

UNIVERSIDADE ESTADUAL DO OESTE DO PARANÁ

CAMPUS DE CASCAVEL

CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA AGRÍCOLA

**EFEITOS DA AMOSTRAGEM NA ANÁLISE DA VARIABILIDADE ESPACIAL DE
VARIÁVEIS GEORREFERENCIADAS**

Leila Ventorin

CASCAVEL – Paraná – Brasil

2017

LEILA VENTORIN

**EFEITOS DA AMOSTRAGEM NA ANÁLISE DA VARIABILIDADE ESPACIAL DE
VARIÁVEIS GEORREFERENCIADAS**

Dissertação apresentada ao programa de Pós-Graduação em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção ao título de Mestre em Engenharia Agrícola, área de concentração em Engenharia de Sistemas Agroindustriais.

Orientadora: Dra. Luciana Pagliosa Carvalho Guedes.

CASCADEL – Paraná – Brasil

MARÇO 2017

LEILA VENTORIN

**EFEITOS DA AMOSTRAGEM NA ANÁLISE DA VARIABILIDADE ESPACIAL DE
VARIÁVEIS GEORREFERENCIADAS**

Dissertação apresentada ao programa de Pós-Graduação em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção ao título de Mestra em Engenharia Agrícola, área de concentração em Engenharia de Sistemas Agroindustriais, linha de pesquisa Geoprocessamento, Estatística Espacial e Agricultura de Precisão, APROVADO(A) pela seguinte banca examinadora:

Orientadora: Dra. Luciana Pagliosa Carvalho Guedes

Centro de Ciências Exatas e Tecnológicas, UNIOESTE

Banca 1: Dr. Miguel Angel Uribe-Opazo

Centro de Ciências Exatas e Tecnológicas, UNIOESTE

Banca 2: Dra. Rosângela Botinha Assumpção

Departamento de Matemática, UTFPR

CASCADEL – Paraná – Brasil

Março 2017

Dados Internacionais de Catalogação-na-Publicação (CIP)

V576e

Ventorin, Leila

Efeitos da amostragem na análise da variabilidade espacial de variáveis georreferenciadas / Leila Ventorin. Cascavel, 2017.

112 f.

Orientadora: Pro^a. Dr^a. Luciana Pagliosa Carvalho Guedes
Revisão Português, Inglês e Normas: Ana Maria Martins Alves
Vasconcelos

Dissertação (Mestrado) – Universidade Estadual do Oeste do Paraná,
Campus de Cascavel, 2017

Programa de Pós-Graduação em Engenharia de Energia na Agricultura

1. Agricultura de precisão. I. Guedes, Luciana Pagliosa Carvalho. II. Vasconcelos, Ana Maria Martins Alves. III. Universidade Estadual do Oeste do Paraná. IV. Título.

CDD 20.ed. 630
CIP-NBR 12899

Ficha catalográfica elaborada por Helena Soterio Bejio – CRB 9^a/965

BIOGRAFIA

Leila Ventorin, Nascida em Medianeira/PR em Fevereiro de 1992, graduada em Engenharia de Produção, pela Universidade Tecnológica Federal do Paraná (UTFPR) no ano de 2014. Experiência profissional em Gestão de processos logísticos na Cooperativa Agroindustrial LAR. Professora de matemática na ONG MediAres, campus UTFPR-medianeira. Em Março de 2015 ingressou no Mestrado em Engenharia Agrícola, área de concentração Engenharia de Sistemas Agroindustriais, linha de pesquisa Tecnologia da Produção Agrícola na Universidade Estadual do Paraná (UNIOESTE).

“Determinação, coragem e autoconfiança são fatores decisivos para o sucesso. Se estamos possuídos por uma inabalável determinação, conseguiremos superá-los. Independentemente das circunstâncias, devemos ser sempre humildes, recatados e despidos de orgulho”.

Dalai Lama

AGRADECIMENTOS

Mais uma importante etapa de minha vida encerra-se aqui. Portanto, quero, neste momento, retribuir com algumas palavras a minha gratidão pelo apoio, dedicação e carinho aos que me acompanharam neste período.

Primeiramente, agradeço a Deus por minha vida e todas as coisas boas que vivi até hoje.

Agradeço a minha Orientadora Dra. Luciana Pagliosa Carvalho Guedes, por todo conhecimento e sabedoria repassados neste período, além de todo seu apoio e confiança.

A esta Universidade, seu corpo docente, direção e administração.

Ao senhor Agassiz Linhares neto, pela parceria na implantação dos experimentos em sua propriedade.

Aos colegas do grupo de geoestatística aplicada.

Em especial, quero deixar aqui registrado todo meu reconhecimento e agradecimento a minha família. Agradeço a vocês meus pais, sinônimos de heróis, por me ensinarem a viver com dignidade, por iluminarem o meu caminho, pelo amor e compreensão em todos os momentos.

Ao meu esposo Christiano, que de forma especial e carinhosa me deu força e coragem e me apoiou nos momentos de dificuldades.

Enfim, agradeço a todos que, de alguma forma, contribuíram com a realização deste trabalho.

RESUMO

Em agricultura de precisão, esforços têm sido direcionados para caracterizar a variabilidade espacial de atributos do solo, visando estabelecer procedimentos amostrais que garantam a representatividade das amostras georreferenciadas. Assim, o objetivo desse trabalho foi avaliar dados estacionários e isotrópicos e ou com tendência direcional (processos não estacionários) ou anisotrópicos, a influência da configuração amostral na estimação do modelo geoestatístico e na estimação de localizações não amostradas. Para isso, foram simulados os seguintes sistemas de amostragens: aleatória com 100 pontos, sistemática nas versões 10x10, 5x20 e 20x5, e sistemática adicionada de pontos próximos (*lattice plus close pairs*) com pontos próximos adicionados na direção da tendência e da anisotropia e na direção ortogonal à tendência e da anisotropia. Esses resultados servirão como embasamento científico para uma análise mais eficiente da variabilidade espacial de atributos químicos em uma área agrícola, com variáveis isotrópicas, não estacionárias e anisotrópicas. Os resultados dos dados simulados evidenciam que a amostragem *lattice plus close pairs* (em todas as versões simuladas) apresentou os melhores resultados na qualidade da estimativa dos parâmetros do modelo e da predição espacial. Considerando-se os resultados simulados e a análise da variabilidade espacial dos atributos químicos do solo, propõe-se que, em posteriores experimentos nessa área agrícola, considerem-se o aumento da quantidade de pontos próximos e a redução do raio dos pontos próximos. E, ainda para variáveis anisotrópicas e com tendência direcional, a adição dos pontos próximos na direção destes fenômenos e na direção ortogonal a estes.

Palavras-chave: Agricultura de precisão, Configuração amostral, Geoestatística.

SAMPLING EFFECTS IN SPATIAL VARIABILITY ANALYSIS OF GEOREFERENCED VARIABLES

ABSTRACT

In precision agriculture, efforts have been done to characterize the spatial variability of soil attributes, aiming at establishing sampling procedures that guarantee representativeness of georeferenced samples. Thus, this trial aimed at evaluating stationary and isotropic data, with or without directional tendency (non-stationary or anisotropic processes, the influence of sample configuration in geostatistical model estimation and in non-sampled locations estimation. In order to obtain some data, the following sampling systems were simulated: the randomized one with 100 points, the systematic one in 10x10, 5x20 and 20x5 versions, and the lattice plus close pairs with nearby added points toward tendency and anisotropy, and toward orthogonal to the anisotropy. These results will serve as a scientific basis for a more efficient analysis of the spatial variability regarding chemical attributes of an agricultural area with isotropic, non-stationary and anisotropic variables. The results of simulated data have shown that lattice plus close pairs sampling (in all simulated versions) has presented the best results on quality of parameters estimation of model and spatial prediction. Considering the simulated results and the analysis of spatial variability regarding soil chemical attributes, it is proposed that in subsequent experiments in this agricultural area, the increase in number of nearby points and radius reduction of the nearby points should be considered. Also, for anisotropic variables and with directional tendency, the nearby points addition toward these phenomena and in orthogonal direction to these ones.

Keywords: geostatistics, sample configuration, precision agriculture.

SUMÁRIO

LISTA DE TABELAS.....	x
LISTA DE FIGURAS	xi
1 INTRODUÇÃO	14
2 OBJETIVOS.....	16
2.1 Objetivo geral	16
2.2 Objetivos específicos.....	16
3 GEOESTATISTICA.....	18
3.1 Variáveis regionalizadas	19
3.2 Estacionariedade Intrínseca e de Segunda ordem	19
3.3 Semivariograma.....	22
3.4 Anisotropia.....	25
3.5 MODELOS TEÓRICOS.....	27
3.5.1 Modelo Esférico	27
3.5.2 Modelo Exponencial	28
3.5.3 Modelo Gaussiano	29
3.5.4 Família Matérn.....	30
4 ESTIMAÇÃO DOS PARÂMETROS NO AJUSTE DE MODELOS TEÓRICOS.....	31
4.1 Método de Máxima Verossimilhança (MLE)	32
4.1.1 Erros padrão assintóticos das estimativas dos parâmetros.....	32
5 CRITÉRIO DE SELEÇÃO DOS MODELOS.....	33
5.1 Validação cruzada.....	34
5.2 Informação de Akaike.....	35
5.3 Critério de Informação Bayesiano	36
6 KRIGAGEM.....	36
7 AMOSTRAGEM ESPACIAL	39
8 MATERIAIS E METODOS.....	46
8.1 Simulações.....	46
8.2 Medidas para avaliar a qualidade da estimação dos parâmetros e da predição espacial.....	48
8.3 Estudo prático.....	50
8.3.1 Área de estudo.....	50
8.3.2 Análise Descritiva	51
8.3.3 Análise Geoestatística	51

9	Software utilizado	52
10	RESULTADOS E DISCUSSÕES	53
10.1	Estudo de simulações	53
10.1.1	Dados isotrópicos e estacionários	53
10.1.2	Lattice Plus Close Pairs	53
10.1.3	Lattice Plus In-fill	62
10.1.4	Amostragens Sistemática e Aleatória comparadas com as amostragens Lattice plus close pairs e Lattice plus in-fill	68
11.1.2	Dados com tendência direcional	75
11.1.3	Dados Anisotrópicos	86
12	ESTUDO PRÁTICO	94
12.1.1	Análise descritiva	94
12.1.2	Análise geoestatística	98
13	CONCLUSÕES	104
14	REFERÊNCIAS	105

LISTA DE TABELAS

Tabela 1 Estatísticas descritivas dos parâmetros do modelo exponencial estimado por ML e seus respectivos desvios padrões e as medidas de eficiência do estimador VRA(%), VA e REQM para dados isotrópicos.....	54
Tabela 2 Análise descritiva das medidas de qualidade da predição espacial em localizações não amostradas considerando uma amostra teste composta por 25 pontos	57
Tabela 3 Estatísticas descritivas dos parâmetros do modelo exponencial estimado por ML para as grades (8x8,36,1), (7x7,51,1) e (8x9,28,1) com dados isotrópicos.....	59
Tabela 4 Análise descritiva das medidas da qualidade da predição espacial em localizações não amostradas, considerando uma amostra teste composta por 25 pontos	62
Tabela 5 Análise descritiva dos parâmetros do modelo exponencial estimado por ML e seus respectivos desvios padrões e as medidas de eficiência do estimador VRA(%), VA e REQM para dados isotrópicos.....	63
Tabela 6 Análise descritiva das medidas da qualidade da predição espacial em localizações não amostradas, considerando uma amostra teste composta por 25 pontos	67
Tabela 7 Análise descritiva dos parâmetros do modelo exponencial estimado por ML e seus respectivos desvios padrões e as medidas de eficiência do estimador VRA(%), VA e REQM para dados isotrópicos.....	69
Tabela 8 Análise descritiva das medidas da qualidade da predição espacial em localizações não amostradas, considerando uma amostra teste composta por 25 pontos	72
Tabela 9 Estatísticas descritivas dos parâmetros do modelo exponencial estimado por ML para as grades Aleatória, sistemática 10x10, 5x20 e 20x5 e lattice plus close pairs 7x7,51,1 com tendência direcional.....	77
Tabela 10 Análise descritiva das medidas da qualidade da predição espacial do modelo exponencial sem tendência direcional quanto a predição espacial de uma amostra teste composta por 25 pontos	82
Tabela 11 Estatística descritiva dos parâmetros do modelo exponencial estimado por ML para a amostragem lattice plus close pairs 7x7,51,1 anisotrópicos	87
Tabela 12 Análise descritiva das medidas da qualidade da predição espacial do modelo exponencial anisotrópicos quanto a predição espacial da uma amostra teste composta por 25 pontos	90
Tabela 13 Análise exploratória descritiva das variáveis químicas cobre Cu (mg/dm ³), zinco Zn (mg/dm ³), manganês Mn (mg/dm ³), carbono C (g/dm ³), (cmolc/dm ³), cálcio Ca (cmolc/dm ³), magnésio Mg (cmolc/dm ³), Alumínio Al (cmolc/dm ³) e Fósforo P (mg/dm ³).....	95
Tabela 14 estimação dos parâmetros das variáveis cobre, zinco, manganês, carbono, cálcio, magnésio, alumínio e fósforo por MV.....	101
Tabela 15 critérios de validação cruzada, AIC e BIC para a escolha do melhor modelo ajustado.....	100

LISTA DE FIGURAS

Figura 1 Relação entre as funções Semivariância e Covariância	21
Figura 2 Amostragem em duas dimensões de uma variável regionalizada, com dois pontos separados por uma distância h.	22
Figura 3 Exemplo de semivariograma com comportamento ideal	23
Figura 4 Representação gráfica do modelo esférico com $\phi_1 \neq 0$	28
Figura 5 Representação gráfica do modelo exponencial com $\phi_1 \neq 0$	29
Figura 6 Representação gráfica do modelo gaussiano com $\phi_1 \neq 0$	30
Figura 7 (a) Método aleatório de amostragem de solo, coleta aleatória; (b) Método sistemático, em que as amostras são georreferenciadas com distâncias determinadas entre pontos.....	40
Figura 8 Exemplos de amostragem (a) Lattice plus close pairs 9x9,19,5 (b) Lattice plus in-fill 6x6,2,6x6	41
Figura 9 Grades regulares de amostragem	43
Figura 10 Grades com amostragem alinhada sistemática estratificada (AASE): a) AASE 128, b) AASE 64, c) AASE 32. Em cada grade, as parcelas em cinza representam parcelas selecionadas	44
Figura 11 Fluxograma das etapas da pesquisa.....	48
Figura 12 Mapa da área em estudo	50
Figura 13 Gráfico bolxplot dos valores estimados dos seguintes parâmetros (a) Média, (b) Efeito Pepita, (c) Contribuição e (d) Alcance Prático. A linha tracejada indica o valor nominal simulado.....	54
Figura 14 Gráfico bolxplot do desvio padrão das estimativas dos parâmetros estimados (a) Média, (b) Efeito Pepita, (c) Contribuição e (d) Alcance Prático.....	55
Figura 15 Gráficos boxplot: (a) Erro Médio, (b) Erro Médio Relativo, (c) Desvio Padrão do Erro Médio, (d) Desvio Padrão do Erro Médio Relativo, (e) Erro Absoluto, (f) AIC e (g) BIC. A linha tracejada indica o valor ideal.....	56
Figura 16 Gráficos Boxplot: (a) Média da Variância da krigagem e (b) Erro de predição	57
Figura 17 Gráficos boxplot dos valores estimados dos seguintes parâmetros: (a) Média, (b) Efeito Pepita, (c) Contribuição e (d) Alcance Prático, a linha tracejada indica o valor nominal simulado.....	59
Figura 18 Boxplot do desvio padrão das estimativas dos parâmetros: (a) Média, (b) Efeito pepita, (c) Contribuição e (d) Alcance prático.....	60
Figura 19 Gráficos boxplot do (a) Erro Médio, (b) Erro Médio Relativo, (c) Desvio Padrão do Erro Médio, (d) Desvio Padrão do Erro Médio Relativo, (e) Erro Absoluto, (f) AIC e (g) BIC. A linha tracejada indica o valor ideal.....	61
Figura 20 Gráfico boxplot: (a) Média da variância da krigagem e (b) Erro da predição numa amostra teste composta por 25 pontos.....	62
Figura 21 Gráfico bolxplot dos valores estimados dos seguintes parâmetros: (a) Média, (b) Efeito Pepita, (c) Contribuição e (d) Alcance Prático, a linha tracejada indica o valor nominal simulado.....	64
Figura 22 Gráfico bolxplot do desvio padrão das estimativas dos parâmetros: (a) Média, (b) Efeito Pepita, (c) Contribuição e (d) Alcance Prático.....	65

Figura 23 Gráficos boxplot: (a) Erro Médio, (b) Erro Médio Relativo, (c) Desvio Padrão do Erro Médio, (d) Desvio Padrão do Erro Médio Relativo, (e) Erro Absoluto, (f) AIC e (g) BIC. A linha vermelha indica o valor ideal.....	66
Figura 24 Gráfico boxplot: (a) Média da variância da krigagem e (b) Erro da predição	67
Figura 25 Gráfico bolxplot dos valores estimados dos seguintes parâmetros: (a) Média, (b) Efeito Pepita, (c) Contribuição e (d) Alcance Prático.....	69
Figura 26 Gráfico boxplot do desvio padrão das estimativas dos parâmetros: (a) Média, (b) Efeito Pepita, (c) Contribuição e (d) Alcance Prático.....	70
Figura 27 Gráficos boxplot: (a) Erro Médio, (b) Erro Médio Relativo, (c) Desvio Padrão do Erro Médio, (d) Desvio Padrão do Erro Médio Relativo, (e) Erro Absoluto, (f) AIC e (g) BIC. A linha tracejada indica o valor ideal.....	71
Figura 28 Gráfico boxplot: (a) Média da variância da krigagem e (b) Erro da predição	72
Figura 29 Mapa da variância da Krigagem da amostra teste usando as amostragens (a) Lattice Plus close pairs 7x7,51,1 (b) Lattice plus in fill 8x8,3,4x4 (c) Aleatória (d) Sistemática. As esferas em azul representam o valor da variância da krigagem na amostra teste. Seus valores estão representados pelo tamanho da esfera.....	74
Figura 30 Erro da predição espacial feita pela krigagem da amostra teste usando as amostragens: (a) Lattice Plus close pairs 7x7,51,1 (b) Lattice plus in fill 8x8,3,4x4 (c) Aleatória (d) Sistemática. As esfera representam o valor do Erro da predição espacial da amostra teste, e seus valores estão representados pelo seu tamanho.....	75
Figura 31 Gráfico boxplot das estimativas dos seguintes parâmetros: (a) β_0 , (b) β_1 , (c) Efeito pepita (d) Contribuição (e) Alcance prático.....	78
Figura 32 Gráfico boxplot do desvio padrão dos parâmetros estimados: (a) β_0 , (b) β_1 , (c) Efeito pepita (d) Contribuição (e) Alcance prático.....	79
Figura 33 Gráfico de barras do coeficiente de correlação linear de Pearson nas simulações das amostragens (a) 10x10, (b) 5x20, (c) 20x5, (d) Aleatória, (e) 7x7,51,1, (f) 7x7,51,1(a) e (g) 7x7,51,1(b). Em que: Forte ($0,8 \leq r < 1$), Moderado ($0,5 \leq r < 0,8$), Fraca ($0,1 \leq r < 0,5$).....	80
Figura 34 Gráficos boxplot: (a) Erro Médio, (b) Erro Médio Relativo, (c) Desvio Padrão do Erro Médio, (d) Desvio Padrão do Erro Médio Relativo, (e) Erro Absoluto, (f) AIC e (g) BIC. A linha horizontal indica o valor ideal.....	81
Figura 35 Gráfico boxplot: (a) Média da variância da krigagem e (b) Erro da predição	83
Figura 36 Mapa da variância da Krigagem da amostra teste usando as amostragens: Sistemática (a) 10x10 (b) 20x5 (c) 5x20, <i>Lattice plus close pairs</i> (d) 7x7,51,1 (e) 7x7,51,1(a) com adição dos pontos próximos na direção ortogonal a tendência (f) 7x7,51,1(b) com adição dos pontos próximos na direção da tendência (g) Aleatória. As esferas representam o valor da variância da krigagem na amostra teste (25 pontos), e seus valores estão representados pelo tamanho da esfera	84
Figura 37 Erro da predição espacial feita pela Krigagem numa amostra teste de 25 pontos usando as amostragens: Sistemática (a) 10x10 (b) 20x5 (c) 5x20, <i>Lattice plus close pairs</i> (d) 7x7,51,1 (e) 7x7,51,1(a) com adição dos pontos próximos na direção ortogonal a tendência (f) 7x7,51,1(b) com adição dos pontos próximos na direção da tendência (g) Aleatória.....	85
Figura 38 Gráfico boxplot dos parâmetros estimados (a) Média, (b) Efeito pepita, (c) Contribuição (d) Alcance prático e (e) Fator de anisotropia	88
Figura 39 Gráficos boxplot: (a) Erro Médio, (b) Erro Médio Relativo, (c) Desvio Padrão do Erro Médio, (d) Desvio Padrão do Erro Médio Relativo, (e) Erro Absoluto, (f) AIC e (g) BIC. A linha horizontal indica o valor ideal.....	89

Figura 40 Gráfico boxplot: (a) Média da variância da krigagem e (b) Erro da predição da amostra teste composta por 25 pontos.....	91
Figura 41 Mapa da variância da Krigagem da amostra teste usando as amostragens: (a) aleatória, (b) <i>Lattice plus close pairs</i> 7x7,51,1, (c) 7x7,51,1(a) com adição dos pontos próximos na direção ortogonal a tendência , (d) 7x7,51,1(b) com adição dos pontos próximos na direção da tendência e Sistemática (e) 10x10 (f) 5x20 e (g) 20x5.....	92
Figura 42 Erro da predição espacial feita pela Krigagem da amostra teste (25 pontos) usando as amostragens: (a) aleatória, <i>Lattice plus close pairs</i> (b) 7x7,51,1 (c) 7x7,51,1(a) com adição dos pontos próximos na direção ortogonal a anisotropia, (d) 7x7,51,1(b) com adição dos pontos próximos na direção da anisotropia e Sistemáticas (e) 10x10 (f) 5x20 e (g) 20x5	93
Figura 43 Gráfico boxplot das variáveis: a) Cobre Cu (mg/dm ³), (b) Zinco Zn (mg/dm ³), (c) Manganês Mn (mg/dm ³), (d) Carbono C (g/dm ³). (e) Cálcio Ca (cmolc/dm ³), (f) Magnésio Mg (cmolc/dm ³), (g) Alumínio Al (cmolc/dm ³) e (h) Fósforo P (mg/dm ³)	95
Figura 44 Gráfico post-plot das variáveis: (a) Cobre Cu (mg/dm ³), (b) Zinco Zn (mg/dm ³), (c) Manganês Mn (mg/dm ³), (d) Carbono C (g/dm ³). (e) Cálcio Ca (cmolc/dm ³), (f) Magnésio Mg (cmolc/dm ³), (g) Alumínio Al (cmolc/dm ³) e (h) Fósforo P (mg/dm ³)	96
Figura 45 Gráfico de envelopes das variáveis: a) Cobre Cu (mg/dm ³), (b) Zinco Zn (mg/dm ³), (c) Manganês Mn (mg/dm ³), (d) Carbono C (g/dm ³). (e) Cálcio Ca (cmolc/dm ³), (f) Magnésio Mg (cmolc/dm ³), (g) Alumínio Al (cmolc/dm ³) e (h) Fósforo (P) (mg/dm ³).....	97
Figura 46 Semivariograma direcional das variáveis: a) Cobre Cu (mg/dm ³), (b) Zinco Zn (mg/dm ³), (c) Manganês Mn (mg/dm ³), (d) Carbono C (g/dm ³). (e) Cálcio Ca (cmolc/dm ³), (f) Magnésio Mg (cmolc/dm ³), (g) Alumínio Al (cmolc/dm ³) (h) Fósforo P (mg/dm ³)	99
Figura 47 Mapa temático das variáveis: a) Cobre Cu (mg/dm ³), (b) Zinco Zn (mg/dm ³), (d) Manganês Mn (mg/dm ³), (e) Carbono C (g/dm ³). (f) Cálcio Ca (cmolc/dm ³), (g) Magnésio Mg (cmolc/dm ³), (h) Fósforo P (mg/dm ³).....	102

1 INTRODUÇÃO

Pesquisas experimentais objetivam descrever fenômenos ou comparar o comportamento de variáveis em subgrupos de uma população, os quais são chamados de amostra. A utilização de uma amostra em uma pesquisa, ao invés da população se justifica-se porque a pesquisa sob toda população não é acessível ou viável. No entanto, é necessária que a amostra seja representativa para a realização de inferências à população-alvo.

Definem-se o dimensionamento numérico da amostra e a técnica de seleção dos elementos da pesquisa no planejamento amostral do estudo. Esta etapa é fundamental para a elaboração da pesquisa, uma vez que a adoção de técnicas incorretas pode comprometer a interpretação final dos resultados.

No que tange este assunto e baseada na estatística clássica, a geoestatística é um ramo da estatística espacial que tem como objetivo reproduzir a distribuição e a variabilidade espacial de um fenômeno. Também se preocupa com a representatividade da amostra quanto a sua população em estudo, pois, a amostra é um subconjunto de valores do fenômeno espacial que, se representativa, deve reproduzir a distribuição e a variabilidade espacial tanto em tamanho, isto é, número de pontos de dados, como em termos de distribuição dos pontos no domínio a ser estudado (YAMAMOTO & LANDIM, 2013). Neste sentido, o projeto de dissertação aqui apresentado foi motivado pelo interesse que os efeitos da amostragem proporcionam às etapas de uma análise geoestatística.

O planejamento amostral, nesse cenário, assume valor relevante, pois a escolha do tamanho amostral n e das localizações amostrais (S_1, \dots, S_n) implica qualidade do mapeamento da população, uma vez que deve-se ter boa estimativa dos parâmetros do modelo ajustando a função semivariância e uma eficiência na predição espacial, de modo que, as estimativas obtidas a partir da krigagem sejam mais exatas e conseqüentemente mais confiáveis.

Desta forma, especificamente na área agrícola, para que a amostragem do solo represente com exatidão os índices de fertilidade do solo, é necessário o conhecimento dessa variabilidade, pois só assim as recomendações de adubação e calagem não ficariam comprometidas. É pelo conhecimento da variabilidade espacial das propriedades de solo e de planta que se pode contribuir para o planejamento e a otimização dos investimentos em áreas de cultivo.

A variabilidade espacial das propriedades físico-químicas do solo influencia no fator produtividade da cultura. Sendo assim, é importante que no planejamento amostral sejam bem definidos o dimensionamento e a configuração amostral a ser adotada na coleta dos

dados. Desta forma, a variabilidade espacial destas propriedades representará melhor a área estudada.

2 OBJETIVOS

2.1 Objetivo geral

Identificar a influência que a configuração espacial das amostras exerce na qualidade da estimação dos parâmetros do modelo geoestatístico e na predição espacial da variável georreferenciada em localizações não amostradas.

2.2 Objetivos específicos

1. Verificar a influência de fatores (raio e número de pontos próximos) que determinam a configuração amostral sistemática centrada com pares de pontos próximos (*lattice plus close pairs*) quanto à qualidade da estimação do modelo geoestatístico e da estimação espacial de valores da variável georreferenciada em localizações não amostradas.
2. Verificar a influência de fatores (número de pontos, tamanho da *lattice* menor e quantidade de *lattices* menores) que determinam a configuração amostral sistemática centrada com amostragens sistemáticas centradas menores (*lattice plus in-fill*) quanto à qualidade da estimação do modelo geoestatístico e da estimação espacial de valores da variável georreferenciada em localizações não amostradas.
3. Verificar se existe uma influência da adição de pontos na direção da tendência direcional, ou na direção ortogonal, em modelos com tendência direcional no melhor cenário obtido no item 2 quanto à qualidade da estimação dos parâmetros do modelo e da predição espacial feita pela Krigagem.
4. Verificar se existe influência da adição de pontos na direção da anisotropia, ou na direção ortogonal a esta, em modelos anisotrópicos no melhor cenário obtido no item 2 quanto à qualidade da estimação dos parâmetros do modelo e da predição espacial feita pela Krigagem.
5. Verificar, para amostragens regulares e modelos geoestatísticos com tendência direcional (não estacionários), se existe influência de adição de um número maior de pontos na direção da tendência ou na direção ortogonal a esta, quanto à qualidade da estimação dos parâmetros do modelo geoestatístico e da predição espacial feita pela krigagem.

6. Verificar, para amostragens regulares e modelos geoestatísticos anisotrópicos, se existe a influência de adição de um número maior de pontos na direção da anisotropia ou na direção ortogonal a esta, quanto à qualidade da estimação dos parâmetros do modelo geoestatístico e da predição espacial feita pela krigagem.
7. Comparar os resultados anteriores com amostragens aleatória e regular quadrada.

3 GEOESTATÍSTICA

Nos anos 50, na África do Sul, o engenheiro Daniel Krige avaliou jazidas de ouro e observou que para a obtenção de métodos mais eficientes de estimação da concentração de ouro deveria se considerar a existência de variabilidade espacial, ou seja, que somente as informações de variância não seriam suficientes para explicar o fenômeno. Para este fim, seria necessário levar em consideração a distância entre as amostras. Desde então, surgiu o termo geoestatística, que está associado à distribuição estatística dos dados no espaço (REZENDE et al., 2012).

A geoestatística está fundamentada na ideia de que “todas as coisas são parecidas, mas coisas mais próximas se parecem mais que coisas mais distantes” (CÂMARA et al., 2002). Desta forma, um valor observado em determinado ponto mantém relações de dependência com valores observados em pontos próximos, logo, obtém-se uma estrutura de correlação. A partir desta teoria, a geoestatística descreve e modela a relação entre dependência e distância (TEIXEIRA, 2013).

Soares (2014) apresenta a geoestatística como a descrição dos fenômenos espaciais naturais. O autor destaca que, partindo-se de um conjunto discreto e limitado de dados experimentais georreferenciados, a geoestatística permite o delineamento de modelos que visam descrever as distintas realidades de cada estudo, definindo-a como um conjunto de métodos, técnicas e instrumentos estatísticos que caracterizam os fenômenos espaciais naturais.

Landim (2002) destaca que, pelo fato da geoestatística calcular estimativas dentro de um contexto gerido por um fenômeno natural com distribuição no espaço, há uma grande aplicação daquela, em especial para estimativas e simulações de locais não amostrados. Ainda neste sentido, Pontes (2002) ressalta sobre esta metodologia quando diz que a correlação espacial entre as amostras não é considerada um incômodo a ser evitado, mas sim uma fonte de informações que resulta em maior qualidade na análise dos dados.

De modo geral, a geoestatística vem se sobressaindo e trazendo resultados mais eficientes e confiáveis. Sua aplicação destaca-se nas Ciências Humanas (SILVA et al., 2015), Biológicas (PELISSARI et al., 2014) e Exatas. E é popular nos diversos campos da ciência e da indústria, nos quais existe a necessidade de avaliar dados espacialmente ou temporalmente correlacionados (LOURENÇO & LANDIN, 2005; OLIVEIRA JÚNIOR et al., 2011).

Para melhor compreender estes conceitos, é importante entender alguns pressupostos e teorias que envolvem a geoestatística. Na sequência, serão tratados alguns dos principais tópicos englobados pela área.

3.1 Variáveis regionalizadas

O primeiro conceito a ser compreendido é a teoria das variáveis regionalizadas, a qual foi desenvolvida por Matheron (1963). A variável regionalizada é uma função que busca descrever fenômenos que apresentam uma distribuição no espaço como variáveis que dependem da sua posição espacial (GUEDES et al., 2015). Desta forma, Marques et al. (2012) abordam a geoestatística como uma aplicação prática das variáveis regionalizadas.

Na caracterização das variáveis regionalizadas, é aplicada a definição de variáveis aleatórias, em que se assumem distintos valores Z em função da sua posição S dentro de uma área. O conjunto formado pelas variáveis regionalizadas em determinada região pode ser considerado uma função aleatória $Z(S)$, no qual assume-se que a dependência entre elas é definida por uma distribuição de probabilidade (DEL MONEGO et al., 2014).

Assim, a teoria da variável regionalizada implica a soma de três componentes: a) uma componente estrutural, associada a um valor médio constante ou a uma tendência constante; b) uma componente aleatória, espacialmente correlacionada; e c) um ruído aleatório ou erro residual. Se S representa um vetor posição em uma, duas ou três dimensões, então o valor da função aleatória Z , em S , é expresso como (GUEDES et al., 2013):

$$Z(S) = \mu(S) + \varepsilon'(S) + \varepsilon''(S), \quad (1)$$

em que:

- $\mu(S)$ é uma função determinística que descreve a componente estrutural de Z em S ;
- $\varepsilon'(S)$ é um termo estocástico correlacionado, que varia localmente;
- $\varepsilon''(S)$ é um ruído aleatório não correlacionado, com distribuição normal com média zero e variância σ^2 .

3.2 Estacionariedade Intrínseca e de Segunda ordem

A distribuição espacial de um conjunto de valores amostrados de uma variável regionalizada Z pode ser modelada por um processo estocástico (NOGUEIRA, 2013). De acordo com Kestring et al. (2015), um processo estocástico é uma coleção $\{Z(S): S \in R \subset IR^p\}$ de variáveis aleatórias reais, definidas sobre um mesmo espaço de probabilidade, indexadas em um subconjunto R do espaço vetorial P -dimensional IR^P .

Diferente da estatística clássica, que trabalha com n amostras de uma variável aleatória, na teoria das variáveis regionalizadas, em geral, têm-se n variáveis aleatórias com apenas uma observação, ou seja, não existe repetição de uma mesma variável. Logo, em cada local observado, em um mesmo período de tempo, tem-se apenas uma variável aleatória. Desta forma, a restrição do número de repetições impede que se faça o estudo da distribuição dessas variáveis. Por este motivo, é preciso supor que o processo obedeça a algum tipo de estacionariedade (FERREIRA et al., 2013).

Neste cenário, a geoestatística trabalha com duas hipóteses: a hipótese de estacionariedade intrínseca e a de segunda ordem. A hipótese intrínseca é a mais fraca, menos restritiva, já a estacionariedade de segunda ordem possui mais restrições. Desta forma, Rossoni (2011) destaca a necessidade de que a variável em estudo obedeça pelo menos à hipótese intrínseca.

Segundo Cressie (1993), a hipótese de estacionariedade de segunda ordem é satisfeita se:

- a) A esperança matemática de $Z(\mathbf{S})$ existe e não depende da posição \mathbf{S} . Ou seja:

$$E[Z(\mathbf{S})] = \mu = \text{constante}, \quad \forall \mathbf{S} \in R \subset IR^P; \quad \text{e} \quad (2)$$

- b) Para cada par $\{Z(\mathbf{S}), Z(\mathbf{S} + \mathbf{h})\}$, a covariância existe e depende somente de h . Ou seja:

$$\text{Cov}[Z(\mathbf{S}), Z(\mathbf{S} + \mathbf{h})] = E[Z(\mathbf{S} + \mathbf{h}), Z(\mathbf{S})] - \mu^2 = C(\mathbf{h}). \quad (3)$$

em que $h = \|\mathbf{h}\|$ é a distância euclidiana entre duas localizações desconhecidas e \mathbf{h} é o vetor distância entre essas duas localizações.

Desta forma, define-se o covariograma, que consiste no gráfico dos valores da covariância (eixo das ordenadas), em função das distâncias (eixo das abcissas) (Figura 1).

Na hipótese de segunda ordem, observa-se que a variância de uma variável $Z(\mathbf{S})$ é um caso particular da covariância quando $h = 0$:

$$\text{Var}(Z(\mathbf{S})) = C(Z(\mathbf{S} + \mathbf{h}), Z(\mathbf{S})) = C(0). \quad (4)$$

Uma variável é intrinsecamente estacionária se:

- c) A esperança matemática existe e não depende da posição \mathbf{S} :

$$E[Z(\mathbf{S})] = \mu, \quad \forall \mathbf{S} \in R \subset IR^P \quad \text{e} \quad (5)$$

- d) Para todo \mathbf{h} , a variância da diferença $[Z(\mathbf{S}) - Z(\mathbf{S} + \mathbf{h})]$ existe e não depende de \mathbf{S} :

$$\frac{1}{2} \text{Var} [Z(\mathbf{S}) - Z(\mathbf{S} + \mathbf{h})] = \frac{1}{2} E([Z(\mathbf{S}) - Z(\mathbf{S} + \mathbf{h})]^2) = \gamma(\mathbf{h}), \quad (6)$$

em que:

$\gamma(\mathbf{h})$ representa a função semivariância. O prefixo “semi” existe pelo fato de que $\gamma(\mathbf{h})$ representa a metade da variância (Equação 6). Desta forma, o semivariograma consiste em um gráfico dos valores da semivariância em função das distâncias \mathbf{h} (Figura 1).

Se um processo é estacionário de segunda ordem, conseqüentemente é intrínseco. Porém, nem sempre o processo inverso é verdadeiro (NOGUEIRA, 2013). Se a hipótese de estacionariedade de segunda ordem for atendida, é possível estabelecer uma relação entre a função semivariância $\gamma(\mathbf{h})$ e a covariância $C(\mathbf{h})$, descrita como:

$$C(\mathbf{h}) = C(0) - \gamma(\mathbf{h}). \quad (7)$$

Esta relação pode ser representada graficamente (Figura 1). Observa-se que quando \mathbf{h} tende ao infinito, $C(\mathbf{h})$ tende a zero e $\gamma(\mathbf{h})$ tende a $C(0)$.

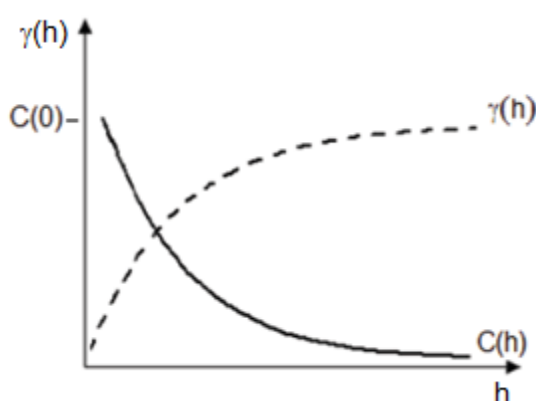


Figura 1 Relação entre as funções Semivariância e Covariância

Observe que a Equação (7) indica que, sob a hipótese de estacionariedade de segunda ordem, a covariância e a semivariância são formas alternativas de se caracterizar a autocorrelação dos pares $Z(\mathbf{S})$ e $Z(\mathbf{S} + \mathbf{h})$ separados pelo vetor \mathbf{h} . Porém, se a condição de segunda ordem não é satisfeita não é possível utilizar a função covariância (MARQUES et al., 2012).

