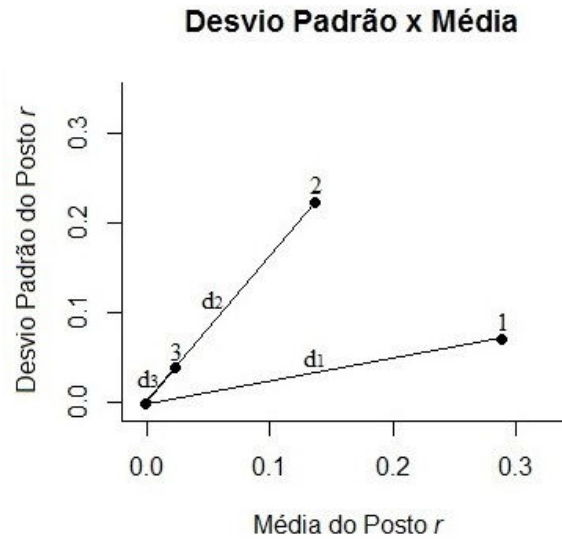
Tabela 1 – Estimativas classificadas de $r = 3$ métodos selecionados, conforme o método de validação apresentado por Bastos (2017), com as médias e os desvios padrão dos postos calculados.

Método r	Modelo i		
	1	2	3
1	4,6140	3,7600	4,0150
Posto	0,2270	0,0000	0,0680
2	5,4720	5,6000	4,0130
Posto	0,3640	0,3950	0,0000
3	5,0980	4,0600	4,0040
Posto	0,2730	0,0140	0,0000
$Média_{Posto}$	0,2880	0,01360	0,0230
DP_{Posto}	0,0700	0,2240	0,0390

Fonte: Adaptado de Bastos (2017).

4. plote um gráfico com os pontos formados pelos desvios padrão em relação às médias dos postos e calcule a distância dos pontos em relação ao ponto mínimo ($Média_{Posto} = 0$, $DP_{Posto} = 0$), conforme a Figura 7:

Figura 7 – Pontos formados pelos desvios padrões em relação às médias dos postos, com as distâncias desses pontos em relação ao ponto (0,0), conforme o método de validação apresentado por Bastos (2017).



Fonte: Adaptado de Bastos (2017).

5. selecione o modelo relacionado à menor distância resultante, ou seja, no exemplo seria selecionado o modelo 3.

4 RESULTADOS E DISCUSSÃO

Nesta seção será apresentada, inicialmente, a construção de uma proposta metodológica para a correção do viés presente na estimação da semivariância na presença de tendência. Em seguida, uma aplicação de tal proposta é feita com o uso dos três conjuntos de dados já descritos.

4.1 Método proposto para a correção de viés na semivariância dos resíduos

Na proposição 4 da seção 2.6 foi visto que, se $\mu(\cdot)$ desconhecida for estimada por $\mu^*(\cdot) = \tilde{\mu}(\cdot)$ que é BLUE, os resíduos $R(u_i) = Z(u_i) - \tilde{\mu}(u_i)$, $\forall i = 1, \dots, n$ da amostra, produzirão um viés que pode ser visto na Equação (7). Pode ser observado que o viés implica em semivariâncias dos resíduos que são sistematicamente subestimadas e o patamar do gráfico de semivariância é menor do que deveria. Um termo de correção para o referido viés deve retificar essa subestimação. Com isso em vista, nesta tese propõe-se um estimador de semivariância $\hat{\gamma}_c(h)$ com correção de viés, dado por:

$$\hat{\gamma}_c(h) = \hat{\gamma}_{Res}(h) + \frac{1}{2} Var[\tilde{\mu}(u) - \tilde{\mu}(u+h)] , \quad (8)$$

em que

$$\frac{1}{2} Var[\tilde{\mu}(u) - \tilde{\mu}(u+h)] = \frac{1}{2} \{Var[\tilde{\mu}(u)] + Var[\tilde{\mu}(u+h)] - 2Cov[\tilde{\mu}(u), \tilde{\mu}(u+h)]\},$$

que, em termos matriciais, é dado por

$$\begin{aligned} \tilde{\mu}(u) &= \mathbf{X}_u \tilde{\boldsymbol{\beta}} \\ Var[\mathbf{X}_u \tilde{\boldsymbol{\beta}}] &= \mathbf{X}_u Var[\tilde{\boldsymbol{\beta}}] \mathbf{X}'_u, \\ Var[\mathbf{X}_u \tilde{\boldsymbol{\beta}}] &= \mathbf{X}_u (\mathbf{X}' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X})^{-1} \mathbf{X}'_u, \end{aligned} \quad (9)$$

sendo \mathbf{X}_u um vetor de dimensão $1 \times (p+1)$, com os monômios (constantes) da função polinomial de tendência na posição u do dado amostral, e \mathbf{X}'_u é a sua transposta, com

$$\mathbf{X}_u = [f_0(u) \ f_1(u) \ \dots \ f_p(u)];$$

$\tilde{\boldsymbol{\beta}}$ é um vetor de dimensão $(p+1) \times 1$, com os estimadores ótimos, isto é, estimadores não-viesados e com variância mínima, dos coeficientes de tendência, ou seja,

$$\tilde{\boldsymbol{\beta}} = \begin{bmatrix} \tilde{\beta}_0 \\ \vdots \\ \tilde{\beta}_p \end{bmatrix};$$

\mathbf{X} é uma matriz de dimensão $n \times (p+1)$, com os monômios da função polinomial de tendência nas posições dos dados amostrais, definida na seção 2.5.2, e \mathbf{X}' é a sua transposta, ou seja,

$$\mathbf{X} = \begin{bmatrix} f_0(u_1) & \cdots & f_p(u_1) \\ \vdots & \ddots & \vdots \\ f_0(u_n) & \cdots & f_p(u_n) \end{bmatrix};$$

$\hat{\Sigma}$ é uma matriz de dimensão $n \times n$, com os estimadores das covariâncias, ou seja,

$$\hat{\Sigma} = \begin{bmatrix} \hat{\sigma}_{11} & \cdots & \hat{\sigma}_{1n} \\ \vdots & \ddots & \vdots \\ \hat{\sigma}_{n1} & \cdots & \hat{\sigma}_{nn} \end{bmatrix}.$$

Além disso,

$$\begin{aligned} Cov[\tilde{\mu}(u), \tilde{\mu}(u+h)] &= Cov[\mathbf{X}_u \tilde{\beta}, \mathbf{X}_{u+h} \tilde{\beta}] = \mathbf{X}_u Var[\tilde{\beta}] \mathbf{X}'_{u+h}, \\ Cov[\tilde{\mu}(u), \tilde{\mu}(u+h)] &= \mathbf{X}_u (\mathbf{X}' \hat{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}'_{u+h}. \end{aligned} \quad (10)$$

Como a variação da variável regionalizada tem erros espacialmente correlacionados, os estimadores ótimos dos coeficientes de tendência, para o cálculo dos resíduos, são obtidos por meio do critério dos mínimos quadrados lineares generalizados.

Considerando que um modelo linear com erros espacialmente correlacionados pode ser representado da seguinte forma (SCHABENBERGER E GOTWAY, 2005):

$$\mathbf{Z} = \mathbf{X}\beta + \varepsilon, \text{ com } \varepsilon \sim (\mathbf{0}, \Sigma),$$

em que \mathbf{Z} é um vetor de observações, de dimensão $n \times 1$, definido na seção 2.5.1, assim como a matriz de covariâncias Σ , de dimensão $n \times n$. \mathbf{X} é uma matriz, de dimensão $n \times (p+1)$, com os monômios da função de tendência nas posições dos dados amostrais, definida na seção 2.5.2 e ε é um vetor, de dimensão $n \times 1$, com os erros espacialmente correlacionados. $\mathbf{0}$ é um vetor, de dimensão $n \times 1$, com valores iguais a zero e β é um vetor, de dimensão $(p+1) \times 1$, com os coeficientes de tendência:

$$\beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix}.$$

O estimador de mínimos quadrados lineares generalizados de β ($\hat{\beta}_{\text{MQLG}}$) é dado por:

$$\hat{\beta}_{\text{MQLG}} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Z}. \quad (11)$$

Para estimar os coeficientes de tendência é necessário o conhecimento dos parâmetros da covariância, os quais, por sua vez, devem ser estimados a partir da semivariância dos

resíduos, isto é, tem-se um problema de circularidade. Diante disso, Schabenberger e Gotway (2005), numa abordagem que se assemelha à proposta por Neuman e Jacobson (1984), sugerem um esquema de estimação por eles chamado de Mínimos Quadrados Generalizados Iterativamente Reponderados (ou IRWGLS, do inglês *Iteratively Re-weighted Generalized Least Squares*), descrito nos seguintes passos:

1. obter uma estimativa inicial de $\boldsymbol{\beta}$ ($\boldsymbol{\beta}^*$), sendo que essa estimativa pode ser obtida por meio do critério dos mínimos quadrados lineares ordinários;
2. calcular os resíduos $\mathbf{R} = \mathbf{Z} - \mathbf{P}\boldsymbol{\beta}^*$;
3. estimar a semivariância dos resíduos e ajustar o modelo de semivariância, usando as técnicas descritas nas seções 2.3 e 2.4;
4. obter uma nova estimativa de $\boldsymbol{\beta}$ usando a Equação (11), mas substituindo $\boldsymbol{\Sigma}$ por $\hat{\boldsymbol{\Sigma}}$, a matriz de covariâncias estimadas dos resíduos, obtida pela relação desta com a semivariância dos resíduos, $\hat{C}_{Res}(h) = \hat{C}_{Res}(0) - \hat{\gamma}_{Res}(h)$, de acordo com a Equação (2), vista na seção 2.3;
5. repetir os passos 2 a 4 até que as mudanças em $\hat{\boldsymbol{\Sigma}}$ ou $\boldsymbol{\beta}^*$ sejam pequenas, segundo algum critério de convergência, o qual, nesta tese, foi o de uma diferença absoluta não maior do que 10^{-3} entre os elementos, em linhas e colunas respectivas, das matrizes $\hat{\boldsymbol{\Sigma}}$ estimadas em duas iterações sucessivas.

Como a semivariância dos resíduos, estimada com o uso do critério dos mínimos quadrados lineares generalizados na modelagem da tendência, é viesada, conforme visto na proposição 4 na seção 2.6, deve-se submeter o modelo de semivariância, estimado dentro do esquema IRWGLS, à correção de viés proposta na Equação (8).

Assim, a implementação da correção do viés na semivariância dos resíduos envolveu a construção do algoritmo com o esquema de estimação IRWGLS e posterior implementação da correção do viés, de acordo com a Equação (8).

O algoritmo com o esquema de estimação IRWGLS é composto de extensa sequência de comandos, e por isso, não será mostrado aqui, podendo ser visto, na íntegra, no Apêndice. Porém, o trecho do *script* com a correção de viés proposta pode ser visto na sequência de comandos abaixo, a qual inclui alguns comentários após o símbolo “#”, conforme a sintaxe da linguagem de programação usada. Os comentários foram inseridos para facilitar a compreensão dos significados dos comandos pelos usuários do *script*. Para implementar a matriz com as variâncias das tendências estimadas $Var[\tilde{\mu}(u)]$ e $Var[\tilde{\mu}(u+h)]$, presentes na

Equação (8) do estimador de semivariância $\hat{\gamma}_c(h)$ com correção de viés, foi utilizada a expressão matricial vista na Equação (9). A matriz com as covariâncias da tendência estimada $Cov[\tilde{\mu}(u), \tilde{\mu}(u+h)]$ foi implementada de acordo com a expressão matricial vista na Equação (10). O trecho do *script* é finalizado pelo símbolo “}”.

Inicialização de matrizes:

```
vartrend = matrix(0,n,1) # Matriz com a variância da tendência estimada
covtrend = matrix(0,n,n) # Matriz com a covariância da tendência estimada
Scorrigid = matrix(0,n,n) # Matriz de semivariâncias estimadas com correção de viés
correc = matrix(0,n,n) # Matriz com correção de viés a ser acrescida à semivariância
```

Aplicação da correção de viés proposta:

```
for(i in 1:n)
{
  for(j in 1:n)
  {
    vartrend[i] = Xt[,i] %*% Ai %*% X[i,] # Equivalente a covtrend[i,i]
    vartrend[j] = Xt[,j] %*% Ai %*% X[j,] # Equivalente a covtrend[j,j]
    covtrend[i,j] = Xt[,i] %*% Ai %*% X[j,]
    correc[i,j] = 0.5*(vartrend[i]+vartrend[j]-2*covtrend[i,j])
    Scorrigid[i,j] = S[i,j] + correc[i,j]
  }
}
```

Para maior compreensão desse trecho do *script*, deve-se ressaltar que a variável n armazena a quantidade de observações amostradas. A variável $S[i,j]$ armazena os valores das semivariâncias estimadas dos resíduos, resultantes do esquema de estimação IRWGLS. A variável $correc[i,j]$ armazena os valores referentes à correção de viés a ser aplicada às semivariâncias estimadas dos resíduos. $Scorrigid[i,j]$ armazena as semivariâncias estimadas dos resíduos com a correção de viés aplicada.