Ainda, se o processo é estacionário de segunda ordem, pode-se utilizar a medida de correlação, definida como a razão entre a covariância dos valores assumidos pela variável Z , nas posições \mathbf{S} e $\mathbf{S} + \mathbf{h}$ e a variância dessa variável, em função da distância \mathbf{h} (GREZEGOZEWSKI, 2012). Dessa forma, tem-se:

$$\rho(h) = \frac{c(h)}{c(0)} = \frac{c(0) - \gamma(h)}{c(0)} = 1 - \frac{\gamma(h)}{c(0)}. \quad (8)$$

Esta função é adimensional e está limitada entre os valores -1 e 1, permitindo-se comparações entre variáveis e também inferências sobre o grau de associação espacial. Porém, na prática, a correlação de uma variável com ela mesma em pontos diferentes varia entre 0 e 1, uma vez que, para $h = 0$, a correlação é máxima, ou seja, $\rho(0) = 1$. No qual decresce até o valor zero, ou seja, até uma distância em que não exista mais relação entre as amostras observadas (ROSSONI, 2011).

Desta forma, se a hipótese de estacionariedade de segunda ordem for satisfeita, as funções semivariância, covariância e correlação são maneiras equivalentes de se caracterizar a dependência espacial. Porém, se o fenômeno apresenta capacidade infinita de dispersão, em que somente a hipótese intrínseca é satisfeita, usa-se apenas a função semivariância. Por este motivo, nos estudos que envolvem a geoestatística, tem-se a preferência de se trabalhar com o semivariograma (SOARES, 2014).

3.3 Semivariograma

O semivariograma experimental é uma ferramenta utilizada na verificação da presença de dependência espacial entre os pontos amostrais georreferenciados espacialmente, e representa um gráfico dos valores de semivariância em função da distância (MORAL et al., 2010).

Os valores da semivariância podem ser calculados experimentalmente, considerando o esquema de amostragem em duas dimensões, apresentado na Figura 2, em que $Z(\mathbf{S}_1)$ representa o valor da variável observada na posição $\mathbf{S}_1 = (x_1, y_1)^T$, e $Z(\mathbf{S}_1 + \mathbf{h})$ é o valor da amostra na posição $\mathbf{S}_1 + \mathbf{h} = (x_2, y_2)$, e \mathbf{h} é o vetor distância que separa os pontos (OPROMOLLA et al., 2006).

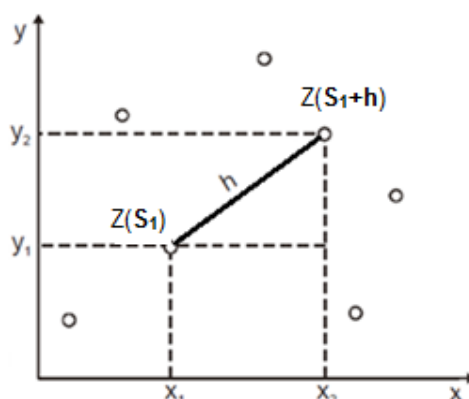


Figura 2 Amostragem em duas dimensões de uma variável regionalizada, com dois pontos separados por uma distância h .

Sendo assim, o semivariograma é representado por uma nuvem de pontos obtidos por um estimador da semivariância. Um dos estimadores mais comumente utilizados pela literatura é o modelo clássico proposto por Matheron (1963), definido como MORAL et al., 2010):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(\mathbf{S}_i) - Z(\mathbf{S}_i + \mathbf{h})]^2, \quad (9)$$

em que:

$\gamma(\mathbf{h})$ é o estimador da semivariância, obtida pelos valores amostrados;

$N(\mathbf{h})$ é o número de pares de valores medidos separados por uma distância h ;

$Z(\mathbf{S}_i)$ e $Z(\mathbf{S}_i + \mathbf{h})$ são os valores da variável na posição \mathbf{S}_i e $\mathbf{S}_i + \mathbf{h}$, de tal modo que esses pontos estão separados por uma distância h .

Contudo, Nogueira (2013) destaca que esse estimador apresenta como desvantagem o fato de ser influenciado com a presença de pontos discrepantes (*outliers*), que pode ser justificado pelo termo ao quadrado que aparece no somatório.

As suposições de média constante e estacionariedade de segunda ordem possibilitam prever um comportamento idealizado para o semivariograma, como ilustrado na Figura 3. Espera-se que amostras próximas geograficamente tenham comportamento mais semelhante do que observações separadas por maiores distâncias (CÂMARA et al., 2002). Deste modo, o valor da semivariância entre pares de observações $Z(\mathbf{S})$ e $Z(\mathbf{S} + \mathbf{h})$ deve crescer à proporção que aumenta a distância h entre elas, até um valor em que ela se mantém constante.

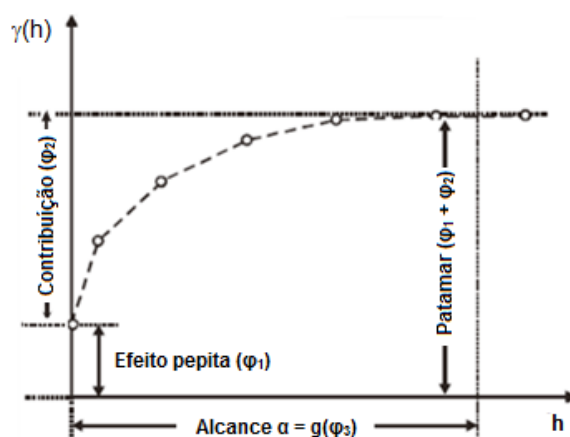


Figura 3 Exemplo de semivariograma com comportamento ideal

Quanto a este comportamento, Landim (2006) explica que se existe dependência espacial, então quanto mais próximos estiverem os pontos observados, maior será a semelhança entre eles e, conseqüentemente, menor será o valor da semivariância. E

quanto maior for a distância entre os pontos observados, menor será a semelhança, portanto, maior será o valor da semivariância.

Na Figura 3, estão representados os parâmetros do modelo que descreverá a função semivariância, no qual, segundo Nogueira (2013), podem ser definidos como:

- Alcance ($a = g(\varphi_3)$): distância máxima da dependência espacial, ou seja, indica que a partir deste ponto não existe mais dependência espacial entre as amostras;
- Efeito Pepita ($\varphi_1 \neq 0$): representa a descontinuidade do semivariograma para distâncias menores do que a menor distância observada na amostra. Teoricamente $\gamma(0) = 0$, porém na prática percebe-se que à medida que h tende a zero, $\gamma(h)$ aproxima-se de um valor positivo chamado efeito pepita. Este parâmetro está associado à variabilidade totalmente aleatória dos dados, ou seja, refere-se à variância do erro experimental;
- Contribuição (φ_2): é denominada como variância de dispersão e representa as diferenças espaciais entre os valores de uma variável tomada em dois pontos separados por distâncias cada vez maiores. É a diferença entre o patamar (C) e o efeito pepita (φ_1);
- Patamar ($C = \varphi_1 + \varphi_2$): é o valor da semivariância correspondente ao seu alcance (a). A partir desse ponto, considera-se que as amostras são independentes, porque a variância da diferença entre pares de amostras ($Var [Z(\mathcal{S}) - Z(\mathcal{S} + \mathbf{h})]$) torna-se aproximadamente constante.

Ainda, para o cálculo dos valores da semivariância, deve ser considerada a configuração das amostras. Ou seja, se o espaçamento entre as amostras é regular ou irregular. Nesta etapa deve ser escolhido o valor de h para os quais os valores da semivariância serão calculados, e por meio quais pontos amostrais as semivariâncias serão estimadas. Se a amostragem é regular, a escolha natural do h é a própria distância entre os pontos amostrais. Nesta situação, cada semivariância pode ser calculada com base em todos os pares de pontos amostrais separados pela mesma distância h (CAMARGO, 2015).

Já em amostragens irregulares, pode existir um número muito pequeno de pares de pontos amostrais espaçados exatamente pela mesma distância h . Desta forma, MCBratney e Webster (1986) recomendam que seja admitida uma tolerância Δh na distância, e uma tolerância $\Delta\Theta$ na direção, de modo a se atingir o requisito mínimo de pares amostrais para a estimativa da semivariância a cada distância. Isaaks e Srivastava (1989) evidenciam a necessidade da utilização de intervalos de abrangência e destacam que a distância h depende muito do tamanho da área experimental e do *cutoff* escolhido, porém, quanto ao ângulo, os autores propõem uma tolerância de 40° .

A somatória para o cálculo da função semivariância $\gamma(h)$ deve ser composta por um número suficiente de pares, que torne o resultado representativo. Como regra prática

adotam-se no mínimo 30 pares de pontos georreferenciados, considerando 50% da distância máxima da área pesquisada, sendo esta chamada de *cutoff* (LANDIM, 2006).

O estimador da função semivariância de Matheron (1963) permite estruturar duas versões do semivariograma, sendo o experimental direcional e o experimental omnidirecional.

O semivariograma direcional permite analisar o comportamento espacial da variável georreferenciada em diferentes direções. Frequentemente, o semivariograma é construído nas direções 0°, 45°, 90° e 135° do eixo adotado pelo pesquisador (YAMAMOTO & LANDIM, 2013). Assim, é possível analisar se o fenômeno é isotrópico ou anisotrópico. Estes aspectos serão discutidos na próxima seção.

Depois de ser analisado o fenômeno nas diferentes direções e concluído que a função semivariância possui apenas dependência espacial com relação à distância entre as amostras e que não dependem da direção analisada, obtém-se então o semivariograma experimental omnidirecional. O semivariograma experimental omnidirecional é determinado com os mesmos procedimentos, porém quando são calculadas as semivariâncias, são utilizadas todas as direções possíveis (SOARES, 2014). Por exemplo, suponha uma grade regular em que a distância entre dois pontos consecutivos seja igual a 100 metros ($h = 100$). Então para qualquer par de observações em todas as direções possíveis, cuja distância seja igual a 100 metros, esse será incluído no cálculo da semivariância, denotada por $\gamma(100)$. Isto feito, os cálculos serão repetidos para as demais distâncias (YAMAMOTO & LANDIM, 2013).

3.4 Anisotropia

A continuidade espacial de um recurso natural pode variar com as diferentes direções do espaço. Uma característica de um recurso natural diz-se que tem uma estrutura de continuidade espacial isotrópica quando o semivariograma tem o mesmo comportamento em todas as direções. Isto é $\gamma(\mathbf{h})$ depende somente do módulo do vetor \mathbf{h} . Todavia quando a variabilidade espacial expressa pelo semivariograma não é a mesma em todas as direções, o fenômeno é chamado de anisotrópico (SOARES, 2014).

Os principais tipos de anisotropia encontrados na natureza são: geométrica, zonal e combinada. A anisotropia geométrica é aquela em que existe uma direção com maior continuidade espacial, isto é, maior valor de alcance (a) no semivariograma experimental em determinada direção. A anisotropia zonal é aquela em que existe uma direção com maior valor de patamar ($\varphi_1 + \varphi_2$) nos semivariogramas experimentais em

relação às demais direções. E a anisotropia combinada existe quando houver determinadas direções com diferentes valores de alcance e patamar nos semivariogramas experimentais (GUEDES et al., 2008).

Ao se detectar a presença de anisotropias, elas devem ser corrigidas para obtenção de um semivariograma isotrópico com parâmetros comuns (efeito pepita, patamar e alcance) em todas as direções (YAMAMOTO & LANDIM, 2013).

A anisotropia geométrica é corrigida por transformações lineares, as quais são usadas na rotação e dilatação das coordenadas espaciais, utilizando-se a notação matricial (DIGGLE & RIBEIRO JUNIOR, 2007):

$$h' = (x, y) \times M_R \times M_D \quad (10)$$

em que, $M_R = \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix}$: Matriz de rotação; $\begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{F_a} \end{bmatrix}$: Matriz de dilatação e F_a o fator de anisotropia e α o ângulo de maior continuidade espacial.

A anisotropia zonal pode ser corrigida por um semivariograma direcional equivalente à distância reduzida, considerando-se como patamar o maior valor de patamar apresentado entre os semivariogramas direcionais construídos (ISAAKS & SRIVASTAVA, 1989). A função semivariância corrigida pode ser descrita como:

$$\gamma'(h) = w\gamma(h') \quad (11)$$

em que $h' = \frac{h}{a'}$, sendo w e a' os valores de patamar e alcance do semivariograma na direção que apresentaram a anisotropia zonal.

Porém, para a anisotropia combinada, que é a combinação das anisotropias geométrica e zonal, Isaaks e Srivastava (1989) propuseram a equação (12) para o estudo da anisotropia combinada, em que a primeira etapa consiste em corrigir a anisotropia geométrica e, a segunda, em corrigir a anisotropia zonal.

$$\gamma'(h) = w_1\gamma_1(h') + w_2\gamma_2(h') \quad (12)$$

em que w_1 representa o patamar do semivariograma direcional que apresentou o maior valor de alcance, e w_2 , o que apresentou menor alcance entre os semivariogramas direcionais; $h' = \sqrt{\left(\frac{x}{a_x}\right)^2 + \left(\frac{y}{a_y}\right)^2}$, em que a_x e a_y representam os alcances nas direções x e y , respectivamente.

3.5 MODELOS TEÓRICOS

O semivariograma experimental elaborado com base nos valores da semivariância estimadas de Matheron (1963), a partir do conjunto amostral tem como objetivo permitir que o pesquisador entenda o comportamento da estrutura de dependência espacial da variável georreferenciada sob estudo. A partir desse, pretende-se ajustar um modelo que represente a função semivariância, posto que é essencial que o modelo reproduza a tendência de $\gamma(h)$ em relação a valores de h (CAMARGO, 2015).

O método de escolha do modelo não é simples e direto, uma vez que esta etapa depende da interpretação do pesquisador, no qual faz um primeiro ajuste e verifica a adequação do modelo teórico. Dependendo do resultado alcançado, pode-se ou não redefinir o modelo e até obter um que melhor represente o fenômeno pesquisado (FARACO et al., 2008).

Os modelos aqui apresentados são considerados isotrópicos e estão divididos em dois grupos: modelos com patamar e modelos sem patamar. Os modelos com patamar são caracterizados pela geoestatística como transitivos. Parte dos modelos transitivos alcança o patamar (C) assintoticamente. Nesses modelos, o alcance (a) é parcialmente determinado como a distância correspondente a 95% do patamar. Já os modelos sem patamar, os valores da semivariância continuam aumentando enquanto a distância aumenta. Tais modelos são aplicados para fenômenos que apresentam capacidade infinita de dispersão, ou seja, atendem somente à hipótese de estacionariedade intrínseca. Os modelos com patamar mais usados são: modelo esférico, modelo exponencial, modelo gaussiano e família Matérn (ISAAKS & SRIVASTAVA, 1989).

3.5.1 Modelo Esférico

O modelo esférico possui uma estrutura de correlação espacial que aumenta com a distância até certo ponto (alcance). A partir de então, os valores da semivariância tornam-se constantes e limitam a área de dependência espacial de cada amostra (PAPANI, 2016).

A equação do modelo esférico da função semivariância com efeito pepita igual a φ_1 , patamar igual a $\varphi_1 + \varphi_2$, e φ_3 função do alcance (a), é dada por:

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[\frac{3}{2} \left(\frac{h}{\varphi_3} \right) - \frac{1}{2} \left(\frac{h}{\varphi_3} \right)^3 \right], & 0 < h \leq \varphi_3 \\ \varphi_1 + \varphi_2 & h > \varphi_3. \end{cases} \quad (13)$$

Sendo assim, o semivariograma do modelo esférico é apresentado da seguinte forma:

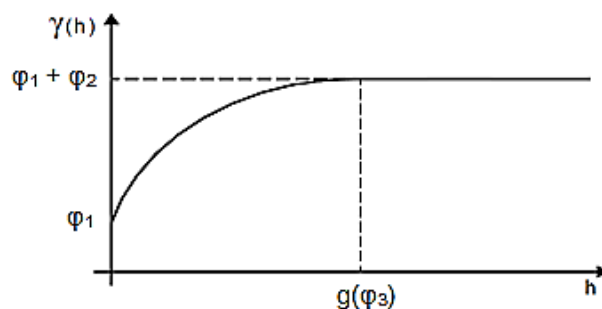


Figura 4 Representação gráfica do modelo esférico com $\varphi_1 \neq 0$

A função covariância é dada por:

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & h = 0 \\ \varphi_2 \left[1 - \frac{3}{2} \left(\frac{h}{\varphi_3} \right) + \frac{1}{2} \left(\frac{h}{\varphi_3} \right)^3 \right], & 0 < h \leq \varphi_3 \\ 0, & h > \varphi_3. \end{cases} \quad (14)$$

A função de correlação espacial é definida por:

$$\rho(h) = \begin{cases} 1, & h = 0 \\ 1 - \frac{3}{2} \left(\frac{h}{\varphi_3} \right) + \frac{1}{2} \left(\frac{h}{\varphi_3} \right)^3, & 0 < h \leq \varphi_3. \end{cases} \quad (15)$$

De acordo com VIERA et al. (1983), o modelo esférico é obtido primeiramente com a seleção dos valores de efeito pepita φ_1 , e do patamar $\varphi_1 + \varphi_2$, na sequência é traçada uma reta que intercepte o eixo-y em φ_1 e que seja tangente aos primeiros pontos próximos de $h = 0$. Essa reta cortará o patamar numa distância $h' = 2/3a$. Deste modo, o alcance será $a = (3/2)h'$. O modelo esférico é linear até aproximadamente $1/3a$.

Papani (2016) ainda acrescenta a respeito do comportamento linear desse modelo da função semivariância para pequenos valores de h . Cressie (1993) complementa ainda que os modelos esféricos são válidos nos espaços R , R^2 e R^3 .

3.5.2 Modelo Exponencial

O modelo Exponencial também descreve a função semivariância. A equação do modelo exponencial para a função semivariância, com efeito pepita, φ_1 , e patamar, $\varphi_1 + \varphi_2$, é dada por:

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[1 - \exp\left(-\frac{h}{\varphi_3}\right) \right], & h > 0. \end{cases} \quad (16)$$

Em que, h é a máxima distância no qual o semivariograma é definido. O alcance (a) é determinado visualmente pela distância na qual o semivariograma se estabiliza. A determinação dos parâmetros φ_1 e φ_2 para o modelo é semelhante à do modelo esférico.

Este modelo atinge o patamar assintoticamente com o alcance prático ou à distância na qual o valor do modelo é 95% de $(\varphi_1 + \varphi_2)$ (ISAAKS; SRIVASTAVA, 1989). Desta forma, o alcance prático é $a = 3\varphi_3$. Sendo assim, o semivariograma do modelo exponencial apresenta a seguinte forma:

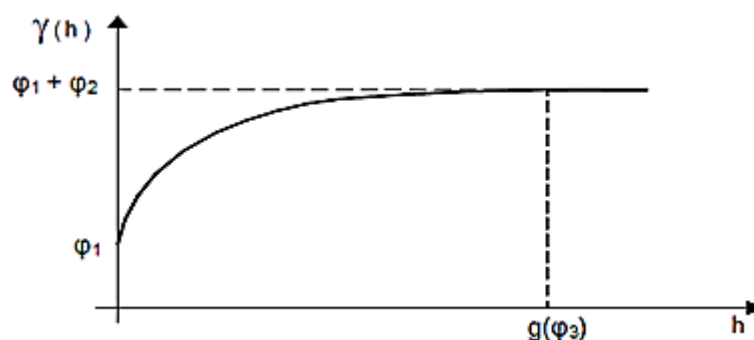


Figura 5 Representação gráfica do modelo exponencial com $\varphi_1 \neq 0$

A função covariância é dada por:

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & h = 0 \\ \varphi_2 \left[\exp\left(-\frac{h}{\varphi_3}\right) \right], & h > 0. \end{cases} \quad (17)$$

A função da correlação espacial é dada por:

$$\rho(h) = \begin{cases} 1, & h = 0 \\ \left[\exp\left(-\frac{h}{\varphi_3}\right) \right], & h > 0. \end{cases} \quad (18)$$

A tangente que intercepta a origem atinge o patamar a um terço do alcance. Do mesmo modo que o modelo esférico, o exponencial é válido nos espaços R , R^2 e R^3 (CRESSIE, 1993).

3.5.3 Modelo Gaussiano

O modelo Gaussiano apresenta comportamento parabólico na origem e é utilizado para modelar um fenômeno extremamente contínuo (ISAAKS & SRIVASTAVA, 1989).

Do mesmo modo que o modelo exponencial, o gaussiano atinge o patamar assintoticamente e o parâmetro a é definido como o alcance prático ou à distância na qual o valor do modelo é 95% do patamar. O que diferencia este modelo dos demais é o seu comportamento parabólico próximo à origem e é o único modelo que apresenta em sua forma um ponto de inflexão (ISAAKS & SRIVASTAVA, 1989).

Este modelo, assim como os demais, é válido apenas nos R , R^2 e R^3 (CRESSIE, 1993), e é dado pela seguinte fórmula:

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left\{ 1 - \exp \left[- \left(\frac{h}{\varphi_3} \right)^2 \right] \right\}, & h > 0. \end{cases} \quad (19)$$

Logo, o semivariograma do modelo gaussiano é apresentado como:

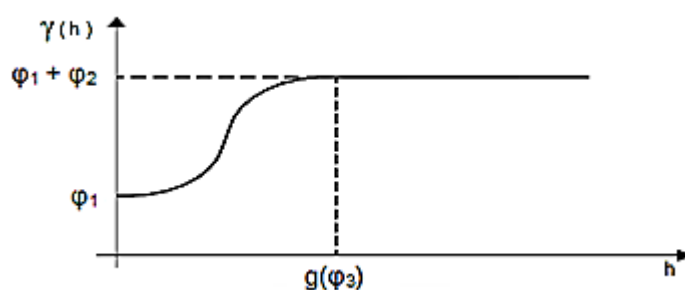


Figura 6 Representação gráfica do modelo gaussiano com $\varphi_1 \neq 0$

A função covariância é dada por:

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & h = 0 \\ \varphi_2 \left\{ \exp \left[- \left(\frac{h}{\varphi_3} \right)^2 \right] \right\}, & h > 0. \end{cases} \quad (20)$$

Os parâmetros φ_1 e φ_2 do modelo são definidos da mesma maneira que o obtido no modelo esférico.

A função correlação é dada por:

$$\rho(h) = \begin{cases} 1, & h = 0 \\ \left[\exp \left(- \frac{h}{\varphi_3} \right)^2 \right], & h > 0. \end{cases} \quad (21)$$

3.5.4 Família Matérn

Matérn apresentou uma função conhecida como família Matérn. E, em termos de função semivariância, pode ser definida por (PAPANI et al., 2016):

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[1 - \left(2^{k-1} \Gamma(k) \right)^{-1} \left(\frac{h}{\varphi_3} \right)^k K_k \left(\frac{h}{\varphi_3} \right) \right], & h > 0. \end{cases} \quad (22)$$

em que: φ_1 , φ_2 , φ_3 e k são os parâmetros da função semivariância e, K_k é a função de Bessel modificada do terceiro tipo e de ordem k , sendo $k_k(u) = \frac{1}{2} \int_0^\infty x^{k-1} e^{-\frac{1}{2}u(x+\frac{1}{x})} dx$. Esta função é válida para φ_3 e k maior que zero. Nesta, o parâmetro k é chamado de ordem e consiste em um parâmetro de forma que determina a suavização analítica do processo subjacente $Z(\mathcal{S})$. Especificamente, $Z(\mathcal{S})$ é $[k - 1]$ vezes diferenciável, em que $[k]$ denota, o menor inteiro, maior ou igual a k (DE BASTIANI, 2012). Sendo assim, define-se a função covariância por:

$$C(h) = \begin{cases} 0, & h = 0 \\ \varphi_2 \left[\left(2^{k-1} \Gamma(k) \right)^{-1} \left(\frac{h}{\varphi_3} \right)^k K_k \left(\frac{h}{\varphi_3} \right) \right], & h > 0. \end{cases} \quad (23)$$

A função correlação é dada por:

$$\rho(h) = \begin{cases} 1 & , se h = 0 \\ \left\{ 2^{k-1} \Gamma(k) \right\}^{-1} \left(\frac{h}{\varphi_3} \right)^k K_k \left(\frac{h}{\varphi_3} \right), & h > 0 \end{cases} \quad (24)$$

4 ESTIMAÇÃO DOS PARÂMETROS NO AJUSTE DE MODELOS TEÓRICOS

Para escolher um modelo que melhor represente a estrutura de dependência espacial do fenômeno estudado entre os modelos válidos, é necessário estimar seus parâmetros. Esta etapa não é simples e direta, uma vez que requer um bom programa interativo, além de exigir do pesquisador experiência e habilidade para sintetizá-lo.

Segundo Barnes (1991), o método mais utilizado é estimar o patamar como a variância dos dados amostrados. Porém, o referido autor ressalva que esse método não é o mais adequado. Pois, apesar de o patamar corresponder a variância dos dados, o verdadeiro disto é incógnito. E a variância da amostra é apresentada somente como uma estimativa de qualidade se a área amostrada abrange múltiplos alcances.

De acordo com McBratney e Webster (1986), o mais adequado é fazer o uso de procedimentos estatísticos para estimar os parâmetros desconhecidos do vetor $\theta = (\varphi_1, \varphi_2, a)^T$ de um modelo geoestatístico. Os modelos estatísticos mais utilizados são: Mínimos Quadrados Ordinários, Mínimos Quadrados Ponderados, Máxima

Verossimilhança e Máxima Verossimilhança Restrita; porém, para realização desta pesquisa, será apresentado apenas o método de Máxima verossimilhança.

4.1 Método de Máxima Verossimilhança (MLE)

A função de verossimilhança de uma amostra aleatória Z_1, Z_2, \dots, Z_n com distribuição gaussiana multivariada, retirada de uma população, é definida como uma função de densidade de probabilidade conjunta $f_{Z_1, Z_2, \dots, Z_n}(Z_1, \dots, Z_n; \theta)$, com um vetor de p parâmetros desconhecidos $\theta = (\beta^T, \phi^T)^T \in \Theta$ (espaço paramétrico) (BORSSOI, 2007). Logo, a função verossimilhança é definida por:

$$L(\theta) = f_{Z_1, Z_2, \dots, Z_n}(Z_1, \dots, Z_n; \theta). \quad (25)$$

Neste método, assume-se que os valores das estimativas do vetor de parâmetros são os que maximizam a função de verossimilhança (BORSSOI, 2011). Para fins de simplicidade dos cálculos, utiliza-se o logaritmo da função verossimilhança, definido por:

$$l(\theta) = -\left(\frac{n}{2}\right) \log(2\pi) - \frac{1}{2} \log|\Sigma| - \frac{1}{2} (\mathbf{Z} - \mathbf{X}\beta)^T \Sigma^{-1} (\mathbf{Z} - \mathbf{X}\beta). \quad (26)$$

em que:

$\theta = (\beta^T, \phi^T)^T$, $\beta = (\beta_1, \dots, \beta_n)^T$ e $\phi = (\varphi_1, \varphi_2, \varphi_3)^T$, assumindo-se um fenômeno isotrópico.

A função máxima verossimilhança se resume em uma função do vetor de parâmetros $\theta \in \Theta$, em que Θ é o espaço paramétrico. Então, a melhor estimativa para o vetor θ de parâmetros será aquela que maximiza o logaritmo da função verossimilhança, ou seja, $l(\hat{\theta}) = \max l(\theta)$, $\theta \in \Theta$, em que o $\hat{\theta}$ é o vetor de estimador da função máxima verossimilhança.

4.1.1 Erros-padrão assintóticos das estimativas dos parâmetros

Em toda análise estatística que envolve a estimação de parâmetros, tem-se a necessidade de quantificar o erro associado às estimativas produzidas, uma vez que auxiliam na avaliação da qualidade do ajuste do modelo. Para isso, é usual considerar as variâncias, ou seja, o desvio padrão dos estimadores.

Utiliza-se o inverso da matriz de informação esperada para a estimação dos erros-padrão assintóticos dos parâmetros estimados θ . A matriz de informação esperada de um modelo isotrópico com distribuição normal de probabilidade multivariada, com vetor de médias zero e matriz de covariância Σ , ou seja, $\epsilon \sim N_n(\mathbf{0}, \Sigma)$ é dada por (URIBE-OPAZO et al., 2012):

$$\mathbf{K}(\theta) = \begin{pmatrix} K(\mu) & \mathbf{0} \\ \mathbf{0} & \mathbf{K}(\varphi) \end{pmatrix} \quad (27)$$

em que:

$\mathbf{K}(\mu) = \mathbf{X}^T \Sigma^{-1} \mathbf{X}$ e a matriz $\mathbf{K}(\varphi) = [(k_{bc}(\varphi))]$. Com elementos, $k_{bc}(\varphi) = \frac{1}{2} \text{tr} \left(\Sigma^{-1} \left[\frac{\partial \Sigma}{\partial \varphi_b} \right] \Sigma^{-1} \left[\frac{\partial \Sigma}{\partial \varphi_c} \right] \right)$ para $b, c = 1, 2, 3$

Visto que, a primeira derivada de Σ com relação a φ_1, φ_2 e φ_3 é respectivamente dada por:

- a) $\frac{\partial \Sigma}{\partial \varphi_1} = \mathbf{I}_n$ e $\frac{\partial \Sigma}{\partial \varphi_2} = \mathbf{R}$,
- b) $\frac{\partial \Sigma}{\partial \varphi_2} = \varphi_2 (\partial \mathbf{R} / \partial \varphi_3)$,
- c) $\partial \mathbf{R} / \partial \varphi_3 = [(\partial r_{ij} / \partial \varphi_3)]$ para $i, j = 1, \dots, n$.

A primeira derivada de r_{ij} , com relação a φ_3 para a função covariância do modelo Exponencial, Gaussiano e Matérn é apresentada respectivamente pelas Equações (28), (29) e (30) (URIBE-OPAZO et al., 2012):

$$\frac{\partial r_{ij}}{\partial \varphi_3} = r_{ij} h_{ij} / \varphi_3^2, \quad (28)$$

$$\frac{\partial r_{ij}}{\partial \varphi_3} = r_{ij} (h_{ij})^2 / \varphi_3^3, \quad (29)$$

$$\frac{\partial r_{ij}}{\partial \varphi_3} = -(1/\varphi_3) \left[k r_{ij} + [2^{k-1} \Gamma(k)]^{-1} \left(h_{ij} / \varphi_3 \right)^{k+1} K'_k \left(h_{ij} / \varphi_3 \right) \right]. \quad (30)$$

Para $i, j = 1, \dots, n$ em que Γ é a função gama. $K'_k = \partial K_k(t) / \partial t = -(1/2)[K_{k+1}(t) + K_{k-1}(t)]$. $K_k(t) = (1/2) \int_0^\infty x^{k-1} \exp[-t(w+w^{-1}/2)] dw$ é a função Bessel modificada de terceira ordem k , com $k > 0$ fixo.

5 CRITÉRIO DE SELEÇÃO DOS MODELOS

A seleção dos modelos teóricos ocorre por validação, em que são realizadas comparações entre valores teóricos de modelos geoestatísticos e valores empíricos. Os resultados da validação são comumente utilizados para comparar a distribuição da estimação de erros ou resíduos dos diferentes procedimentos (FARACO et al., 2008).

A comparação na maioria das vezes não indica qual é a melhor opção, porém os resíduos da validação possuem importantes informações que, se forem adequadamente estudados, podem prever indícios de problemas em um processo de estimação. Sendo assim, na sequência, apresentam-se os critérios de Validação Cruzada, Informação de Akaike e Informação Bayesiana.

5.1 Validação cruzada

Segundo Grzegozewski (2012), a validação cruzada é uma técnica que consiste em avaliar os erros de estimativa da variável aleatória, pela comparação de valores estimados e amostrados. Tal comparação ajuda a escolher, dentre os distintos modelos de estimação, aquele que mais se aproxima da função semivariância (RAGAGNING et al., 2010).

Neste método, uma amostra $Z(s_i)$ é temporariamente retirada da amostra de dados. Então, é feita uma estimação de $Z(s_i)$, denotada por $Z^*(s_i)$ usando um interpolador e as informações dos demais elementos amostrais. Após esta estimação, o valor real da amostra é reintroduzido nos dados e o processo se repete para cada uma das demais amostras (SILVA et al., 2013).

Conhecendo-se os valores amostrados e estimados, é possível obter a variância total da estimativa, para aferir a qualidade do processo. Assim, espera-se que os erros de estimação (Equação 31) apresentem média nula, variância constante e distribuição normal de probabilidade (SOARES, 2014).

$$\varepsilon(S_i) = Z(S_i) - Z^*(S_i), \quad (31)$$

em que $Z^*(S_i)$ é o valor predito por krigagem sem a i -ésima observação, sem $Z(S_i)$.

Portanto, de acordo com Grzegozewski (2012), o erro de estimação não significa apenas a eficácia do ajuste dos modelos teóricos a semivariogramas experimentais, mas também afere em relação à estacionariedade e à presença de valores atípicos. A comparação dos modelos é realizada por erro médio (EM), erro médio reduzido (ER), desvio padrão dos erros médios (S_{EM}), desvio padrão dos erros reduzidos (S_{ER}) e do erro

absoluto (EA), apresentados pelas Equações (32), (33), (34), (35) e (36), respectivamente (FARACO et al., 2008):

$$EM = \frac{1}{n} \sum_{i=1}^n (Z(S_i) - Z^*(S_i)), \quad (32)$$

$$ER = \frac{1}{n} \sum_{i=1}^n \frac{Z(S_i) - Z^*(S_i)}{\sigma(Z^*(S_i))}, \quad (33)$$

$$S_{EM} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Z(S_i) - Z^*(S_i))^2}, \quad (34)$$

$$S_{ER} = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{(Z(S_i) - Z^*(S_i))^2}{\sigma(Z^*(S_i))^2}}, \quad (35)$$

$$EA = \sum_{i=1}^n |Z(S_i) - Z^*(S_i)|, \quad (36)$$

em que:

n é o número de observações amostrais; $Z(S_i)$ é o valor observado no ponto S_i ; $Z^*(S_i)$ é o valor estimado por krigagem no ponto S_i , sem considerar a observação $Z(S_i)$ na estimação do modelo para a função semivariância e no processo da krigagem; e $(\sigma(Z^*(S_i)))$ é o desvio padrão da krigagem no ponto S_i , sem considerar a observação $Z(S_i)$.

Segundo Mello et al. (2005) e Faraco et al. (2008), ao se aplicar a restrição de não tendenciosidade, o valor populacional para o erro médio reduzido deve ser zero e do desvio padrão do erro reduzido deve ser igual a um. Portanto, os valores de EM e ER mais próximos de zero, os valores de S_{EM} e EA menor e o valor de S_{ER} mais próximo de um são os critérios para escolha do melhor modelo ajustado.

5.2 Informação de Akaike

A Informação de Akaike, segundo Sobral e Barreto (2011), pode ser explicada como um critério que dá pontuação ao modelo, baseado em sua adequação e na ordem do modelo. Este critério busca escolher o modelo mais simples entre dois modelos que apresentam dados igualmente satisfatórios (AKAIKE, 1973). A Informação de Akaike (Equação 37) é definida como:

$$AIC = -2l(\theta) + 2r, \quad (37)$$

em que:

$l(\theta)$ é o valor da função log-verossimilhança, obtido na otimização por ML e,

r é o número de parâmetros do modelo estimado.

O critério de informação de Akaike avalia a qualidade da estimação do modelo paramétrico pelo método máxima verossimilhança, com uma medida relativa das informações perdidas, quando determinado modelo é utilizado para descrever a realidade (EMILIANO, 2013).

O modelo que apresentar a menor estimativa de AIC é considerado o modelo de melhor ajuste (EMILIANO, 2013).

Grzegozewski (2012) ainda destaca que, pelo motivo de a informação de Akaike basear-se na função verossimilhança, o número de observações não pode ser pequeno, ($n/r < 40$). Sendo então, neste caso, recomendado utilizar o critério de Akaike de segunda ordem, definido como:

$$CAIC = -2l(\theta) + 2r + \frac{2r(r+1)}{n-r-1}. \quad (38)$$

Para um menor valor de r , o valor de $CAIC$ tende para AIC , quando o número de observações (n) cresce. Da mesma forma, o modelo que apresentar a menor estimativa de $CAIC$ é considerado o modelo de melhor ajuste.

5.3 Critério de Informação Bayesiano

O critério de Informação Bayesiano (BIC), assim como AIC, penaliza a função verossimilhança para que um modelo mais parcimonioso seja selecionado (EMILIANO, 2013). O BIC é expresso por:

$$BIC = \frac{(-2l(\theta))}{n} + \frac{(k \log n)}{n}, \quad (39)$$

em que: k é o número de restrições; n é o tamanho da amostra; e $l(\theta)$ é o valor da função log-verossimilhança obtido na otimização por ML.

O modelo que apresentar a menor estimativa de BIC é considerado o modelo de melhor ajuste (EMILIANO, 2013).

6 KRIGAGEM

A Krigagem é o método de interpolação da geoestatística que utiliza o valor de uma variável regionalizada e sua posição geográfica para interpolar, ou seja, faz a estimação de valores da variável georeferenciada em locais não amostrados, a partir de valores

conhecidos (amostrados) e do modelo estimado para a função semivariância (que expressa a estrutura de dependência espacial da variável georreferenciada em estudo. Esta técnica distingue-se dos demais algoritmos, pelo fato de fornecer além dos valores estimados, o “erro” associado a tal estimativa (CARDOSO et al., 2016). A krigagem utiliza a função semivariância para encontrar os pesos corretos associados às amostras que irão estimar um ponto.