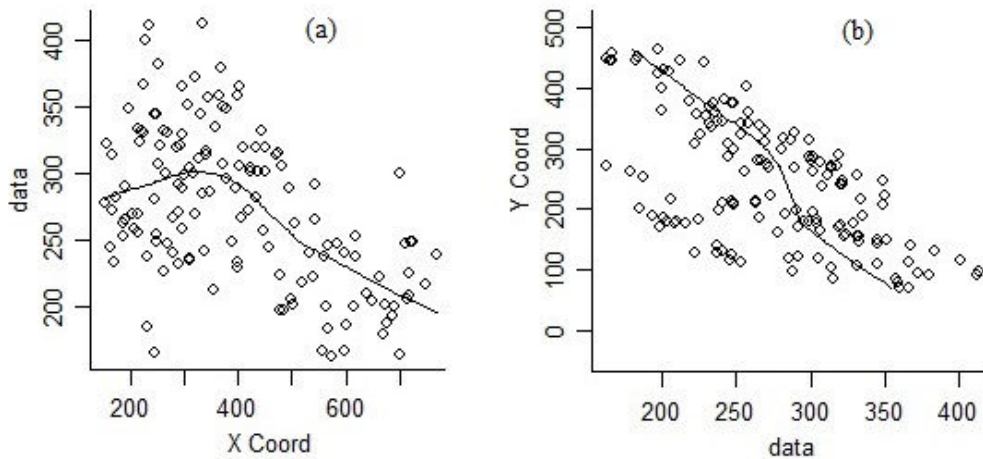
4.2 Aplicações do método proposto para a correção de viés na semivariância dos resíduos

Para avaliar o desempenho do método proposto em questão, foi feita uma análise com os três conjuntos de dados descritos.

Uma análise exploratória de dados preliminar foi realizada. Os gráficos de tendência, usados para a identificação da presença de tendência na análise preliminar, mostraram a existência de tendência nos três conjuntos de dados observados. A presença de tendência pode ser vista pelas linhas não retas nos gráficos de tendência, apresentados na Figura 8. Se não houvesse tendência, no gráfico da Figura 8a apareceria uma linha reta horizontal, paralela ao eixo das coordenadas X, e no gráfico da Figura 8b, uma linha reta vertical, paralela ao eixo das coordenadas Y.

Optou-se, então, por submeter os conjuntos de dados à modelagem de tendência de segunda ordem, pois essa modelagem, quando aplicada aos dados, se apresentou como a mais adequada, resultando em linhas horizontal e vertical, aproximadamente retas, e paralelas aos eixos das coordenadas X e Y, nos gráficos de tendência.

Figura 8 – Gráficos de tendência para o conjunto de dados Parana.



Fonte: Do autor (2017).

Os conjuntos de dados foram, inicialmente, submetidos à modelagem de tendência de segunda ordem por meio do critério dos mínimos quadrados lineares ordinários (MQLO), com o qual se supõe que a variação de uma variável em estudo envolve erros independentes, isto é, não correlacionados. Assim, a estimação da tendência $\mu(\cdot)$ desconhecida, por meio desse critério, envolve um estimador que não é BLUE. O critério MQLO foi aqui considerado por estar disponível para uso em pacotes de software para análise geoestatística, como, por exemplo, geoR (RIBEIRO JÚNIOR E DIGGLE, 2001) e gstat (PEBESMA, 2004).

Em seguida, a estimação da semivariância dos resíduos foi feita com o uso dos coeficientes de tendência estimados por meio do critério MQLO. A semivariância estimada foi, então, submetida ao ajuste com os modelos de semivariância exponencial, gaussiano e esférico, os quais são comumente usados na literatura (GOOVAERTS, 1997), para se averiguar qual modelo de semivariância se apresenta como o mais apropriado para descrever a variabilidade espacial do fenômeno, por meio do critério MQLO.

Um procedimento similar foi realizado, modelando a tendência de segunda ordem por meio do critério dos mínimos quadrados lineares generalizados com IRWGLS, doravante chamado critério IRWGLS, e dos mínimos quadrados lineares generalizados com IRWGLS com correção de viés, doravante chamado critério IRWGLS com correção de viés.

Posteriormente, foi feita a estimação da semivariância dos resíduos, considerando os modelos exponencial, gaussiano e esférico.

Convém ressaltar que a estimação da semivariância dos resíduos foi realizada utilizando o estimador clássico de Matheron. Na estimação foi usada a distância máxima no semivariograma correspondente a 70% da maior distância entre os pontos amostrados no domínio sob estudo. Essa distância máxima não está longe do limite sugerido por Journel e Huijbregts (1978), de cerca de 50% da maior distância. Para os critérios IRWGLS e IRWGLS com correção de viés, o critério de convergência usado no passo 5 do algoritmo IRWGLS implicou que duas matrizes $\hat{\Sigma}$, estimadas em iterações sucessivas, foram consideradas iguais se a diferença absoluta entre os seus elementos, em linhas e colunas respectivas, não foram maiores do que 10^{-3} .

No ajuste do modelo de semivariância foi usado sempre o mesmo critério: o de mínimos quadrados não-lineares ponderados proposto por Cressie (1985), que dá mais peso aos menores *lags* de distância no cálculo do estimador, nos quais há mais pares de dados envolvidos e o viés na semivariância dos resíduos é menor. O critério de mínimos quadrados não-lineares ponderados proposto por Cressie (1985) é definido por:

$$\sum_{k=1}^K \omega_k \left[\frac{\hat{\gamma}(h_k) - \gamma(h_k)}{\gamma(h_k)} \right]^2,$$

em que o peso ω_k , referente a cada *lag*, é proporcional ao número $N(h_k)$ de pares de dados que contribuem para a estimativa $\hat{\gamma}(h_k)$.

Além disso, no ajuste do modelo foi usado como função de otimização o método de Nelder-Mead, que é a função default no pacote geoR (RIBEIRO JÚNIOR E DIGGLE, 2001) e é considerado um método robusto (NASH, 2014).

Para avaliar e comparar os resultados obtidos nos três casos, isto é, usando o critério dos mínimos quadrados lineares ordinários, dos mínimos quadrados lineares generalizados com IRWGLS e dos mínimos quadrados lineares generalizados com IRWGLS com correção de viés, recorreu-se ao método de validação proposto por Bastos (2017), descrito na seção 3.2.

4.2.1 Aplicação do método proposto usando o conjunto de dados Parana

Para o conjunto de dados Parana, na Tabela 2 são vistos os resultados obtidos com o uso dos critérios MQLO, IRWGLS e IRWGLS com correção de viés, com ajuste do modelo de semivariância por meio do critério de mínimos quadrados não-lineares ponderados (MQNP) proposto por Cressie (1985), em relação aos parâmetros estimados de efeito pepita

(τ^2), contribuição (ϕ^2) e alcance (ϕ), para os modelos de semivariância exponencial, gaussiano e esférico. Também é apresentado, para comparação, o resultado obtido com o uso do critério IRWGLS com correção de viés, com ajuste do modelo de semivariância por meio do critério de mínimos quadrados não-lineares ordinários (MQNO), definido na Equação (4).

Pode ser observado um aumento a partir do valor do patamar estimado ($\tau^2 + \phi^2$) com MQLO, para os valores estimados com IRWGLS e IRWGLS com correção de viés, com ajuste por meio de MQNP. O aumento percentual do patamar estimado com MQLO para o estimado com IRWGLS com correção de viés, foi de 1,30% para o modelo de semivariância exponencial, de 4,42% para o modelo gaussiano e de 12,60% para o modelo esférico.

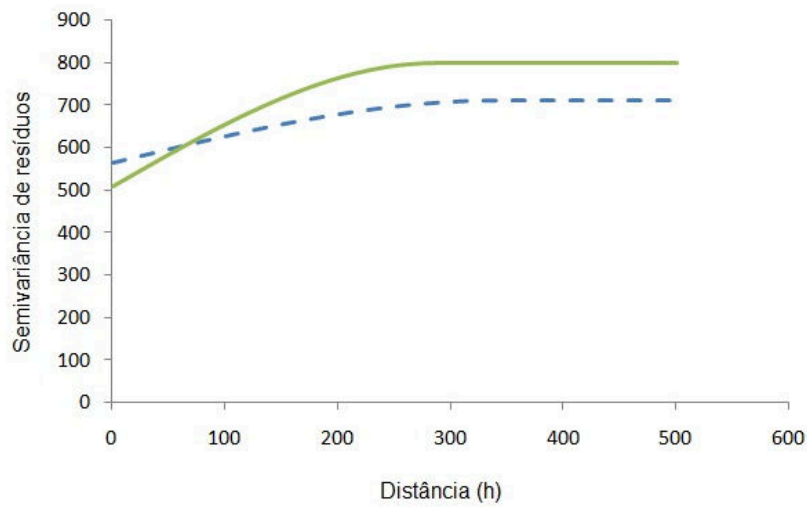
Tabela 2 – Estimativas dos parâmetros de pepita, contribuição e alcance, com MQLO, IRWGLS e IRWGLS com correção de viés, para os modelos exponencial, gaussiano e esférico, usando o conjunto de dados Parana.

Critério	Parâmetros estimados	Modelos		
		Exponencial	Gaussiano	Esférico
MQLO com ajuste MQNP	τ^2	555,8513	591,8259	565,3892
	ϕ^2	189,1160	123,7124	145,4311
	ϕ	197,7982	182,8532	341,6401
IRWGLS com ajuste MQNP	τ^2	557,0421	557,0421	546,1953
	ϕ^2	184,8839	184,8847	284,5314
	ϕ	125,8949	125,8953	152,3310
IRWGLS com correção de viés e ajuste MQNP	τ^2	485,4989	551,0034	507,8699
	ϕ^2	269,1246	196,1851	292,5068
	ϕ	88,0108	116,9815	285,7951
IRWGLS com correção de viés e ajuste MQNO	τ^2	519,9493	560,5268	535,4973
	ϕ^2	225,1460	169,0018	192,6851
	ϕ	108,1822	117,5969	236,4410

Fonte: Do autor (2017).

Na Figura 9 é apresentado o gráfico de semivariâncias para o modelo esférico, ajustado com o critério MQNP, cuja semivariância dos resíduos foi estimada com o critério MQLO (com linha tracejada), e o gráfico de semivariâncias para o modelo esférico, ajustado com o critério MQNP, cuja semivariância dos resíduos foi estimada com o critério IRWGLS com correção de viés (com linha contínua), que se destacou por apresentar o maior aumento percentual para o patamar, de 12,60%. Esse aumento é visível pela diferença na altura entre os patamares dos dois gráficos.

Figura 9 – Gráfico de semivariâncias para o modelo esférico ajustado com MQNP e MQLO (com linha tracejada), e com MQNP e IRWGLS com correção de viés (com linha contínua), usando o conjunto de dados Parana.



Fonte: Do autor (2018).

Para avaliar o efeito da correção de viés da Equação (8) sobre os valores de semivariâncias dos resíduos que foram estimados por meio do critério IRWGLS, procedeu-se ao cálculo da média dos valores dessas semivariâncias para os modelos ajustados com MQNP e IRWGLS $[\bar{\gamma}_{Res}(h)]$, e com MQNP e IRWGLS com correção de viés $[\bar{\gamma}_c(h)]$. Como o termo de correção para o referido viés deve retificar a subestimação presente na semivariância dos resíduos estimada com IRWGLS, é esperado um aumento nos valores das semivariâncias estimadas com IRWGLS com correção de viés. O aumento na média das semivariâncias estimadas dos resíduos foi de aproximadamente 3,21% para os modelos exponencial e gaussiano, e de cerca de 6,03% para o modelo esférico.

Na Tabela 3 têm-se as médias dos valores das semivariâncias estimadas dos resíduos para os modelos ajustados com os dois critérios em questão e com o critério MQLO $[\bar{\gamma}_{Res}^{MQLO}(h)]$, com o qual se ignora a correlação presente nos dados, mas que é considerado aqui para comparação com os outros dois critérios. Também pode ser visto o aumento percentual na média das semivariâncias estimadas dos resíduos com a aplicação da correção de viés da Equação (8). Cabe ressaltar que, se for considerada a diferença entre a média das semivariâncias estimadas com o critério IRWGLS com correção de viés e a média das semivariâncias estimadas com o critério MQLO, o aumento percentual é maior. Esse aumento, é de 5,84% para o modelo exponencial, de 7,28% para o modelo gaussiano e de 27,52% para o modelo esférico.

Tabela 3 – Médias das semivariâncias estimadas com MQLO, IRWGLS e IRWGLS com correção de viés, para os modelos ajustados com MQNP, e aumento percentual envolvido, para o conjunto de dados Parana.

Critério	Média das semivariâncias	Modelos		
		Exponencial	Gaussiano	Esférico
MQLO com ajuste MQNP	$\bar{\gamma}_{Res}^{MQLO}(h)$	714,9809	705,3892	675,9278
IRWGLS com ajuste MQNP	$\bar{\gamma}_{Res}(h)$	733,2371	733,2378	812,9555
IRWGLS com correção de viés e ajuste MQNP	$\bar{\gamma}_c(h)$	756,7544	756,7564	861,9362
Aumento percentual de $\bar{\gamma}_{Res}(h)$ para $\bar{\gamma}_c(h)$		3,21%	3,21%	6,03%

Fonte: Do autor (2018).

Os modelos ajustados com MQNP e MQLO, IRWGLS e IRWGLS com correção de viés, e com MQNO e IRWGLS com correção de viés, cujas estimativas dos parâmetros podem ser vistas na Tabela 2, foram, então, submetidos ao teste de validação. Os resultados são apresentados na Tabela 4. De acordo com o método de validação de Bastos (2017), para o caso do ajuste com MQNP, o modelo que apresentou a menor distância e, portanto, o melhor resultado, foi o modelo esférico ajustado com o uso do critério IRWGLS com correção de viés. É importante observar que os resultados para o critério IRWGLS com correção de viés e ajuste com MQNO apontam o modelo exponencial, e não o esférico, como o que teve a menor distância, logo, o melhor resultado, entre os modelos com ajuste MQNO, cujas estimativas dos parâmetros pepita, alcance e contribuição não foram mostradas na Tabela 2, mas que, também, foram submetidos ao teste de validação.

Tabela 4 – Resultados [em unidades de distância estabelecida pelo método de Bastos (2017)] do teste de validação para os modelos ajustados com MQLO, IRWGLS e IRWGLS com correção de viés (utilizando MQNP e MQNO), usando o conjunto de dados Parana.