Esta técnica cumpre dois critérios em relação ao erro de estimação ($(\varepsilon(S_i) = [Z^*(S_i)] - Z(S))$): o erro é uma variável aleatória com esperança matemática igual a zero, ($E[\varepsilon(S_i)] = 0$), e a variância de estimação mínima ($\min\{\text{var}[\varepsilon(S_i)]\}$), em que, $i = 1, \dots, n$ (SOARES, 2014).

Fazio (2013) destaca que, pelo fato da krigagem estar fundamentada nestes dois princípios, este possui grande capacidade de produzir estimativas de qualidade em termos de interpolação. Estas duas características fazem da krigagem um estimador BLUE (*Best Linear Unbiased Estimator* - melhor estimador linear não-viciado).

A interpolação por Krigagem permite, a partir da estrutura de dependência espacial das variáveis regionalizadas em estudo, estimar valores em locais não amostrados e construir mapas temáticos com alta precisão (GRZEGOZEWSKI, 2012). O termo krigagem abrange um conjunto de métodos de estimação, cujos mais usuais são a krigagem simples, ordinária, universal, indicatriz e co-krigagem (LANDIM, 2002). Porém, neste estudo, para efeito de comparação, serão discutidas apenas a krigagem simples, ordinária e universal.

De acordo com Saghafian e Bondarabadi (2009) e Bettini (2007), o que basicamente diferencia estes métodos é a maneira como os pesos são atribuídos, e para encontrar os pesos é necessário saber qual a tendência utilizada.

Na Krigagem simples (KS), admite-se que a média local μ é constante e conhecida e com tendência igual a zero. Desta forma, a média da população é utilizada para cada estimação local, em conjunto com os pontos vizinhos estabelecidos como necessários para a estimação (ANDRIOTTI, 2002). A KS é expressa como:

$$Z^*(S_0) = \sum \lambda_i Z(S_i) + \mu(1 - \sum \lambda_i) \quad (40)$$

Já a Krigagem ordinária (KO) concebe uma forma de estimação linear para uma variável regionalizada que atende à hipótese intrínseca, ou seja, não requer o conhecimento prévio da média e assume-se a hipótese de estacionariedade local, logo, utilizam-se apenas os pontos vizinhos para estimação (ROBINSON et al., 2013).

A KO foi formulada na ideia de regressão linear, em que a predição de um valor desconhecido $Z^*(S_i)$ é obtido pela combinação linear de n valores conhecidos $Z(S_i)$ adicionada a um parâmetro λ_0 (AGUIRRE et al., 2013):

$$Z^*(S_i) = \lambda_0 + \sum_{i=1}^n \lambda_i Z(S_i) \quad (41)$$

Este processo requer que a soma dos pesos seja igual a um $\sum_{i=1}^n \lambda_i = 1$, e que o parâmetro λ_0 seja igual a zero (LANDIM, 2002). Assim, a soma dos pesos das amostras igual a um garante a não tendenciosidade na estimativa de $Z^*(S)$. E, para que a variância seja mínima, sob a condição da soma dos pesos, faz-se necessária a introdução do interpolador de Lagrange para dedução das equações (OLIVEIRA et al., 2013). O sistema resultante pode ser expresso por:

$$\begin{aligned} \sum_{j=1}^n \lambda_j (S_i, S_j) - \eta &= \gamma(S_0, S_i) \\ \sum_{i=1}^n \lambda_i &= 1 \end{aligned} \quad (42)$$

O sistema expresso em (42), constituído de n+1 equações e n+1 incógnitas, é conhecido como sistema de krigagem ordinária. A incógnita η é um multiplicador de Lagrange, introduzido ao minimizar a variância do erro. Na forma matricial, o sistema pode ser escrito como:

$$\gamma \lambda = A \quad (43)$$

em que:

$$\begin{aligned} \lambda &= [\lambda_1 \ \lambda_2 \ \lambda_3 \ \dots \ \lambda_n \ \eta]^T \\ \gamma &= \begin{bmatrix} \gamma(S_1, S_1) & \gamma(S_1, S_2) & \dots & \gamma(S_1, S_n) & 1 \\ \gamma(S_2, S_1) & \gamma(S_2, S_2) & \dots & \gamma(S_2, S_n) & 1 \\ & & \vdots & & \\ \gamma(S_n, S_1) & \gamma(S_n, S_2) & \dots & \gamma(S_n, S_n) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \\ A &= [\gamma(S_1, S_0) \ \gamma(S_2, S_0) \ \dots \ \gamma(S_n, S_0) \ 1] \end{aligned}$$

Assim, os pesos λ podem ser obtidos mediante o produto da matriz inversa de γ pela matriz A , isto é, $\lambda = \gamma^{-1}A$. Uma vez obtidos os pesos λ e o valor de η , pode-se também calcular a variância do erro que é dada por:

$$\sigma^2 E = \lambda_i \gamma(S_i, S_0) + \eta \quad (44)$$

A krigagem universal (KU) é utilizada quando é observada realmente nos dados uma tendência relacionada a uma direção ou qualquer outro tipo de covariável (FAZIO, 2013). Além disso, modeliza-se a estrutura de correlação espacial quando a média não é mais constante e a semivariância ou a covariância dos dados originais não são mais apropriados. Vale ressaltar que são necessários um semivariograma dos resíduos e um modelo para descrever a tendência (ANDRIOTTI, 2002).

Logo, a krigagem universal segundo Diggle e Ribeiro Junior (2007) consiste em extrair da variável original $Z(\mathcal{S})$ a parte não estacionária a partir de uma componente determinística $\mu(\mathcal{S})$ que representa a tendência até encontrar a parte estacionária do fenômeno, obtendo-se um processo estocástico $e(\mathcal{S})$, relacionado pela Equação (1). Para o componente determinístico, é utilizada uma função polinomial das coordenadas para modelar a tendência (SANTOS et al., 2011), isto é:

$$\mu(\mathcal{S}) = \sum_{i=1}^p \alpha_i f_i(\mathcal{S}) \quad (45)$$

em que:

α_i são os coeficientes;

p é o número de termos usados para aproximar $\mu(\mathcal{S})$;

f_i é a função que descreve a tendência.

Quanto à utilização desta técnica, Fazio (2013) destaca que, no caso da inexistência de tendência direcional nos dados, a krigagem universal gera resultados piores que a krigagem ordinária.

7 AMOSTRAGEM ESPACIAL

O objetivo principal de uma pesquisa experimental é compreender características específicas de um grupo de interesse, seja ele um grupo de animais, pessoas ou uma área agrícola. Este grupo, denominado população, consiste em um conjunto de indivíduos com uma característica em comum. Como existem limitações tais como disponibilidade de tempo e recursos financeiros, torna-se inviável e inexecutável a utilização de toda a população. Desta forma, utiliza-se apenas uma parte da população, denominada de amostra. A maneira de se escolher as unidades amostrais na população de interesse é realizada pela técnica de amostragem (GUEDES, 2008). O método de escolha da amostra permite fazer o controle inicial da pesquisa quando as características estranhas à amostra são irrelevantes e constantes (MACHADO et al., 2005).

Para o caso bidimensional, os métodos de amostragem espacial são uma maneira racional de se empregarem as ferramentas estatísticas para a definição do tamanho amostral n e das localizações $(\mathcal{S}_1, \mathcal{S}_2 \dots \mathcal{S}_n)$ na região de interesse $D \subset R^d$ ($d = 2$) e, por conseguinte, definir a configuração da amostra (GUEDES, 2008).

Walvoort et al. (2010) destacam a importância do desenvolvimento de esquemas de amostragens mais eficientes, no sentido de que a amostra represente as condições reais da área observada. O propósito dessa eficiência pode ser a previsão dos valores de Z em locais não amostrados, ou ainda, a estimação dos parâmetros do modelo ajustado à

semivariância em função da distância. Existem vários planos de amostragem espacial, e sua escolha além de estar restrita à disponibilidade de tempo e de recursos econômicos também depende das características da área em estudo e da distribuição espacial da variável em estudo. Aqui serão apresentados alguns planos amostrais disponíveis na literatura.

Guedes (2008) discute três planos amostrais clássicos: a amostragem aleatória simples e sistemática determinística. A amostragem aleatória simples consiste em escolher pontos aleatoriamente por uma distribuição uniforme sobre a região D . A amostragem sistemática consiste em escolher a localização de um ponto amostral de forma aleatória e, em seguida, especificar as $(n - 1)$ localizações dos pontos amostrais, de tal forma que todos os pontos estejam localizados segundo algum padrão regular. Quando a localização do ponto inicial não é escolhida de forma aleatória, então o plano de amostragem é denominado determinístico ou regular. Sua principal característica é a estrutura geométrica da localização das amostras, cuja área de interesse é dividida em polígonos com todos os ângulos internos iguais e com os elementos da amostra posicionados nos vértices dos polígonos. Nas Figuras 7(a) e 7(b) estão representados exemplos dos métodos de amostragem aleatória simples e sistemática, respectivamente.

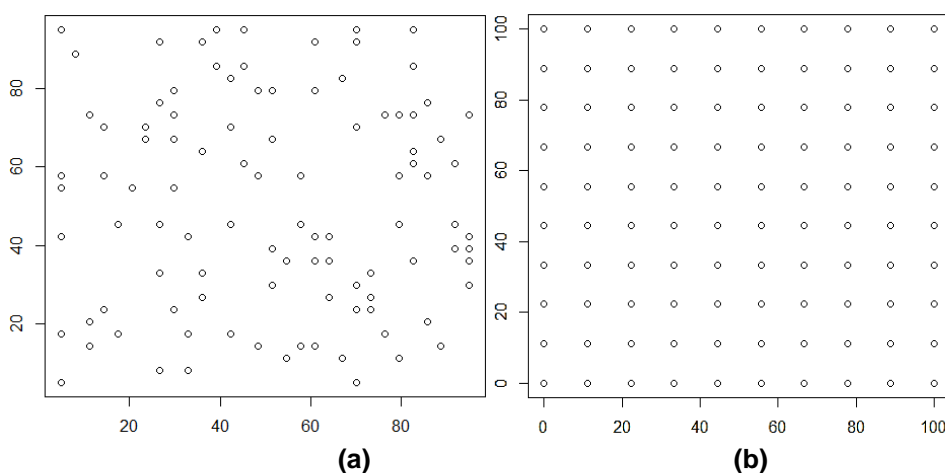


Figura 7 (a) Método aleatório de amostragem de solo, coleta aleatória; (b) Método sistemático, em que as amostras são georreferenciadas com distâncias determinadas entre pontos

A amostragem estratificada consiste em dividir a região D em subgrupos homogêneos denominados estratos. As localizações dos pontos amostrais são escolhidas em cada estrato por uma amostra aleatória uniforme. Este tipo de amostragem é utilizado quando existem áreas, dentro de uma região, heterogêneas entre si, mas que dentro de cada estrato existe uma homogeneidade (AZEVEDO, 2011).

De modo geral e ampliado, Quenouilli (1949) propôs novas configurações amostrais a partir de combinações de amostragens nas duas direções da área de interesse das amostragens: aleatória, sistemática e estratificada, assim, surgem diversos tipos de

amostragens. Logo, por exemplo, pode-se obter na direção x a amostragem aleatória e na direção y a amostragem estratificada. Considera-se, ainda, se as amostras são alinhadas ou independentes (desalinhadas) em uma ou em ambas as direções (GUEDES, 2008).

Segundo Diggle e Lophaven (2006), os métodos de amostragem espacial aleatório, sistemático e estratificado, não consideram importantes aspectos. Como o efeito pepita relevante e as amostras mensuradas muito próximas, em localizações praticamente idênticas, que podem influenciar na estimação dos parâmetros desconhecidos do modelo. Desta forma, segundo esses autores, outros dois métodos de amostragem propostos são: amostragem sistemática centrada com pares de pontos próximos (*lattice plus close pairs*) e amostragem sistemática centrada com amostragens sistemáticas centradas menores (*lattice plus in-fill*). Tais amostragens também consideraram a regularidade espacial de cobertura de toda área de interesse D , inclusive de pares de pontos próximos.

A amostragem *lattice plus close pairs* ($k \times k, m, \delta$) é formada por uma grade regular de tamanho $k \times k$ e f pontos escolhidos aleatoriamente em um círculo de raio δ centrado nos m pontos da grade, os quais foram escolhidos aleatoriamente na grade regular (Figura 8-a). A amostragem *lattice plus in-fill* ($k \times k, m, k_0 \times k_0$) consiste em uma grade regular $k \times k$, em que são selecionadas m células aleatoriamente dentro da grade, e em cada uma delas é criada uma nova grade de tamanho $k_0 \times k_0$ (Figura 8-b) (DIGGLE e LOPHAVEN, 2006; OSÓRIO, 2013).

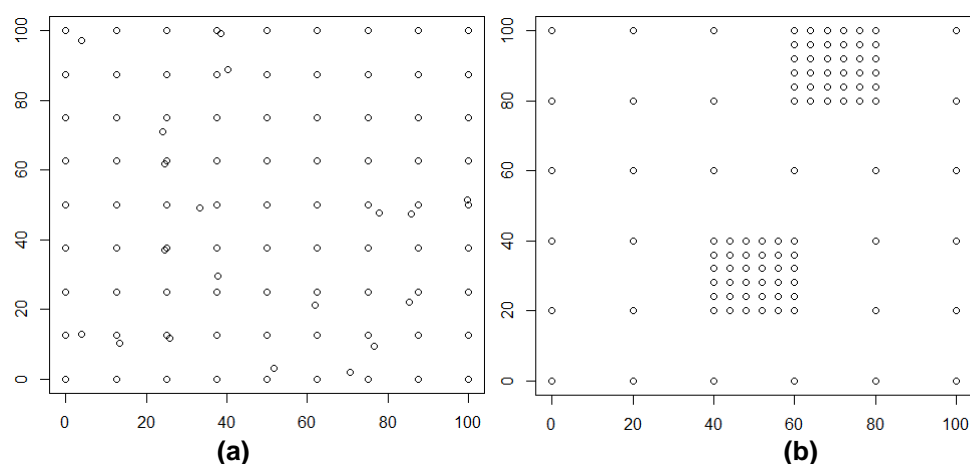


Figura 8 Exemplos de amostragem (a) Lattice plus close pairs $9 \times 9, 19,5$ (b) Lattice plus in-fill $6 \times 6, 2, 6 \times 6$

Os autores supracitados pesquisam essas técnicas de amostragem e afirmam que a amostragem *lattice plus close pairs* é mais eficiente do que a *lattice plus in-fill*. Isto significa que as predições são computadas de forma mais precisa a partir do projeto de estrutura de pontos próximos. E que a malha *lattice plus in-fill* é apenas ligeiramente melhor do que a malha regular quanto à qualidade da estimação dos parâmetros do modelo e da predição espacial.

Contudo, as variações dos parâmetros e de características que evidenciam o processo espacial fazem com que as amostragens descritas acima e as amostragens

aleatória simples, sistemática e estratificada apresentem resultados altamente variáveis (DIGGLE e RIBEIRO JUNIOR, 2007). Logo, pode-se determinar, em um experimento que envolva, na análise da variabilidade espacial, um esquema de amostragem que minimize os custos operacionais com a sua obtenção e maximize a qualidade dos resultados obtidos para a predição espacial em localizações não amostradas. Na literatura, existem propostas de metodologias de otimização que podem ser usadas para a escolha de uma configuração amostral mais eficiente na predição espacial. Essas metodologias consistem em, a partir de uma malha amostral inicial (considerada como a discretização da área sob estudo), escolher a melhor configuração amostral, que minimize as perdas quanto à acurácia dos resultados da predição espacial (GUEDES, 2015).

Além disso, existem as metodologias que buscam uma configuração amostral que otimize critérios geométricos. Uma destas metodologias é a *space filling design*, ou seja, uma configuração amostral de preenchimento de espaço, que são planos de amostragem espaciais que otimizam um critério de cobertura baseado na distância (BEAL et al., 2013).

Tomando como parâmetro essa metodologia, Johnson et al. (1990) desenvolveram o sistema de amostragem de máxima cobertura (*cover.design*), que tem como objetivo desenvolver um projeto com boa cobertura global da área, sem grandes regiões com pontos não amostrados. Existem dois destes tipos de amostragens: o projeto de mínima distância e de máxima distância entre os pontos. Segundo os autores, pontos amostrados em estreita proximidade proporcionam informação redundante, assim é desejável que existam pontos tão longe quanto possível um do outro. Os mesmos ainda destacam que projeto de mínima distância é melhor para a predição espacial, já o projeto de máxima distância é melhor para estimar parâmetros de regressão.

Além da configuração ideal das amostras, os pesquisadores buscam encontrar qual o tamanho ideal da amostra, ou até mesmo a associação entre o tamanho da área, o número de pontos amostrais e a configuração dos pontos na área pesquisada (GUEDES et al., 2016; BERNARDI et al., 2014). Com o propósito de se entender essa relação, Oliveira et al. (2011) fizeram um levantamento do tipo de malha amostral, tamanho de área e número de pontos utilizados em 65 trabalhos de geoestatística entre os anos de 1997 e 2010. A pesquisa mostrou que, entre os pesquisadores de ciência do solo, não existe uma definição clara em relação ao número ideal de pontos a ser utilizado na pesquisa em diversas condições de solo. Uma vez que as variáveis químicas e físicas podem requerer diferentes números de pontos amostrais, quando se busca conhecer a variabilidade em relação à área específica. Também não foi verificada uma associação entre o tamanho da área e o número de pontos amostrais. Isso indica que existem outros fatores que explicam tal variabilidade em supressão ao tamanho de área. Como o uso de manejo do solo, o relevo e as formas de paisagem. Porém, quanto ao tipo de malha, 82,54% dos artigos

publicados e avaliados utilizaram a grade regular e 17,46% utilizaram malhas irregulares de amostragem do solo.

Neste contexto, Souza et al. (2014) e Mondo et al. (2012) destacam a dificuldade dos pesquisadores em determinar o espaçamento ideal da amostragem. O que, em algumas circunstâncias, inviabiliza a aplicação das técnicas de Agricultura de Precisão (AP). Assim, o estudo dos aspectos da amostragem do solo, a fim de auxiliar nas definições sobre a utilização e recomendação destas técnicas em ambientes diferentes, é preocupação constante dos pesquisadores.

Na busca por encontrar uma relação entre o número ótimo de amostras utilizadas e a melhor representatividade da variabilidade espacial do atributo químico pesquisado, Silva et al. (2014) estudaram dois atributos químicos: pH e Carbono em cinco grades amostrais. Sendo elas: 70,8 x 70,8m (2 amostras/ha), 100 x 100m (1 amostra/ha), 141,4 x 141,4m (1 amostra/2ha), 173,3 x 173,3m (1 amostra/3ha) e 200 x 200m (1 amostra/4ha) (Figura 9). E concluíram que em grades com menor densidade de amostras (dentre as estudadas), o atributo pode ser caracterizado de forma satisfatório e representativo pela geoestatística, pois o atributo pH mostrou-se com maior qualidade em grade regular com uma amostra por hectare, e o carbono em grade regular de uma amostra a cada três hectares.

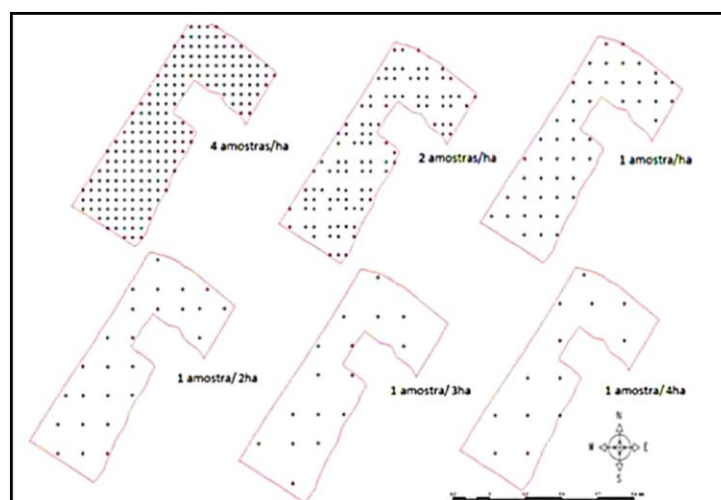


Figura 9 Grades regulares de amostragem

Fonte: Silva et. al (2014)

No entanto, Coelho et al. (2009) investigaram a influência da densidade amostral de três grades com amostragem sistemática estratificada (AASE) e do tipo de interpolador (krigagem, inverso da distância ao quadrado e polinomial) na elaboração de mapas temáticos da produtividade de soja. As três grades foram construídas a partir de 256 parcelas originais, sendo uma grade com 128 amostras, uma com 64 amostras e uma com 32 amostras (Figura 10). Com isto, por meio do coeficiente de desvio relativo (CDR), proposto no trabalho, os autores concluíram que, para se utilizar todo potencial da

krigagem e recomendar o método, são necessários muitos pontos para que um bom semivariograma seja construído.

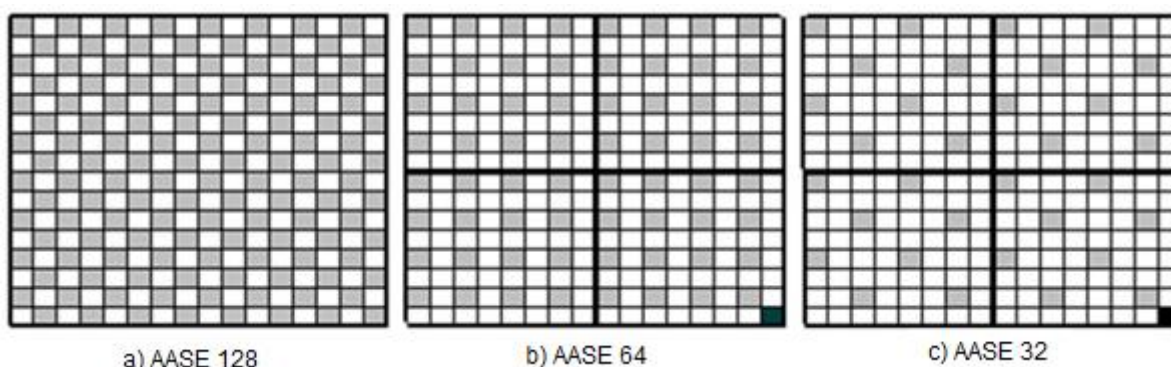


Figura 10 Grades com amostragem alinhada sistemática estratificada (AASE): a) AASE com 128 pontos amostrais, b) AASE com 64 pontos amostrais, c) AASE com 32 pontos amostrais. Em cada grade, as parcelas em cinza representam parcelas selecionadas

Fonte: Coelho et al. (2009)

Da mesma forma, Anchieta (2012) analisou, em relação ao aspecto da AP, as semelhanças e diferenças estatísticas e geoestatísticas de atributos químicos do solo em três áreas com históricos distintos amostrados sob a mesma grade amostral. E concluiu que a AP não pode limitar-se apenas em modelos de grades universais para o levantamento químico do solo. Pois, em cada contexto, é preciso introduzir ações que se adaptam às características da paisagem, histórico do manejo e tipo de solo a fim de não seguir o tratamento uniforme, já que os atributos químicos do solo possuem grande variabilidade espacial. Portanto, o objetivo não pode ser a busca por grades amostrais universais de solo, mas melhorar a prática no momento do levantamento amostral, a qual pode variar em função da localidade, do manejo e histórico agrícola, dos aspectos morfológicos, dos aspectos da paisagem, etc. Assim, deve-se atuar segundo um pensamento estratégico no levantamento amostral da fertilidade do solo e que seja coerente para cada situação.

Neste sentido, Helle e Pebesma (2015) destacam a importância de se conhecer o fenômeno pesquisado, já que a amostragem pode fazer o uso desse conhecimento e se concentrar em locais onde informações mais relevantes podem ser esperadas.

Como pode ser observado, a amostragem com malhas adensadas fornece claramente a visão da variabilidade espacial de uma variável regionalizada. Porém, com custos mais elevados quando comparados com malhas menos densas. Desta forma, Montanari et al. (2005) destacam a importância de se combinar um número mínimo de pontos amostrados com uma máxima representação do local amostrado (mínima variância) a fim de se otimizar o esquema de amostragem e reduzirem-se custos.

Além disso, segundo Anchieta (2012), outros aspectos devem ser considerados na definição das grades amostrais agrícolas, pois as etapas de levantamento da fertilidade do solo também influenciam na variabilidade espacial, ou seja, dependendo da metodologia

adotada, modos de confecção de amostras e laboratórios de análises químicas, os resultados e diagnósticos de avaliação podem diferir. Há, portanto, a influência nas quantidades necessárias de fertilizantes e corretivos em uma área, não cumprindo com os objetivos da agricultura de precisão.

8 MATERIAL E MÉTODOS

Esta pesquisa foi dividida em duas etapas. Primeiramente, foram realizadas simulações pelo método de Monte Carlo em sistemas de amostragem *lattice plus close pairs*, sistemática, aleatória e *lattice plus in-fill*, e posteriormente um estudo prático.

8.1 Simulações

As simulações foram realizadas pelo experimento de Monte Carlo para processos espaciais isotrópicos, ou anisotrópicos, ou não estacionários pelo método de decomposição de Cholesky (CRESSIE, 1993).

Para executar o método de Monte Carlo, dois passos foram adotados (MULLER, 2008). O primeiro passo foi definir as populações de interesse, a qual possui parâmetros do modelo geoestatístico exponencial com valores pré-definidos: Média ($\mu = 20$), Efeito Pepita ($\varphi_1 = 0$), Contribuição ($\varphi_2 = 10$), Patamar ($\varphi_1 + \varphi_2 = 10$) e Alcance ($\varphi_3 = 20$; $a = 60$).

O segundo passo foi a obtenção de amostras aleatórias dessas populações e o cálculo das estatísticas de interesse. Para a obtenção das amostras, foi usado o gerador de números aleatórios Cholesky. A decomposição de Cholesky é uma operação matricial que, aplicada ao vetor de números aleatórios sorteados, produz um vetor de números aleatórios que têm a característica de obedecer uma matriz de correlação entre eles (GUEDES et al., 2011).

Seja $\mathbf{Y} = (Y(\mathcal{S}_1), \dots, Y(\mathcal{S}_n))^T$ o vetor $n \times 1$ dos dados simulados, os quais representam a realização de um processo estocástico ou função de variáveis aleatórias $Y(\mathcal{S}_i)$, com ($i = 1, \dots, n$), $\mathcal{S} \in R$, em que $R \in \mathbb{R}^2$ e \mathbb{R}^2 é um espaço euclidiano bi-dimensional. Admite-se que o vetor de médias do processo estocástico satisfaz a hipótese de estacionariedade de segunda ordem, logo, tem-se que:

$$\mu = E[Y(\mathcal{S}_i)], \quad (46)$$

para $i = 1, \dots, n$, e

$$C(Y(\mathcal{S}_i), Y(\mathcal{S}_u)) = C(h_{iu}), \quad (47)$$

em que $h_{iu} = \|\mathcal{S}_i - \mathcal{S}_u\|$ é a distância euclidiana. Cada elemento do vetor é igual a um valor constante μ e cada (i, u) -ésimo elemento da matriz Σ , $n \times n$, é igual a $C(h_{iu})$.

Assim, quando são pré-definidos o valor de μ e a função covariância $C(h)$ (Equações 46 e 47), então o vetor \mathbf{Y} pode ser simulado pela relação (GUEDES et al., 2011):

$$Y = \mu \mathbf{1} + L \epsilon, \quad (48)$$

em que L é uma matriz triangular inferior $n \times n$ obtida mediante a decomposição de Σ no produto LL^T , chamada de decomposição de Cholesky; e $\epsilon = (\epsilon(S_1), \dots, \epsilon(S_n))^T$ é um vetor de variáveis aleatórias não correlacionadas.

Para cada simulação, o tamanho amostral considerado foi igual a 100 e a área em que foram realizadas as simulações é quadrada com coordenadas x e y que variaram de 0 a 100. Para as simulações dos dados isotrópicos e não estacionários, considerou-se a tendência direcional na direção 0° (eixo y) (sistema Azimute). Neste modelo, o parâmetro média é dado da seguinte forma $\mu(s_i) = \beta_0 + \beta_1 \cdot y_i$, sendo $s_i = (x_i, y_i)$. Os valores utilizados das variáveis que compõem o parâmetro média $\mu(s_i)$ foram $\beta_0 = 10$ e $\beta_1 = 0,06$, os quais fornecem uma moderada correlação linear. Já para o modelo geoestatístico anisotrópico, mais dois parâmetros foram definidos: a direção de maior continuidade espacial ($\alpha = 0^\circ$) (sistema Azimuth) e o fator de anisotropia ($Fa = 3$).

Três diferentes grades amostrais foram simuladas para avaliar a influência dos fatores que definem o sistema de amostragem *Lattice Plus Close Pairs*. Em uma configuração sistemática 9×9 com adição de 19 pontos próximos, posicionados dentro uma circunferência, cujo centro será sempre um ponto da amostragem regular, escolhido de forma aleatória, os valores de raio adotados foram 5, 3 e 1, no qual correspondem respectivamente a 50%, 27% e 9% da menor distância entre os pontos.

As escolhas do tamanho amostral e da configuração, usados nas simulações devem-se ao fato do tipo de configuração utilizada na área agrícola em estudo, quanto à análise da variabilidade espacial dos atributos químicos do solo. A partir do melhor raio encontrado na etapa anterior, foram simuladas três novas grades: 8×8 com 36 pontos próximos, 8×9 com 28 pontos próximos e 7×7 com 51 pontos próximos, conforme fluxograma descrito na Figura 11.

Para avaliar a influência da tendência direcional e da anisotropia na qualidade da estimação dos parâmetros do modelo e na predição espacial feita pela Krigagem, foi escolhido, a partir dos cenários pesquisados, aquele que apresentou os melhores resultados, para então simular a grade amostral escolhida entre elas, com tendência direcional e anisotropia na direção 0° (Sistema Azimute), e com os pontos próximos direcionados no sentido da tendência e da anisotropia e na direção ortogonal a esses (Figura 11).

Desta forma, será possível avaliar se tais fatores influenciam de forma positiva ou negativa na qualidade da estimação dos parâmetros do modelo e na qualidade da predição espacial quando o fenômeno pesquisado possui tendência direcional e anisotropia.

Para a amostragem *lattice plus in-fill* ($k \times k, m, k_0 \times k_0$), dois diferentes cenários foram simulados: (8x8,3,4x4) e (6x6,2,6x6) para dados isotrópicos e estacionários (Figura 11).

A amostragem sistemática foi simulada na configuração 10x10, com 100 pontos espaçados de forma regular. Para o caso da tendência direcional e da anisotropia, serão simuladas três grades, a amostragem sistemática 10x10, 5x20 e 20x5 (Figura 11).

A amostragem aleatória será simulada com tamanho igual a 100, com pontos amostrais dispostos em uma área regular, com limite máximo das coordenadas x e y igual a 100 m, considerando-se o modelo exponencial isotrópico, ou isotrópico com tendência direcional, ou anisotrópico (Figura 11).

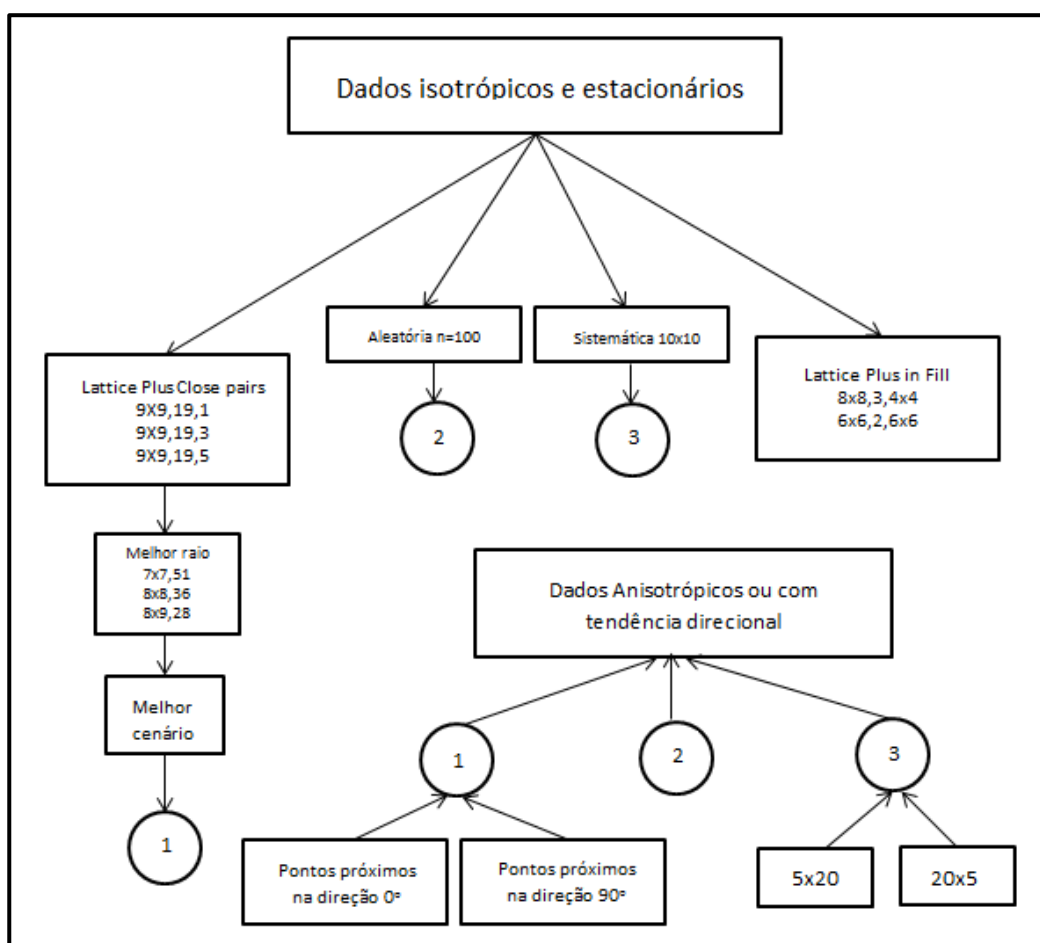


Figura 11 - Fluxograma das etapas da pesquisa

8.2 Medidas para avaliar a qualidade da estimação dos parâmetros e da predição espacial

Foram realizados 100 conjuntos de simulações em cada ensaio realizado (Figura 11). Para cada simulação, estimaram-se os parâmetros da função semivariância pelo método de Máxima Verossimilhança e o erro padrão assintótico das estimativas dos

parâmetros (URIBE OPAZO et al., 2012). Foram também calculadas as medidas de validação cruzada (Equações 32 a 36) e de qualidade de estimação do modelo (Equações 37 e 39).

Além disso, ainda para avaliar a qualidade de estimação dos parâmetros do modelo geoestatístico, foram calculadas as seguintes medidas para cada parâmetro do modelo geoestatístico: a Raiz do Erro Quadrático Médio (Equação 44), Viés Absoluto e Viés Relativo Absoluto (Equação 45) (WILKS, 2006).

A Raiz do Erro Quadrático Médio (REQM) é comumente usada para expressar a acurácia dos resultados numéricos, sendo que a REQM é sempre positiva. Assim, quando $REQM = 0$ indica uma estimação perfeita em todas as simulações e essa medida é calculada por (HALLAK e PEREIRA FILHO, 2011):

$$REQM = \left[\frac{1}{p} \sum_{j=1}^p (\hat{\theta}_{ij} - \theta_i)^2 \right]^{1/2} \quad (49)$$

em que p é o número de simulações, $\hat{\theta}_{ij}$ a estimativa do i -ésimo parâmetro obtido pelo método de máxima verossimilhança na j -ésima simulação. Sendo que, $\theta_i \in (\alpha, \mu, \varphi_1, \varphi_2, \varphi_3)$, com $p = 1, \dots, 100$.

O Viés Relativo Absoluto (VRA) e Viés Absoluto (VA) são definidos como (PAPANI et al., 2016):

$$VRA(\hat{\theta}_i) = \left| \frac{\bar{\theta}_i - \theta_i}{\theta_i} \right| \times 100, \quad VA(\hat{\theta}_i) = |\bar{\theta}_i - \theta_i|, \quad (50)$$

em que $\bar{\theta}_i = (1/p) \sum_{j=1}^p \hat{\theta}_{ij}$, com $\hat{\theta}_{ij}$ sendo a estimativa obtida pelo método Máxima Verossimilhança para o i -ésimo parâmetro, em que $\theta_i \in (\alpha, \mu, \varphi_1, \varphi_2, \varphi_3)$, e na j -ésima simulação de Monte Carlo, com $p = 1, \dots, 100$.

Algumas medidas de incerteza da estimação dos valores da variável georreferenciada foram utilizadas em localizações não amostradas para medir a qualidade da predição espacial. São elas: o erro calculado pela diferença entre o valor estimado $Z^*(x_0)$ e o valor real $Z(x_0)$, e a média da variância da predição espacial (Equação 44). O erro foi calculado com base na proposta de Rossiter (2012), que consiste em simular na grade pesquisada a adição de pontos extras, chamada de amostra-teste. Para elas, posteriormente, foi realizada a interpolação por meio da krigagem usando a amostra original, sem a adição de pontos extras, chamada amostra-treino. Por fim, calcula-se a diferença entre o valor obtido na simulação com a amostra teste e o obtido pela estimação, permitindo assim calcular a diferença d_i entre o valor predito do ponto i no espaço ($i = 1, 2, 3, \dots, n$) e o observado para uma mesma variável, no mesmo ponto no espaço. Além disso, $d_i = 0$ indica predição perfeita naquele ponto i , enquanto $d_i > 0$ ou $d_i < 0$, indica predição imperfeita. Quanto mais distante de 0 o valor de d_i , mais imperfeita a predição (HALLAK e PEREIRA FILHO, 2011). A amostra teste utilizada foi de 25 pontos.

Estudo prático

8.2.1 Área de estudo

A parte prática desta pesquisa foi realizada em uma área comercial de produção de grãos, no município de Cascavel, com 167.35 hectares, cuja localização geográfica é aproximadamente 24.95° Sul de latitude, 53.37° Oeste de longitude e altitude média de 650 metros. Esses dados referem-se ao ano agrícola de 2015/2016, obtidos em experimentos conduzidos por pesquisadores do grupo de pesquisa do Laboratório de Estatística Espacial (LEE) e Laboratório de Estatística Aplicada (LEA) da Universidade Estadual do Oeste do Paraná – UNIOESTE, campus de Cascavel.

Nesta área, foi realizada uma amostragem *lattice plus close pairs*, com distância máxima de 141 metros entre os pontos. Em dezenove (19) locais, escolhidos de forma aleatória, a amostragem foi realizada com distâncias menores (75 e 50 m entre pontos), obtendo-se no total 102 pontos amostrais. Todas as amostras foram georreferenciadas e localizadas com auxílio de um aparelho receptor de sinal com o sistema de posicionamento global (GPS) GEOEXPLORE 3, em um sistema espacial de coordenadas UTM. Na Figura 12, estão ilustradas a área experimental e a grade de amostragem. Cada círculo desta figura representa um ponto de amostragem.

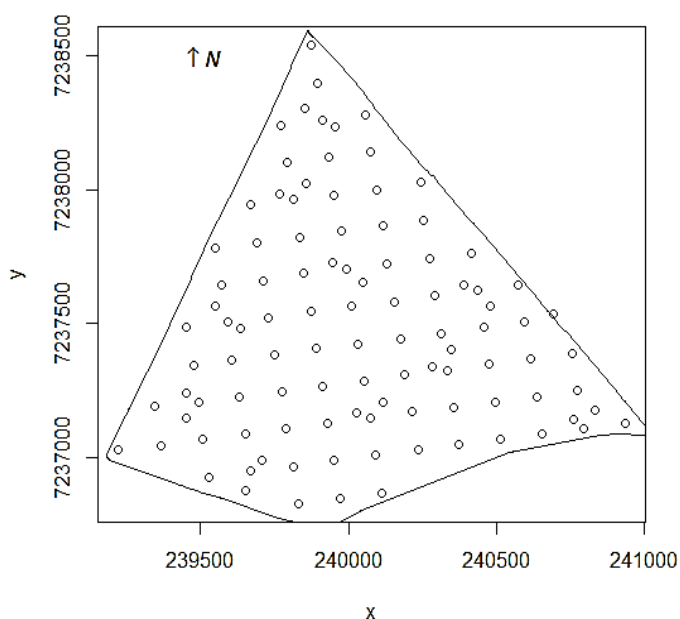


Figura 12 - Mapa da área em estudo

A área foi cultivada com soja e os dados utilizados nesse trabalho são referentes aos seguintes atributos químicos do solo: carbono (g dm^{-3}), cálcio (cmol dm^{-3}), magnésio ($\text{cmol}_c \text{dm}^{-3}$), manganês (mg dm^{-3}), potássio ($\text{cmol}_c \text{dm}^{-3}$), Cobre (mg/dm^3), Zinco (mg/dm^3), Ferro (mg/dm^3), e Alumínio (cmolc/dm^3).

O conjunto de dados sob pesquisa foi obtido pela realização de análise química de rotina, a partir de uma amostragem obtida em cada ponto demarcado. Foram coletadas cinco subamostras de solo, de 0,0 a 0,2 m de profundidade nas proximidades dos pontos, misturadas e colocadas em sacos plásticos, com aproximadamente 500 g, compondo-se assim a amostra representativa da parcela. As amostras foram encaminhadas ao laboratório de análise do solo da COODETEC (Cooperativa Central de Pesquisa Agrícola).

O estudo prático foi realizado em dois momentos: primeiro realizou-se a análise descritiva dos dados e em seguida realizou-se o estudo geoestatístico.

8.2.2 Análise Descritiva

A primeira etapa em qualquer estudo de dados é a análise descritiva, a qual permite ao pesquisador organizar, resumir e descrever os aspectos importantes do conjunto das características observadas. Além disso, possibilita melhor detecção de erros e incoerências nos dados, análises e resultados.

Nesta etapa inicial, as estatísticas descritivas calculadas foram medidas de localização, medidas de dispersão e medidas de forma.

Outra análise realizada foi a identificação de pontos discrepantes (*outliers*). Esta técnica possibilitará escolher qual o melhor estimador da função semivariância, uma vez que os pontos discrepantes afetam a análise dos dados e mascaram as estatísticas calculadas. A técnica que aqui será utilizada para identificação destes *outliers* é o gráfico *boxplot*.

Por fim, foi analisada a presença de tendência nos dados observados, uma vez que esta característica influencia na estimativa da função semivariância e dificulta o ajuste de modelos adequados, além de mascarar o verdadeiro comportamento espacial da variável. E, para identificar a tendência dos dados, serão utilizadas duas ferramentas, o mapa da distribuição das observações e o gráfico de dispersão dos dados *versus* suas coordenadas (TEIXEIRA, 2013).

8.2.3 Análise Geoestatística

Após a análise exploratória dos dados, foi feito o estudo variográfico pela construção semivariograma para identificar o comportamento espacial dos dados. Esta construção foi feita utilizando o estimador robusto de Matheron (1963) (Equação 9) e,

posteriormente, o ajuste dos modelos teóricos Esférico (Equação 13), Exponencial (Equação 16), Gaussiano (Equação 19) e Família Matérn com $k=1,5$ (Equação 22) pelo Método de Máxima Verossimilhança.

Assim, para verificar a qualidade do ajuste do modelo, foram utilizados três critérios: validação cruzada, informação de Akaike e informação Bayesiano. E por fim, de acordo com a estrutura de variabilidade fornecida pela função semivariância estimada, foram estimados pontos não amostrados e construído o mapa de contorno a partir do método de krigagem ordinária para dados sem tendência direcional e da krigagem universal para as variáveis com tendência direcional.

9 Software utilizado

Todas as simulações e análises dos dados serão feitas no software R (R DEVELOPMENT CORE TEAM, 2015) – pacote geoR (RIBEIRO JÚNIOR & DIGLLE, 2001).

10 RESULTADOS E DISCUSSÕES

10.1 Estudo de simulações

10.1.1 Dados isotrópicos e estacionários

10.1.2 Lattice Plus Close Pairs

Na Tabela 1 são apresentadas as medidas descritivas dos parâmetros do modelo Exponencial, estimado por ML e seus respectivos desvios padrão, e também as medidas de eficiência do estimador para os três cenários pesquisados sem tendência direcional. Veja que os parâmetros média (μ), contribuição (φ_2) e alcance (a) apresentaram, em média, valores estimados mais próximos ao nominal nos três cenários (Tabela 1 e Figura 13-a, 13-c, 13-d). O efeito pepita teve, em média, a estimativa mais próxima ao valor nominal para o raio 1 (Tabela 1 e Figura 13-b). Ainda, observou-se que com o aumento do raio, os valores do efeito pepita aumentaram assim como a presença de pontos discrepantes. Os raios 1, 3 e 5 correspondem respectivamente a 9%, 27% e 50% da menor distância entre os pontos. Note que o CV do parâmetro média em todos os cenários foi baixo, o da contribuição e alcance prático foi elevado, e do efeito pepita muito elevado (PIMENTEL GOMES, 2000).

O desvio padrão dos parâmetros, média e alcance prático foram similares para os três cenários (Tabela 1 e Figuras 14-a e 14-d). Já para o efeito pepita e a contribuição, a redução do raio de localização dos pontos próximos proporcionou menores valores do desvio padrão das estimativas dos parâmetros (Tabela 1 e Figuras 14-b e 14-c).

Quanto às medidas de eficiência do estimador (Tabela 1), para todos os raios simulados, foram obtidas medidas similares de VA e REQM para o parâmetro média. Com o aumento do raio, observou-se um aumento de VA, VRA e REQM para a contribuição, do VA para o efeito pepita e da REQM para o alcance; e um decaimento da REQM para o efeito pepita, e do VA e VRA para o alcance.

Em resumo, com base nesses resultados, observa-se que as melhores estimativas do efeito pepita e da contribuição ocorrem quando foram escolhidos os menores raios de pontos próximos na amostragem *lattice plus close pairs*. Isso mostra a ausência de erros de variabilidade de pequena escala captada pela amostragem (OLIVEIRA JÚNIOR et al., 2011). Os parâmetros média e alcance não tiveram influência relevante nas suas estimativas, quanto à mudança do raio nesta amostragem.

Tabela 1 - Estatísticas descritivas dos parâmetros do modelo exponencial estimado por ML e seus respectivos desvios padrões e as medidas de eficiência do estimador VRA (%), VA e REQM para dados isotrópicos

Análise Descritiva	Raio	Média ($\mu=20$)	Efeito Pepita ($\varphi_1=0$)	Contribuição ($\varphi_2=10$)	Alcance Prático ($\alpha=60$)	DP (μ)	DP (φ_1)	DP (φ_2)	DP (α)
Média		19,98	0,033	8,92	54,81	0,9116	0,1523	1,498	0,0813
Desvio Padrão		1,054	0,079	2,486	18,2	0,3074	0,0362	0,411	0,055
CV (%)	1	5,28	240,01	27,85	33,21	33,72	23,77	27,43	67,72
VRA (%)		0,0725	-	10,751	8,655				
VA		0,0145	0,033	1,0751	5,193				
REQM		0,551	0,003	3,638	177,51				
Média		19,99	0,1008	8,82	55,01	0,908	0,4601	1,819	0,0829
Desvio Padrão		1,053	0,247	2,49	18,75	0,311	0,098	0,49	0,057
CV (%)	3	5,27	245,9	28,29	34,08	34,33	21,49	26,95	69,52
VRA (%)		0,0735	-	11,78	8,33				
VA		0,0147	0,1008	1,1781	4,99				
REQM		0,549	0,035	3,779	186,48				
Média		19,99	0,1081	8,89	55,86	0,919	0,497	1,864	0,0866
Desvio Padrão		1,052	0,208	2,49	19,68	0,312	0,098	0,486	0,065
CV (%)	5	5,26	192,45	28,08	35,24	33,97	19,75	26,09	76,05
VRA (%)		0,057	-	11,054	6,89				
VA		0,0114	0,1081	1,1054	4,1394				
REQM		0,548	0,027	3,698	200,39				

Valores nominais: μ é a média, φ_1 é o efeito pepita; φ_2 é a contribuição; α é o alcance prático.

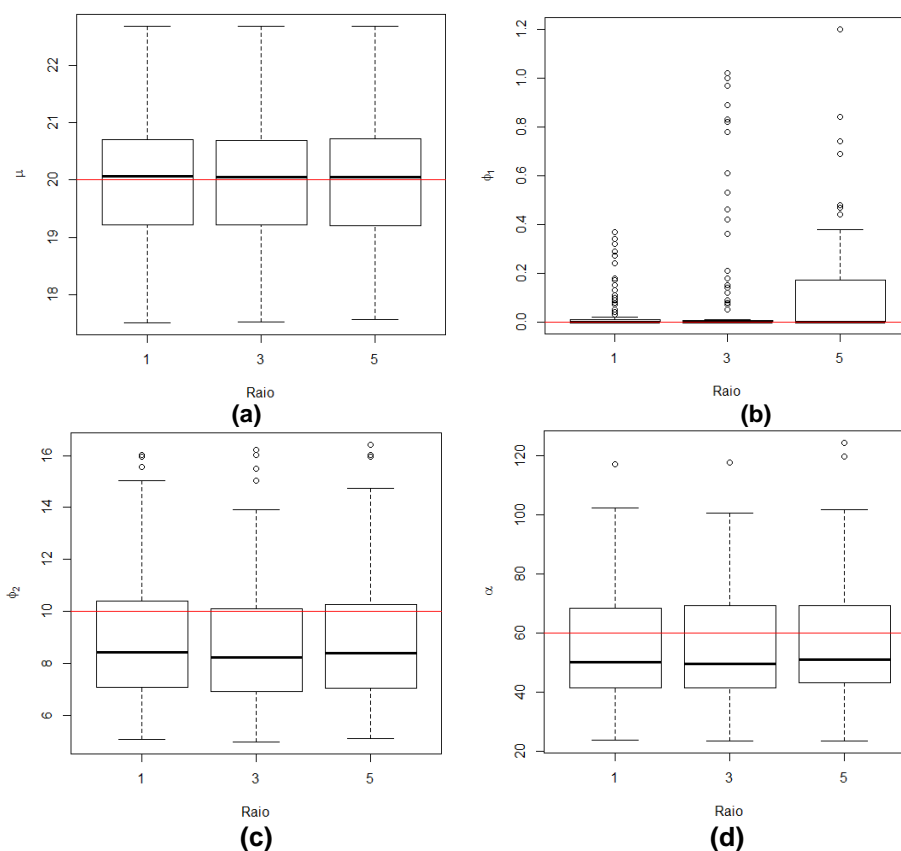


Figura 13 Gráfico boxplot dos valores estimados dos seguintes parâmetros (a) Média, (b) Efeito Pepita, (c) Contribuição e (d) Alcance Prático. A linha tracejada indica o valor nominal simulado

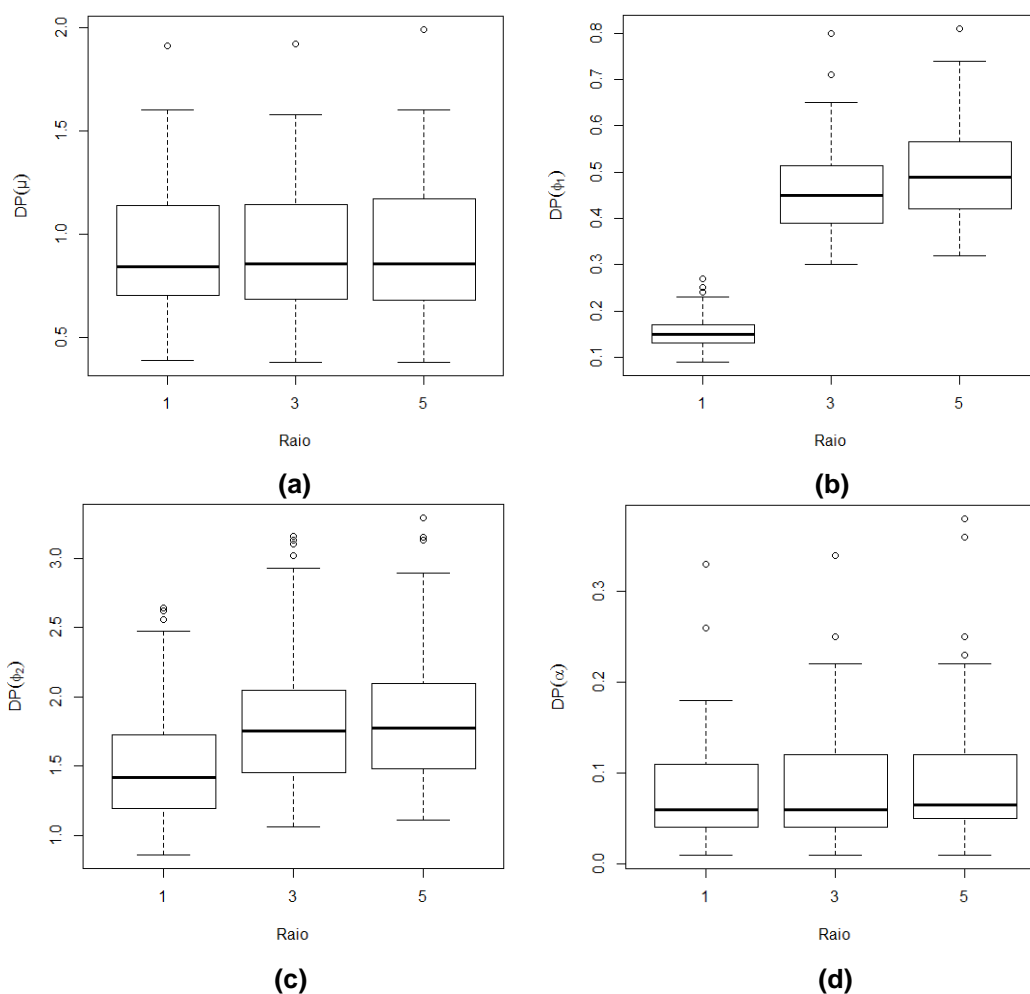


Figura 14 Gráfico boxplot do desvio padrão das estimativas dos parâmetros estimados (a) Média, (b) Efeito Pepita, (c) Contribuição e (d) Alcance Prático

As medidas de validação cruzada, usadas para avaliar a qualidade da estimação do modelo, são exibidas na Figura 15. Observa-se por essas medidas que os valores de EM e EMR foram mais próximos de zero, e os valores do SEMR foram similares e próximos a 1, para todos os raios simulados. Observou-se também que, com o aumento do raio, houve também um aumento do SEM, EA, AIC e BIC. Sendo assim, considerando a maioria dessas medidas, conclui-se que a *lattice plus close pairs* com o raio de pontos próximos igual a 1 (menor raio) forneceu melhores estimativas das medidas associadas a validação cruzada e portanto, uma melhor qualidade na estimação dos parâmetros do modelo exponencial.

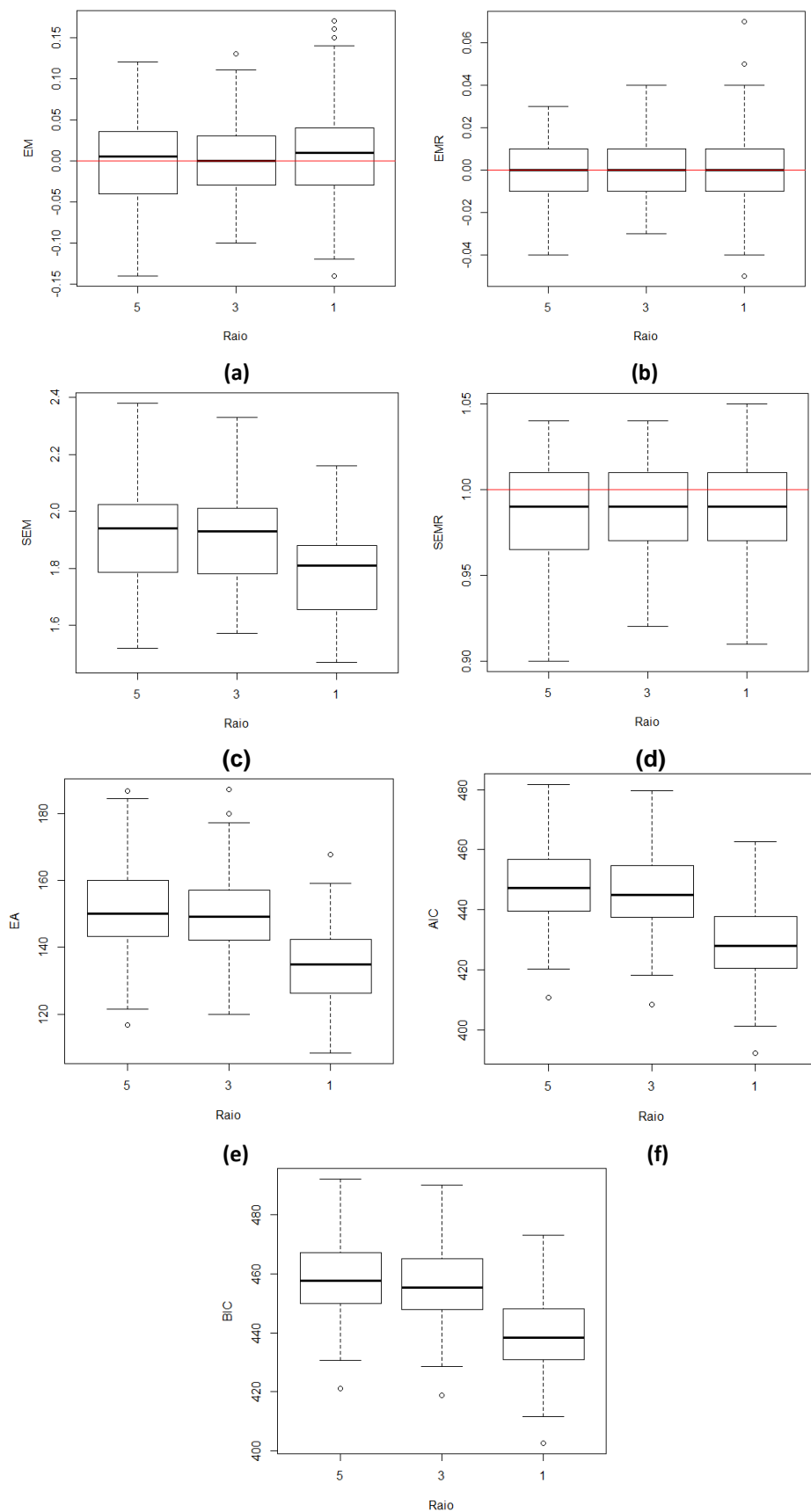


Figura 15 Gráficos boxplot: (a) Erro Médio, (b) Erro Médio Relativo, (c) Desvio Padrão do Erro Médio, (d) Desvio Padrão do Erro Médio Relativo, (e) Erro Absoluto, (f) AIC e (g) BIC. A linha tracejada indica o valor ideal

As medidas de qualidade da predição espacial (Tabela 2 e Figura 16), na análise descritiva da média da variância da grade pesquisada e do erro da amostra teste, mostraram que o valor do raio não influencia de forma relevante na qualidade da predição espacial. Veja que, em média, o erro da predição espacial na amostra teste e a média da variância foram semelhantes nos três cenários (Tabela 2). Apenas a Raiz do Erro Quadrático médio apresentou leve queda com o aumento do raio (Tabela 2). O Coeficiente de Variação do erro da variância na amostra-teste foi elevado para todos os cenários, porém, a média da variância teve média dispersão nos três cenários (Tabela 2). Além disso, os gráficos Boxplots evidenciam que os valores da média da variância da krigagem (Figura 16-a) e do Erro na amostra teste (Figura 16-b) foram similares nos três cenários.

Tabela 2 Análise descritiva das medidas de qualidade da predição espacial em localizações não amostradas considerando uma amostra teste composta por 25 pontos

Estadística Descritiva	Raio	Erro de predição	Média da variância da Krigagem
Média		0,041	2,79
Desvio Padrão	1	0,33	0,46
CV (%)		810,11	16,75
REQM		1,362	-
Média		0,041	2,73
Desvio Padrão	3	0,32	0,46
CV (%)		782,94	17,00
REQM		1,310	-
Média		0,034	2,76
Desvio Padrão	5	0,33	0,45
CV (%)		945,38	16,38
REQM		1,311	-

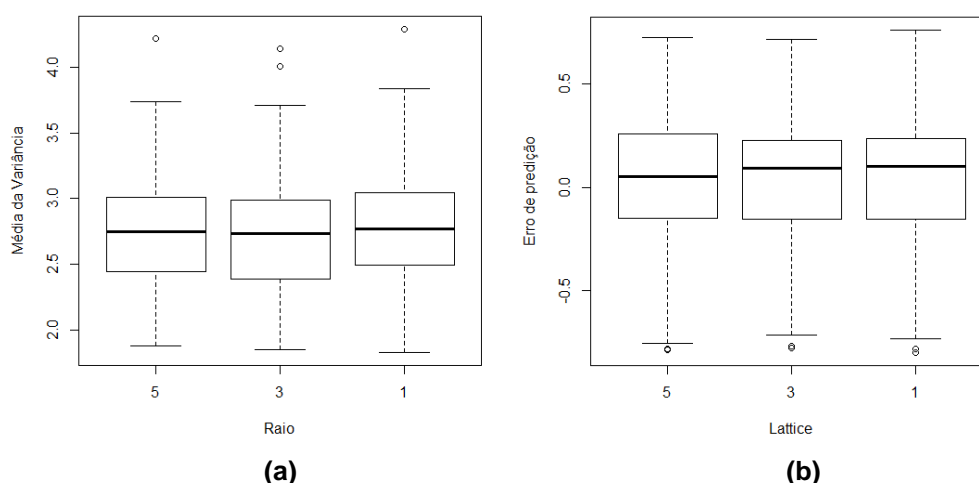


Figura 16 - Gráficos Boxplot: (a) Média da Variância da krigagem e (b) Erro de predição

Vale ressaltar que, na etapa anterior, foi discutido, que o raio de pontos próximos igual a um (menor raio simulado) produz melhor qualidade da estimativa dos parâmetros do modelo geoestatístico, principalmente do efeito pepita e contribuição. E ainda que, não

foi obtida diferença relevante nas medidas de qualidade da predição espacial quando foi variada a localização dos pontos próximos.

Assim, na sequência, foi discutida a influência da quantidade de pontos próximos usados na amostragem *lattice plus close pairs* quando o raio de pontos próximos é igual a 1. Foram testadas as grades (8x9,28,1), (8x8,36,1) e (7x7,51,1), e, nesta sequência de cenários pesquisados, foi reduzido o número de pontos da grade regular e aumentado o número de pontos próximos que formam a amostragem *lattice plus close pairs*.

A Tabela 3 apresenta a análise descritiva dos parâmetros do modelo exponencial estimado por ML, seus respectivos desvios-padrão e as medidas de eficiência do estimador para os três cenários. Note, no efeito pepita, na contribuição e no alcance prático que, quando se aumenta o número de pontos próximos, os valores estimados para esses parâmetros apresentam valores mais próximos ao valor nominal (Tabela 3 e Figuras 17-b,17-c,17-d), e apenas a média manteve-se semelhante em todos os cenários (Figura 17-a). Tal resultado complementa o obtido nos três cenários anteriores, pois no cenário (9x9,19,1) (Tabela 1 e Figura 14-b), foi registrada a melhor estimativa do efeito pepita, e aqui, quando houve aumento no número de pontos próximos, também houve, em média, valores estimados mais próximos ao nominal na estimativa desse parâmetro.

Os desvios-padrão da média, contribuição e alcance foram similares em todos os cenários (Figuras 18-a, 18-c, 18-d). No entanto, os valores do desvio padrão do efeito pepita estimado decresceram quando foi aumentada a quantidade de pontos próximos na grade (Figura 18-b).

Quanto às medidas de eficiência do estimador, foram obtidos valores maiores de VRA para o parâmetro média, e um decaimento de VRA, VA e REQM para os parâmetros alcance prático e contribuição com o aumento da quantidade de pontos próximos.

Em resumo, com base nesses resultados, observa-se que as melhores estimativas do efeito pepita, contribuição e alcance prático ocorrem com o aumento da quantidade de pontos próximos na amostragem *lattice plus close pairs* (7x7,51,1). O parâmetro média não teve influência relevante na sua estimativa quando a mudança da quantidade de pontos próximos.

Tabela 3 Estatísticas descritivas dos parâmetros do modelo exponencial estimado por ML para as grades (8x9,36,1), (7x7,51,1) e (8x9,28,1) com dados isotrópicos

Análise Descritiva	Grade	Média ($\mu=20$)	Efeito Pepita ($\varphi_1=0$)	Contribuição ($\varphi_2=10$)	Alcance Prático ($\alpha=60$)	DP (μ)	DP (φ_1)	DP (φ_2)	DP (α)
Média		20,02	0,0252	9,07	56,904	0,941	0,0894	1,51	0,09
Desvio Padrão		1,027	0,046	2,68	21,26	0,342	0,019	0,436	0,077
CV (%)		5,13	184,88	29,55	37,36	36,34	22,13	28,85	86,49
VRA (%)	(8x9,28,1)	0,0885	-	9,298	5,1595				
VA		0,0177	0,0252	0,9298	3,0957				
REQM		0,522	0,0013	3,989	228,59				
Média		20,01	0,024	9,101	57,76	0,9543	0,0775	1,56	0,0918
Desvio Padrão		1,038	0,043	2,61	20,42	0,332	0,016	0,444	0,066
CV (%)		5,19	182,24	28,63	35,35	34,81	21,49	28,44	71,91
VRA (%)	(8x8,36,1)	0,027	-	8,982	3,723				
VA		0,0054	0,024	0,8982	2,2538				
REQM		0,5334	0,0012	3,76	208,94				
Média		19,95	0,007	9,41	60,06	0,9989	0,0225	1,539	0,1012
Desvio Padrão		1,091	0,014	2,801	20,49	0,339	0,006	0,451	0,0804
CV (%)		5,46	191,7	29,76	34,12	34,02	28,51	29,31	79,49
VRA (%)	(7x7,51,1)	0,21	-	5,893	0,1021				
VA		0,042	0,0077	0,5893	0,0613				
REQM		0,5905	0,00013	4,05	207,88				

Valores nominais: μ é a média, φ_1 é o efeito pepita; φ_2 é a contribuição; α é o alcance prático.

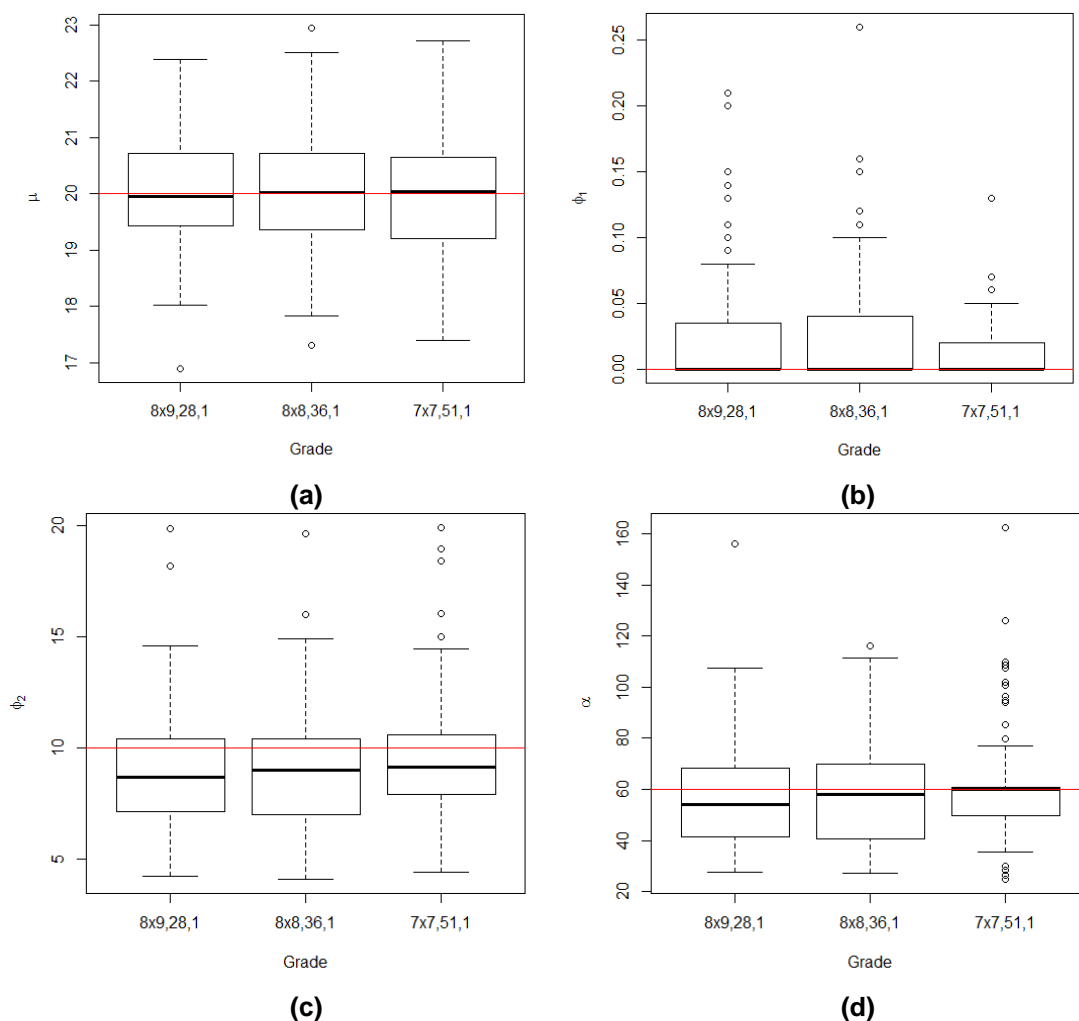


Figura 17 Gráficos boxplot dos valores estimados dos seguintes parâmetros: (a) Média, (b) Efeito Pepita, (c) Contribuição e (d) Alcance Prático, a linha tracejada indica o valor nominal simulado

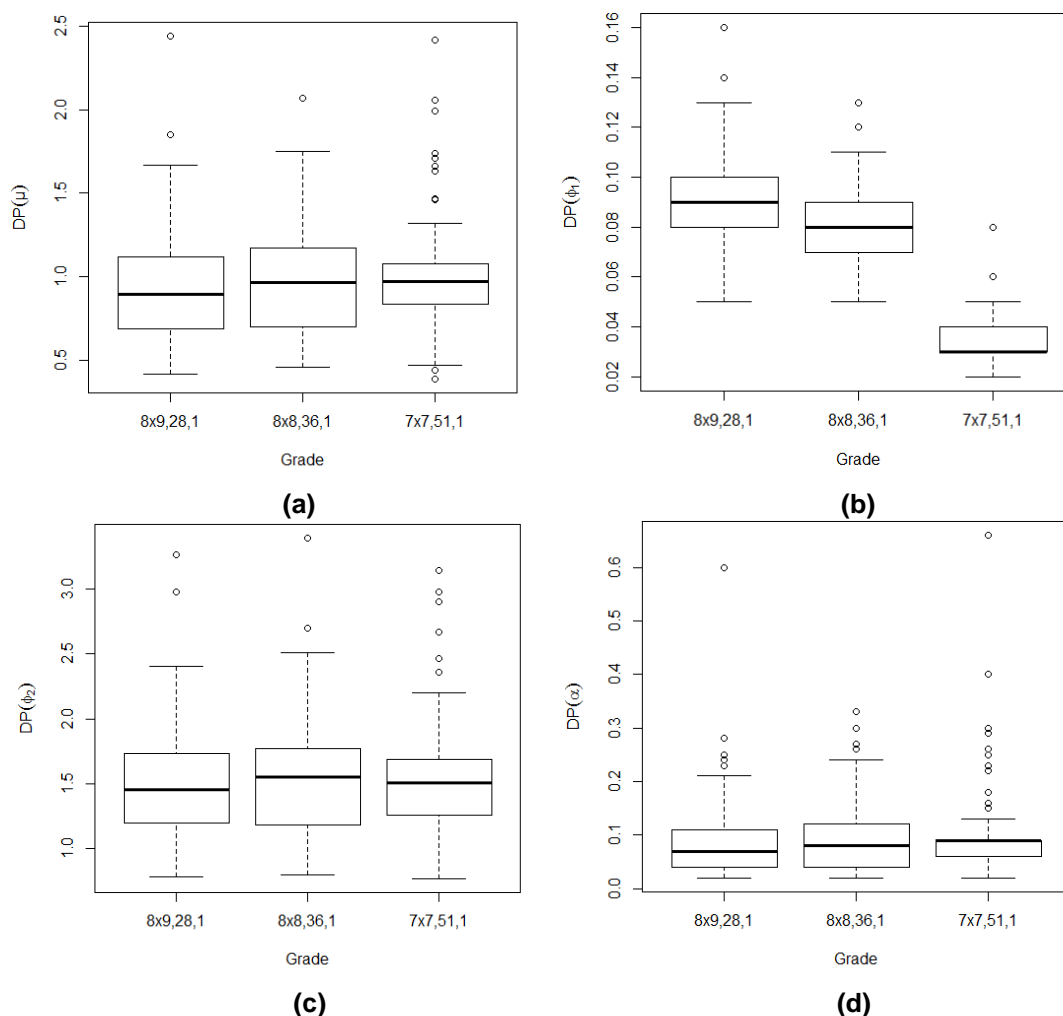


Figura 18 Boxplot do desvio padrão das estimativas dos parâmetros: (a) Média, (b) Efeito pepita, (c) Contribuição e (d) Alcance prático

A Figura 19 mostra as medidas de validação cruzada, usadas para avaliar a qualidade da estimação dos parâmetros do modelo. A partir dessas medidas, observa-se que o EM e o EMR tiveram valores mais próximos de zero, SEMR próximo de um e EA, AIC e BIC menores com o aumento da quantidade de pontos próximos. Apenas os valores do SEM foram menores para a grade (8x8,36,1). Portanto, ao se considerar a maioria dessas medidas, conclui-se que o aumento da quantidade de pontos próximos na amostragem *lattice plus close pairs* forneceu melhores estimativas das medidas associadas à validação cruzada e às medidas quanto à qualidade da estimação dos parâmetros do modelo exponencial.

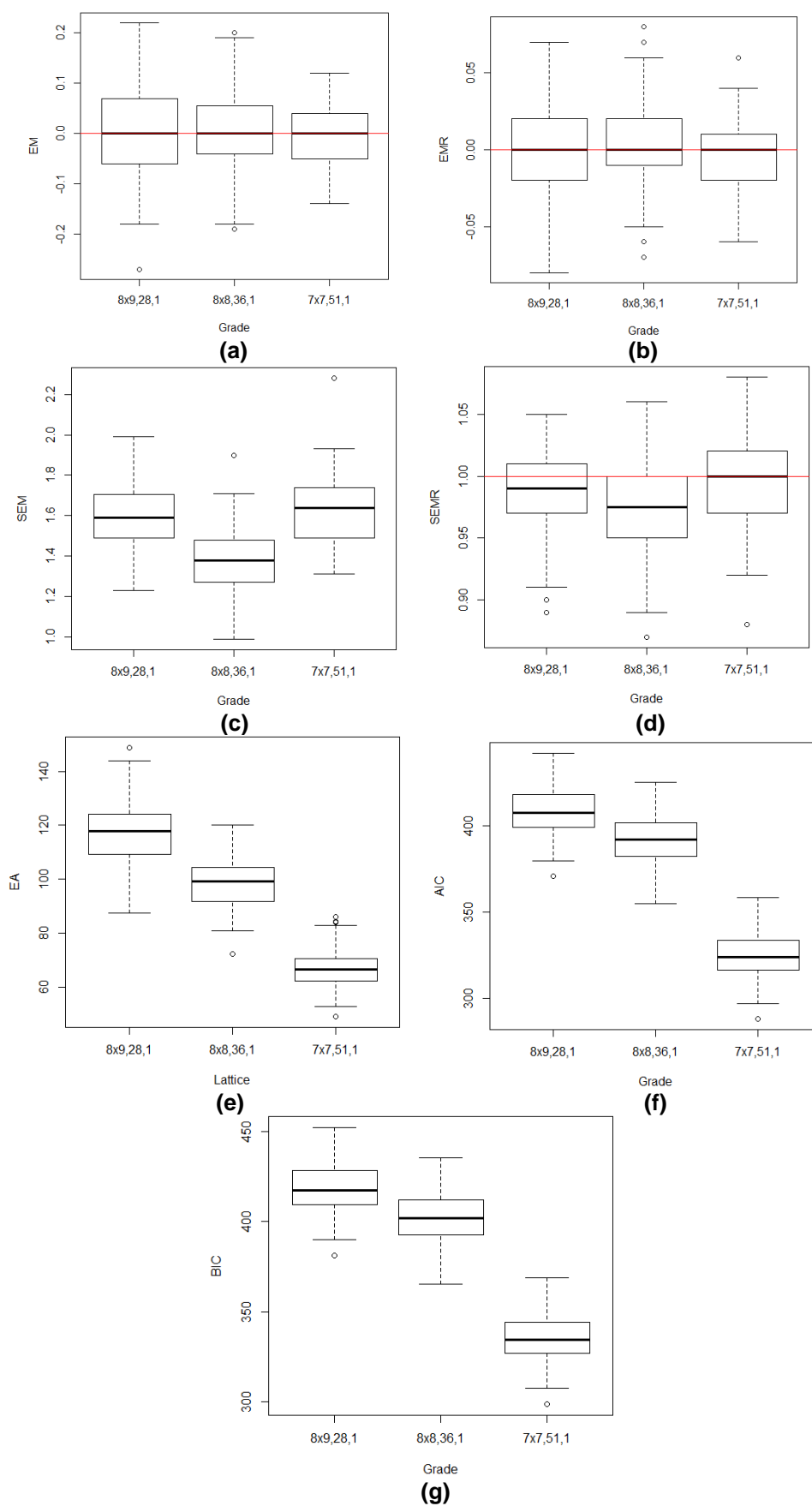


Figura 19 Gráficos boxplot do (a) Erro Médio, (b) Erro Médio Relativo, (c) Desvio Padrão do Erro Médio, (d) Desvio Padrão do Erro Médio Relativo, (e) Erro Absoluto, (f) AIC e (g) BIC. A linha tracejada indica o valor ideal

Na avaliação da qualidade da predição espacial (Tabela 4 e Figura 20), a média da variância foi em média menor para o *grid* (7x7,51,1), assim como a REQM do erro da predição. O Coeficiente de Variação do erro da predição nos três cenários foi muito elevado (PIMENTEL GOMES, 2000). Todavia, em ambos os cenários, a média da variância da krigagem teve dispersão média. Assim, conclui-se que o aumento da quantidade de pontos próximos melhorou a qualidade da predição espacial feita pela krigagem.