Critério	Modelos		
	Exponencial	Gaussiano	Esférico
MQLO com ajuste MQNP	0,0756	0,1105	0,0994
IRWGLS com ajuste MQNP	0,0781	0,0781	0,0042
IRWGLS com correção de viés e ajuste MQNP	0,0471	0,0767	0,0009
IRWGLS com correção de viés e ajuste MQNO	0,0000	0,0417	0,0422

Fonte: Do autor (2017).

Pode-se ver, pelos resultados na Tabela 4, que, no caso do ajuste com MQNP, o uso do critério IRWGLS com correção de viés no modelo esférico fez com que a distância fosse reduzida a uma razão maior do que cem vezes (0,0009 equivale a 0,0900/100), quando comparado ao critério MQLO, com o qual se desconsidera que a variação da variável regionalizada envolve erros correlacionados. Isto indica que o método proposto produz um efeito não desprezível na correção do viés em questão, para o conjunto de dados Parana. Além disso, a distância zero observada para o modelo exponencial com IRWGLS com correção de viés e ajuste com MQNO sugere que este deve ser o modelo de semivariância escolhido para o fenômeno.

4.2.2 Aplicação do método proposto usando o conjunto de dados Meuse

Para o conjunto de dados Meuse, na Tabela 5 são apresentados os resultados obtidos com o uso dos critérios MQLO, IRWGLS e IRWGLS com correção de viés, os três com ajuste do modelo de semivariância com MQNP, em relação aos parâmetros estimados de pepita (τ^2), contribuição (ϕ^2) e alcance (ϕ), para os modelos exponencial, gaussiano e esférico. Também é visto o resultado obtido com o uso do critério IRWGLS com correção de viés, com ajuste do modelo de semivariância por meio do critério MQNO.

Houve um aumento a partir do valor do patamar estimado ($\tau^2 + \phi^2$) com MQLO, para os valores estimados com IRWGLS e IRWGLS com correção de viés, e ajuste com MQNP. O aumento percentual do patamar estimado com MQLO para o estimado com IRWGLS com

correção de viés, foi de 6,35% para o modelo exponencial, de 6,27% para o modelo gaussiano e de 6,25% para o esférico.

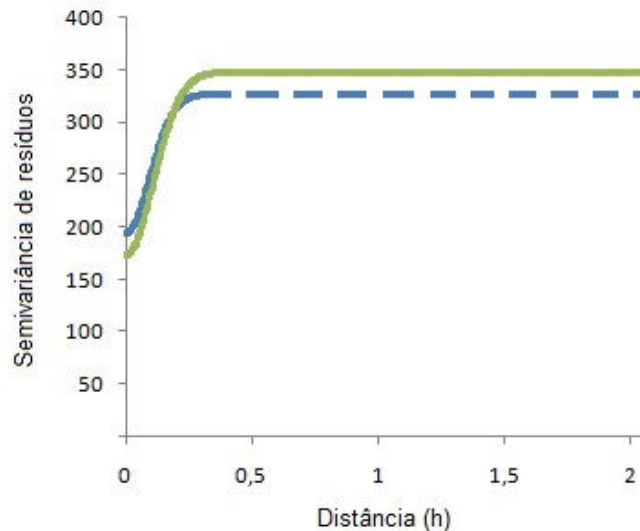
Tabela 5 – Estimativas dos parâmetros de pepita, contribuição e alcance, com MQLO, IRWGLS e IRWGLS com correção de viés, para os modelos exponencial, gaussiano e esférico, usando o conjunto de dados Meuse.

Critério	Parâmetros estimados	Modelos		
		Exponencial	Gaussiano	Esférico
MQLO com ajuste MQNP	τ^2	194,0966	247,4484	235,9223
	ϕ^2	133,5183	80,5415	92,3555
	ϕ	0,2295	0,4168	0,8898
IRWGLS com ajuste MQNP	τ^2	235,9550	236,1477	235,9523
	ϕ^2	112,8130	112,3937	112,8258
	ϕ	0,4562	0,4561	0,4562
IRWGLS com correção de viés e ajuste MQNP	τ^2	173,0878	236,1473	217,6144
	ϕ^2	175,3183	112,3967	131,1937
	ϕ	0,2657	0,456138	0,9441
IRWGLS com correção de viés e ajuste MQNO	τ^2	172,0718	235,1260	220,1692
	ϕ^2	163,6195	101,1721	116,0382
	ϕ	0,2240	0,4109	0,8691

Fonte: Do autor (2017).

Na Figura 10 é exibido o gráfico de semivariâncias para o modelo exponencial, ajustado com o critério MQNP e MQLO (com linha tracejada), e o gráfico de semivariâncias para o modelo exponencial, ajustado com o critério MQNP e IRWGLS com correção de viés (com linha contínua), que apresentou o maior aumento percentual para o patamar, de 6,35%. O aumento pode ser visto pela diferença na altura entre os patamares de ambos os gráficos.

Figura 10 – Gráfico de semivariâncias para o modelo exponencial ajustado com MQNP e MQLO (com linha tracejada) e com MQNP e IRWGLS com correção de viés (com linha contínua), usando o conjunto de dados Meuse.



Fonte: Do autor (2018).

Na avaliação do efeito da correção de viés proposta, verificou-se o aumento na média das semivariâncias para os modelos ajustados com MQNP. Na Tabela 6 são vistas as médias dos valores das semivariâncias estimadas dos resíduos para os modelos ajustados com os critérios considerados. Também pode ser visto o aumento percentual na média das semivariâncias estimadas dos resíduos com a aplicação da correção de viés da Equação (8). Esse aumento foi de aproximadamente 2,68% para os modelos exponencial e esférico, e de cerca de 2,67% para o modelo gaussiano. Se for considerada a diferença entre a média das semivariâncias estimadas com o critério IRWGLS com correção de viés e a média das semivariâncias estimadas com o critério MQLO, o aumento percentual é maior, de 9,02% para o modelo exponencial, de 8,78% para o modelo gaussiano e de 10,86% para o modelo esférico.

Tabela 6 – Médias das semivariâncias estimadas com MQLO, IRWGLS e IRWGLS com correção de viés, para os modelos ajustados com MQNP, e aumento percentual envolvido, para o conjunto de dados Meuse.

Critério	Média das semivariâncias	Modelos		
		Exponencial	Gaussiano	Esférico
MQLO com ajuste MQNP	$\bar{\gamma}_{Res}^{MQLO}(h)$	325,5599	326,0421	320,1719
IRWGLS com ajuste MQNP	$\bar{\gamma}_{Res}(h)$	345,6476	345,4339	345,6573
IRWGLS com correção de viés e ajuste MQNP	$\bar{\gamma}_c(h)$	354,9275	354,6700	354,9360
Aumento percentual de $\bar{\gamma}_{Res}(h)$ para $\bar{\gamma}_c(h)$		2,68%	2,67%	2,68%

Fonte: Do autor (2018).

Os resultados do teste de validação, ao qual foram submetidos os modelos, cujas estimativas dos parâmetros foram apresentadas na Tabela 5, podem ser vistos na Tabela 7. De acordo com o método de validação de Bastos (2017), o modelo que apresentou a menor distância e, portanto, o melhor resultado, foi o modelo exponencial ajustado com o uso do critério IRWGLS com correção de viés. A distância desse modelo em relação ao ponto (0,0) foi igual a zero, o que significa que ele teve o melhor desempenho nos quatro métodos que compõem o método de Bastos (2017), quando comparado com os modelos concorrentes. Isso ocorreu tanto para o caso do ajuste com MQNP como para o caso com MQNO. Cabe ressaltar que as estimativas dos parâmetros pepita, alcance e contribuição, para os modelos ajustados com MQNO, que também foram submetidos ao teste de validação, não foram mostradas na Tabela 5.

Tabela 7 – Resultados [em unidades de distância estabelecida pelo método de Bastos (2017)] do teste de validação para os modelos ajustados com MQLO, IRWGLS e IRWGLS com correção de viés, usando o conjunto de dados Meuse.

Critério	Modelos		
	Exponencial	Gaussiano	Esférico
MQLO com ajuste MQNP	0,0818	0,1044	0,1007
IRWGLS com ajuste MQNP	0,0473	0,0478	0,0473
IRWGLS com correção de viés e ajuste MQNP	0,0000	0,0478	0,0312
IRWGLS com correção de viés e ajuste MQNO	0,0000	0,0330	0,0276

Fonte: Do autor (2017).

Os resultados na Tabela 7 mostram que, para o ajuste com MQNP, o uso do critério IRWGLS com correção de viés no modelo exponencial fez com que a distância tivesse uma redução “infinita”, quando comparado ao critério MQLO. A distância passou de 0,0818, com o critério MQLO, para 0,0000, com o critério IRWGLS com correção de viés. Resultado similar ocorreu para o ajuste com MQNO, cujos valores de distâncias dos demais modelos não foram exibidos na Tabela 7. Essa redução na distância indica que a proposta metodológica contribuiu para uma melhoria não desprezível na correção do viés em questão, para o conjunto de dados considerado. Mais uma vez, a distância zero mantida para o modelo exponencial no critério IRWGLS com correção de viés e ajuste com MQNO sugere que este deve ser o modelo escolhido.

4.2.3 Aplicação do método proposto usando o conjunto de dados Castanha

Para o conjunto de dados Castanha, na Tabela 8 são exibidos os resultados obtidos com os critérios MQLO, IRWGLS e IRWGLS com correção de viés, e ajuste do modelo de semivariância com MQNP, em relação aos parâmetros estimados de pepita (τ^2), contribuição (ϕ^2) e alcance (ϕ), para os modelos gaussiano e esférico. O resultado obtido com o uso do critério IRWGLS com correção de viés, e ajuste do modelo de semivariância por meio do critério MQNO, também é mostrado.

Para o modelo exponencial, o ajuste com MQNP resultou em efeito pepita puro, o que significa que esse modelo não conseguiu captar a dependência espacial entre os dados observados na escala considerada, e por isso, esse modelo não será considerado. Para o

modelo esférico, houve um aumento a partir do valor do patamar estimado ($\tau^2 + \phi^2$) com MQLO, para os valores estimados com IRWGLS e IRWGLS com correção de viés, e ajuste com MQNP. O aumento percentual do patamar estimado com MQLO para o estimado com IRWGLS com correção de viés, foi de 0,18%. Para o modelo gaussiano houve uma diminuição percentual do patamar estimado com MQLO para o estimado com IRWGLS com correção de viés, de 0,01%.

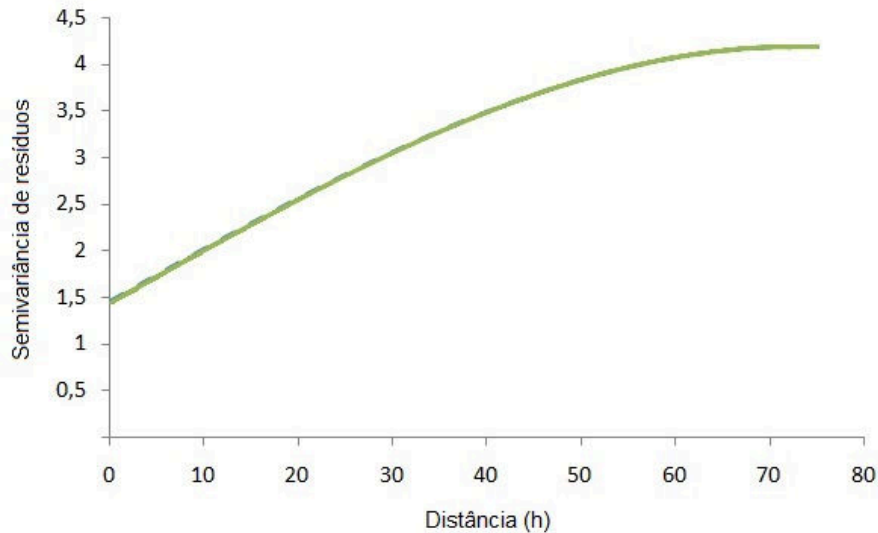
Tabela 8 – Estimativas dos parâmetros de pepita, contribuição e alcance, com MQLO, IRWGLS e IRWGLS com correção de viés, para os modelos gaussiano e esférico, usando o conjunto de dados Castanha.

Critério	Parâmetros estimados	Modelos	
		Gaussiano	Esférico
MQLO com ajuste MQNP	τ^2	1,6656	1,4519
	ϕ^2	2,5103	2,7312
	ϕ	32,9488	72,2814
IRWGLS com ajuste MQNP	τ^2	1,6655	1,4338
	ϕ^2	2,5098	2,7571
	ϕ	32,9562	72,3457
IRWGLS com correção de viés e ajuste MQNP	τ^2	1,6655	1,4337
	ϕ^2	2,5098	2,7572
	ϕ	32,9562	72,3458
IRWGLS com correção de viés e ajuste MQNO	τ^2	1,4648	1,7745
	ϕ^2	2,5291	2,02427
	ϕ	29,0797	70,8958

Fonte: Do autor (2017).

Na Figura 11 pode ser visto o gráfico de semivariâncias para o modelo esférico, ajustado com o critério MQNP e MQLO (com linha tracejada), e o gráfico de semivariâncias para o modelo esférico, ajustado com o critério MQNP e IRWGLS com correção de viés (com linha contínua), que apresentou aumento percentual para o patamar de 0,18%. O aumento é pouco perceptível porque, praticamente, não houve diferença de altura entre os patamares, na escala usada nos gráficos.

Figura 11 – Gráfico de semivariâncias para o modelo esférico ajustado com MQNP e MQLO (com linha tracejada) e com MQNP e IRWGLS com correção de viés (com linha contínua), usando o conjunto de dados Castanha.



Fonte: Do autor (2018).

O efeito da correção de viés proposta foi avaliado pelo aumento na média das semivariâncias. Na Tabela 9 são vistas as médias dos valores das semivariâncias estimadas dos resíduos para os modelos ajustados com MQNP. Também pode ser visto o aumento percentual na média das semivariâncias estimadas dos resíduos com a aplicação da correção de viés da Equação (8). Esse aumento foi de aproximadamente 5,54% para o modelo gaussiano e de cerca de 9,96% para o modelo esférico. Ao se considerar a diferença entre a média das semivariâncias estimadas com o critério IRWGLS com correção de viés e a das semivariâncias estimadas com o critério MQLO, tem-se um aumento percentual de 5,53%, para o modelo gaussiano, e de 10,15%, para o modelo esférico.