Tabela 4 Análise descritiva das medidas da qualidade da predição espacial em localizações não amostradas, considerando uma amostra teste composta por 25 pontos

Estadística Descritiva	Grade	Erro da predição	Média da variância
Média		0,041	2,8896
Desvio Padrão	(8x9,28,1)	0,342	0,5138
CV (%)		832,3	17,78
REQM		1,4534	-
Média		0,043	2,9951
Desvio Padrão	(8x8,36,1)	0,35	0,5222
CV (%)		801,55	17,43
REQM		1,5287	-
Média		0,021	0,5426
Desvio Padrão	(7x7,51,1)	0,2782	0,092
CV (%)		1311,7	17,06
REQM		0,2701	-

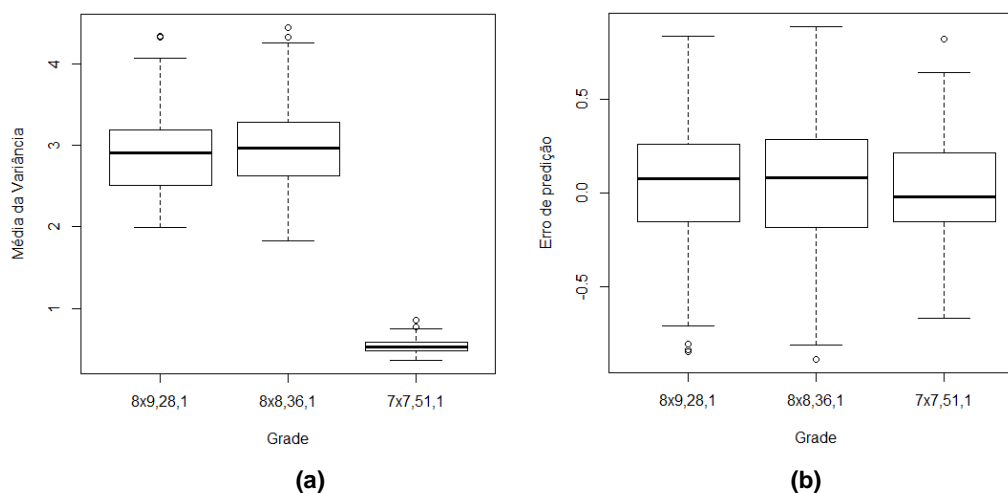


Figura 20 Gráfico boxplot: (a) Média da variância da krigagem e (b) Erro da predição numa amostra teste composta por 25 pontos

10.1.3 Lattice Plus In-fill

Na Tabela 5, estão apresentadas as medidas descritivas dos parâmetros do modelo Exponencial, estimado por ML, o desvio padrão das estimativas dos parâmetros do modelo

e as medidas de eficiência do estimador para os dois cenários pesquisados da amostragem *lattice plus in fill*, considerando o fato de a variável georreferenciada não ter tendência direcional. Pode-se observar que todos os parâmetros do modelo não tiveram, em média, diferença relevante em suas estimativas nos dois cenários da amostragem sistemática centrada com amostragens sistemáticas centradas menores (Tabela 5 Figuras 21-a, 21-b, 21-c, 21-d). O coeficiente de variação do parâmetro média nos dois cenários foi baixo, da contribuição e alcance prático foi alto, e o CV efeito pepita muito alto (PIMENTEL GOMES, 2000).

Os desvios-padrão dos parâmetros média, contribuição e alcance prático foram similares nos dois cenários (Tabela 5 e Figuras 22-a, 22-c e 22-d). Já a redução da quantidade de amostragens centradas menores e o aumento do tamanho das mesmas (6x6,2,6x6) proporcionaram menores valores de desvio padrão dos estimadores do efeito pepita (Tabela 5 e Figura 22-b).

Quanto às medidas de eficiência do estimador (Tabela 5), os dois cenários pesquisados apresentaram diferença relevante quanto à estimação dos parâmetros contribuição e alcance prático. Enquanto o aumento da quantidade de amostragens sistemática centradas menores e a redução do tamanho das mesmas (8x8,3,4x4) proporcionaram redução de VA, VRA e REQM para o parâmetro alcance. No entanto, no outro cenário (6x6,2,6x6), obtiveram-se valores menores de VRA e VA para o parâmetro contribuição.

Em resumo, com base nesses resultados, observa-se que a alteração dos fatores que determinam a configuração amostral *lattice plus in fill* não influenciam de forma relevante na estimação dos parâmetros do modelo. Note que os resultados são conflitantes, pois, a REQM teve menores valores para os parâmetros estimados na amostragem (6x6,2,6x6), e as medidas VRA e VA menores valores para a grade (8x8,3,4x4).

Tabela 5 Análise descritiva dos parâmetros do modelo exponencial estimado por ML, respectivos desvios-padrão e as medidas de eficiência do estimador VRA(%), VA e REQM para dados isotrópicos

Análise Descritiva	Grade	Média ($\mu=20$)	Efeito Pepita ($\varphi_1=0$)	Contribuição ($\varphi_2=10$)	Alcance Prático ($\alpha = 60$)	DP (μ)	DP (φ_1)	DP (φ_2)	DP (α)
Média		20,01	0,1431	8,92	57,86	0,9431	0,5991	2,1661	0,0978
Desvio Padrão		1,03	0,29	2,73	22,32	0,34	0,13	0,572	0,079
CV (%)	(8x8,3,4x4)	5,16	203,22	30,58	38,58	36,68	20,84	26,42	81,26
VRA (%)		0,665	-	10,717	3,5503				
VA		0,0133	0,1431	1,0717	2,1302				
REQM		0,528	0,052	4,26	249,01				
Média		19,92	0,1444	9,25	63,35	1,0261	0,4082	2,2453	0,1378
Desvio Padrão		1,02	0,25	3,03	27,61	0,39	0,079	0,69	0,1374
CV (%)	(6x6,2,6x6)	5,13	174,2	32,73	43,57	38,26	19,53	30,85	99,74
VRA (%)		0,388	-	7,408	5,597				
VA		0,0776	0,1444	0,7408	3,3586				
REQM		0,519	0,041	4,823	383,02				

Valores nominais: μ é a média, φ_1 é o efeito pepita; φ_2 é a contribuição; α é o alcance prático.

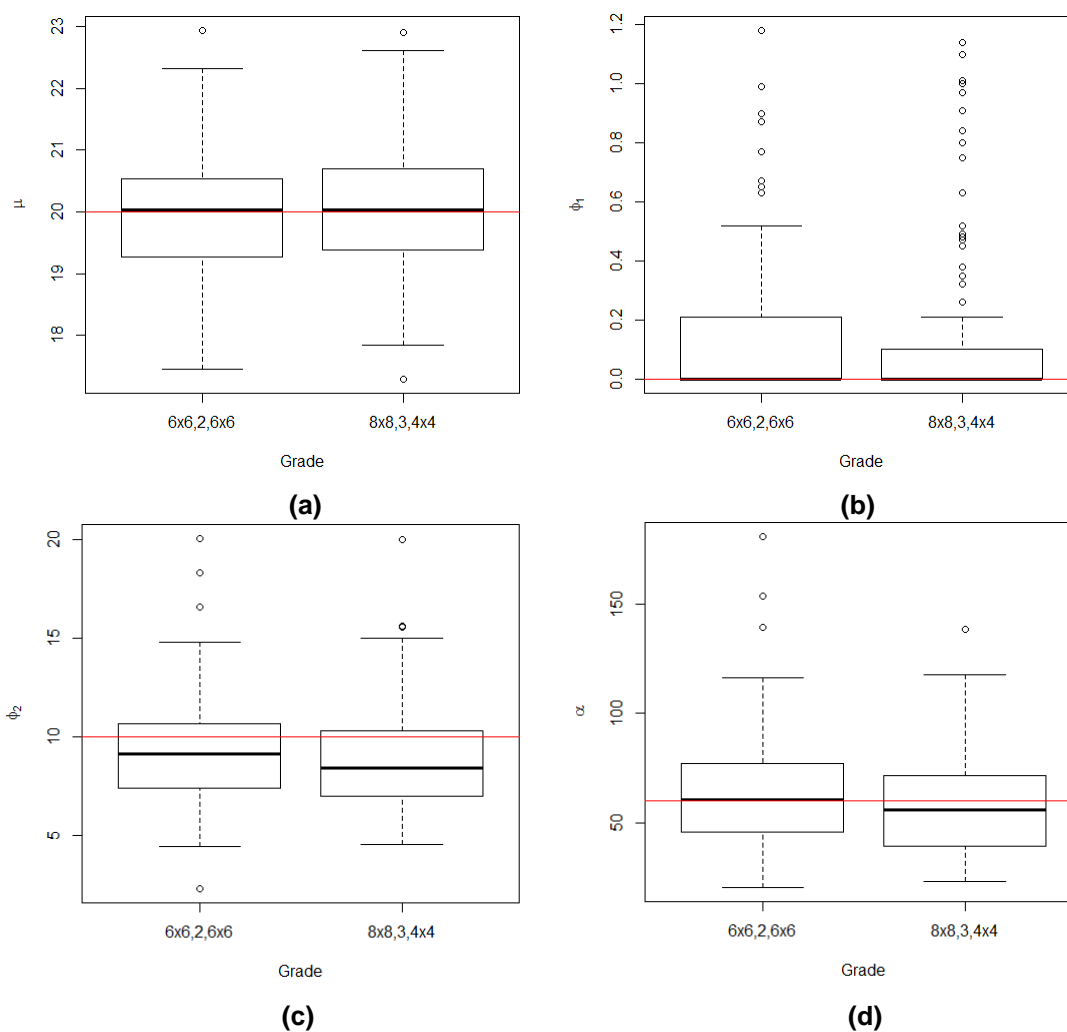


Figura 21 Gráfico boxplot dos valores estimados dos seguintes parâmetros: (a) Média, (b) Efeito Pepita, (c) Contribuição e (d) Alcance Prático, a linha tracejada indica o valor nominal simulado

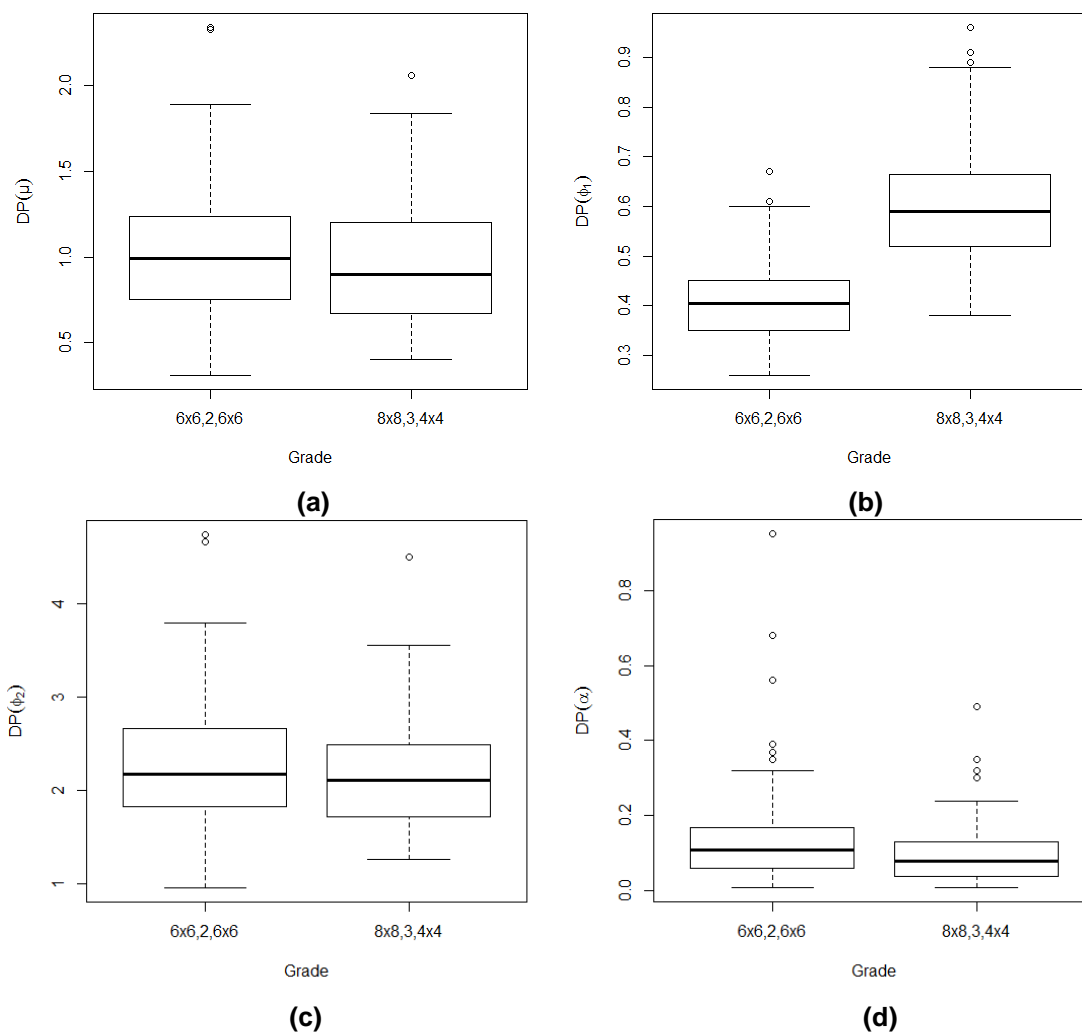


Figura 22 Gráfico boxplot do desvio padrão das estimativas dos parâmetros: (a) Média, (b) Efeito Pepita, (c) Contribuição e (d) Alcance Prático

As medidas de validação Cruzada usadas para avaliar a qualidade da estimação do modelo exponencial são exibidas na Figura 23. A partir dessas medidas, observa-se que os valores de SEM, EA, AIC e BIC foram menores para a grade (6x6,2,6x6) e os valores de EM mais próximos de zero e os valores de SEMR foram mais próximos de um para a grade (8x8,3,4x4). Logo, ao se considerar a maioria dessas medidas, conclui-se que a amostragem sistemática centrada 6x6 com 2 amostragens sistemáticas centradas menores 6x6 forneceu melhores estimativas das medidas associadas à validação cruzada e as medidas quanto à qualidade da estimação dos parâmetros do modelo exponencial.

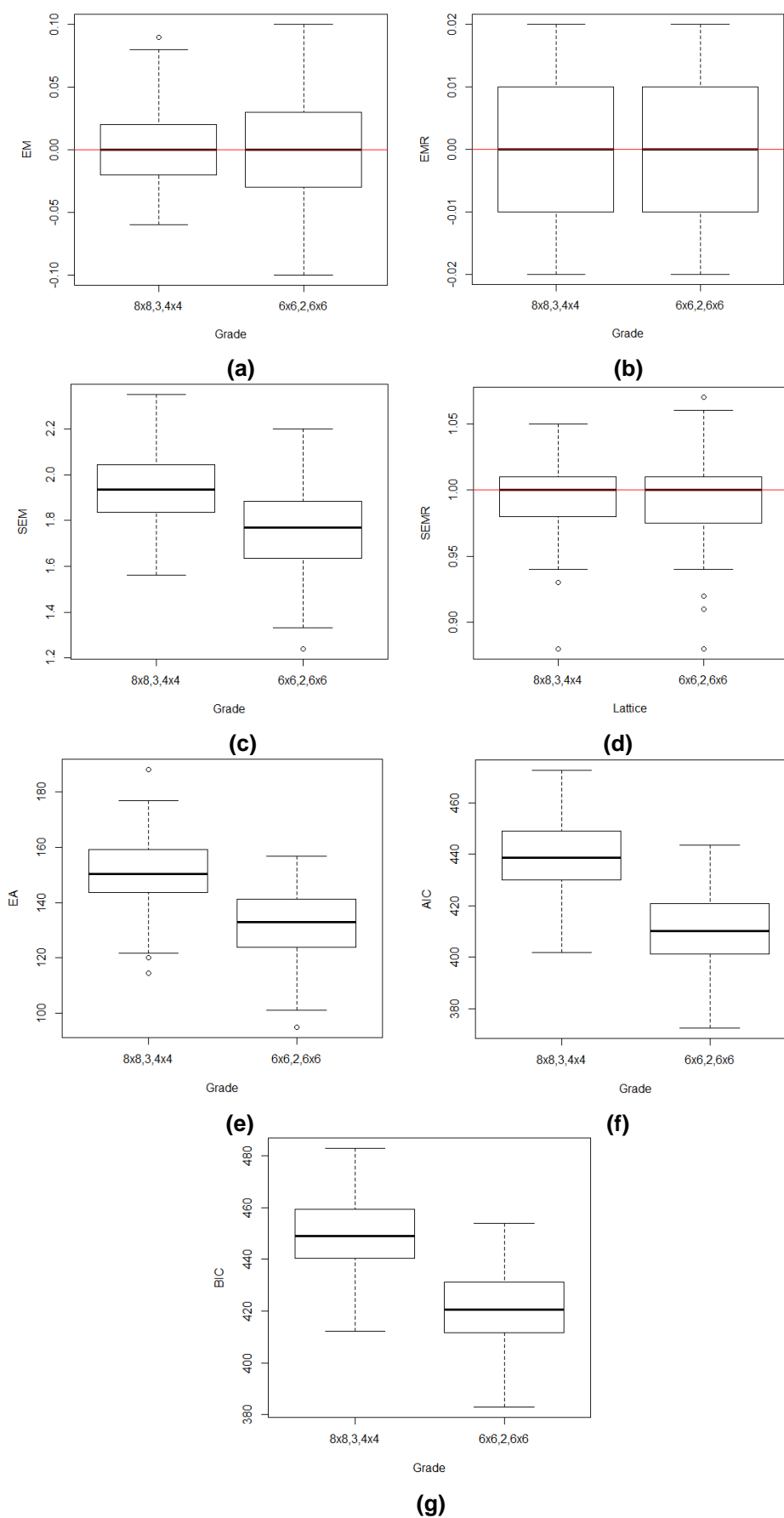


Figura 23 Gráficos boxplot: (a) Erro Médio, (b) Erro Médio Relativo, (c) Desvio Padrão do Erro Médio, (d) Desvio Padrão do Erro Médio Relativo, (e) Erro Absoluto, (f) AIC e (g) BIC. A linha vermelha indica o valor ideal

Porém, nas medidas da qualidade da predição espacial (Tabela 6 e Figura 24), a análise descritiva do erro da predição espacial feita pela krigagem e a média da variância mostraram resultado oposto. A média da variância apresentou, em média, valores menores para a grade (8x8,3,4x4), assim como a REQM, que mede a eficiência do estimador.

Tabela 6 Análise descritiva das medidas da qualidade da predição espacial em localizações não amostradas, considerando uma amostra teste composta por 25 pontos

Estatística Descritiva	Grade	Erro da predição	Média da variância
Média		0,045	3,0545
Desvio Padrão	(8x8,3,4x4)	0,357	0,513
CV (%)		769,2	16,79
REQM		1,493	-
Média		0,046	4,1306
Desvio Padrão	(6x6,2,6x6)	0,436	0,67
CV (%)		935,9	16,37
REQM		2,155	-

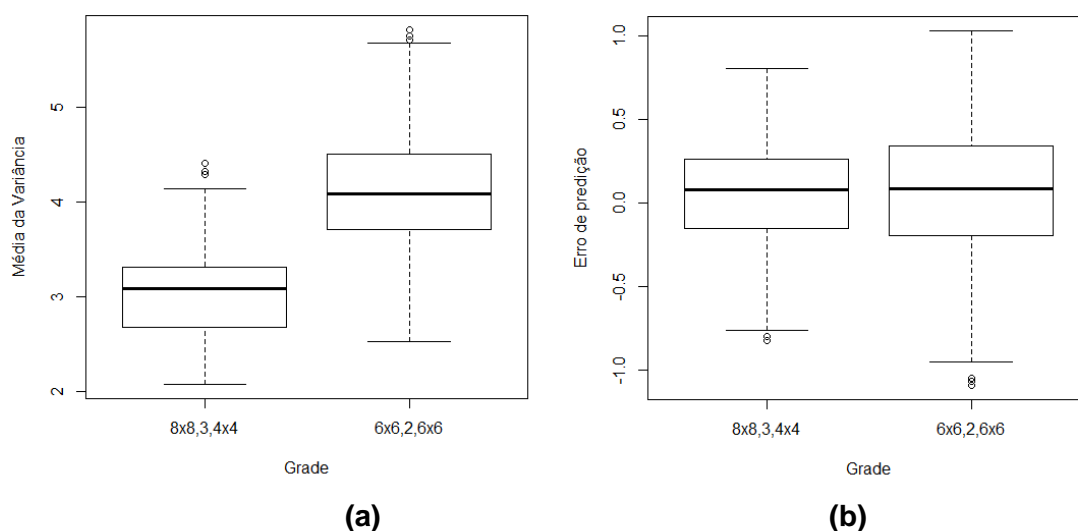


Figura 24 Gráfico boxplot: (a) Média da variância da krigagem e (b) Erro da predição

Como pode ser visto, obtiveram-se os piores resultados quanto à qualidade da predição espacial para a amostragem in fill (6x6,2,6x6). Segundo Soares (2014), isto acontece pelo fato de a variância de estimação aumentar com a covariância entre as amostras, ou seja, para a mesma distância ao ponto a se estimar e que, quanto mais próximas as amostras estiverem umas das outras, maior é efeito de redundância de informação, logo pior é a estimação em localizações não amostradas.

10.1.4 Amostragens Sistemática e Aleatória comparadas com as amostragens *Lattice plus close pairs* e *Lattice plus in-fill*

Nesta seção, foi feita a comparação entre os melhores cenários pesquisados quanto à qualidade da estimação dos parâmetros do modelo geoestatístico e à qualidade da predição espacial de valores de uma variável georreferenciada em localizações não amostradas. Como a *lattice plus close pairs* (7x7,51,1) apresentou a melhor estimativa dos parâmetros efeito pepita, contribuição e alcance prático, e ainda as melhores estimativas de valores em locais não amostrados, e a *lattice plus in fill* (8x8,3,4x4) apresentou as melhores medidas quanto à predição espacial bem como foi realizada a comparação das duas amostragens com a amostragem sistemática (10x10) e aleatória.

Na Tabela 7, estão apresentadas as medidas descritivas dos parâmetros do modelo Exponencial estimado por ML, os desvios-padrão das estimativas dos parâmetros desse modelo, e as medidas de eficiência do estimador para as amostragens sistemática (10x10) e aleatória. Os parâmetros média e contribuição não apresentaram diferença relevante em suas estimativas nas duas amostragens, porém, os valores estimados do efeito pepita foram mais próximos ao seu valor nominal na amostragem aleatória, e o alcance com valores estimados mais próximos ao nominal na amostragem sistemática.

Quando os resultados das amostragens Aleatória e Sistemática (Tabela 7) foram comparados com a *Lattice plus close pairs* (7x7,51,1) (Tabela 3) e a *Lattice plus in fill* (8x8,3,4x4) (Tabela 5), tem-se que os valores estimados da média mantiveram-se similares e próximos ao valor nominal nas quatro amostragens (Figura 25-a). Os valores estimados do efeito pepita (Figuras 25-b) foram mais próximo ao nominal para a grade (7x7,51,1). Todavia, de forma geral, os valores estimados dos parâmetros contribuição e alcance prático foram similares em todas as amostragens (Figuras 25-c, 25-d). No entanto, observa-se que a amostragem sistemática teve um maior número de simulações, nas quais houve subestimação do alcance, quando comparada às demais amostragens (Figura 25-d).

Os desvios-padrão das estimativas dos parâmetros média e alcance prático apresentaram valores similares para todas as amostragens (Figuras 26-a, 26-d). No entanto, os desvios-padrão das estimativas do efeito pepita e da contribuição apresentaram valores menores para a *Lattice plus close pairs* (7x7,51,1) e os maiores valores para a amostragem sistemática (Figuras 26-b, 26-c).

As medidas de eficiência do estimador (Tabelas 3, 5 e 7) para as amostragens aleatória, sistemática (10x10), *lattice plus close pairs* (7x7,51,1) e *lattice plus in fill* (8x8,3,4x4) mostram que a amostragem *lattice plus close pairs* (7x7,51,1) obteve os

menores valores de VRA, VA e REQM dos valores estimados dos parâmetros efeito pepita, contribuição e alcance prático, quando comparadas com as demais.

Tabela 7 Análise descritiva dos parâmetros do modelo exponencial estimado por ML e seus respectivos desvios padrões e as medidas de eficiência do estimador VRA(%), VA e REQM para dados isotrópicos

Análise Descritiva	Grid	Média ($\mu=20$)	Efeito Pepita ($\varphi_1=0$)	Contribuição ($\varphi_2=10$)	Alcance Prático ($a=60$)	DP (μ)	DP (φ_1)	DP (φ_2)	DP (a)
Média		19,92	0,1403	8,7	56,53	1,0271	0,5194	2,16	0,107
Desvio Padrão		1,16	0,23	2,81	22,23	0,38	0,093	0,66	0,091
CV (%)		5,82	164,05	32,35	39,32	37,19	18,06	30,55	84,78
VRA (%)	Aleatória	0,3875	-	12,93	5,77				
VA		0,077	0,1403	1,293	3,46				
REQM		0,669	0,036	4,76	250,61				
Média		19,95	0,2973	8,95	60,09	0,957	1,1861	2,89	0,1073
Desvio Padrão		1,01	0,67	3,42	26,84	0,404	0,245	0,903	0,12
CV (%)		5,06	225,66	38,23	44,67	42,23	20,74	31,24	113,9
VRA (%)	(10x10)	0,2425	-	10,464	0,154				
VA		0,0485	0,2973	1,0464	0,0924				
REQM		0,506	0,267	6,348	356,77				

Valores nominais: μ é a média, φ_1 é o efeito pepita; φ_2 é a contribuição; a é o alcance prático.

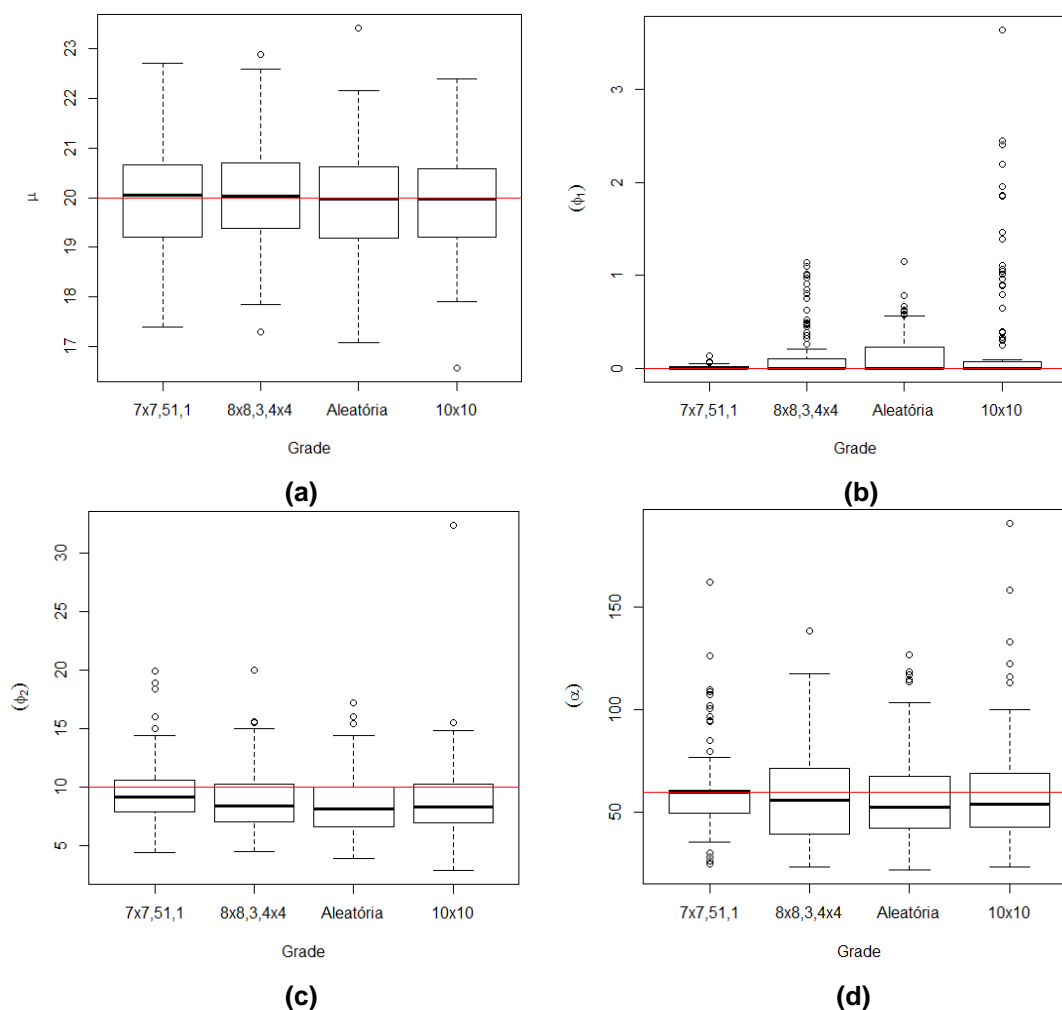


Figura 25 Gráfico boxplot dos valores estimados dos seguintes parâmetros: (a) Média, (b) Efeito Pepita, (c) Contribuição e (d) Alcance Prático

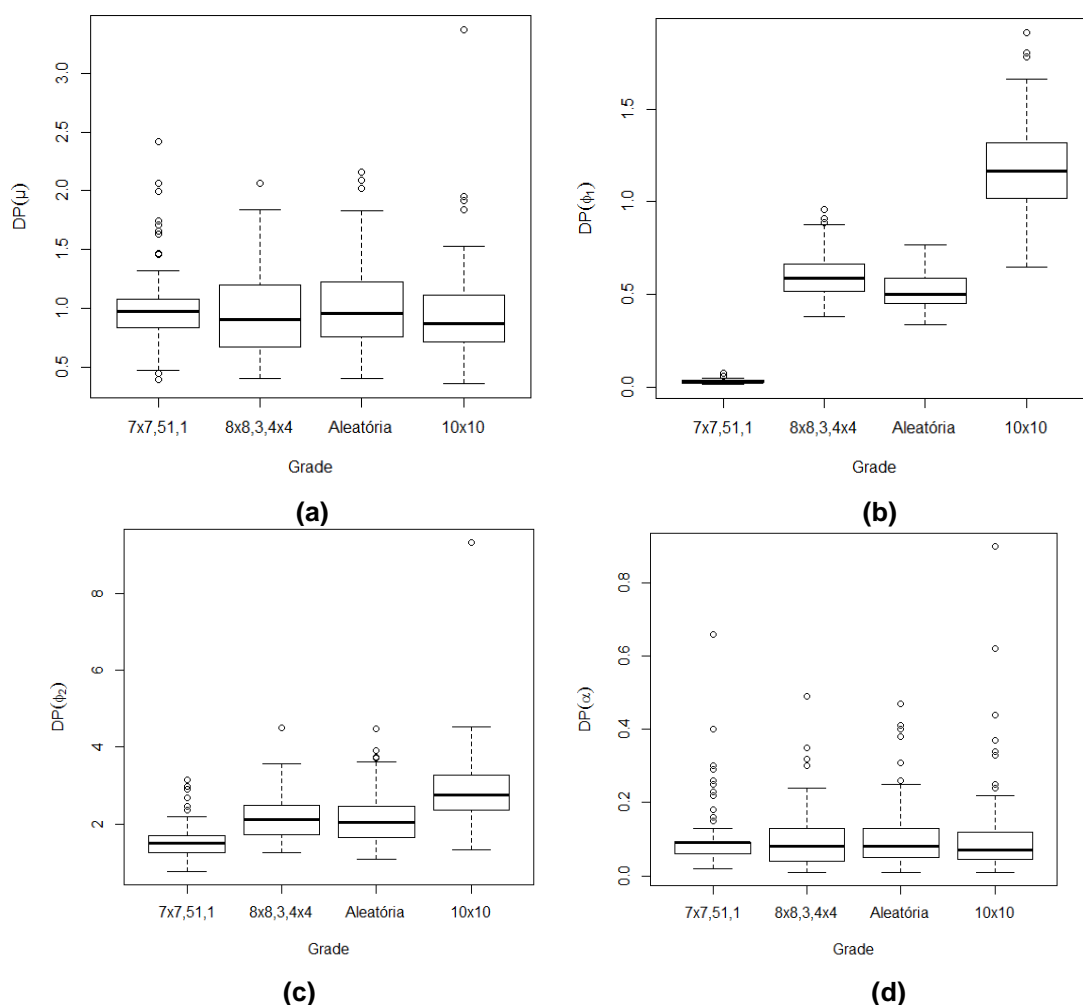


Figura 26 Gráfico boxplot do desvio padrão das estimativas dos parâmetros: (a) Média, (b) Efeito Pepita, (c) Contribuição e (d) Alcance Prático

As medidas de validação usadas para avaliar a qualidade da estimação do modelo são exibidas na Figura 27. Observa-se, de acordo com essas medidas, que o EM e o EMR foram mais próximos de zero. Assim, o SEMR foi mais próximo de um para a amostragem sistemática (10x10), quando comparado com as demais amostragens. As demais medidas (SEM, EA, AIC e BIC) apresentaram valores menores para a *lattice plus close pairs* (7x7,51,1), quando comparadas com as demais amostragens. Desta forma, considerando-se a maioria dessas medidas, conclui-se que a amostragem *lattice plus close pairs* 7x7 com 51 pontos próximos e o raio de pontos próximos igual a 1 forneceram melhores estimativas das medidas associadas à validação cruzada e as medidas quanto à qualidade da estimação dos parâmetros do modelo exponencial.

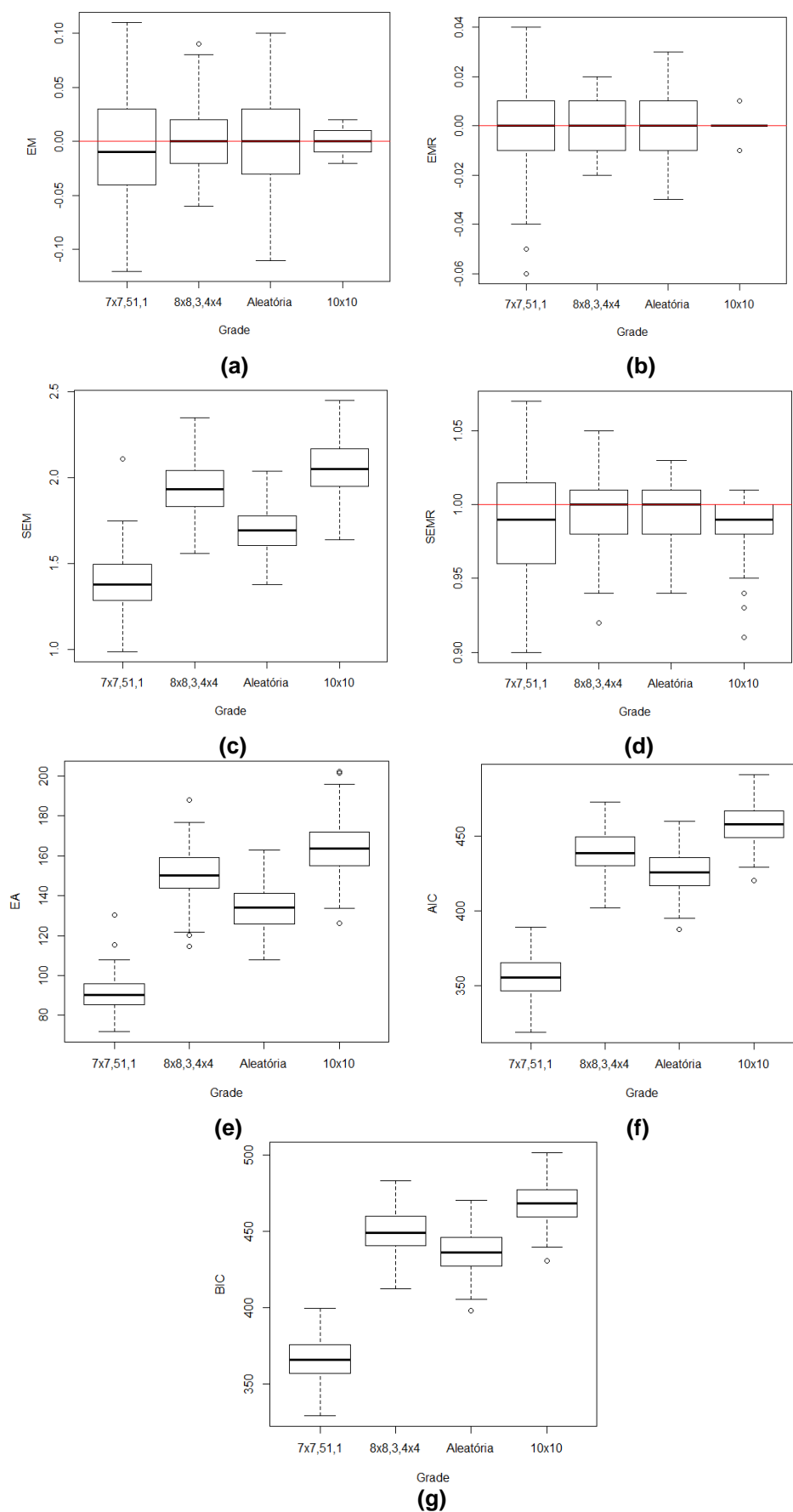


Figura 27 Gráficos boxplot: (a) Erro Médio, (b) Erro Médio Relativo, (c) Desvio Padrão do Erro Médio, (d) Desvio Padrão do Erro Médio Relativo, (e) Erro Absoluto, (f) AIC e (g) BIC. A linha tracejada indica o valor ideal

As medidas de qualidade da predição espacial, erro da predição, REQM e da média variância da krigagem (Tabela 8 e Figura 28) não apresentaram diferença relevante nas amostragens sistemática (10x10) e aleatória. E quando são comparadas com a *lattice plus close pairs* e a *lattice plus in fill* (Tabelas 4 e 6), nota-se que a amostragem *lattice plus close pairs* (7x7,51,1) forneceu melhores estimativas da variável georreferenciada em localizações não amostradas.