Tabela 9 – Médias das semivariâncias estimadas com MQLO, IRWGLS e IRWGLS com correção de viés, para os modelos ajustados com MQNP, e aumento percentual envolvido, para o conjunto de dados Castanha.

Critério	Média das semivariâncias	Modelos	
		Gaussiano	Esférico
MQLO com ajuste MQNP	$\bar{\gamma}_{Res}^{MQLO}(h)$	4,1293	4,0839
IRWGLS com ajuste MQNP	$\bar{\gamma}_{Res}(h)$	4,1287	4,0907
IRWGLS com correção de viés e ajuste MQNP	$\bar{\gamma}_c(h)$	4,3576	4,4982
Aumento percentual de $\bar{\gamma}_{Res}(h)$ para $\bar{\gamma}_c(h)$		5,54%	9,96%

Fonte: Do autor (2018).

Os resultados do teste de validação para os modelos, cujas estimativas dos parâmetros foram apresentadas na Tabela 8, podem ser vistos na Tabela 10. De acordo com o método de validação de Bastos (2017), o modelo que apresentou a menor distância e, portanto, o melhor resultado, foi o modelo esférico ajustado por meio do critério IRWGLS com correção de viés. A distância desse modelo foi igual a zero, tanto no ajuste com MQNP, como no ajuste com MQNO, o que significa que ele teve o melhor desempenho em todos os métodos que compõem o método de Bastos (2017), quando comparado com os demais modelos. Observe que o valor da distância para o modelo esférico ajustado por meio do critério IRWGLS foi apresentado com mais do que quatro casas decimais, fugindo da convenção adotada nesta tese, a qual, se aplicada, geraria o valor 0,0000, o que levaria à confusão com o zero absoluto do modelo esférico ajustado por meio do critério IRWGLS com correção de viés. Convém salientar que as estimativas dos parâmetros pepita, alcance e contribuição, para os modelos ajustados com MQNO, também submetidos ao teste de validação, não foram mostradas na Tabela 8.

Tabela 10 – Resultados [em unidades de distância estabelecida pelo método de Bastos (2017)] do teste de validação para os modelos ajustados com MQLO, IRWGLS e IRWGLS com correção de viés, usando o conjunto de dados Castanha.

Critério	Modelos	
	Gaussiano	Esférico
MQLO com ajuste MQNP	0,0486	0,0046
IRWGLS com ajuste MQNP	0,0484	$1,5275 \times 10^{-5}$
IRWGLS com correção de viés e ajuste MQNP	0,0484	0,0000
IRWGLS com correção de viés e ajuste MQNO	0,0737	0,0000

Fonte: Do autor (2017).

Pode-se ver, pelos resultados na Tabela 10, que, no caso do ajuste com MQNP, o uso do critério IRWGLS com correção de viés no modelo esférico ocasionou uma redução “infinita” na distância, quando comparado ao critério MQLO. A distância passou de 0,0046, com o critério MQLO, para 0,0000, com o critério IRWGLS com correção de viés. Resultado semelhante ocorreu para o ajuste com MQNO, cujos valores de distâncias dos demais modelos não foram exibidos na Tabela 10. Essa redução na distância indica que houve uma melhoria não desprezível com a aplicação da metodologia proposta, para o conjunto de dados usado. O modelo esférico deve ser o escolhido por causa da manutenção da distância zero no método IRWGLS com correção de viés, tanto no ajuste com MQNP quanto no ajuste com MQNO.

4.2.4 Mapas de krigagem e de erro de krigagem dos conjuntos de dados

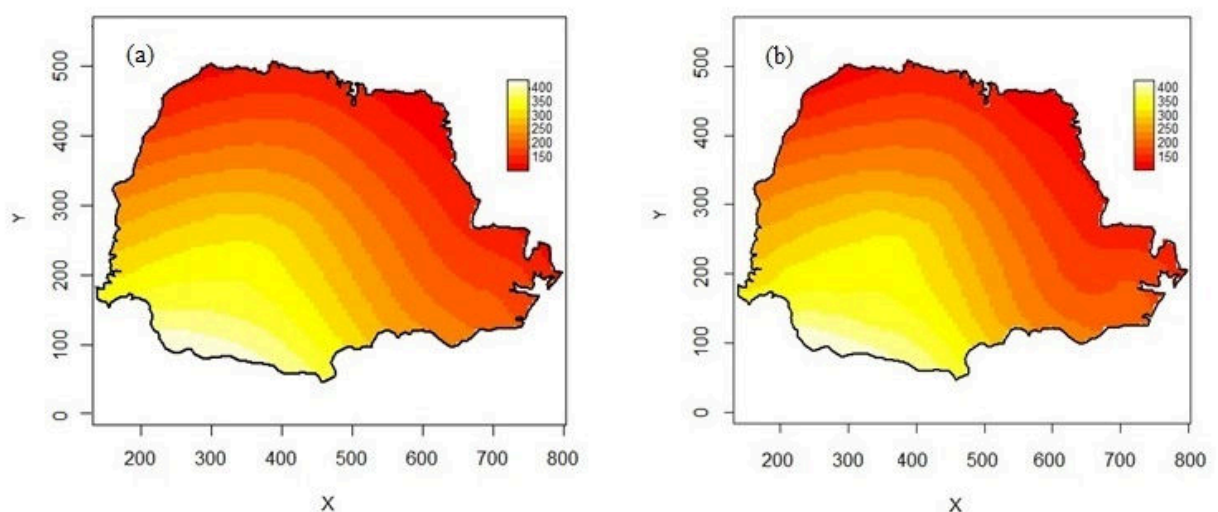
Para os conjuntos de dados usados, foram implementados mapas de krigagem ou de predição e mapas de erro de krigagem. Por causa de restrições na implementação de funcionalidade no pacote geoR (RIBEIRO JÚNIOR E DIGGLE, 2001), a krigagem foi realizada com vizinhança global, isto é, considerando todos os dados amostrados, vizinhos ao ponto que se quer interpolar. Para uma inspeção visual do impacto do método proposto na krigagem, um modelo de semivariância, ajustado por meio de dois critérios na estimação da semivariância dos resíduos, IRWGLS com correção de viés e MQLO, foi considerado na predição. Assim, foram produzidos dois mapas de krigagem e dois mapas de erro de

krigagem, para cada conjunto de dados. Um mapa de krigagem e um mapa de erro de krigagem foram feitos com o modelo de semivariância ajustado com o critério MQNP e IRWGLS com correção de viés, que foi tido como o mais adequado para descrever a variabilidade espacial do fenômeno, por meio do método de validação (poderia ser utilizado tanto o ajuste com MQNP como o com MQNO, nesta tese foi escolhido o ajuste com MQNP). Um segundo mapa de krigagem e um segundo mapa de erro de krigagem foram feitos com o mesmo modelo de semivariância, porém ajustado com o critério MQNP e MQLO.

Para o conjunto de dados Parana, na Figura 12a é apresentado o mapa de krigagem, implementado com o modelo de semivariância esférico, ajustado com o critério MQLO, e na Figura 12b, o mapa de krigagem com o modelo esférico, ajustado com o critério IRWGLS com correção de viés.

O mapa de krigagem na Figura 12a apresenta, em geral, curvas de nível um pouco mais suaves do que o mapa da Figura 12b, indicando uma menor variabilidade na predição dos dados na krigagem realizada com o modelo de semivariância ajustado com o uso do critério MQLO. Isto está de acordo com a teoria, que diz que o viés na semivariância dos resíduos afeta a variância estimada de krigagem, de modo que esta será menor do que a real variância de krigagem (CRESSIE, 1993; KIM E BOOS, 2004). O mapa da Figura 12b caracteriza, então, de maneira mais fidedigna, a variabilidade espacial do fenômeno, pois o uso do critério IRWGLS com correção de viés retifica a variância estimada de krigagem, tornando-a maior, mais próxima da verdadeira variância de krigagem.

Figura 12 – Mapas de krigagem do conjunto de dados Parana, com modelo de semivariância ajustado (a) com MQLO e (b) com IRWGLS com correção de viés.

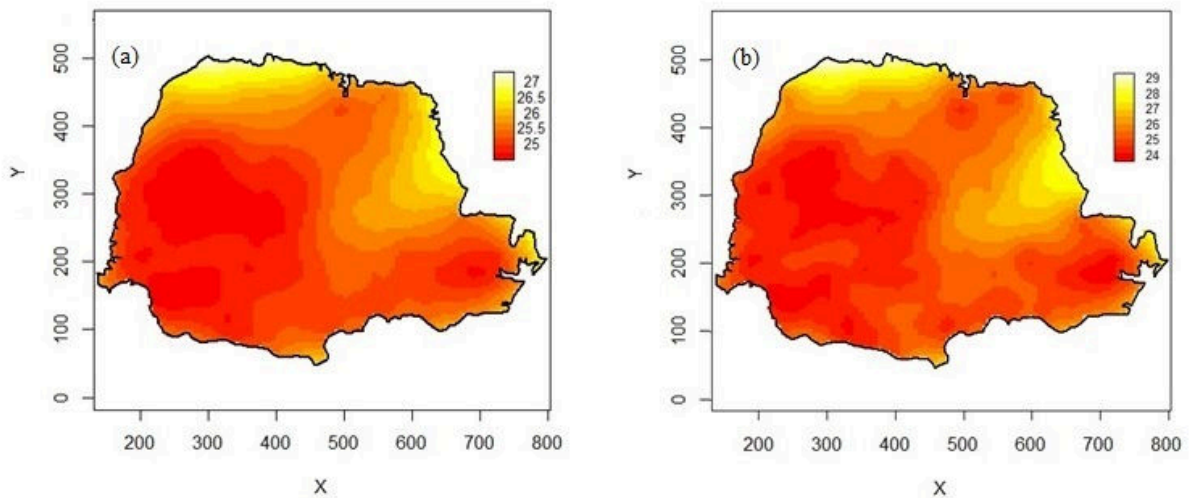


Fonte: Do autor (2018).

Para o mesmo conjunto de dados, na Figura 13a é apresentado o mapa de erro de krigagem, implementado com o modelo de semivariância esférico, ajustado com o critério MQLO, e na Figura 13b, o mapa de erro de krigagem com o modelo esférico, ajustado com o critério IRWGLS com correção de viés.

O mapa de erro de krigagem na Figura 13a exibe aspecto menos entrecortado e com valores menores dos desvios padrões de krigagem, do que o mapa da Figura 13b, indicando um menor erro de krigagem para a predição realizada com o modelo de semivariância ajustado com o uso do critério MQLO. Isto confirma o que a teoria diz acerca de que o viés na semivariância dos resíduos afeta a variância estimada de krigagem, fazendo com que esta variância seja menor do que deveria. Assim, o mapa da Figura 13b atesta que o uso do critério IRWGLS com correção de viés retifica a variância estimada de krigagem, tornando-a maior, mais próxima da real variância de krigagem.

Figura 13 – Mapas de erro de krigagem do conjunto de dados Parana, com modelo de semivariância ajustado (a) com MQLO e (b) com IRWGLS com correção de viés.



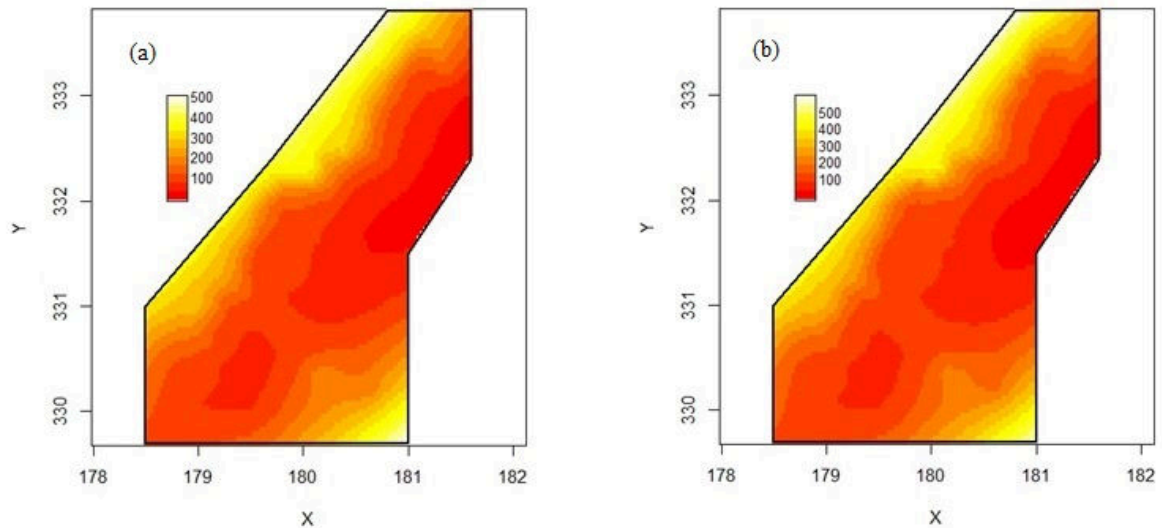
Fonte: Do autor (2018).

Para o conjunto de dados Meuse, na Figura 14a é exibido o mapa de krigagem, implementado com o modelo de semivariância exponencial, ajustado com o critério MQLO, e na Figura 14b, o mapa de krigagem com o modelo exponencial, ajustado com o critério IRWGLS com correção de viés.