Tabela 8 Análise descritiva das medidas da qualidade da predição espacial em localizações não amostradas, considerando uma amostra-teste composta por 25 pontos

Estatística Descritiva	Grade	Erro da predição	Média da variância
Média	(10x10)	0,041	3,023
Desvio Padrão		0,337	0,58
CV (%)		818,7	19,45
REQM		1,41	-
Média	Aleatória	0,054	2,8148
Desvio Padrão		0,332	0,425
CV (%)		604,97	15,13
REQM		1,389	-

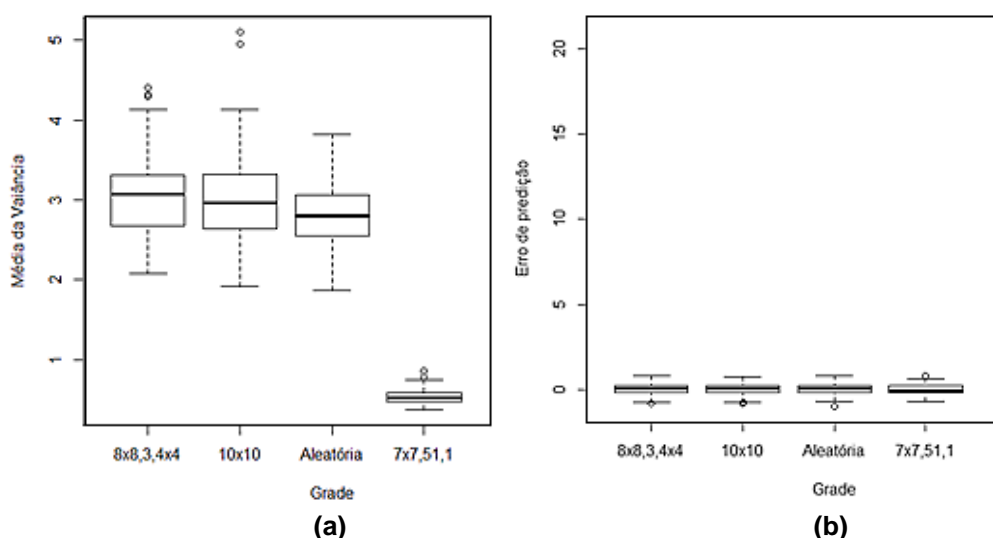


Figura 28 Gráfico boxplot: (a) Média da variância da krigagem e (b) Erro da predição

Os resultados aqui obtidos são semelhantes aos de Diggle e Lophaven (2006), os quais simularam três amostragens: a sistemática (8x8), a *Lattice plus close pairs* (7x7,15,0.5) e *in fill* (7x7,3,3x3) com 64 pontos. Os autores concluíram que a amostragem sistemática centrada com pares de pontos próximos (*lattice plus close pairs*) forneceu as melhores estimativas dos parâmetros do modelo geoestatístico e da predição espacial feita pela krigagem, seguidas pela amostragem *lattice plus in fill* e sistemática.

É importante ressaltar que a amostragem *lattice plus close pairs* (7x7,51,1) teve melhores resultados nas duas etapas da análise geoestatística, tanto na estimação dos parâmetros do modelo como na predição espacial (krigagem). Vale ainda destacar os bons resultados obtidos nessa amostragem quanto à estimativa do efeito pepita, da contribuição

e do alcance prático. Visto que o efeito pepita desempenha um papel importante nos resultados obtidos pela estimação espacial de localizações não amostradas, usando como estimador a krigagem ordinária. Pois, o efeito pepita exerce uma influência negativa na estabilidade desse estimador, ou seja, quanto maior é o efeito pepita, menor será a eficiência da krigagem ordinária, quanto a estimação espacial (ANDRIOTTI, 2002).

Para a análise dos fatores que qualificam a predição espacial feita pela krigagem, foi feita uma comparação entre os mapas da variância (Figura 29) e do erro da predição espacial (Figura 30) feito pela krigagem na amostra teste. Para isso, foram usadas as melhores configurações espaciais pesquisadas.

O mapa da variância da krigagem (Figura 29) ilustra que, quando há muitos pontos próximos ao ponto a ser estimado, maior é a variância da krigagem. Estes resultados corroboram com a discussão feita por Soares (2014). Veja como exemplo nas amostragens pesquisadas que os pontos estimados que estão circulados na *Lattice plus close pairs 7x7,51,1* (Figura 29-a), e na aleatória (Figura 29-c) tiveram valores menores de variância da krigagem pelo fato de ter menor quantidade de pontos próximos que circundam o ponto a se estimar. Já no caso da *Lattice plus in fill (8x8,3,4x4)* (Figura 29-b) e da Sistemática (Figura 29-d), houve maior quantidade de pontos próximos ao ponto a se estimar e, por conseguinte, maior foi a variância da krigagem. Note que as amostragens *lattice plus close pairs* e aleatória apresentaram de modo geral os menores valores de variância da krigagem.

Quanto à variância da krigagem, Yamamoto e Landim (2013) e Soares (2014) destacam o cuidado que deve-se tomar ao usar a medida de incerteza, uma vez que ela reflete somente as relações da geometria das amostras e do domínio a se estimar. Além disso, não se leva em conta a variabilidade das próprias amostras, ou seja, a mesma configuração espacial entre amostras e o ponto a se estimar produz rigorosamente a mesma variância de estimação, independente da variabilidade local das amostras.

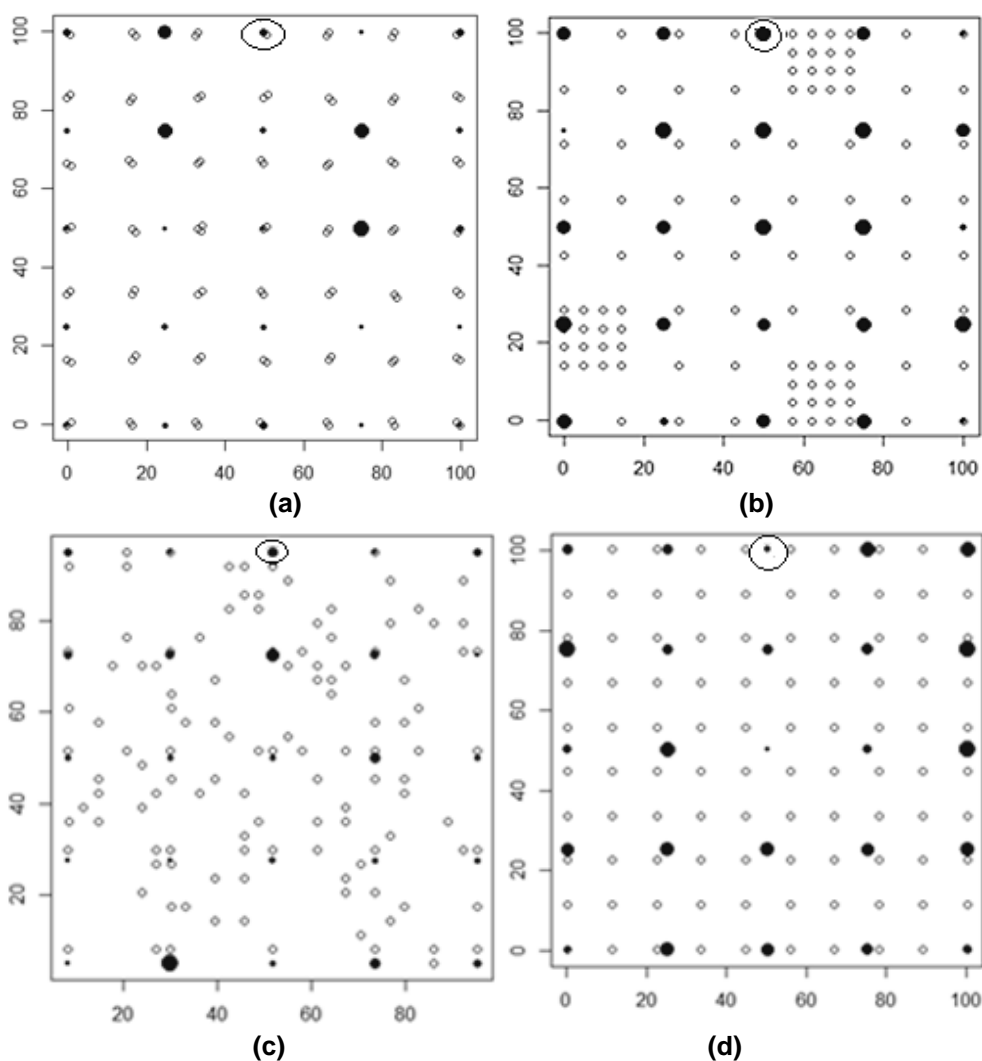


Figura 29 Mapa da variância da Krigagem da amostra teste usando as amostragens (a) *Lattice Plus close pairs* 7x7,51,1 (b) *Lattice plus in fill* 8x8,3,4x4 (c) Aleatória (d) Sistemática. As esferas em azul representam o valor da variância da krigagem na amostra-teste. E os valores estão representados pelo tamanho da esfera

O mapa do erro da predição espacial feito pela krigagem na amostra teste (Figura 30) ilustra que, quanto mais próximas as amostras estiverem umas das outras, maior é o efeito de redundância da informação, logo pior foi a estimativa. Os resultados corroboram com a discussão feita por Soares (2014). Por exemplo, na amostragem simulada *lattice plus in fill* (8x8,3,4x4), (Figura 30-b), o agrupamento de amostras próximas ao ponto a se estimar proporcionou maior erro de estimativa espacial da variável georreferenciada em localizações não amostradas. Já nas amostragens *Lattice plus close pairs* (7x7,51,1) (Figura 30-a), Aleatória (Figura 30-c) e Sistemática 10x10 (Figura 30-d), no ponto usado como exemplo, não foi observado agrupamento de amostras tão próximas ao ponto a se estimar como nas demais amostragens, logo, a estimativa feita pela krigagem foi melhor.

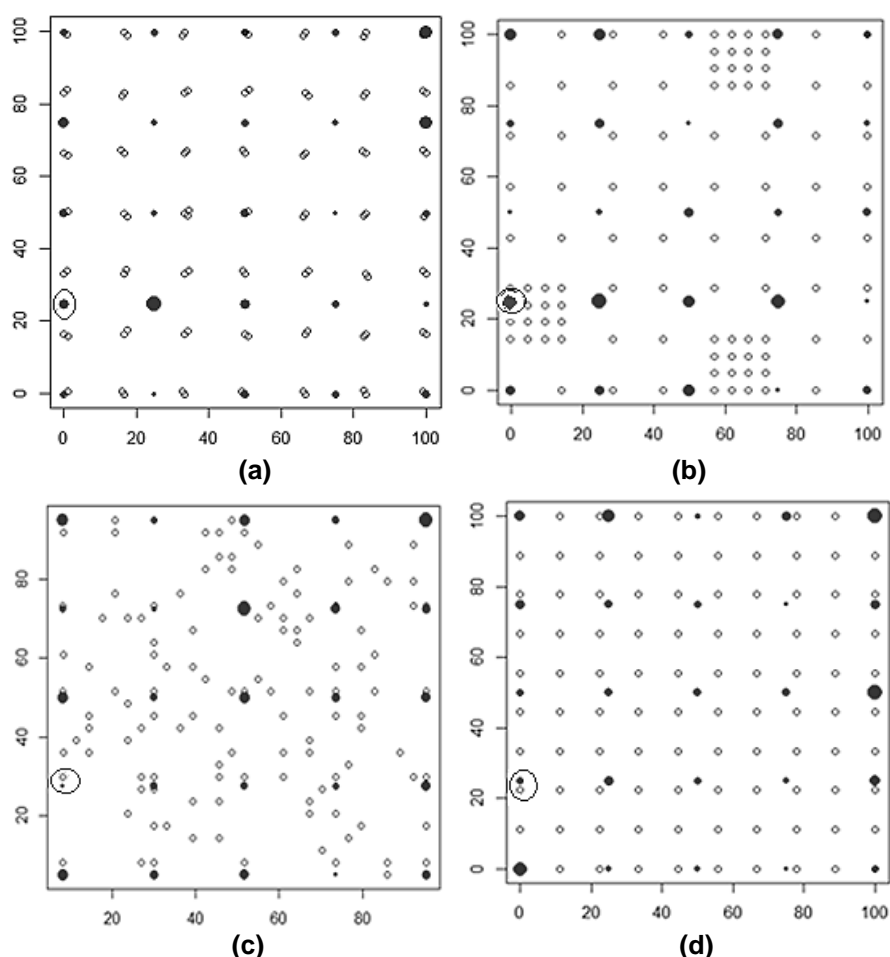


Figura 30 Erro da predição espacial feita pela krigagem da amostra teste usando as amostragens: (a) *Lattice Plus close pairs* 7x7,51,1 (b) *Lattice plus in fill* 8x8,3,4x4 (c) Aleatória (d) Sistemática. As esferas representam o valor do Erro da predição espacial da amostra-teste, cujos valores estão representados pelo seu tamanho.

11.1.2 Dados com tendência direcional

Nesta seção, estão os resultados obtidos nas simulações das amostragens sistemática 10x10, aleatória e *lattice plus close pairs* (7x7,51,1) com tendência direcional na direção 0° (sistema Azimuth). Não foram simulados dados com a amostragem *lattice plus in fill* com tendência direcional, pelo fato dessa não ter apresentado bons resultados nas etapas da análise geoestatística. A amostragem sistemática e a *lattice plus close pairs* ainda foram trabalhadas em outras duas versões. Na sistemática, as versões foram 5x20 e 20x5, portanto houve maior concentração de pontos na direção da tendência (no primeiro cenário) e maior concentração de pontos na direção ortogonal a tendência (no segundo cenário). A amostragem *lattice plus close pairs* (7x7,51,1) foi trabalhada com adição dos pontos próximos na direção da tendência e na direção ortogonal a essa. A amostragem aleatória foi trabalhada apenas na sua versão original.

Na Tabela 9, são apresentadas as medidas descritivas dos valores estimados dos parâmetros do modelo exponencial. Também são apresentados os desvios-padrão das estimativas dos parâmetros, as medidas de eficiência do estimador e a análise descritiva do coeficiente linear de correlação linear de Pearson dos valores da variável simulada versus a coordenada y (direção 90° na área) para todas as amostragens consideradas.

Veja que os parâmetros estimados dos parâmetros β_0 , β_1 , alcance e contribuição não apresentaram diferenças relevantes entre as amostragens quando comparados entre si (Tabela 9 e Figuras 31-a, 31-b, 31-d, 31-e). Já os valores estimados do parâmetro efeito pepita apresentaram melhores resultados nas amostragens (7x7,51,1), (7x7,51,1) com adição dos pontos próximos na direção ortogonal à tendência e (7x7,51,1) com adição dos pontos próximos na direção da tendência (Tabela 9 e Figuras 31-c). Ou seja, valores estimados mais próximos ao valor nominal.

Dados referentes ao desvio padrão das estimativas dos parâmetros β_0 , β_1 , à contribuição e ao alcance não apresentaram diferenças relevantes entre as amostragens pesquisadas e quando comparadas entre si (Tabela 9 e Figuras 32-a, 32-b, 32-d, 32-e). Já o desvio padrão das estimativas do efeito pepita apresentou uma ligeira queda nos seus valores nas versões pesquisadas da amostragem *lattice plus close pairs* (Tabela 9 e Figura 32-c).

As medidas de eficiência do estimador VRA, VA e REQM para a estimativa de todos os parâmetros diminuem nas versões da amostragem *lattice plus close pairs*, com exceção da REQM do parâmetro β_1 que foi similar em todas as amostragens pesquisadas.

Em todas as amostragens pesquisadas as simulações apresentam, em média, coeficiente de correlação linear moderada positiva, com exceção da amostragem aleatória, que apresentou, em média, correlação linear fraca (Tabela 9) (MUKAKA, 2012). Veja no gráfico de colunas (Figura 33) que, em todas as amostragens, as simulações tiveram valor de coeficiente linear de Pearson moderado positivo ($0,5 \leq r < 0,8$), no qual corresponde à maior frequência (60% das simulações), com exceção da amostragem aleatória que teve 50% das suas simulações com coeficiente de correlação linear fraco ($0 \leq r < 0,5$) e 42% moderado positivo ($0,5 \leq r < 0,8$).

Com base nas medidas de validação cruzada, observa-se nas amostragens pesquisadas que a tendência influencia na estimativa dos parâmetros do modelo. Quando se utilizam dados com tendência direcional na amostragem sistemática esse fenômeno reduz a qualidade da estimativa do alcance prático quando comparado com a sistemática (10x10) sem tendência direcional (Tabela 7). Embora não tenha apresentado influência relevante na estimativa dos parâmetros efeito pepita e contribuição, ainda se verificou que a adição de pontos na direção da tendência (5x20) e na direção ortogonal a esta (20x5) não apresentou melhora relevante na qualidade da estimativa destes parâmetros (Tabela 9).

Tabela 9 Estatísticas descritivas dos parâmetros do modelo exponencial estimado por ML para as grades Aleatória, sistemática 10x10, 5x20 e 20x5 e *lattice plus close pairs* 7x7,51,1 com tendência direcional

Análise Descritiva	Grid	Beta 0 ($\beta_0=10$)	Beta 1 ($\beta_1=0,06$)	Efeito Pepita ($\varphi_1=0$)	Contribuição ($\varphi_2=10$)	Alcance Prático ($\alpha=60$)	DP (β_0)	DP(β_1)	DP (φ_1)	DP (φ_2)	DP (φ_3)	Coef. Linear de Pearson	
Média	Aleatória	10,16	0,057	0,1317	7,83	49,65	1,6	0,025	0,51	1,9	0,08	0,448	
Desv. P.		2,19	0,032	0,275	2,58	18,24	0,47	0,007	0,1	0,57	0,07	0,2352	
CV (%)		21,58	56,17	209,15	33,02	36,75	29,1	27,79	19,6	29,2	80,6	52,47	
VRA (%)		1,644	4,66	-	21,68	17,24							
VA		0,1644	0,0028	0,1317	2,1683	10,34							
REQM		2,39	0,0005	0,046	5,66	218,37							
Média	(10x10)	9,905	0,0601	0,2392	8,19	52,72	1,3	0,0196	1,24	2,7	0,08	0,5334	
Desv. P.		1,62	0,024	0,622	2,84	22,36	0,43	0,006	0,27	0,73	0,08	0,1825	
CV (%)		16,34	40,85	260,25	34,76	42,42	33,2	30,69	21,9	27,1	101	34,23	
VRA (%)		0,942	0,166	-	18,047	12,126							
VA		0,0942	0,0001	0,2392	1,8047	7,276							
REQM		1,301	0,0002	0,22	5,64	274,04							
Média	(5x20)	9,91	0,0606	0,1165	8,03	50,46	1,3	0,0205	0,57	2,06	0,07	0,5209	
Desv. P.		1,6	0,23	0,2425	2,83	19,78	0,42	0,0057	0,1	0,66	0,06	0,1873	
CV (%)		16,14	39,3	208,15	35,23	39,19	32,2	28,05	18,3	32,1	91,5	35,96	
VRA (%)		0,0841	1	-	19,67	15,88							
VA		0,0841	0,0006	0,1165	1,9669	9,53							
REQM		1,27	0,0002	0,035	5,9	239,13							
Média	(20x5)	9,917	0,0604	0,1128	8,19	51,48	1,29	0,0196	0,57	2,09	0,07	0,5808	
Desv. P.		1,56	0,023	0,263	2,51	18,38	0,39	0,006	0,11	0,58	0,05	0,176	
CV (%)		15,8	38,44	234,07	30,69	35,71	30,3	31,54	19,3	27,6	82,7	30,44	
VRA (%)		0,821	0,667	-	18,09	14,19							
VA		0,0821	0,0004	0,1128	1,81	8,51							
REQM		1,22	0,0002	0,0408	4,766	203,57							
Média	7x7,51,1	9,98	0,059	0,0349	8,69	56,03	1,39	0,0215	0,15	1,82	0,09	0,54	
Desv. P.		1,49	0,021	0,063	1,15	27,42	0,51	0,006	0,03	0,65	0,12	0,183	
CV (%)		14,97	36,71	182,38	36,24	48,93	36,3	29,09	18,8	35,9	127	33,9	
VRA (%)		0,161	0,667	-	13,088	6,604							
VA		0,0161	0,0004	0,0349	1,308	3,96							
REQM		1,105	0,0002	0,0026	5,76	379,92							
Média	7x7,51,1(a)	9,99	0,059	0,02	8,72	55,8	1,39	0,023	0,09	1,72	0,09	0,541	
Desv. P.		1,49	0,022	0,042	3,11	25,72	0,48	0,006	0,02	0,61	0,11	0,18	
CV (%)		14,91	36,82	212,42	35,63	46,09	35,1	29,58	18,9	35,6	119	33,52	
VRA (%)		0,004	1	-	12,7	6,99							
VA		0,0004	0,0006	0,0201	1,27	4,19							
REQM		1,1	0,0002	0,0011	5,59	336,37							
Média	7x7,51,1(b)	9,98	0,059	0,012	8,75	58,89	1,39	0,02	0,06	1,56	0,08	0,5414	
Desv. P.		1,48	0,022	0,028	2,99	22,44	0,46	0,006	0,01	0,52	0,08	0,179	
CV (%)		14,91	37,2	222,37	34,18	40,88	33	28,33	21,6	33,6	102	33,17	
VRA (%)		0,185	0,33	-	12,4	8,51							
VA		0,018	0,0002	0,0126	1,24	5,11							
REQM		1,09	0,0002	0,0004	5,2	262,37							

Valores nominais: μ é a média, φ_1 é o efeito pepita; φ_2 é a contribuição; α é o alcance prático; (a) *Lattice plus close pairs* com adição dos pontos próximos na direção ortogonal a tendência (b) *Lattice plus close pairs* com adição dos pontos próximos na direção da tendência.

Na *lattice plus close pairs* (7x7,51,1), observa-se que a tendência direcional piorou a estimativa dos parâmetros efeito pepita, contribuição e alcance quando comparados com essa sem tendência direcional (Tabela 3). Além disso, a adição dos pontos próximos na direção da tendência e na direção ortogonal a esta não apresentou melhora relevante na estimativa dos parâmetros. Da mesma forma, na amostragem aleatória, a tendência direcional piorou a estimativa do alcance prático e da contribuição. E, apenas o efeito

pepita não apresentou diferenças relevantes de estimativas quando comparado com a amostragem aleatória sem tendência direcional (Tabela 7).

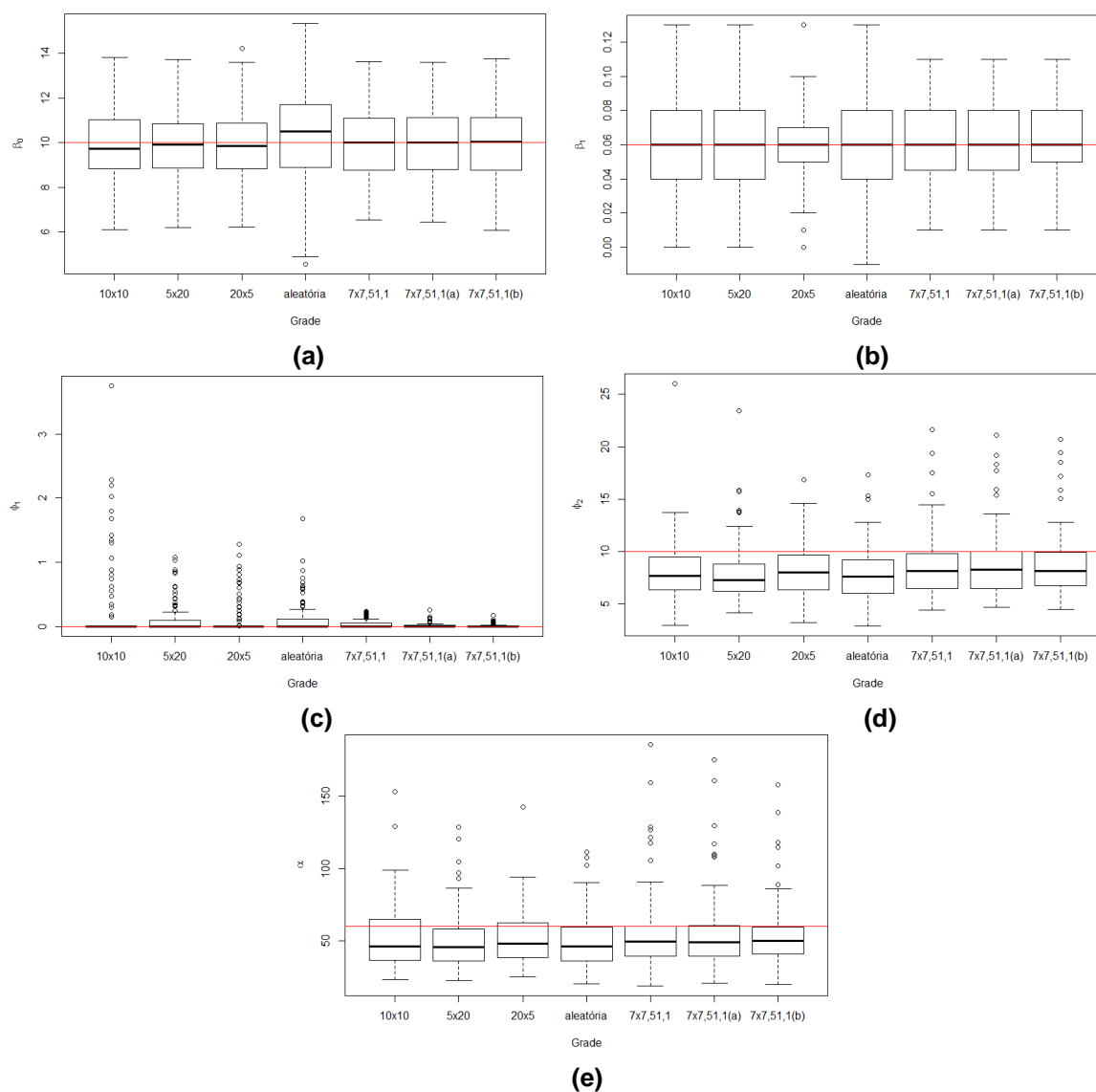


Figura 31 Gráfico boxplot das estimativas dos seguintes parâmetros: (a) β_0 , (b) β_1 , (c) Efeito pepita (d) Contribuição (e) Alcance prático

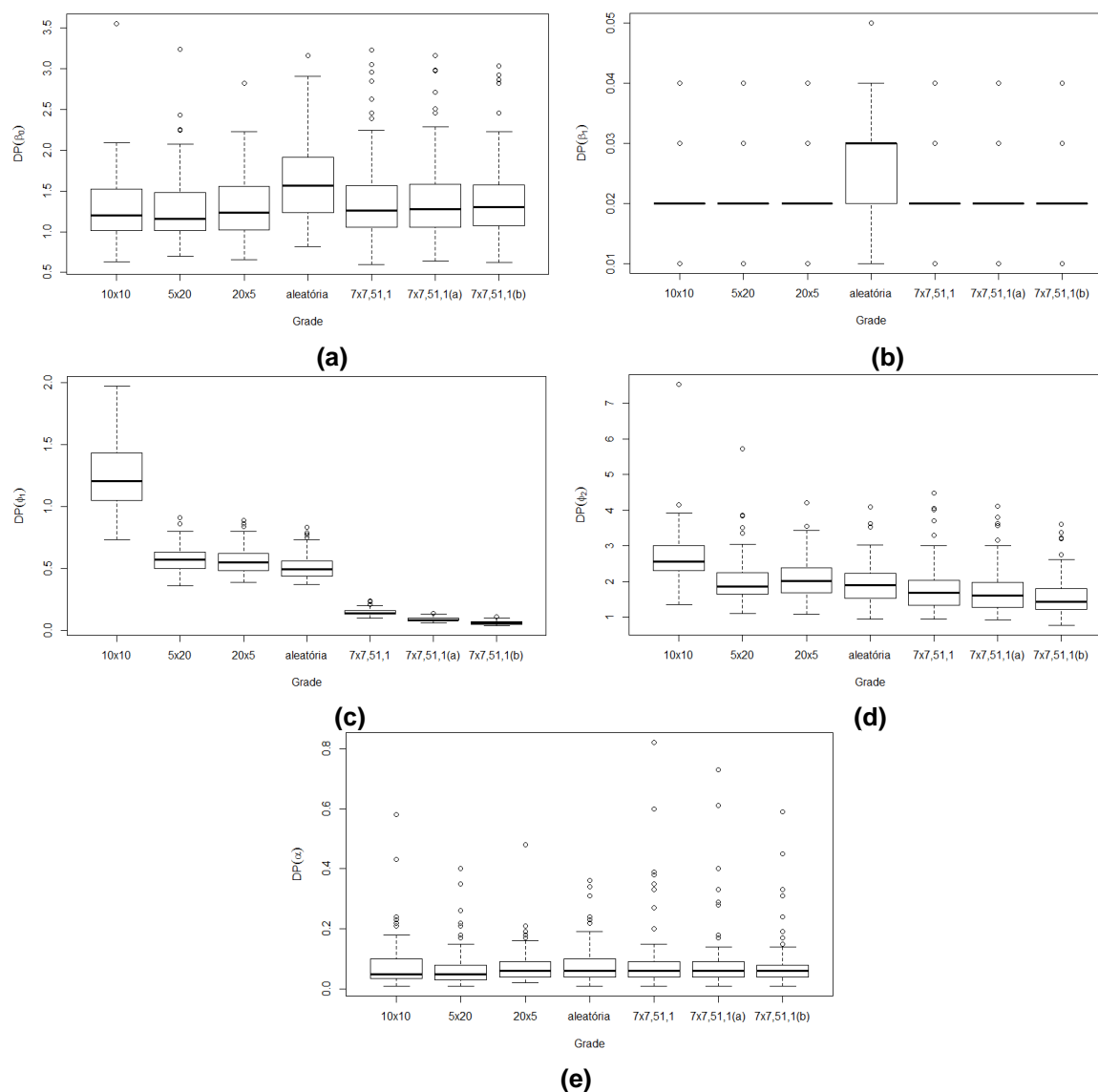


Figura 32 Gráfico boxplot do desvio padrão dos parâmetros estimados: (a) β_0 , (b) β_1 , (c) Efeito pepita (d) Contribuição (e) Alcance prático

As medidas de validação cruzada, usadas para avaliar a qualidade da estimação do modelo, estão exibidas na Figura 34. Observa-se por meio dessas medidas que, em ambos os cenários com tendência, os valores de EM, EMR e SEMR foram semelhantes (Figuras 34-a, 34-b, 34-d). Já os valores de SEM, EA, AIC e BIC foram menores para a amostragem 7x7,51,1, na qual observa-se ainda uma diminuição desses valores para as amostragens 7x7,51,1 com adição dos pontos próximos na direção da tendência e na direção ortogonal à tendência (Figuras 34-c, 34-e, 34-f, 34-g). Sendo assim, considerando-se a maioria dessas medidas, conclui-se que a *lattice plus close pairs* com tendência direcional, em sua versão original e com a adição de pontos próximos na direção da tendência e na direção ortogonal a essa, proporcionou melhores estimativas das medidas associadas à validação cruzada e à qualidade da estimativa dos parâmetros do modelo geoestatístico.

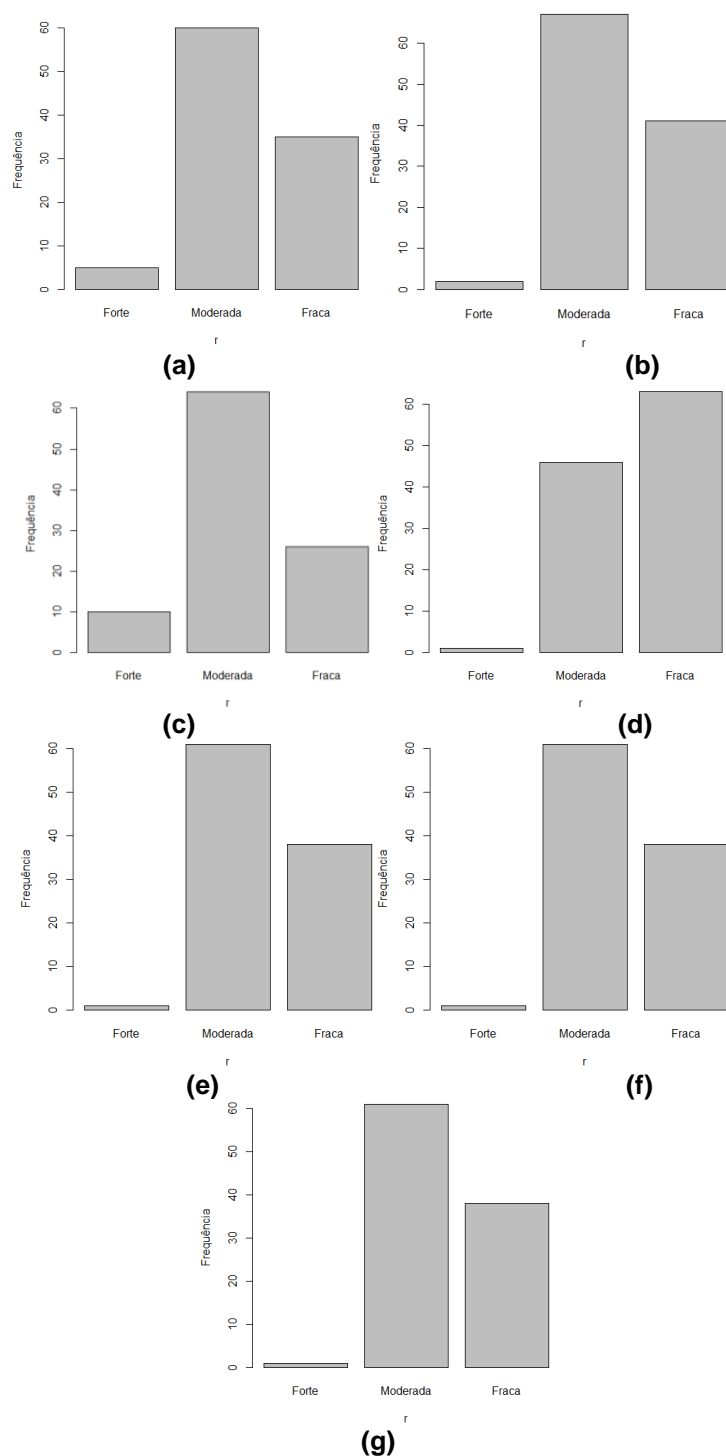


Figura 33 Gráfico de barras do coeficiente de correlação linear de Pearson nas simulações das amostragens (a) 10x10, (b) 5x20, (c) 20x5, (d) Aleatória, (e) 7x7,51,1, (f) 7x7,51,1(a) e (g) 7x7,51,1(b). Em que: Forte ($0,8 \leq r < 1$), Moderado ($0,5 \leq r < 0,8$) e Fraca ($0,1 \leq r < 0,5$)

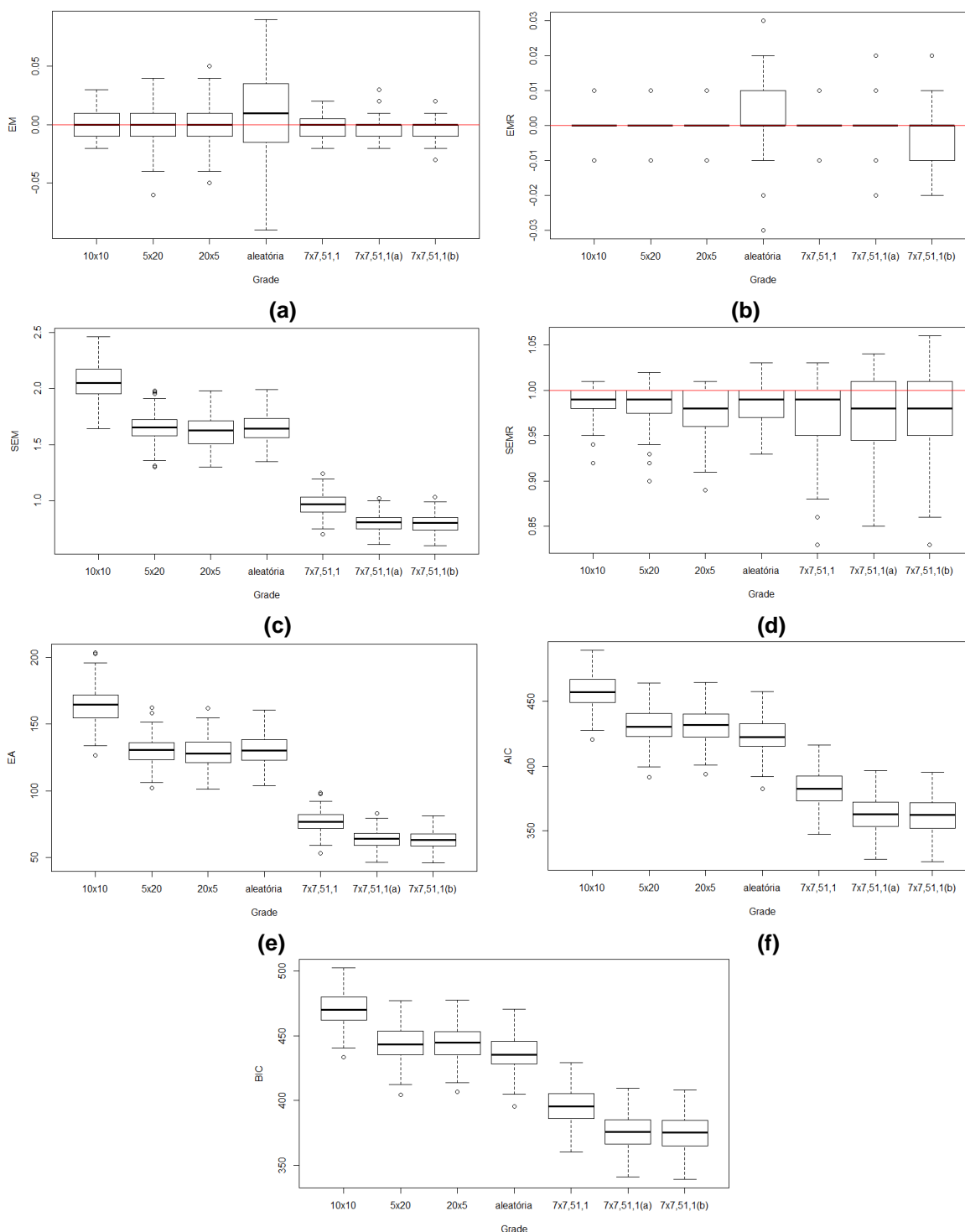


Figura 34 Gráficos boxplot: (a) Erro Médio, (b) Erro Médio Relativo, (c) Desvio Padrão do Erro Médio, (d) Desvio Padrão do Erro Médio Relativo, (e) Erro Absoluto, (f) AIC e (g) BIC. A linha horizontal indica o valor ideal.

A análise descritiva das medidas de qualidade da predição espacial, média da variância e do erro da predição na amostra teste (Tabela 10 e Figura 35) mostrou que a tendência direcional reduz a qualidade da predição espacial feita pela krigagem das amostragens pesquisadas quando comparadas com as amostragens sistemática (10x10), aleatória (Tabela 8) e *lattice plus close pairs* (7x7,51,1) (Tabela 4 e Figura 28) sem tendência direcional. E que a adição de pontos na direção da tendência e na direção

ortogonal a essa não proporcionou melhores estimativas da variável em localizações não amostradas.

Nota-se que as combinações da tendência direcional e das diferentes configurações amostrais não apresentaram influência relevante na qualidade das estimativas dos parâmetros e, como consequência, na qualidade da predição espacial. Isto ocorre porque, para efetuar a krigagem, esses parâmetros precisam ser estimados com precisão, de modo que as estimativas obtidas a partir da krigagem sejam mais exatas e consequentemente mais confiáveis (FERREIRA et al., 2013).

Tabela 10 Análise descritiva das medidas da qualidade da predição espacial do modelo exponencial sem tendência direcional quanto à predição espacial de uma amostra-teste composta por 25 pontos

Estatística Descritiva	Grade	Erro da predição	Média da variância
Média	Aleatória	-0,006	2,579
Desvio Padrão		0,33	0,393
CV (%)		-4992,31	15,26
REQM		1,409	-
Média	10x10	0,038	3,037
Desvio Padrão		0,35	0,586
CV (%)		923,7	19,32
REQM		1,407	-
Média	5x20	0,046	2,77
Desvio Padrão		0,32	0,415
CV (%)		702,72	14,98
REQM		1,35	-
Média	20x5	0,028	2,76
Desvio Padrão		0,33	0,434
CV (%)		1189,53	15,71
REQM		1,38	-
Média	7x7,51,1	0,054	3,373
Desvio Padrão		0,37	0,569
CV (%)		689,44	16,89
REQM		1,72	-
Média	7x7,51,1(a)	0,055	3,42
Desvio Padrão		0,38	0,596
CV (%)		687,48	17,42
REQM		1,765	-
Média	7x7,51,1(b)	0,056	3,44
Desvio Padrão		0,38	0,57
CV (%)		679,44	16,56
REQM		1,76	-

(a) *Lattice plus close pairs* com adição dos pontos próximos na direção ortogonal à tendência (b) *Lattice plus close pairs* com adição dos pontos próximos na direção da tendência.

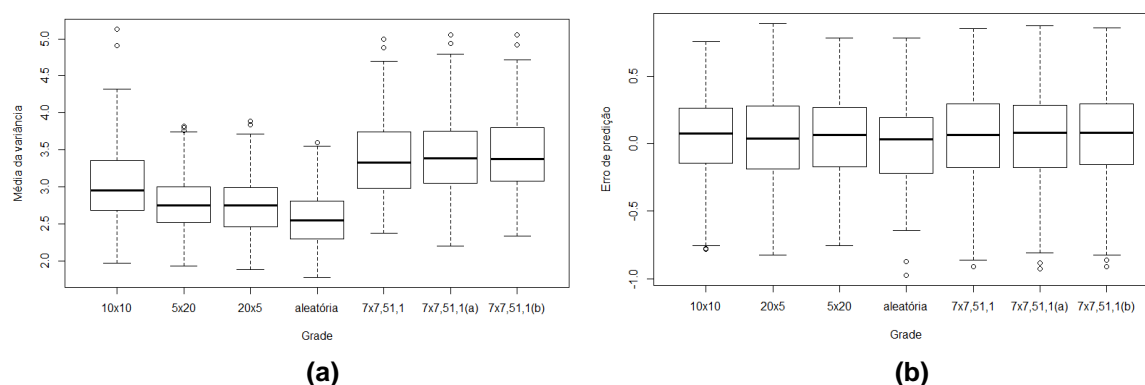


Figura 35 Gráfico boxplot: (a) Média da variância da krigagem e (b) Erro da predição

O mapa da variância da krigagem das amostragens pesquisadas com tendência direcional apresentado na Figura 36, para exemplo de simulação, proporcionou valores maiores de variância da krigagem quando comparado com estas sem tendência direcional (Figuras 29-a, 29-c, 20-d). Isto ocorre pelo fato de a variância da krigagem depender da distribuição geométrica dos pontos e do modelo de variograma (LANDIM & YAMAMOTO, 2013).

Nas amostragens com e sem tendência direcional, há diferentes configurações amostrais combinadas com diferentes valores que determinam o termo determinístico: um modelo com média constante e outro com média variável. Por exemplo, na Figura 36, o ponto circulado nas diferentes configurações apresentou diferentes valores de variância da krigagem, já que a distribuição geométrica dos pontos não é a mesma. Com exceção das versões da *lattice plus close pairs* que teve valores de variância da krigagem muito similares, pelo fato de a distribuição geométrica dos pontos ser muito semelhante.

Porém, é válida a teoria de que pontos muito próximos ao ponto a se estimar aumenta a validade da variância da krigagem, pois as amostragens com tendência direcional 10x10 (Figura 36-a), 5x20 (Figura 36-c) e aleatória (Figura 36-g) apresentaram valor menor de variância da krigagem pelo fato de apresentarem menor quantidade de pontos próximos que circundam o ponto a se estimar. Todavia, as amostragens sistemática 20x5 (Figura 36-b) e *lattice plus close pairs* 7x7,51,1 (Figura 36-c), 7x7,51,1 (a) (Figura 36-d) e 7x7,51,1(b) tiveram maior quantidade de pontos próximos ao ponto a se estimar, logo a variância da krigagem foi maior.

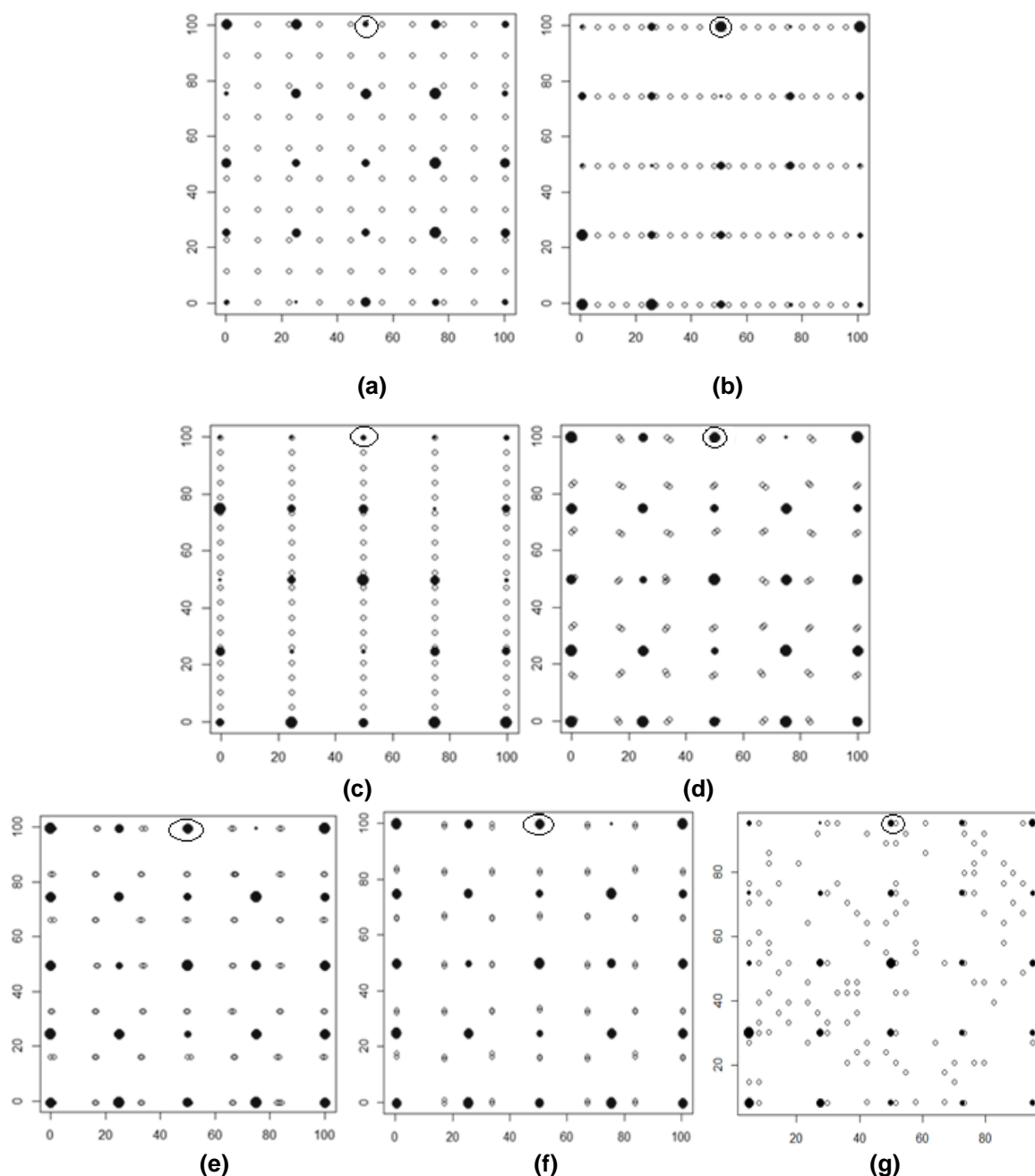


Figura 36 Mapa da variância da Krigagem da amostra-teste usando as amostragens: Sistemática (a) 10x10 (b) 20x5 (c) 5x20, *Lattice plus close pairs* (d) 7x7,51,1 (e) 7x7,51,1(a) com adição dos pontos próximos na direção ortogonal à tendência (f) 7x7,51,1(b) com adição dos pontos próximos na direção da tendência (g) Aleatória. As esferas representam o valor da variância da krigagem na amostra-teste (25 pontos), e seus valores estão representados pelo tamanho da esfera.

O Erro da predição espacial feito pela krigagem nas amostragens com tendência direcional (Figura 37-a, 37-d, 37-g) foi maior quando comparado com as amostragens sem tendência direcional (30-a, 30-c, 30-d). A adição de pontos na direção da tendência e na direção ortogonal a essa nas amostragens sistemática e *lattice plus close pairs* não proporcionou menores valores de erro da predição espacial feita pela krigagem (Figuras 37-b, 37-c, 37-e, 37-f). Porém, é válida a teoria de que o agrupamento de amostras, próximo ao ponto a se estimar, piora a estimação. Nas amostragens sistemática 10x10 (Figura 37-a), 20x5 (Figura 37-b) e *lattice plus close pairs* 7x7, 51,1 (Figura 37-d) e

7x7,51,1(a) com adição dos pontos próximos na direção ortogonal à tendência (Figura 37-e), no ponto usado como exemplo, não foi observado agrupamento de amostras próximo ao ponto a se estimar, logo menor foi o erro. Já nas amostragens sistemática 5x20 (Figura 37-c), *lattice plus close pairs* 7x7,51,1(b) com adição dos pontos próximos na direção da tendência (Figura 37-f) e aleatória (Figura 37-g) o agrupamento de amostras próximo ao ponto a se estimar proporcionou maior erro de estimação feita pela krigagem.

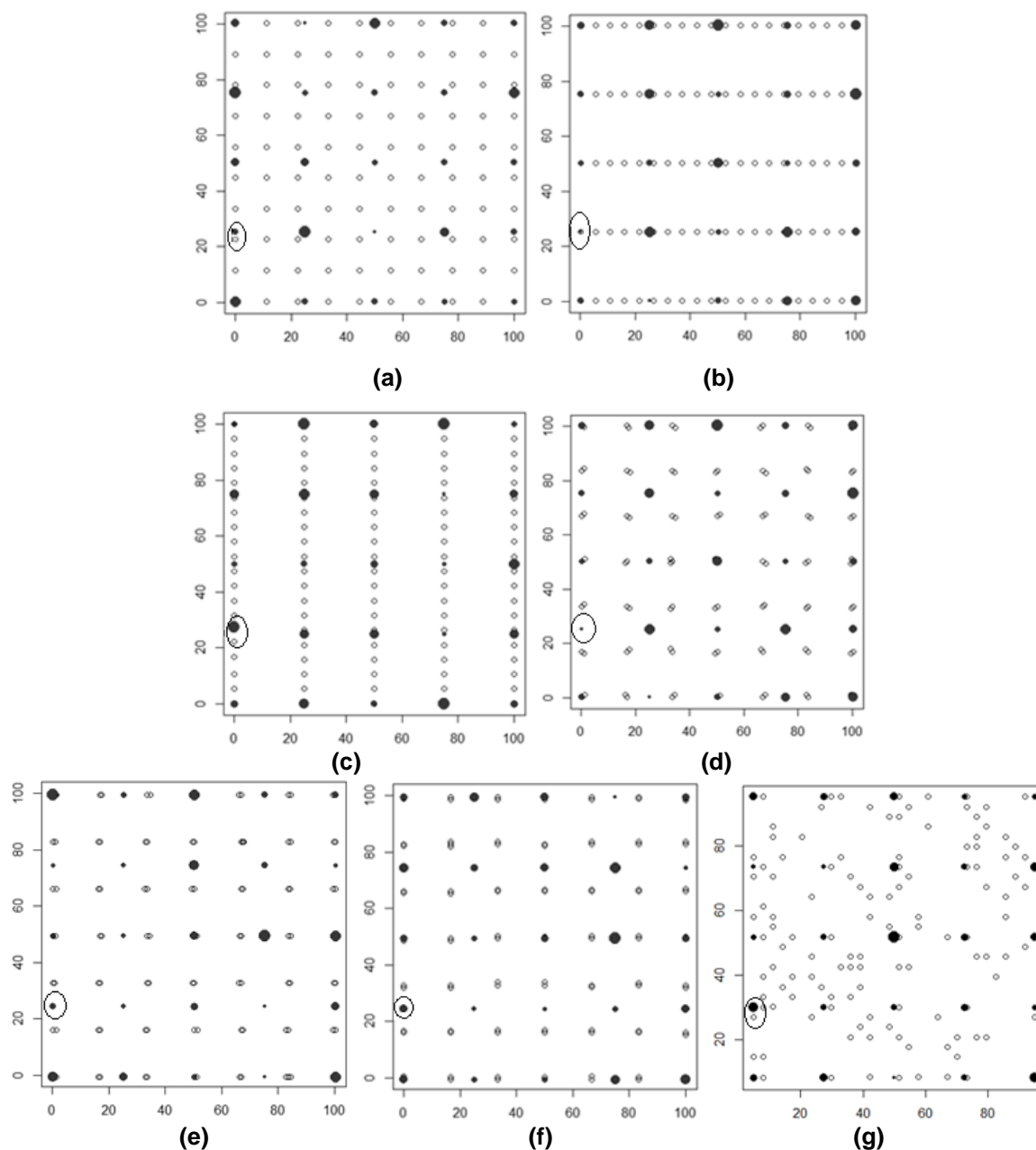


Figura 37 Erro da predição espacial feita pela Krigagem numa amostra teste de 25 pontos usando as amostragens: Sistemática (a) 10x10 (b) 20x5 (c) 5x20, *Lattice plus close pairs* (d) 7x7,51,1 (e) 7x7,51,1(a) com adição dos pontos próximos na direção ortogonal à tendência (f) 7x7,51,1(b) com adição dos pontos próximos na direção da tendência (g) Aleatória. As esferas representam o valor da variância da krigagem na amostra-teste e seus valores estão representados pelo tamanho da esfera.

11.1.3 Dados Anisotrópicos

Nesta seção, os resultados obtidos estão nas simulações das amostragens sistemática, aleatória e *lattice plus close pairs* (7x7,51,1) com anisotropia na direção 0° (sistema Azymuth). As amostragens sistemática e *lattice plus close pairs* foram realizadas em três versões. Na sistemática, as versões foram 10x10, 5x20 e 20x5, assim, foi maior a concentração de pontos na direção da anisotropia, no segundo caso, e maior concentração de pontos na direção ortogonal a anisotropia, no terceiro caso. E a *lattice plus close pairs* (7x7,51,1) foi trabalhada com adição dos pontos próximos em qualquer direção, na direção da anisotropia e na direção ortogonal a esta. A amostragem aleatória foi trabalhada apenas na sua versão original.

Na Tabela 11 são apresentadas as medidas descritivas das estimativas dos parâmetros do modelo exponencial com anisotropia. Também são apresentados os desvios padrão das estimativas dos parâmetros, as medidas de eficiência do estimador e a análise descritiva do fator de anisotropia para todas as amostragens consideradas. A estimativa dos parâmetros média e fator de anisotropia foram semelhantes em todos os cenários (Tabela 11 e Figuras 38-a, 38-e), já as estimativas do efeito pepita, contribuição e alcance prático foram mais próximas ao nominal na amostragem *lattice plus close pairs*, principalmente na 7x7,51,1(a). Ou seja, com a adição dos pontos próximos na direção da anisotropia (Tabela 11 e Figuras 38-b, 38-c, 38-d).

A maioria das medidas de eficiência do estimador VRA, VA e REQM calculadas para as estimativas de todos os parâmetros diminuem nas versões da amostragem *lattice plus close pairs*, principalmente para a versão 7x7,51,1(a) com a adição dos pontos próximos na direção ortogonal à anisotropia (Tabela 11).

Além disso, com base nessas medidas, observa-se nas amostragens pesquisadas que a anisotropia influencia a estimativa dos parâmetros do modelo. Na amostragem sistemática (Tabela 11), a anisotropia reduz a qualidade da estimativa do alcance prático quando comparada com a sistemática (10x10) sem anisotropia (Tabela 7). Embora não tenha apresentado influência relevante na estimativa dos parâmetros efeito pepita e contribuição, ainda verificou-se que a adição de pontos na direção da anisotropia (5x20) e na direção ortogonal a essa (20x5) não apresentou melhora relevante na qualidade da estimativa destes parâmetros.

Na *lattice plus close pairs* (7x7,51,1) observa-se que a anisotropia piorou a estimativa dos parâmetros efeito pepita, contribuição e alcance quando comparada com a amostragem com dados isotrópicos e estacionários (Tabela 3). Isto ocorre pela complexidade do modelo, pois, de modo geral, quanto maior é o número de parâmetros e o número de cálculos envolvidos, menor é a qualidade do modelo ajustado (FLORIANO et

al., 2006). Ainda verificou-se que a adição dos pontos próximos na direção ortogonal, a anisotropia apresentou pequena melhora na estimativa destes parâmetros. Já na amostragem aleatória, a anisotropia não apresentou influencia relevante na estimativa dos parâmetros do modelo geoestatístico quando comparada com a amostragem aleatória com dados isotrópicos e estacionários (Tabela 7).

Tabela 11 Estatística descritiva dos parâmetros do modelo exponencial estimado por ML para a amostragem *lattice plus close pairs* 7x7,51,1 anisotrópicos

Análise Descritiva	Grade	Média ($\mu=0$)	Efeito Pepita ($\phi_1=0$)	Contribuição ($\phi_2=10$)	Alcance Prático ($a=60$)	Fator de anisotropia ($Fa=3$)
Média	Aleatória	20,055	0,1161	8,83	57,95	1,13
Desv. P.		1,13	0,226	3,159	31,27	0,19
CV (%)		5,64	195,01	35,77	53,97	17,6
VRA (%)		0,2765	-	11,678	3,413	62,32
VA		0,055	0,1161	1,1678	2,04	1,86
REQM		0,636	0,032	5,623	486,34	1,767
Média	(10x10)	19,96	0,334	8,64	57,32	1,15
Desv. P.		1,008	0,762	3,18	25,27	0,307
CV (%)		5,05	228,34	36,84	44,09	26,68
VRA (%)		0,194	-	13,54	4,475	28,85
VA		0,038	0,334	1,354	2,68	0,865
REQM		0,504	0,342	5,938	319,7	365,24
Média	5x20	19,95	0,334	7,89	43,18	2,64
Desv. P.		1,05	0,3	2,49	27,74	3,11
CV (%)		5,26	90,02	31,58	64,25	117,47
VRA (%)		0,21	-	21,07	28,03	8,68
VA		0,042	0,1835	2,107	16,82	0,26
REQM		0,546	0,0614	5,29	522,52	5,18
Média	20x5	19,93	0,1835	8,93	57,22	1,16
Desv. P.		1,02	0,2517	2,93	21,48	0,252
CV (%)		5,12	137,21	32,89	37,54	22,41
VRA (%)		0,309	-	10,66	4,62	8,68
VA		0,0619	0,1167	1,066	2,77	1,868
REQM		0,517	0,0381	4,84	232,36	1,77
Média	7x7,51,1	19,93	0,1167	9,02	58,39	1,12
Desv. P.		1,02	0,041	2,99	26,63	0,22
CV (%)		5,12	35,46	33,23	45,6	20,08
VRA (%)		0,1265	-	9,74	2,66	62,17
VA		0,0253	0,0376	0,974	1,601	1,86
REQM		0,607	0,002	4,92	352,31	1,76
Média	7x7,51,1(a)	19,98	0,013	9,31	59,001	1,08
Desv. P.		1,09	0,027	2,89	23,12	0,17
CV (%)		5,47	214,38	31,08	39,19	16,35
VRA (%)		0,092	-	6,823	1,66	63,45
VA		0,0184	0,013	0,6823	0,998	1,903
REQM		0,1	0,0004	4,38	265,16	0,015
Média	7x7,51,1(b)	19,97	0,028	8,66	55,15	1,61
Desv. P.		1,083	0,046	3,004	27,55	2,06
CV (%)		5,42	163,67	34,65	49,94	127,6
VRA (%)		0,1275	-	13,31	8,06	46,37
VA		0,025	0,028	1,33	4,85	1,39
REQM		0,581	0,001	5,35	387,42	3,05

Valores nominais: μ é a média, ϕ_1 é o efeito pepita; ϕ_2 é a contribuição; a é o alcance prático. (a) *Lattice plus close pairs* com adição dos pontos próximos na direção ortogonal à anisotropia (b) *Lattice plus close pairs* com adição dos pontos próximos na direção da anisotropia.

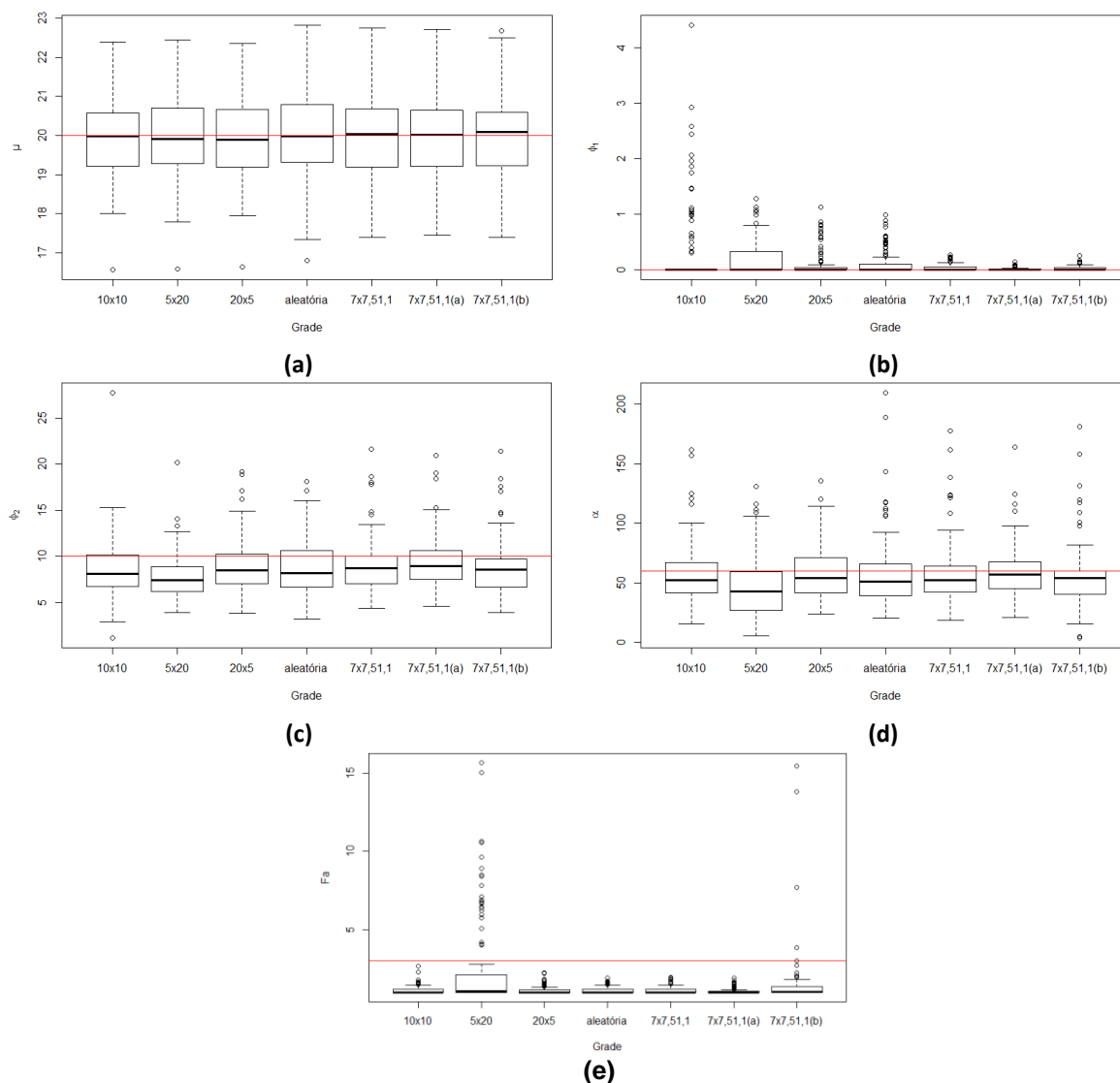


Figura 38 Gráfico boxplot dos parâmetros estimados (a) Média, (b) Efeito pepita, (c) Contribuição (d) Alcance prático e (e) Fator de anisotropia

As medidas de validação cruzada, usadas para avaliar a qualidade da estimação do modelo geoestatístico, estão exibidas na Figura 39. Observa-se que os valores de EM e EMR foram semelhantes e próximos de zero em todos os cenários (Figura 39-a, 39-b). Enquanto os valores do SEMR foram próximos de um na amostragem sistemática (10x10) (Figura 39-d). Os valores do SEM, EA, AIC e BIC foram menores nas amostragens *lattice plus close pairs* (7x7,51,1), principalmente quando se considera a adição dos pontos próximos na direção da anisotropia e na direção ortogonal a essa (Figuras 39-c, 39-e, 39-f, 39-g). Logo, de acordo com tais medidas, conclui-se que a amostragem *lattice plus close pairs* 7x7,51,1 com a adição dos pontos próximos na direção da anisotropia e na direção ortogonal a essa, forneceu as melhores estimativas das medidas associadas à validação cruzada e à qualidade do modelo exponencial.

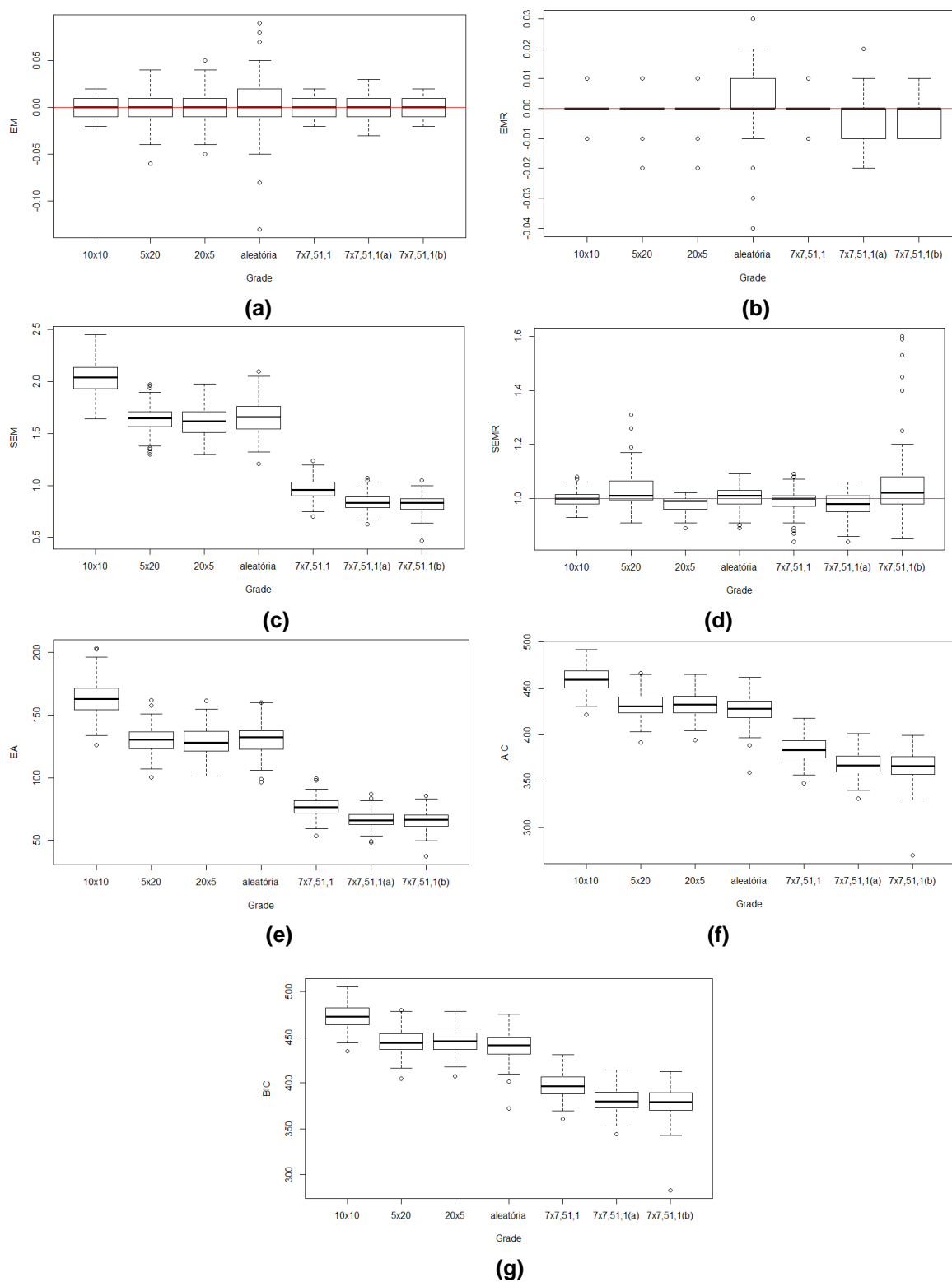


Figura 39 Gráficos boxplot: (a) Erro Médio, (b) Erro Médio Relativo, (c) Desvio Padrão do Erro Médio, (d) Desvio Padrão do Erro Médio Relativo, (e) Erro Absoluto, (f) AIC e (g) BIC. A linha horizontal indica o valor ideal.

As medidas de qualidade da predição espacial (Tabela 12 e Figura 40) mostraram na análise descritiva da média da variância e do erro da amostra-teste nas amostragens pesquisadas que a anisotropia apresentou influência relevante na qualidade da predição espacial, quando comparadas com tais amostragens utilizando variáveis isotrópicas

(Tabelas 4 e 8). E que a adição de pontos na direção da anisotropia e na direção ortogonal a esta nas amostragens sistemática e *lattice plus close pairs* (7x7,51,1) não proporcionou melhores estimativas da variável em localizações não amostradas.

A amostragem *lattice plus close pairs* 7x7,51,1 com adição dos pontos próximos na direção da anisotropia e na direção ortogonal a essa, apresentou pequena melhora na qualidade da estimativa dos parâmetros (Tabela 11), porém não proporcionou os mesmos resultados na predição espacial (Tabela 12 e Figura 40).

Tabela 12 Análise descritiva das medidas da qualidade da predição espacial do modelo exponencial anisotrópicos quanto à predição espacial da uma amostra-teste composta por 25 pontos

Estatística Descritiva	Grade	Erro da predição	Média da variância
Média		0,023	2,57
Desvio Padrão	Aleatória	0,35	0,44
CV (%)		1490,39	17,11
REQM		1,33	-
Média		0,041	2,98
Desvio Padrão	(10x10)	0,33	0,59
CV (%)		822,02	20,01
REQM		1,408	-
Média		0,036	3,57
Desvio Padrão	5x20	0,34	1,8
CV (%)		953,37	50,36
REQM		1,672	-
Média		0,0414	2,58
Desvio Padrão	20x5	0,336	0,44
CV (%)		812,22	17,12
REQM		1,379	-
Média		0,055	3,26
Desvio Padrão	7x7,51,1	0,379	0,601
CV (%)		747,95	18,41
REQM		1,72	-
Média		0,048	3,35
Desvio Padrão	7x7,51,1(a)	0,382	0,62
CV (%)		793,51	18,5
REQM		1,756	-
Média		0,049	3,32
Desvio Padrão	7x7,51,1(b)	0,385	0,658
CV (%)		772,87	19,83
REQM		1,77	-

(a) *Lattice plus close pairs* com adição dos pontos próximos na direção ortogonal à anisotropia (b) *Lattice plus close pairs* com adição dos pontos próximos na direção da anisotropia.

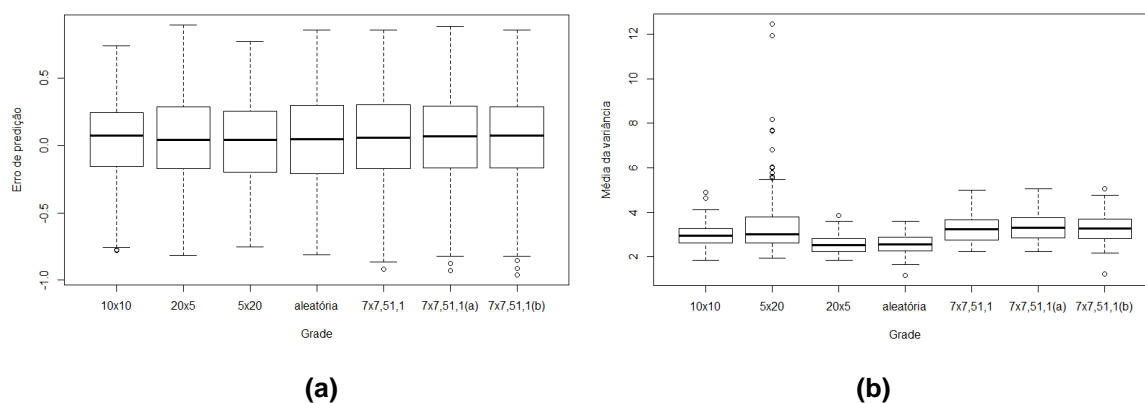


Figura 40 Gráfico boxplot: (a) Média da variância da krigagem e (b) Erro da predição da amostragem, composta por 25 pontos

O mapa da variância da krigagem das amostragens pesquisadas com dados anisotrópicos, considerando um exemplo de simulação (Figura 41), apresentou valores maiores de variância da krigagem quando comparado com as amostragens simuladas com dados isotrópicos (Figuras 29-a, 29-c, 20-d). Isto ocorre pelo fato de a variância da krigagem depender da distribuição geométrica dos pontos e do modelo de semivariograma (LANDIM & YAMAMOTO, 2013). Em tais amostragens, têm-se diferentes configurações amostrais combinadas com diferentes modelos de semivariograma, um semivariograma isotrópico (simulado com dados isotrópicos e estacionários) e um anisotrópico (simulado com dados anisotrópicos). Por exemplo, na Figura 41, o ponto circulado nas diferentes configurações apresenta diferentes comportamentos de variância da krigagem, já que a distribuição geométrica dos pontos não é a mesma. Com exceção das versões da *lattice plus close pairs*, a qual teve valores de variância da krigagem muito semelhantes, pelo fato de a distribuição geométrica dos pontos serem muito semelhantes.