No mapa de krigagem na Figura 14a podem ser vistas, em alguns locais, curvas de nível um pouco mais suaves do que no mapa da Figura 14b. Tem-se, então, uma menor variabilidade na predição dos dados na krigagem realizada com o modelo de semivariância, ajustado com o uso do critério MQLO. O uso do critério IRWGLS com correção de viés

retifica a variância estimada de krigagem, elevando o seu valor, para que fique mais próxima da variância de krigagem real. Assim, o mapa na Figura 14b, implementado com o uso desse segundo critério, caracteriza mais adequadamente a variabilidade espacial.

Figura 14 – Mapas de krigagem do conjunto de dados Meuse, com modelo de semivariância ajustado (a) com MQLO e (b) com IRWGLS com correção de viés.

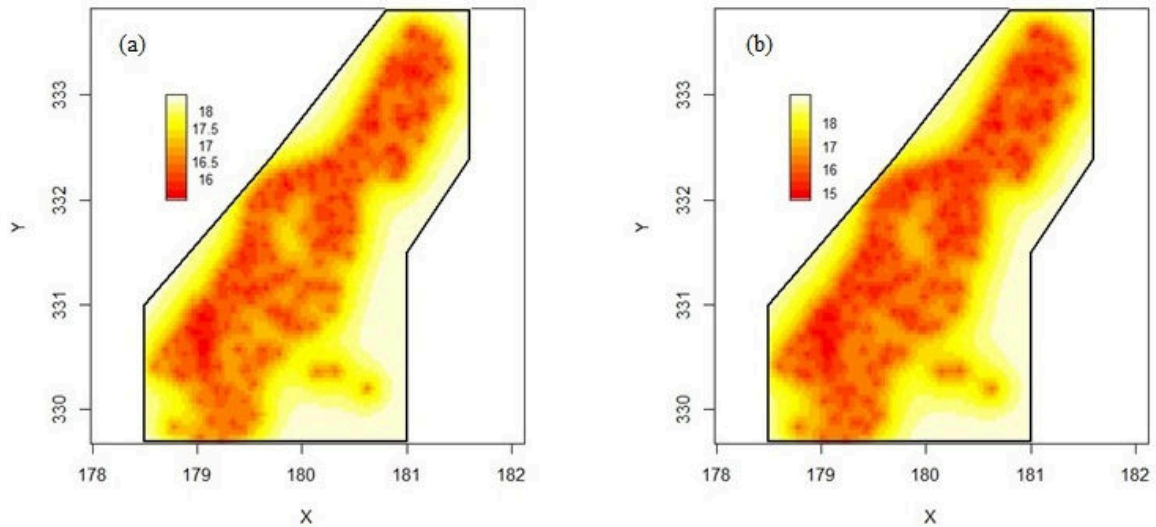


Fonte: Do autor (2018).

Para o conjunto de dados Meuse, na Figura 15a é exibido o mapa de erro de krigagem, implementado com o modelo de semivariância exponencial, ajustado com o critério MQLO, e na Figura 15b, o mapa de erro de krigagem com o modelo exponencial, ajustado com o critério IRWGLS com correção de viés.

O mapa na Figura 15a apresenta, em geral, menores valores de desvio padrão, do que o mapa da Figura 15b, o que indica um menor erro de krigagem para a predição realizada com o modelo de semivariância ajustado com o uso do critério MQLO. Como o uso do critério IRWGLS com correção de viés retifica a variância estimada de krigagem, elevando o seu valor, o mapa da Figura 15b exibe um retrato mais fidedigno do erro de krigagem envolvido.

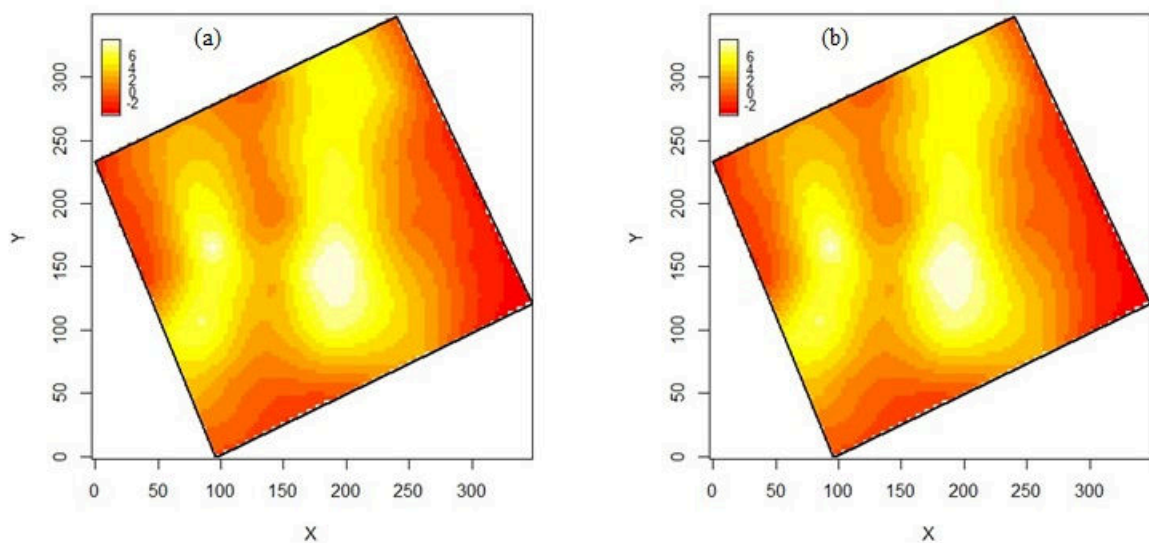
Figura 15 – Mapas de erro de krigagem do conjunto de dados Meuse, com modelo de semivariância ajustado (a) com MQLO e (b) com IRWGLS com correção de viés.



Fonte: Do autor (2018).

Na Figura 16a é apresentado o mapa de krigagem, correspondente ao conjunto de dados Castanha, implementado com o modelo de semivariância esférico, ajustado com o critério MQLO. Na Figura 16b, é visto o mapa de krigagem com o mesmo modelo de semivariância, mas ajustado com o critério IRWGLS com correção de viés. Ambos os mapas apresentam curvas de níveis parecidas, com algumas diferenças pontuais, que ocorrem por causa da maior variância estimada de krigagem com o uso do critério IRWGLS com correção de viés.

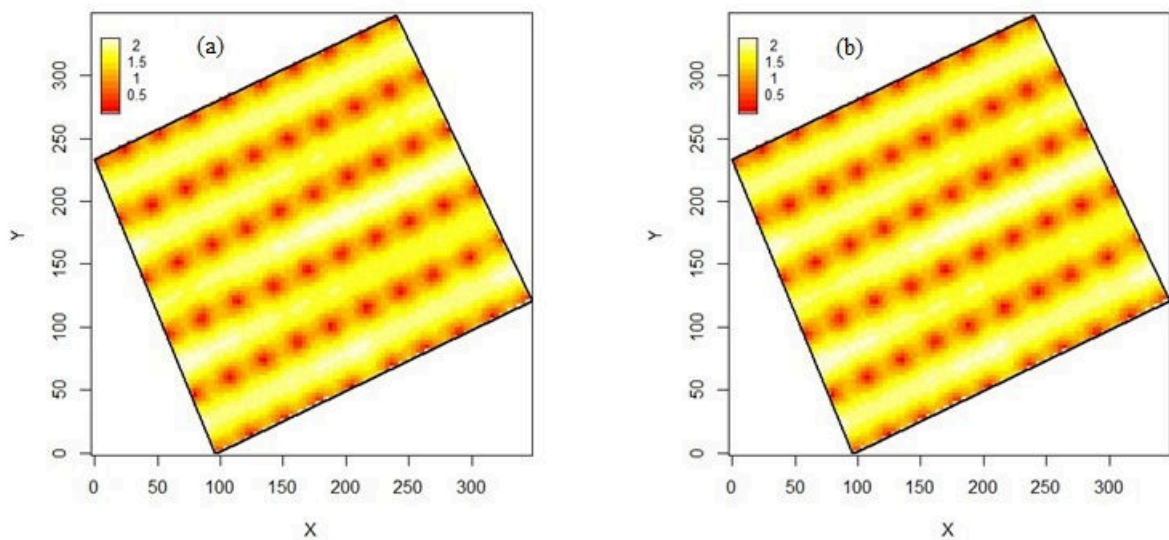
Figura 16 – Mapas de krigagem do conjunto de dados Castanha, com modelo de semivariância ajustado (a) com MQLO e (b) com IRWGLS com correção de viés.



Fonte: Do autor (2018).

Para o conjunto de dados Castanha, na Figura 17a é apresentado o mapa de erro de krigagem, implementado com o modelo de semivariância esférico, ajustado com o critério MQLO, e na Figura 17b, o mapa de erro de krigagem com o mesmo modelo de semivariância, porém ajustado com o critério IRWGLS com correção de viés. Os mapas são muito parecidos, apresentando diferenças somente em alguns pontos, nos quais o mapa na Figura 17b mostra um maior erro de krigagem. Isto atesta que o critério IRWGLS com correção de viés está retificando a variância estimada de krigagem, aumentando-a, a fim de que seu valor fique mais próximo da verdadeira variância de krigagem.

Figura 17 – Mapas de erro de krigagem do conjunto de dados Castanha, com modelo de semivariância ajustado (a) com MQLO e (b) com IRWGLS com correção de viés.



Fonte: Do autor (2018).

4.2.5 Ponderações sobre as aplicações do método proposto

Ao observar as Tabelas 4, 7 e 10, referentes aos resultados do teste de validação para os conjuntos de dados Parana, Meuse e Castanha, pôde-se perceber uma sequência efetiva de melhoria crescente nos resultados de cada modelo de semivariância a partir do critério MQLO para o critério IRWGLS com correção de viés. Isto é, as distâncias vão diminuindo até chegarem a valores mínimos, indicando melhores resultados, no critério IRWGLS com correção de viés. Essa melhoria nos resultados é coerente com a teoria na qual está baseado o método proposto.

Nas Tabelas 3, 6 e 9 observou-se um aumento nos valores das médias das semivariâncias estimadas dos resíduos com o uso do critério IRWGLS com correção de viés, em comparação com o critério IRWGLS. Esse aumento é um dos objetivos do método

proposto para a correção do viés visto na Equação (7), na seção 2.6, pois esse viés implica na subestimação dos valores das semivariâncias dos resíduos.

Para o conjunto de dados Parana, verificou-se, a partir dos resultados da Tabela 2, que o aumento percentual no valor do patamar estimado ($\tau^2 + \phi^2$), após a aplicação da correção de viés com o método proposto, foi maior para o modelo esférico ajustado por meio do critério IRWGLS com correção de viés, um aumento de 12,60%. Na Tabela 3, pode ser visto que o aumento percentual no valor médio das semivariâncias estimadas dos resíduos, após a aplicação da correção de viés com o método proposto, também foi maior para o modelo esférico ajustado por meio do critério IRWGLS com correção de viés, um aumento de 6,03%. Esse modelo foi considerado o mais adequado para descrever a estrutura da variabilidade espacial do fenômeno, por meio do método de validação de Bastos (2017), como foi visto pelo menor valor de distância do modelo, em comparação com os valores de distâncias dos demais modelos, nos resultados na Tabela 4.

Também, para o conjunto de dados Castanha, como se verificou pelos resultados da Tabela 8 e como pode ser visto na Tabela 9, respectivamente, o aumento percentual no valor do patamar estimado e na média das semivariâncias estimadas dos resíduos, após a aplicação da correção de viés com o método proposto, foi maior para o modelo de semivariância que foi considerado o mais adequado para descrever a variabilidade espacial do fenômeno, por meio do método de validação usado. O aumento no valor do patamar foi de 0,18% e na média das semivariâncias foi de 9,96%. O modelo mais adequado foi o modelo esférico ajustado por meio do critério IRWGLS com correção de viés, como foi indicado pelo seu menor valor de distância, em comparação com os valores de distâncias dos modelos concorrentes, nos resultados na Tabela 10.

Assim como ocorreu para os conjuntos de dados Parana e Castanha, para o conjunto de dados Meuse, como foi verificado a partir dos resultados da Tabela 5 e como pode ser visto na Tabela 6, respectivamente, o aumento percentual no valor do patamar estimado e na média das semivariâncias estimadas dos resíduos, após a aplicação da correção de viés com o método proposto, foi maior para o modelo de semivariância que foi considerado o mais adequado para descrever a variabilidade espacial, por meio do método de validação. Esse modelo foi o modelo exponencial ajustado por meio do critério IRWGLS com correção de viés, como foi visto pelo seu menor valor de distância, em comparação com os valores dos demais modelos, na Tabela 7. O aumento no valor do patamar foi de 6,35%. Cabe ressaltar que o aumento percentual na média das semivariâncias para esse modelo, de 2,68%, foi igual ao aumento percentual do modelo esférico ajustado por meio do critério IRWGLS com

correção de viés, o qual foi o segundo melhor modelo, com o segundo menor valor de distância.

Pelo que foi visto, para os conjuntos de dados considerados, o maior valor no aumento percentual no patamar e na média das semivariâncias estimadas dos resíduos, após a aplicação da correção de viés com o método proposto, se mostra sempre ligado ao modelo de semivariância que foi considerado o mais adequado para descrever a estrutura da variabilidade espacial do fenômeno. Pode-se concluir que o modelo de semivariância que melhor capta a dependência espacial nos dados amostrais tende a ser mais sensível à correção de viés por meio do método proposto, isto é, o método proposto potencializa a correção de viés, expressada pelo aumento percentual no valor do patamar e da média das semivariâncias estimadas dos resíduos, no modelo de semivariância que pode ser considerado como o mais correto para descrever a estrutura da variabilidade espacial do fenômeno.

Em suma, com base nas aplicações e considerando as ponderações referentes às melhorias nos pontos destacados nos parágrafos anteriores desta subseção, pode-se dizer que o método proposto nesta tese para a correção de viés na semivariância dos resíduos mostrou desempenho satisfatório ao avançar na qualidade da estimação da semivariância na presença de tendência em, pelo menos, três aspectos:

1. o método proposto, conforme o método de validação usado (BASTOS, 2017), diminuiu as distâncias para os três conjuntos de dados, isto é, o teste de validação confirma que o método proposto nesta tese melhora a qualidade da estimação da semivariância;
2. o método proposto aumenta os valores das semivariâncias estimadas, o que indica que ele caminha na direção do que a teoria prevê, que é a correção da subestimação;
3. o método proposto privilegia o modelo de semivariância tido como o mais adequado para descrever a estrutura da variabilidade espacial do fenômeno.

Em função desses três aspectos, pode-se recomendar sua utilização na análise geoestatística.

4.2.6 Considerações finais

O método proposto nesta tese, cabe ressaltar, apresenta a vantagem de não estar limitado a qualquer pressuposição adicional, além das pressuposições padrões da Geoestatística. Nesse aspecto, ele se iguala aos métodos já existentes que não exigem pressuposições.

Uma característica do método proposto é que ele pode ser adaptado às necessidades do usuário, no que tange à modelagem da semivariância e implementação de mapas de krigagem e de erro de krigagem.

Com relação à modelagem da semivariância, o usuário pode ajustar um modelo específico de semivariância (por exemplo, o exponencial), executando, apenas, o trecho do *script* relacionado a esse modelo. O restante do *script*, que está relacionado com os outros modelos, pode ser desconsiderado na execução do programa ou pode ser executado posteriormente.

Com relação à implementação de mapas de krigagem e de erro de krigagem, o modelo de semivariância, com o viés corrigido pelo método proposto, pode ser empregado na predição de valores com o uso de vizinhança local, diferentemente do que se fez nesta tese, na qual se utilizou a vizinhança global. Para isso, basta usar um comando existente ou implementar um novo comando na linguagem do ambiente do *software* R, que permita escolher o número de vizinhos a serem considerados na krigagem.

É importante destacar que a implementação do método proposto foi feita com o uso do recurso de ajuste “a sentimento” do pacote *geoR*, na modelagem da semivariância, como etapa inicial, que ocorre no passo 3 do algoritmo do esquema de estimação IRWGLS. Isto exige a intervenção ativa e contínua do usuário no processo de estimação dos parâmetros de efeito pepita, contribuição e alcance. Essa intervenção é repetida a cada iteração no esquema de estimação, imediatamente antes do ajuste automático do modelo de semivariância por meio do critério de mínimos quadrados não-lineares ponderados, proposto por Cressie (1985).

Vislumbra-se, então, como trabalho futuro, o desenvolvimento de um ajuste automático do modelo de semivariância, que dispense a intervenção do usuário, na etapa inicial da modelagem, e que resulte num processo de estimação dos parâmetros mais eficiente e mais rápido.

5 CONCLUSÃO

A análise com conjuntos de dados de diferentes tamanhos, sendo um destes conjuntos, de tamanho pequeno, serviu para demonstrar que o uso do método proposto para a correção do viés na estimação da semivariância na presença de tendência, denominado critério IRWGLS com correção de viés, produziu uma melhoria efetiva na qualidade da estimação dos parâmetros do modelo de semivariância, como foi indicado pelo método de validação usado. Em todos os três conjuntos de dados utilizados observou-se uma diminuição da distância do teste de validação, sendo que, em dois dos três conjuntos de dados esta distância tornou-se zero. Essa melhoria está de acordo com a teoria na qual se fundamentou o desenvolvimento do método proposto.

Um dos objetivos do método proposto é o aumento no valor das semivariâncias estimadas dos resíduos, a fim de retificar a subestimação deste valor, a qual ocorre quando é feita a estimação por meio do critério dos mínimos quadrados generalizados. Esse objetivo foi alcançado para todos os conjuntos de dados usados, como atestam os valores das médias das semivariâncias estimadas dos resíduos para os modelos ajustados antes e depois da correção de viés com o método proposto. O aumento percentual nos valores das médias dessas semivariâncias foi de, no mínimo, 2,67%, chegando até 9,96%, conforme o conjunto de dados usado.

Verificou-se que o maior valor para o aumento percentual no patamar e na média das semivariâncias estimadas dos resíduos com o uso do método proposto está sempre ligado ao modelo de semivariância considerado, por meio do método de validação usado, como o mais adequado para captar a dependência espacial dos dados amostrados. Isso indica que o uso do método proposto potencializa a correção de viés, expressada pelo aumento percentual nos valores do patamar e da média das semivariâncias estimadas dos resíduos, no modelo de semivariância que pode ser considerado como o mais correto para descrever a estrutura da variabilidade espacial do fenômeno.

Cabe destacar que o método proposto tem a vantagem de não estar limitado a qualquer pressuposição adicional, além das pressuposições padrões da Geoestatística, sendo, portanto, de validade geral.

Finalmente, os argumentos desenvolvidos nesta tese demonstram que o método proposto para a correção do viés na estimação da semivariância na presença de tendência se constitui numa importante contribuição para o avanço na qualidade da estimação da

semivariância, da qual depende a qualidade de todos os produtos da análise geoestatística, tais como: mapa de krigagem, simulações, entre outros.

Como a etapa de ajuste do modelo de semivariância tem um papel fundamental e determinante para o sucesso das etapas posteriores da análise geoestatística, convém ressaltar que, na escolha deste modelo, a experiência de um especialista com conhecimento do aspecto físico do fenômeno aliada à experiência de um estatístico que saiba interpretar esse conhecimento, em função das pressuposições da modelagem espacial, é de importância vital. É neste contexto que o método proposto nesta tese é recomendável e pode servir como uma relevante ferramenta auxiliar para que o objetivo da análise geoestatística seja alcançado com mais acurácia e precisão, isto é, a implementação de um mapa de krigagem que caracterize, de modo fidedigno, a variabilidade espacial inerente ao fenômeno em estudo.

REFERÊNCIAS

- BASTOS, R. L. **Métodos para seleção de modelos de semivariograma em campos aleatórios gaussianos**. 2017. 136 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, 2017.
- BATISTA, A.P.B. Comunicação pessoal, 2017.
- BECKERS, F.; BOGAERT, P. Nonstationarity of the mean and unbiased variogram estimation: extension of the weighted least-squares method. **Mathematical Geology**, v. 30, n. 2, p. 223-240, 1998. Disponível em: <<https://link.springer.com/article/10.1023/A:1021780731393>>. Acesso em: 03 nov. 2017.
- BLIZNYUK, N.; CARROLL, R.J.; GENTON, M.G.; WANG, Y. Variogram estimation in the presence of trend. **Statistics and Its Interface**, v. 5, n. 2, p. 159-168, 2012. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3378336/>>. Acesso em: 03 nov. 2017.
- BURROUGH, P.A. **Principles of geographical information systems for land resources assessment**. Oxford: Clarendon, 1987. 193 p.
- CHILÈS, J.P.; DELFINER, P. **Geostatistics: modeling spatial uncertainty**. 2nd ed. New York: J. Wiley, 2012. 734 p.
- CRESSIE, N.A.C. Fitting variogram models using weighted least squares. **Journal of the Mathematical Association for Mathematical Geology**, v. 17, p. 563-586, 1985.
- CRESSIE, N.A.C. **Statistics for spatial data**. 2nd ed. New York: J. Wiley, 1993. 900 p.
- CRESSIE, N.; HAWKINS, D.M. Robust estimation of the variogram: I. **Journal of the Mathematical Association for Mathematical Geology**, v. 2, n. 2, p. 115-125, 1980.
- DESASSIS, N.; RENARD, D. **Experimental variogram of the residuals in the universal kriging (UK) model**. Fontainebleau: MINES Paris Tech, 2014. Disponível em: <http://cg.ensmp.fr/bibliotheque/public/DESASSIS_Rapport_02305.pdf>. Acesso em: 03 nov. 2017.
- GOOVAERTS, P. **Geostatistics for natural resources evaluation**. New York: Oxford University, 1997. 483 p.
- ISAAKS, E.H.; SRIVASTAVA, R.M. **An introduction to applied geostatistics**. New York: Oxford University, 1989. 561 p.
- JOURNEL, A.G. **Fundamentals of geostatistics in five lessons**. Washington D.C.: American Geophysical Union, 1989. 40 p.
- JOURNEL, A.G. **Geostatistics for the environmental sciences: an introduction**. Las Vegas: Environmental Systems Laboratory, 1987. 134 p.

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

KIM,H.; BOOS, D.D. Variance estimation in spatial regression using a non-parametric semivariogram based on residuals. **Scandinavian Journal of Statistics**, v. 31, p. 387-401, 2004. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9469.2004.02-025.x/epdf>>. Acesso em: 03 nov. 2017.

KRIGE, D.G. **A statistical approach to some mine valuation and allied problems on the Witwatersrand**. 1951. 136f. Thesis (M.Sc. in Engineering) – Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg. 1951.

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

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

MATHERON, G. The theory of regionalized variables and its applications. **Les Cahiers du Centre de Morphologie Mathématique**, n. 5. Fontanebleau: École Nationale Supérieure des Mines de Paris, 1971. Disponível em: <http://cg.ensmp.fr/bibliotheque/public/MATHERON_Ouvrage_00167.pdf>. Acesso em: 03 nov. 2017.

NASH, J.C. On best practice optimization methods in R. **Journal of Statistical Software**, v. 60, n. 2, p. 1-14, 2014. Disponível em: <<https://www.jstatsoft.org/v060/i02>>. Acesso em: 17 abr. 2017.

NEUMAN, S.P.; JACOBSON, E.A. Analysis of nonintrinsic spatial variability by residual kriging with application to regional groundwater levels. **Journal of the International Association for Mathematical Geology**, v. 16, n. 5, p. 499-521, 1984. Disponível em: <<https://link.springer.com/article/10.1007/BF01886329>>. Acesso em: 03 nov. 2017.

OLIVEIRA, M.S. **Planos amostrais para variáveis espaciais utilizando geoestatística**. 1991. 100 p. Dissertação (Mestrado em Estatística) – Instituto de Matemática, Estatística e Ciência da Computação, Universidade de Campinas, Campinas, 1991.

PEBESMA, E.J. Multivariable geostatistics in S: the gstat package. **Computers & Geosciences**, v. 30, n. 7, p. 683-691, 2004.

PELLETIER, B.; DUTILLEUL, P.; LAROCQUE, G.; FYLES, J.W. Corregionalization analysis with a drift for multi-scale assessment of spatial relationships between ecological variables 1. Estimation of drift and random components. **Environmental and Ecological Statistics**, v. 16, p. 439-466, 2009. Disponível em: <<https://link.springer.com/article/10.1007/s10651-008-0090-z>>. Acesso em: 03 nov. 2017.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna: R foundation for Statistical Computing, 2017. Disponível em: <<https://www.r-project.org/>>. Acesso em: 19 dez. 2017.

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

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

SHERMAN, M. **Spatial statistics and spatio-temporal data: covariance functions and directional properties**. Chichester: J. Wiley, 2011. 268 p.

TOBLER, W.R. A computer movie simulating urban growth in the Detroit region. **Economic Geography**, v. 46, p. 234-240, 1970.

VIEIRA, S.R. Geoestatística em estudos de variabilidade do solo. In: NOVAIS, R.F.; ALVAREZ, V.; V.H. & SCHAEFFER, G.R., eds. **Tópicos em ciência do solo**. Viçosa, Sociedade Brasileira de Ciência do Solo, v. 1, p. 1-54, 2000.

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

APÊNDICE – *Script* na linguagem de programação do R

```
#####
# Algoritmo escrito na linguagem de programação do software R,
# por meio do qual foi implementada a proposta metodológica desta tese
#####

# Para limpar a memória:
rm(list=ls())

# Mudando para o diretório no qual estão os arquivos a serem usados:
setwd ("diretório_em_que_se_encontra_o_arquivo_de_dados")

# Carregando o pacote geoR, o qual deve ter sido instalado previamente:
library(geoR)

# Transformando arquivo_de_dados em objeto do tipo geodata do pacote geoR
# Observação: substitua arquivo_de_dados pelo nome e extensão do arquivo que
# contém os dados amostrais e respectivas coordenadas de localização
Dados<-read.geodata("arquivo_de_dados", head=T, coords.col = 1:2,data.col = 3 )

#####
# Modelagem da tendência por meio do critério dos mínimos quadrados lineares ordinários:
#####

Dadosres <- Dados

# Modelagem da tendência de grau 2 pela função lm
x <- Dadosres$coords[,1]
y <- Dadosres$coords[,2]
x2 <- Dadosres$coords[,1]*Dadosres$coords[,1]
y2 <- Dadosres$coords[,2]*Dadosres$coords[,2]
xy <- Dadosres$coords[,1]*Dadosres$coords[,2]
trend <- lm(Dadosres$data~x+y+x2+y2+xy)
beta0 <- trend$coefficients[1]
beta1 <- trend$coefficients[2]
beta2 <- trend$coefficients[3]
beta3 <- trend$coefficients[4]
beta4 <- trend$coefficients[5]
beta5 <- trend$coefficients[6]

Dadosres$data <- Dadosres$data-beta0+beta1*x+beta2*y+beta3*x2+beta4*y2+beta5*xy)
plot(Dadosres, low=T)
```

```
#####
# Estimação da semivariância dos resíduos usando 70% da maior distância entre os
# pontos:
#####
hmax70 <- summary(Dadosres)[[3]][[2]]*0.7
semivarres70 <- variog(Dadosres, estimator="classical", max.dist=hmax70)
plot(semivarres70, xlab="Distância (h)", ylab="Semivariância de resíduos", col="blue",
pch=16,main="Matheron (70% da maior distância entre pontos)")

#####
# Validação inicial dos modelos de semivariância ajustados com os coeficientes de
# tendência resultantes do uso do critério de mínimos quadrados lineares
# ordinários (MQLO).
# Variáveis resultantes da validação inicial dos modelos serão submetidas ao método
# de validação de Bastos (2017).
#####

#####
# Modelo exponencial
#####
# Ajuste a sentimento inicial com modelo exponencial:
semivarresas1=eyefit(semivarres70)
contrib1=semivarresas1[[1]]$cov.pars[1] # Contribuição
alcance1=semivarresas1[[1]]$cov.pars[2] # Alcance
pepita1=semivarresas1[[1]]$nugget      # Pepita

# Ajuste do modelo de semivariância por meio do método dos mínimos quadrados
# não-lineares ponderados:
var.wls1=variofit(semivarres70,cov.model = "exponential",ini=c(contrib1,alcance1),
nugget=pepita1,weights="cressie",minimisation.function="optim")
plot(semivarres70, xlab="Distância (h)", ylab="Semivariância")
lines(var.wls1,lwd=1,lty=1)

# Autovalidação:
xv.semivarresas1=xvalid(Dados,model=var.wls1)
summary(xv.semivarresas1)

# Variável a ser usada dentro do método de validação de Bastos (2017):
vo <- xv.semivarresas1$data
vpexp1 <- xv.semivarresas1$predicted

# Parametros do modelo de semivariância ajustado:
C0_exp<-var.wls1$nugget      # Pepita
C1_exp<-var.wls1$cov.pars[1] # Contribuição
a_exp<-var.wls1$cov.pars[2] # Alcance
C_exp<-C0_exp+C1_exp       # Patamar
```



```
#####
# Modelo gaussiano
#####
# Ajuste a sentimento inicial com modelo gaussiano:
semivarresas2=eyefit(semivarres70)
contrib2=semivarresas2[[1]]$cov.pars[1]
alcance2=semivarresas2[[1]]$cov.pars[2]
pepita2=semivarresas2[[1]]$nugget

# Ajuste do modelo de semivariância por meio do método dos mínimos quadrados
# não-lineares ponderados:
var.wls2=variofit(semivarres70,cov.model = "gaussian", ini=c(contrib2,alcance2),
nugget=pepita2,weights="cressie",minimisation.function="optim")
plot(semivarres70, xlab="Distância (h)", ylab="Semivariância")
lines(var.wls2,lwd=1,lty=1)

# Autovalidação:
xv.semivarresas2=xvalid(Dados,model=var.wls2)
summary(xv.semivarresas2)

# Variável a ser usada dentro do método de validação de Bastos (2017):
vpgaul <- xv.semivarresas2$predicted

# Parametros do modelo de semivariância ajustado:
C0_gau<-var.wls2$nugget
C1_gau<-var.wls2$cov.pars[1]
a_gau<-var.wls2$cov.pars[2]
C_gau<-C0_gau+C1_gau

#####
# Modelo esférico
#####
# Ajuste a sentimento inicial com modelo esférico:
semivarresas3=eyefit(semivarres70)
contrib3=semivarresas3[[1]]$cov.pars[1]
alcance3=semivarresas3[[1]]$cov.pars[2]
pepita3=semivarresas3[[1]]$nugget

# Ajuste do modelo de semivariância por meio do método dos mínimos quadrados
# não-lineares ponderados:
var.wls3=variofit(semivarres70,cov.model = "sph", ini=c(contrib3,alcance3),
nugget=pepita3, weights="cressie",minimisation.function="optim")
plot(semivarres70, xlab="Distância (h)", ylab="Semivariância")
lines(var.wls3,lwd=1,lty=1)

# Autovalidação:
xv.semivarresas3=xvalid(Dados,model=var.wls3)
summary(xv.semivarresas3)
```

```

# Variável a ser usada dentro do método de validação de Bastos (2017):
vpesf1 <- xv.semivarresas3$predicted

# Parametros do modelo de semivariância ajustado:
C0_esf<-var.wls3$nugget
C1_esf<-var.wls3$cov.pars[1]
a_esf<-var.wls3$cov.pars[2]
C_esf<-C0_esf+C1_esf

#####
# O usuário deve escolher o modelo de semivariância ajustado por meio do critério MQLO
# a ser submetido ao critério dos mínimos quadrados generalizados por meio do
# algoritmo iterativo de Mínimos Quadrados Generalizados Iterativamente
# Reponderados (IRWGLS), antes de digitar uma das próximas três sequências de
# comandos a seguir.
# Observação: o método de validação de Bastos (2017) só pode ser usado depois que todos os
# modelos: exponencial, gaussiano e esférico, forem submetidos aos critérios IRWGLS,
# e IRWGLS com correção de viés, um modelo de cada vez.
#####

# Se o modelo a ser submetido, desta vez, ao critério IRWGLS é o exponencial, digite o
# comando a seguir:
semivarreswls= var.wls1

# Se o modelo a ser submetido, desta vez, ao critério IRWGLS é o gaussiano, digite o
# comando a seguir:
semivarreswls= var.wls2

# Se o modelo a ser submetido, desta vez, ao critério IRWGLS é o esférico, digite o
# comando a seguir:
semivarreswls= var.wls3

#####
# Estimativas dos parâmetros do modelo escolhido ajustado por meio do critério MQLO a ser
# submetido ao critério dos mínimos quadrados lineares generalizados com IRWGLS:
#####

if (semivarreswls$cov.model == "exponential"){
  pepita_model_MQLO = C0_exp
  contrib_model_MQLO = C1_exp
  alcance_model_MQLO = a_exp
  covariancia_model_MQLO = semivarreswls$cov.model
}

if (semivarreswls$cov.model == "gaussian"){
  pepita_model_MQLO = C0_gau
  contrib_model_MQLO = C1_gau
  alcance_model_MQLO = a_gau
  covariancia_model_MQLO = semivarreswls$cov.model
}

```

```

}

if (semivarreswls$cov.model == "spherical"){
  pepita_model_MQLO = C0_esf
  contrib_model_MQLO = C1_esf
  alcance_model_MQLO = a_esf
  covariancia_model_MQLO = semivarreswls$cov.model
}

#####
# Modelagem da tendência por meio do critério dos mínimos quadrados lineares
# generalizados com IRWGLS e ajuste do modelo de semivariância com os coeficientes
# de tendência resultantes do uso do critério IRWGLS:
#####

# Tamanho do conjunto de dados:
n = summary(Dados)[[1]]

# Inicialização da matriz de distâncias e covariâncias:
H = matrix(0,n,n)
D = matrix(0,n,n)

# Matriz de distâncias:
for(i in 1:n)
{
  for(j in 1:n)
  {
    H[i,j] = sqrt(((x[i]-x[j])^2)+((y[i]-y[j])^2))
  }
}

D = H

# Inicialização da matriz de semivariâncias:
S = matrix(0,n,n)

# Estimativas dos parâmetros do modelo de semivariância ajustado que está sendo submetido
# ao método IRWGLS:
pepita_model = semivarreswls$nugget           # Pepita
contrib_model = semivarreswls$cov.pars[1]    # Contribuição
alcance_model = semivarreswls$cov.pars[2]    # Alcance
covariancia_model = semivarreswls$cov.model # Modelo de covariância

if (covariancia_model == "exponential"){
  for(i in 1:n)
  {
    for(j in 1:n)
    {
      # Modelo exponencial
      if (D[i,j]<alcance_model) {

```

```

        S[i,j] = pepita_model + contrib_model*(1-exp(-3*((D[i,j]/alcance_model)^2)))
    } else {
        S[i,j] = pepita_model + contrib_model
    }
}
}
S_exp = matrix(0,n,n)
S_exp = S # Matriz de semivariâncias estimadas com MQLO para modelo exponencial
S_exp_Barra = mean(S_exp) # Média das semivariâncias para o modelo exponencial
} # end-if

if(covariancia_model == "gaussian"){
  for(i in 1:n)
  {
    for(j in 1:n)
    {
      # Modelo gaussiano
      if (D[i,j]<alcance_model) {
        S[i,j] = pepita_model + contrib_model*(1-exp(-3*(D[i,j]/alcance_model)))
      } else {
        S[i,j] = pepita_model + contrib_model
      }
    }
  }
}
S_gau = matrix(0,n,n)
S_gau = S # Matriz de semivariâncias estimadas com MQLO para modelo gaussiano
S_gau_Barra = mean(S_gau) # Média das semivariâncias para o modelo gaussiano
} # end-if

if(covariancia_model == "spherical"){
  for(i in 1:n)
  {
    for(j in 1:n)
    {
      # Modelo esférico
      if (D[i,j]<alcance_model) {
        S[i,j] = pepita_model + contrib_model*((1.5*(D[i,j]/alcance_model))-
          0.5*((D[i,j]/alcance_model)^3))
      } else {
        S[i,j] = pepita_model + contrib_model
      }
    }
  }
}
S_esf = matrix(0,n,n)
S_esf = S # Matriz de semivariâncias estimadas com MQLO para modelo esférico
S_esf_Barra = mean(S_esf) # Média das semivariâncias para o modelo esférico
} # end-if

```

```

# Patamar:
varres = pepita_model + contrib_model

# Verificando se há semivariância dos resíduos > patamar:
if (max(S) > varres) varres = max(S)

# Atualização da matriz de covariâncias:
for(i in 1:n)
{
  for(j in 1:n)
  {
    H[i,j] = varres-S[i,j]
  }
}

# Inicialização da matriz de coordenadas (polinômio de grau 2):
X = matrix(0,n,6)

for(i in 1:n)
{
  X[i,1] = 1
  X[i,2] = Dadosres$coords[i,1] # Coordenadas x
  X[i,3] = Dadosres$coords[i,2] # Coordenadas y
  X[i,4] = Dadosres$coords[i,1]*Dadosres$coords[i,1] # x2
  X[i,5] = Dadosres$coords[i,2]*Dadosres$coords[i,2] # y2
  X[i,6] = Dadosres$coords[i,1]*Dadosres$coords[i,2] # xy
}

# Xt recebe a transposta de X:
Xt = t(X)

# Inicialização da matriz de observações:
Z = matrix(0,n,1)

obs = as.matrix(Dados$data)

# Matriz Z recebe valores das observações amostrais Z:
Z = obs

repete = TRUE

# A cada iteração a matriz de covariâncias é atualizada, até que haja convergência,
# no caso, uma diferença absoluta menor ou igual a 0.001 nos elementos correspondentes
# de matrizes calculadas em iterações sucessivas

# Inicialização de matrizes:
C = matrix(0,n,n)
A = matrix(0,6,6)
B = matrix(0,6,1)

```

```
##### Início da iteração IRWGLS #####

while (repete == TRUE)
{
  # C recebe a inversa da matriz H de covariâncias com tolerância na detecção de
  # dependência linear nas colunas da matriz:
  C = solve(H, tol = 1e-30)

  A = Xt %*% C %*% X      # Multiplicação matricial
  Ai = solve(A, tol = 1e-30) # Ai recebe a inversa da matriz A
  B = Ai %*% Xt %*% C %*% Z # Matriz B dos coeficientes estimados da tendência

  # Atualização dos coeficientes beta estimados da tendência:
  beta0 <- B[1]
  beta1 <- B[2]
  beta2 <- B[3]
  beta3 <- B[4]
  beta4 <- B[5]
  beta5 <- B[6]

  # Dados para resíduos permanecem os mesmos a cada iteração,
  # os coeficientes estimados da tendência é que são atualizados:
  Dadosres <- Dados

  # Resíduos atualizados:
  Dadosres$data <- Dadosres$data-(beta0+beta1*x+beta2*y+beta3*x2+beta4*y2+beta5*xy)

#####
# Estimação da semivariância dos resíduos usando 70% da maior distância entre
# os pontos:
#####
semivarres70 <- variog(Dadosres, estimator="classical", max.dist=hmax70)
plot(semivarres70, xlab="Distância (h)", ylab="Semivariância de resíduos", col="blue",
     pch=16,main="Matheron (70% da maior distância entre pontos)")
semivarresas <- eyefit(semivarres70)
covres <- semivarresas[[1]]$cov.model
contribuições <- semivarresas[[1]]$cov.pars[1]
alcanceres <- semivarresas[[1]]$cov.pars[2]
pepitares <- semivarresas[[1]]$nugget
semivarreswls <- variofit(semivarres70, cov.model = covres,
ini=c(contribuições,alcanceres), nugget=pepitares, weights="cressie")
plot(semivarres70, xlab="Distância (h)", ylab="Semivariância", col='blue', pch=16)
lines(semivarreswls, lwd=1, lty=1)

# Atualização das estimativas dos parâmetros do modelo de semivariância:
pepita_model = semivarreswls$nugget
contrib_model = semivarreswls$cov.pars[1]
alcance_model = semivarreswls$cov.pars[2]
covariancia_model = semivarreswls$cov.model
```

```

if (covariancia_model == "exponential"){
  for(i in 1:n)
  {
    for(j in 1:n)
    {
      # Modelo exponencial
      if (D[i,j]<alcance_model) {
        S[i,j] = pepita_model + contrib_model*(1-exp(-3*((D[i,j]/alcance_model)^2)))
      } else {
        S[i,j] = pepita_model + contrib_model
      }
    }
  }
} # end-if

```

```

if (covariancia_model == "gaussian"){
  for(i in 1:n)
  {
    for(j in 1:n)
    {
      # Modelo gaussiano
      if (D[i,j]<alcance_model) {
        S[i,j] = pepita_model + contrib_model*(1-exp(-3*(D[i,j]/alcance_model)))
      } else {
        S[i,j] = pepita_model + contrib_model
      }
    }
  }
} # end-if

```

```

if (covariancia_model == "spherical"){
  for(i in 1:n)
  {
    for(j in 1:n)
    {
      # Modelo esférico
      if (D[i,j]<alcance_model) {
        S[i,j] = pepita_model + contrib_model*((1.5*(D[i,j]/alcance_model))-
          0.5*((D[i,j]/alcance_model)^3))
      } else {
        S[i,j] = pepita_model + contrib_model
      }
    }
  }
} # end-if

```

```

# H_anterior recebe os valores da matriz de covariâncias:
H_anterior = H

```

```

# Patamar dos resíduos:
varres = pepita_model + contrib_model

if (max(S) > varres) varres = max(S)

# Atualização da matriz de covariâncias:
for(i in 1:n)
{
  for(j in 1:n)
  {
    H[i,j] = varres-S[i,j]
  }
}

# Diferença entre as matrizes de covariâncias de duas iterações sucessivas:
Dif = abs(H_anterior - H)

# Verificação da convergência das matrizes de covariâncias:
for(i in 1:n)
{
  for(j in 1:n)
  {
    # Se houver alguma diferença > 0.001 então repete iteração
    if (any(Dif[i,j]>0.001)) repete = TRUE else repete = FALSE
  }
}

} # end-while

##### Fim da iteração IRWGLS #####

# Parametros do modelo de semivariância ajustado com o uso do critério IRWGLS:
pepita_model_IRWGLS = semivarreswls$nugget          # Pepita
contrib_model_IRWGLS = semivarreswls$cov.pars[1]   # Contribuição
alcance_model_IRWGLS = semivarreswls$cov.pars[2]   # Alcance
covariancia_model_IRWGLS = semivarreswls$cov.model # Modelo de covariância
patamar_model_IRWGLS = pepita_model_IRWGLS + contrib_model_IRWGLS

# Atribuição de valores da matriz de semivariâncias estimadas por meio do
# critério IRWGLS, de acordo com o modelo ajustado por meio do critério MQLO:

if (covariancia_model_MQLO == "exponential"){
  S_exp2 = matrix(0,n,n)
  S_exp2 = S # Matriz de semivariâncias para modelo exponencial
  S_exp2_Barra = mean(S_exp2) # Média das semivariâncias para o modelo exponencial
}

```



```

if (covariancia_model_MQLO == "gaussian"){
  S_gau2 = matrix(0,n,n)
  S_gau2 = S # Matriz de semivariâncias para modelo gaussiano
  S_gau2_Barra = mean(S_gau2) # Média das semivariâncias para o modelo gaussiano
}

if (covariancia_model_MQLO == "spherical"){
  S_esf2 = matrix(0,n,n)
  S_esf2 = S # Matriz de semivariâncias para modelo esférico
  S_esf2_Barra = mean(S_esf2) # Média das semivariâncias para o modelo esférico
}

#####
# Validação inicial do modelo de semivariância ajustado com os coeficientes de
# tendência resultantes do uso do critério IRWGLS, de acordo com o modelo
# ajustado por meio do critério MQLO.
# Variáveis resultantes da validação inicial do modelo serão submetidas ao método de
# validação de Bastos (2017).
#####

if (covariancia_model_MQLO == "exponential"){
  #####
  # Modelo exponencial
  #####
  # Autovalidação:
  xv.semivarresas4=xvalid(Dados,model=semivarreswls)

  # Variável a ser usada dentro do método de validação de Bastos (2017):
  vpxp2 <- xv.semivarresas4$predicted

  # Parametros do modelo de semivariância ajustado:
  C0_exp2<-semivarreswls$nugget
  C1_exp2<-semivarreswls$cov.pars[1]
  a_exp2<-semivarreswls$cov.pars[2]
  C_exp2<-C0_exp2+C1_exp2
}

if (covariancia_model_MQLO == "gaussian"){
  #####
  # Modelo gaussiano
  #####
  # Autovalidação:
  xv.semivarresas5=xvalid(Dados,model=semivarreswls)

  # Variável a ser usada dentro do método de validação de Bastos (2017):
  vpgau2 <- xv.semivarresas5$predicted

```

```

# Parametros do modelo de semivariância ajustado:
C0_gau2<-semivarreswls$nugget
C1_gau2<-semivarreswls$cov.pars[1]
a_gau2<-semivarreswls$cov.pars[2]
C_gau2<-C0_gau2+C1_gau2
}

if (covariância_model_MQLO == "spherical"){
#####
# Modelo esférico
#####
# Autovalidação:
xv.semivarresas6=xvalid(Dados,model=semivarreswls)

# Variável a ser usada dentro do método de validação de Bastos (2017):
vpsf2 <- xv.semivarresas6$predicted

# Parametros do modelo de semivariância ajustado:
C0_esf2<-semivarreswls$nugget
C1_esf2<-semivarreswls$cov.pars[1]
a_esf2<-semivarreswls$cov.pars[2]
C_esf2<-C0_esf2+C1_esf2
}

#####
# Estimativas dos parâmetros do modelo ajustado por meio do critério IRWGLS,
# de acordo com o modelo ajustado por meio do critério MQLO:
#####

if (covariância_model_MQLO == "exponential"){
#####
# Modelo exponencial
#####
pepita_model_IRWGLS = C0_exp2
contrib_model_IRWGLS = C1_exp2
alcance_model_IRWGLS = a_exp2
S = S_exp2
}

if (covariância_model_MQLO == "gaussian"){
#####
# Modelo gaussiano
#####
pepita_model_IRWGLS = C0_gau2
contrib_model_IRWGLS = C1_gau2
alcance_model_IRWGLS = a_gau2
S = S_gau2
}

```

```

if (covariância_model_MQLO == "spherical"){
#####
# Modelo esférico
#####
pepita_model_IRWGLS = C0_esf2
contrib_model_IRWGLS = C1_esf2
alcance_model_IRWGLS = a_esf2
S = S_esf2
}

#####
# Semivariância dos resíduos com correção de viés
#####

# Inicialização de matrizes:
vartrend = matrix(0,n,1) # Matriz com a variância da tendência estimada
covtrend = matrix(0,n,n) # Matriz com a covariância da tendência estimada
Scorrigid = matrix(0,n,n) # Matriz de semivariâncias estimadas com correção de viés
correc = matrix(0,n,n) # Matriz com correção de viés a ser acrescida à semivariância

# Aplicação da correção de viés proposta:
for(i in 1:n)
{
  for(j in 1:n)
  {
    vartrend[i] = Xt[,i] %*% Ai %*% X[i,] # Equivalente a covtrend[i,i]
    vartrend[j] = Xt[,j] %*% Ai %*% X[j,] # Equivalente a covtrend[j,j]
    covtrend[i,j] = Xt[,i] %*% Ai %*% X[j,]
    correc[i,j] = 0.5*(vartrend[i]+vartrend[j]-2*covtrend[i,j])
    Scorrigid[i,j] = S[i,j] + correc[i,j]
  }
}

# Média das semivariâncias estimadas com correção de viés:
ScorrigidBarra = mean(Scorrigid)

# Verificando se há semivariância dos resíduos > Patamar dos resíduos:
if (max(Scorrigid) > varres) varres = max(Scorrigid)

# Atualização da matriz de covariâncias:
for(i in 1:n)
{
  for(j in 1:n)
  {
    H[i,j] = varres-Scorrigid[i,j]
  }
}
}

```

```
#####
# Modelagem da tendência por meio do critério dos mínimos quadrados lineares
# generalizados com IRWGLS com correção de viés e ajuste do modelo de
# semivariância com os coeficientes de tendência resultantes do uso do critério IRWGLS
# com correção de viés:
#####
C = solve(H, tol = 1e-30)      # C recebe a inversa da matriz H de covariâncias
A = Xt %*% C %*% X            # Multiplicação matricial
Ai = solve(A, tol = 1e-30)    # Ai recebe a inversa da matriz A
B = Ai %*% Xt %*% C %*% Z    # Matriz B dos coeficientes estimados da tendência

# Atualização dos coeficientes beta estimados:
beta0 <- B[1]
beta1 <- B[2]
beta2 <- B[3]
beta3 <- B[4]
beta4 <- B[5]
beta5 <- B[6]

# Dados para resíduos permanecem os mesmos;
# os coeficientes estimados da tendência é que são atualizados
Dadosres <- Dados

# Resíduos atualizados:
Dadosres$data <- Dadosres$data-(beta0+beta1*x+beta2*y+beta3*x2+beta4*y2+beta5*xy)

#####
# Estimação da semivariância dos resíduos usando 70% da maior distância entre
# os pontos:
#####
semivarres70 <- variog(Dadosres, estimator="classical", max.dist=hmax70)
plot(semivarres70, xlab="Distância (h)", ylab="Semivariância de resíduos", col="blue",
pch=16,main="Matheron (70% da maior distância entre pontos)")

#####
# Validação inicial do modelo de semivariância ajustado com os coeficientes de
# tendência resultantes do uso do critério IRWGLS com correção de viés.
# Este modelo deve ser igual ao modelo ajustado com o critério MQLO. Por exemplo,
# se o modelo exponencial por meio do critério MQLO foi o escolhido, o modelo exponencial
# por meio do critério IRWGLS com correção de viés deve ser escolhido para
# validação inicial nos comandos a seguir.
# As variáveis resultantes da validação inicial do modelo serão submetidas ao método
# de validação de Bastos (2017).
#####
```

```

if (covariancia_model_MQLO == "exponential"){
#####
# Modelo exponencial
#####
# Ajuste a sentimento inicial com modelo exponencial:
semivarresas7=eyefit(semivarres70)
contrib7=semivarresas7[[1]]$cov.pars[1]
alcance7=semivarresas7[[1]]$cov.pars[2]
pepita7=semivarresas7[[1]]$nugget

# Ajuste do modelo de semivariância por meio do método dos mínimos
# quadrados não-lineares ponderados:
var.wls7=variofit(semivarres70,cov.model = "exponential",ini=c(contrib7,alcance7),
nugget=pepita7,weights="cressie",minimisation.function="optim")
plot(semivarres70, xlab="Distância (h)", ylab="Semivariância", main="Exponencial")
lines(var.wls7,lwd=1,lty=1)

# Autovalidação:
xv.semivarresas7=xvalid(Dados,model=var.wls7)
summary(xv.semivarresas7)

# Variável a ser usada dentro do método de validação de Bastos (2017):
vpexp3 <- xv.semivarresas7$predicted

# Parametros do modelo de semivariância ajustado:
C0_exp3<-var.wls7$nugget
C1_exp3<-var.wls7$cov.pars[1]
a_exp3<-var.wls7$cov.pars[2]
C_exp3<-C0_exp3+C1_exp3
model_exp3<-var.wls7$cov.model
}

if (covariancia_model_MQLO == "gaussian"){
#####
# Modelo gaussiano
#####
# Ajuste a sentimento inicial com modelo gaussiano:
semivarresas8=eyefit(semivarres70)
contrib8=semivarresas8[[1]]$cov.pars[1]
alcance8=semivarresas8[[1]]$cov.pars[2]
pepita8=semivarresas8[[1]]$nugget

# Ajuste do modelo de semivariância por meio do método dos mínimos
# quadrados não-lineares ponderados:
var.wls8=variofit(semivarres70,cov.model = "gaussian", ini=c(contrib8,alcance8),
nugget=pepita8,weights="cressie",minimisation.function="optim")
plot(semivarres70, xlab="Distância (h)", ylab="Semivariância")
lines(var.wls8,lwd=1,lty=1)

```

```

# Autovalidação:
xv.semivarresas8=xvalid(Dados,model=var.wls8)
summary(xv.semivarresas8)

# Variável a ser usada dentro do método de validação de Bastos (2017):
vpgau3 <- xv.semivarresas8$predicted

# Parametros do modelo de semivariância ajustado:
C0_gau3<-var.wls8$nugget
C1_gau3<-var.wls8$cov.pars[1]
a_gau3<-var.wls8$cov.pars[2]
C_gau3<-C0_gau3+C1_gau3
model_gau3<-var.wls8$cov.model
}

if (covariancia_model_MQLO == "spherical"){
#####
# Modelo esférico
#####
# Ajuste a sentimento inicial com modelo esférico:
semivarresas9=eyefit(semivarres70)
contrib9=semivarresas9[[1]]$cov.pars[1]
alcance9=semivarresas9[[1]]$cov.pars[2]
pepita9=semivarresas9[[1]]$nugget

# Ajuste do modelo de semivariância por meio do método dos mínimos
# quadrados não-lineares ponderados:
var.wls9=variofit(semivarres70,cov.model = "sph", ini=c(contrib9,alcance9),
nugget=pepita9, weights="cressie",minimisation.function="optim")
plot(semivarres70, xlab="Distância (h)", ylab="Semivariância")
lines(var.wls9,lwd=1,lty=1)

# Autovalidação:
xv.semivarresas9=xvalid(Dados,model=var.wls9)
summary(xv.semivarresas9)

# Variável a ser usada dentro do método de validação de Bastos (2017):
vpesf3 <- xv.semivarresas9$predicted

# Parametros do modelo de semivariância ajustado:
C0_esf3<-var.wls9$nugget
C1_esf3<-var.wls9$cov.pars[1]
a_esf3<-var.wls9$cov.pars[2]
C_esf3<-C0_esf3+C1_esf3
model_esf3<-var.wls9$cov.model
}

```

```

#####
# Estimativas dos parâmetros do modelo ajustado por meio do critério IRWGLS com correção
# de viés, de acordo com o modelo ajustado por meio do critério MQLO:
#####

if(covariancia_model_MQLO == "exponential"){
#####
# Modelo exponencial
#####
pepita_model_IRWGLS_corrigeid = C0_exp3
contrib_model_IRWGLS_corrigeid = C1_exp3
alcance_model_IRWGLS_corrigeid = a_exp3
covariancia_model_IRWGLS_corrigeid = model_exp3
}

if(covariancia_model_MQLO == "gaussian"){
#####
# Modelo gaussiano
#####
pepita_model_IRWGLS_corrigeid = C0_gau3
contrib_model_IRWGLS_corrigeid = C1_gau3
alcance_model_IRWGLS_corrigeid = a_gau3
covariancia_model_IRWGLS_corrigeid = model_gau3
}

if(covariancia_model_MQLO == "spherical"){
#####
# Modelo esférico
#####
pepita_model_IRWGLS_corrigeid = C0_esf3
contrib_model_IRWGLS_corrigeid = C1_esf3
alcance_model_IRWGLS_corrigeid = a_esf3
covariancia_model_IRWGLS_corrigeid = model_esf3
}

```