A teoria de que pontos muito próximos ao ponto a se estimar aumenta a variância da krigagem ainda é válida, pois, na amostragem aleatória (Figura 41-a), teve valor menor de variância da krigagem pelo fato de ter menor quantidade de pontos próximos que circundam o ponto a se estimar. Porém, as amostragens *lattice plus close pairs* 7x7,51,1 (Figura 41-b) 7x7,51,1 (a) (Figura 41-c) e 7x7,51,1(b) (Figura 41-d) e sistemáticas 10x10 (Figura 41-e), 5x20 (Figura 41-f) 20x5 (Figura 41-g) tiveram maior quantidade de pontos próximos ao ponto a serem estimados, logo, a variância da krigagem foi maior.

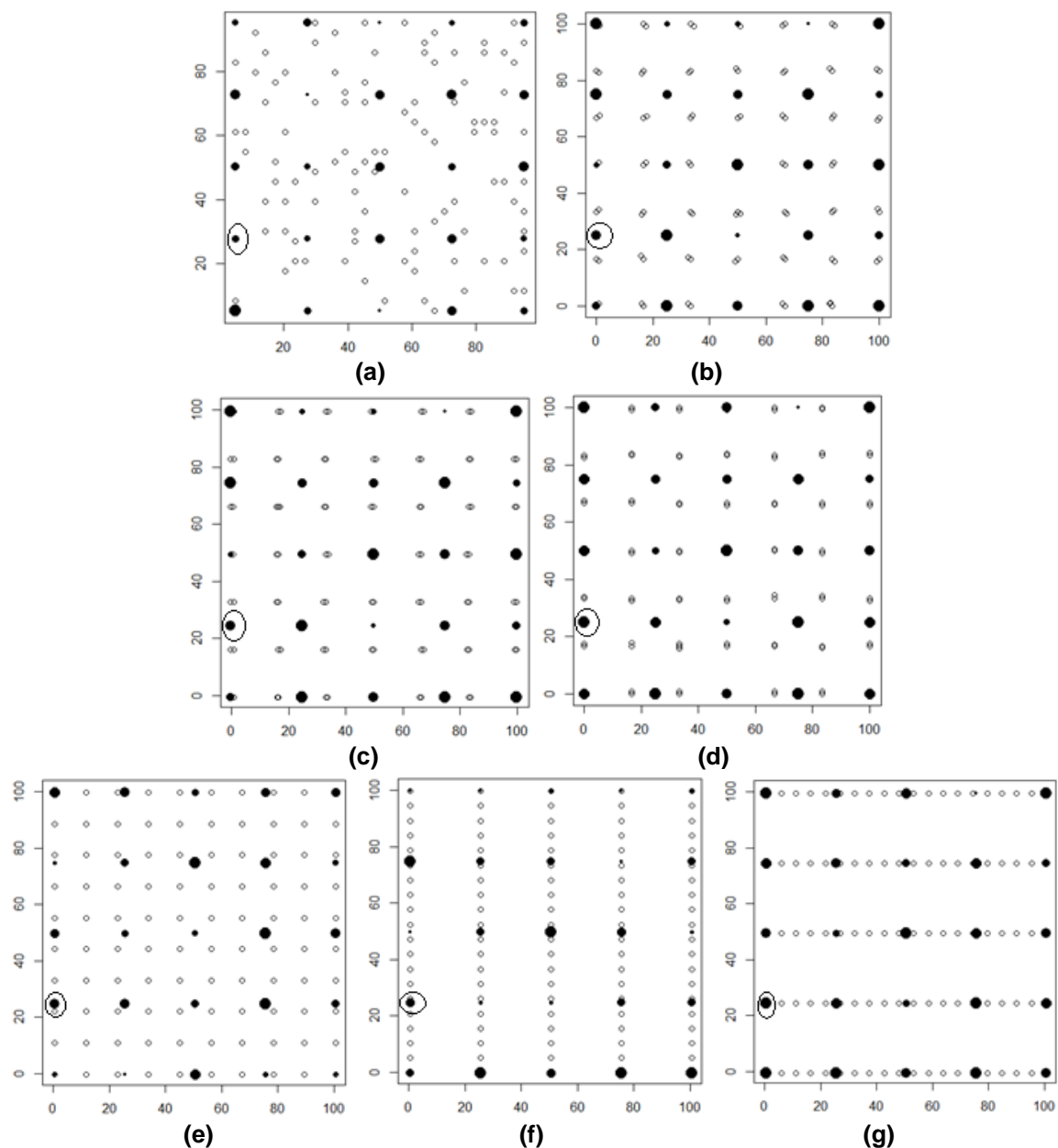


Figura 41 Mapa da variância da Krigagem da amostra teste usando as amostragens: (a) aleatória, (b) *Lattice plus close pairs* $7 \times 7,51,1$, (c) $7 \times 7,51,1$ (a) com adição dos pontos próximos na direção ortogonal à tendência, (d) $7 \times 7,51,1$ (b) com adição dos pontos próximos na direção da tendência e Sistemática (e) 10×10 (f) 5×20 e (g) 20×5 . As esferas representam o valor da variância da krigagem na amostra teste (25 pontos), e seus valores estão representados pelo tamanho da esfera.

O erro da predição espacial feito pela krigagem nas amostragens com dados anisotrópicos (Figura 42-a, 42-d, 42-g) também foi maior quando comparado com as amostragens com dados isotrópicos (30-a, 30-c, 30-d). A adição de pontos na direção da anisotropia e na direção ortogonal a essa nas amostragens sistemática e *lattice plus close pairs* não proporcionou menores valores de erro da predição espacial feita pela krigagem (Figuras 42-c, 42-d, 42-f, 42-g). Porém, é válida a teoria de que o agrupamento de amostras próximo ao ponto a se estimar piora a estimação. Nas amostragens aleatória (Figura 42-a) *lattice plus close pairs* $7 \times 7,51,1$ com adição dos pontos próximos na direção

da anisotropia e na ortogonal a essa (Figura 42-d, 42-c) e sistemática 20x5 (Figura 42-g) no ponto usado como exemplo, não foi observado agrupamento de amostras, próximo ao ponto a se estimar, logo, o erro foi menor. Já nas amostragens *lattice plus close pairs* 7x7,51,1 (Figura 43-b), e sistemáticas 10x10 (Figura 42-e) e 5x20 (Figura 42-f), o agrupamento de amostras, próximo ao ponto a se estimar, proporcionou maior erro de estimação feita pela krigagem.

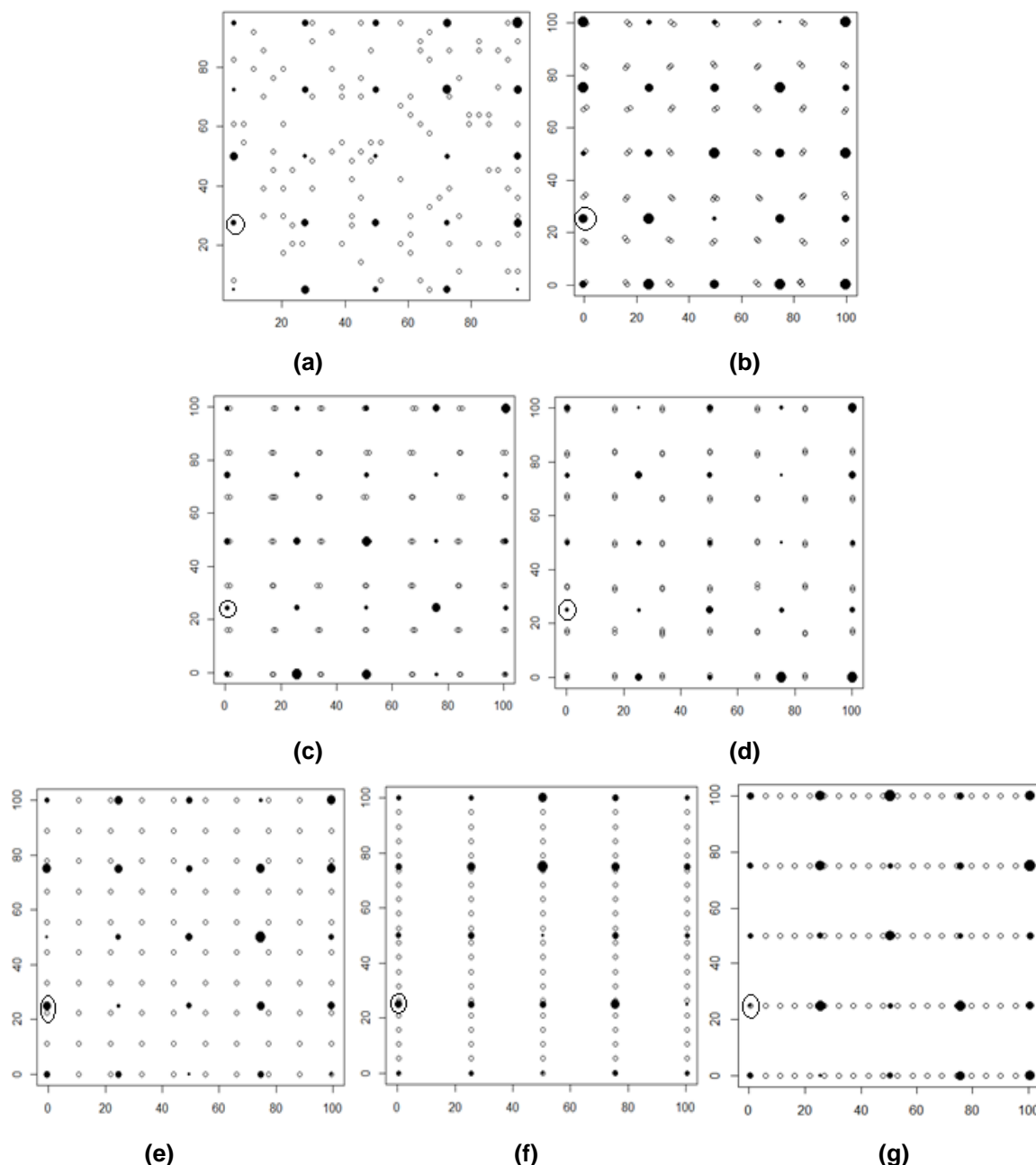


Figura 42 Erro da predição espacial feita pela Krigagem da amostra-teste (25 pontos) das amostragens: (a) aleatória, *Lattice plus close pairs* (b) 7x7,51,1 (c) 7x7,51,1(a) com adição dos pontos próximos na direção ortogonal à anisotropia (d) 7x7,51,1(b) com adição dos pontos próximos na direção da anisotropia e Sistemáticas (e) 10x10 (f) 5x20 e (g) 20x5. As esferas representam o valor da variância da krigagem na amostra-teste (25 pontos), e seus valores estão representados pelo tamanho da esfera.

12 ESTUDO PRÁTICO

Nesta seção, serão estudadas as análises descritivas e espaciais das variáveis: carbono (g dm^{-3}), cálcio (cmol dm^{-3}), magnésio ($\text{cmol}_c \text{ dm}^{-3}$), manganês (mg dm^{-3}), Cobre (mg/dm^3), Zinco (mg/dm^3), fósforo (mg/dm^3) e Alumínio ($\text{cmol}_c/\text{dm}^3$).

12.1.1 Análise descritiva

Na Tabela 13, estão as estatísticas descritivas e os valores do coeficiente de correlação linear de Pearson das variáveis carbono (C), cálcio (Ca), magnésio (Mg), manganês (Mn), cobre (Cu), zinco (Zn), alumínio (Al) e fósforo (P). O carbono apresentou média dispersão dos seus valores em relação a sua média ($10\% \leq CV \leq 20\%$). Todavia, o manganês e o cálcio apresentaram elevada dispersão ($20\% \leq CV \leq 30\%$). As demais variáveis cobre, zinco, magnésio e fósforo apresentaram variação muito elevada ($CV > 30\%$) dos seus valores em relação às próprias médias (PIMENTEL GOMES, 2000). Os valores calculados para o coeficiente de correlação linear (Tabela 13) mostraram apenas uma correlação linear forte positiva do cobre (Cu) em relação ao eixo y ($r=0,73$), indicando aumento do seu valor com o aumento do valor da sua ordenada. As demais variáveis apresentaram correlação linear fraca.

Os gráficos boxplot (Figura 43) mostraram que todas as variáveis apresentaram pontos discrepantes, com exceção do cobre (Cu). Quanto à distribuição dos dados, nota-se que as variáveis cobre (Cu), carbono (C), cálcio (Ca) e fósforo (P) apresentaram distribuição simétrica enquanto as demais variáveis, zinco (Zn), magnésio (Mg), manganês (Mn) e alumínio (Al) apresentaram distribuição assimétrica negativa.

Tabela 13 Análise exploratória descritiva das variáveis químicas cobre Cu (mg/dm^3), zinco Zn (mg/dm^3), manganês Mn (mg/dm^3), carbono C (g/dm^3), (cmolc/ dm^3), cálcio Ca (cmolc/ dm^3), magnésio Mg (cmolc/ dm^3), Alumínio Al (cmolc/ dm^3) e Fósforo P (mg/dm^3)

Análise Descritiva	Cu	Zn	Mn	C	Ca	Mg	Al	P
Min.	0,5	0,5	38,79	22,4	2,25	0,49	0	3,4
Max.	7,52	11,04	136,8	45,22	8,76	4,73	2,01	60
1º Q.	2,902	1,752	62,29	27,48	4,465	1,325	0,065	11,52
3º Q.	5,12	3,37	88,4	31,32	6,112	2,062	0,367	23,82
Mediana	4,07	2,355	71,4	29,33	5,32	1,745	0,16	16,9
Média	4,063	2,818	76,54	29,42	5,387	1,817	0,282	19,29
VAR	2,341	3,014	440,97	13,901	6,112	0,569	0,127	123,3
DP	1,53	1,736	20,999	3,728	1,353	0,754	0,356	11,1
CV (%)	37,66	61,61	27,43	12,67	25,12	41,54	126,28	57,57
r(x)	-0,03	0,307	0,077	0,23	0,222	0,11	-0,15	-0,09
r(y)	0,73	0,005	-0,02	-0,11	0,031	-0,17	0,094	0,01

Nota: r (x) coeficiente de correlação linear de Pearson dos dados em relação ao eixo x; r(y) coeficiente de correlação linear dos dados em relação ao eixo y.

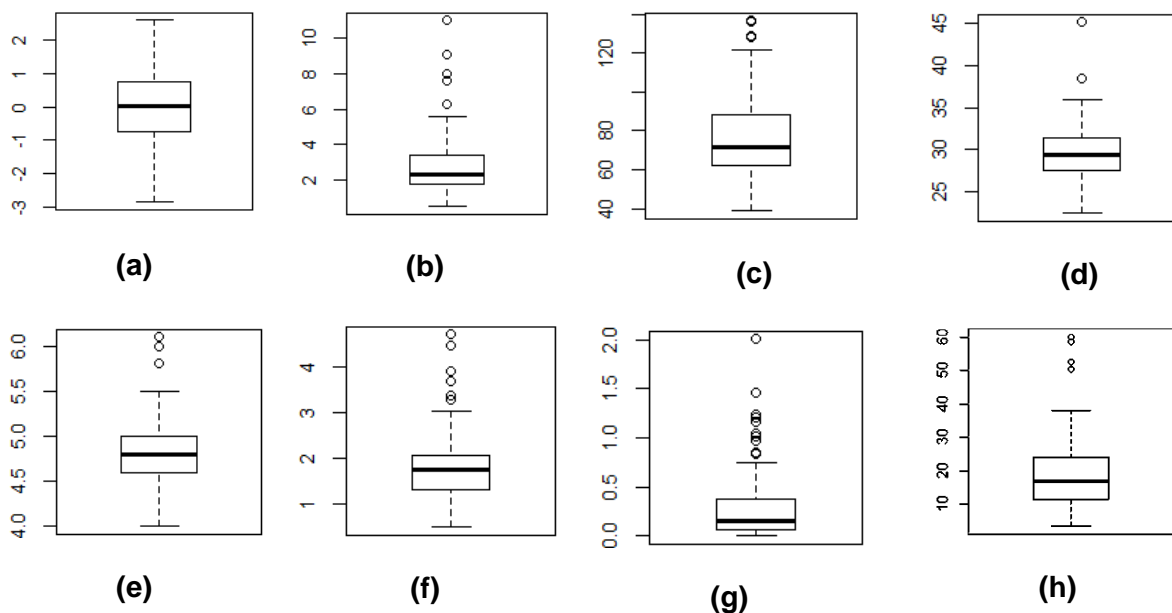


Figura 43 Gráfico boxplot das variáveis: a) Cobre Cu (mg/dm^3), (b) Zinco Zn (mg/dm^3), (c) Manganês Mn (mg/dm^3), (d) Carbono C (g/dm^3). (e) Cálcio Ca (cmolc/ dm^3), (f) Magnésio Mg (cmolc/ dm^3), (g) Alumínio Al (cmolc/ dm^3) e (h) Fósforo P (mg/dm^3)

Os gráficos espaciais da área em estudo estão apresentados na Figura 44 e classificados segundo os quartis para as variáveis estudadas. Nesses gráficos, pode-se observar que no caso da variável Cu (Figura 44-a), existe aumento gradativo no seu valor na direção 0° (sistema azimute) indicando a presença de tendência direcional. Nas variáveis C e Ca ocorre um agrupamento de valores semelhantes na direção de 90° (Figuras 44-d, 44-e) e nas demais variáveis, Zn, Mn, Mg, Al e P, não foi observado qualquer comportamento tendencioso (Figuras 44-b, 44-c, 44-f, 44-g, 44-h).

Como para atender à hipótese intrínseca não deve ser possível a identificação de padrões de tendência em qualquer direção, ajustou-se a essa tendência um modelo de regressão múltipla entre os valores da variável e as localizações geográficas. Por isso, no caso da variável cobre, a tendência foi eliminada e, assim, trabalhou-se com o resíduo resultante.

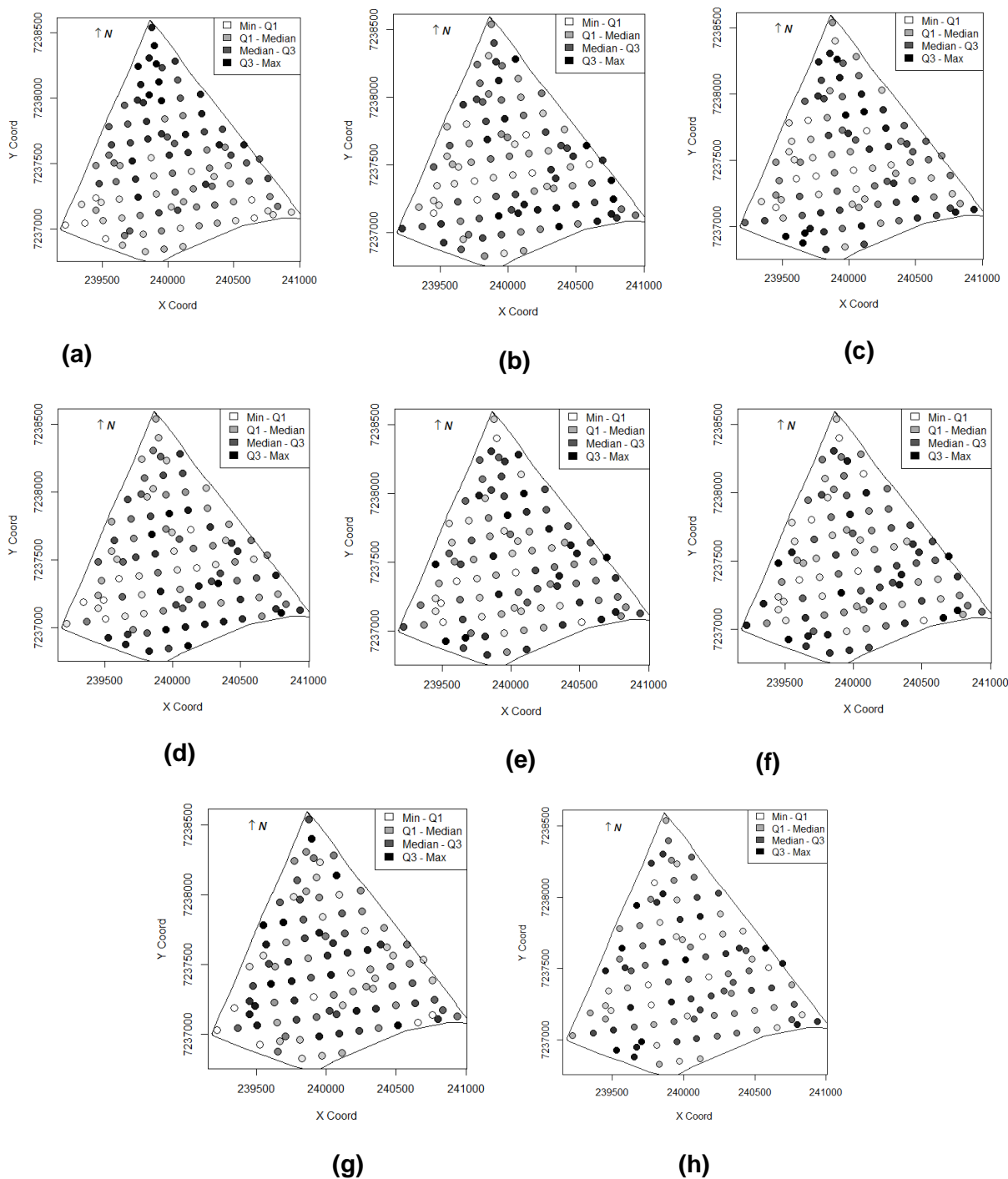


Figura 44 Gráfico post-plot das variáveis: (a) Cobre Cu (mg/dm³), (b) Zinco Zn (mg/dm³), (c) Manganês Mn (mg/dm³), (d) Carbono C (g/dm³). (e) Cálcio Ca (cmolc/dm³), (f) Magnésio Mg (cmolc/dm³), (g) Alumínio Al (cmolc/dm³) e (h) Fósforo P (mg/dm³)

Para identificar, de forma exploratória, a existência da continuidade espacial de cada variável regionalizada, utilizou-se a técnica dos envelopes, proposta por Diggle & Ribeiro Junior (2007). O envelope é construído pelos valores máximos e mínimos de todos os semivariogramas dos conjuntos de dados modificados, definindo-se assim uma região de independência espacial na variável em estudo. O semivariograma experimental omnidirecional para cada uma das variáveis em estudo está apresentado na Figura 45, com os respectivos envelopes. No caso das variáveis Carbono e Cálcio (Figuras 45-d, 45-e) não encontraram-se pontos fora dos envelopes. Este fato sugere a não existência de correlação espacial nos dados dessas variáveis, segundo esse critério. Porém, os modelos serão ajustados para que se consiga identificar se existe alguma dependência espacial.

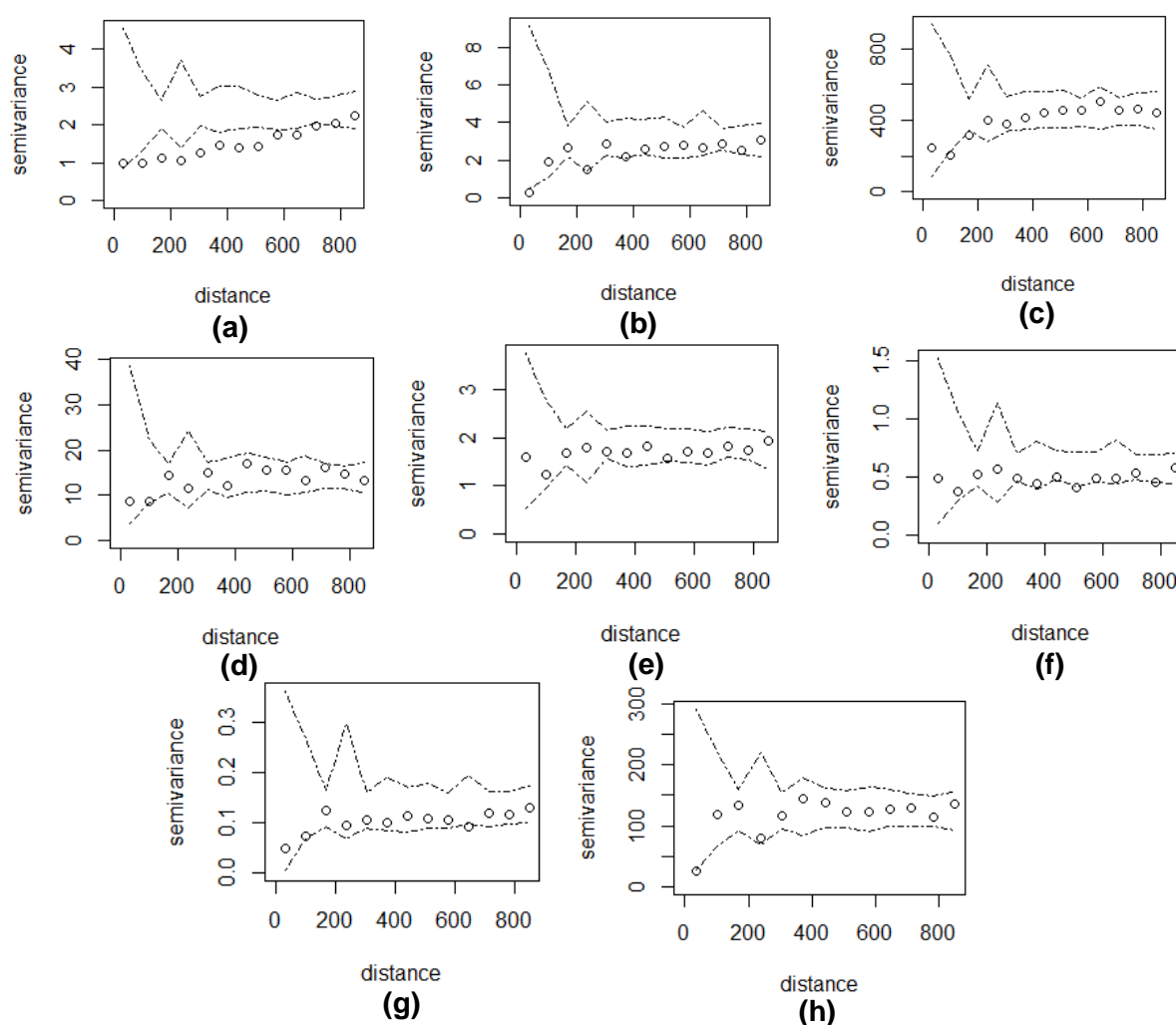


Figura 45 Gráfico de envelopes das variáveis: a) Cobre Cu (mg/dm^3), (b) Zinco Zn (mg/dm^3), (c) Manganês Mn (mg/dm^3), (d) Carbono C (g/dm^3). (e) Cálcio Ca (cmolc/dm^3), (f) Magnésio Mg (cmolc/dm^3), (g) Alumínio Al (cmolc/dm^3) e (h) Fósforo (P) (mg/dm^3)

12.1.2 Análise geoestatística

Os semivariogramas direcionais apresentados na Figura 46 mostram que, em algumas variáveis, a estrutura de dependência espacial expressa pelo semivariograma não é a mesma em todas as direções. Somente fósforo, manganês, magnésio, alumínio, e o resíduo do cobre (Figura 46-h, 46-c, 46-f, 46-g, 46-b) apresentaram o mesmo comportamento nas quatro direções analisadas, caracterizando-se assim uma estrutura isotrópica. Devido à não similaridade das semivariâncias nas direções 0° , 45° , 90° e 135° das variáveis carbono e cálcio (Figuras 46-d, 46-e), assume-se que a distribuição é anisotrópica. Pode-se observar nas variáveis que o patamar foi semelhante em todas as direções, apenas o alcance foi diferente, caracterizando-se assim uma estrutura de anisotropia geométrica. Nas variáveis carbono e cálcio verificou-se uma estrutura anisotrópica na direção de 90° com ângulos de maior continuidade espacial $F_a = 4,85$ e $F_a = 3,94$ respectivamente (Figura 46-d, 46-e).

Depois de identificada a presença de anisotropia geométrica nas variáveis carbono e cálcio, ela foi corrigida mediante transformações lineares nas coordenadas espaciais com o modelo proposto por Diggle e Ribeiro Junior (2007). Portanto, depois de corrigida a anisotropia, foi possível utilizar um único semivariograma que representa todas as direções, o semivariograma omnidirecional. Os modelos exponencial, esférico, gaussiano e Matérn $k=1,5$ foram ajustados para cada semivariograma experimental das variáveis em estudo, utilizando-se o método de máxima verossimilhança. Para cada modelo ajustado, foram obtidos os valores da média (μ), Efeito pepita (φ_1), Contribuição (φ_2), Alcance prático (a). Ainda para o cobre (variável não estacionária), foram estimados os parâmetros Beta 0 (β_0) e Beta 1 (β_1). Nesta Tabela, foi incluído o coeficiente de efeito pepita ($EPR = \varphi_1 / (\varphi_1 + \varphi_2) \times 100$), que mede o grau de dependência da variável na área em estudo.

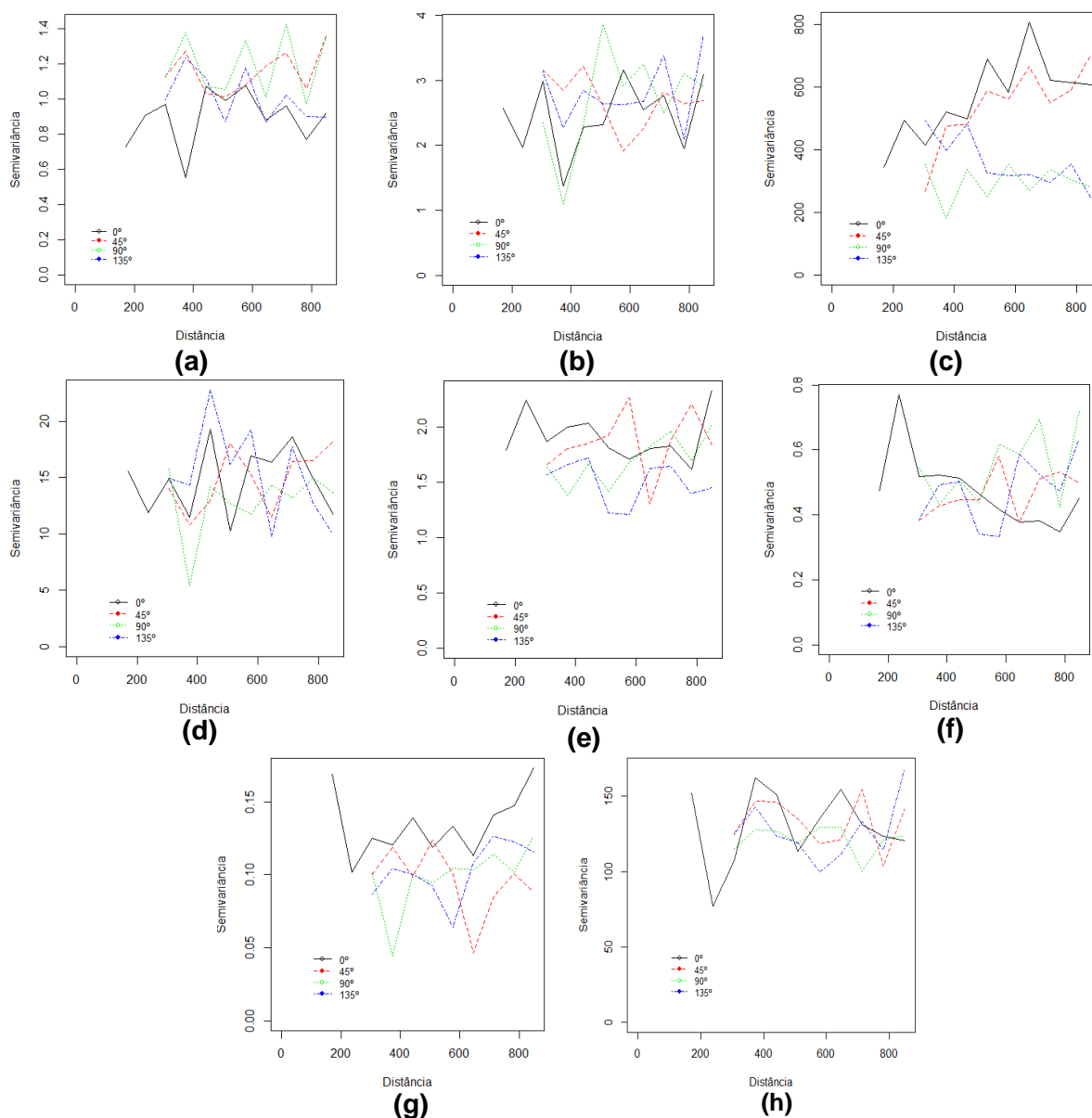


Figura 46 Semivariograma direcional das variáveis: a) resíduo do Cobre Cu (mg/dm^3), (b) Zinco Zn (mg/dm^3), (c) Manganês Mn (mg/dm^3), (d) Carbono C (g/dm^3). (e) Cálcio Ca (cmolc/dm^3), (f) Magnésio Mg (cmolc/dm^3), (g) Alumínio Al (cmolc/dm^3) (h) Fósforo P (mg/dm^3)

Para a escolha do melhor modelo ajustado, foram utilizados o critério de validação cruzada (FARACO, 2008), a Informação de Akaike (AIC) e a Informação Bayesiano (BIC) (EMILIANO, 2013). O modelo escolhido foi aquele que apresentou os menores Erro Médio (EM) e Erro Médio Reduzido (EMR), Desvio padrão do erro médio (SEM), menor possível, Desvio padrão do erro médio reduzido (SEMR) mais próximo de um e Erro Absoluto (EA) seja o menor possível. E AIC e BIC os menores valores. Os parâmetros calculados pelo método de validação cruzada de informação de akaike e bayesiana estão apresentados na Tabela 14.

Com base nessas medidas, os melhores modelos ajustados para as variáveis pesquisadas apresentaram grau de dependência espacial classificado como de moderado a forte. Assim, chega-se à conclusão de que, para as variáveis Mn, C, Ca, Mg e resíduo do

Cu, apenas 64,02%, 33,21%, 29,58%, 34,62% e 26,05% da variação total dos dados respectivamente, são explicados pela dependência espacial. Já para as variáveis Zn, Al e P, 100% da variabilidade foi explicada pela correlação espacial (Tabela 15).

A partir dos parâmetros obtidos para o semivariograma omnidirecional, foi possível a construção de mapas que expressam a variabilidade das variáveis avaliadas na área em estudo, representando uma superfície contínua que caracteriza o comportamento das variáveis no campo (Figura 47). Observou-se que, para as variáveis C (Figura 47-d) e Ca (Figura 47-e), existe maior continuidade das sub-regiões na direção de 90°. As regiões mais escuras nestes mapas indicam maior teor desses elementos, assim como as regiões mais claras indicam teores mais baixos destes atributos químicos no solo.

Tabela 14 critérios de validação cruzada, AIC e BIC para a escolha do melhor modelo ajustado

Variáveis	Modelos	EM	EMR	SEM	SEMR	EA	AIC	BIC
Cu	Exponencial	-0,0011	-0,0004	1,319	1,017	105,9	305,6	318,7
	Esférico	-0,0015	-0,0005	1,314	1,019	105,6	305,7	318,9
	Gaussiano	-0,0019	-0,0007	1,311	1,019	105,3	305,7	318,9
	Matérn k=1,5	-0,0015	-0,0005	1,314	1,019	105,6	305,7	318,8
Zn	Exponencial	0,0299	0,0097	1,668	1,029	114,1	399,0	409,5
	Esférico	0,0383	0,0130	1,690	1,049	117,7	399,1	409,6
	Gaussiano	0,0382	0,0120	1,694	1,048	117,5	399,1	409,6
	Matérn k=1,5	0,0400	0,0130	1,692	1,049	116,6	398,4	408,9
Mn	Exponencial	-0,1800	-0,0050	16,950	0,990	1342,9	890,9	901,4
	Esférico	-0,0580	-0,0010	1,900	1,001	1356,2	890,4	900,9
	Gaussiano	-0,0250	-0,0007	17,250	1,001	1394,5	892,1	902,6
	Matérn k=1,5	-0,1200	-0,0030	17,040	0,990	1357,1	891,0	901,5
C	Exponencial	-0,0080	-0,0010	3,370	1,015	267,5	555,6	568,7
	Esférico	-0,0060	-0,0009	3,390	1,020	267,3	556,8	569,9
	Gaussiano	-0,0110	-0,0010	3,370	1,006	265,8	555,1	568,2
	Matérn k=1,5	-0,0090	-0,0013	3,370	1,013	266,9	555,3	568,4
Ca	Exponencial	-0,0011	-0,0004	1,319	1,017	105,9	355,8	368,8
	Esférico	-0,0015	-0,0005	1,313	1,018	105,4	355,9	369,0
	Gaussiano	-0,0019	-0,0007	1,311	1,019	105,3	356,5	369,6
	Matérn k=1,5	-0,0015	-0,0005	1,314	1,019	105,6	356,1	369,2
Mg	Exponencial	-0,0003	-0,0002	0,760	1,009	55,7	238,8	249,3
	Esférico	-0,0004	-0,0003	0,757	1,011	55,6	238,3	248,8
	Gaussiano	-0,0002	-0,0001	0,758	1,010	55,7	238,4	248,9
	Matérn k=1,5	-0,0003	-0,0002	0,759	1,010	55,7	238,8	249,3
Al	Exponencial	0,0008	0,0011	0,3590	1,013	25,1	85,3	95,8
	Esférico	0,0018	0,0026	0,3607	1,023	25,2	84,5	95,0
	Gaussiano	0,0016	0,0023	0,3611	1,022	25,3	84,7	95,2
	Matérn k=1,5	0,0013	0,0020	0,3612	1,020	25,3	85,0	95,5
P	Exponencial	0,0015	0,000069	11,130	1,007	846,7	786,7	797,2
	Esférico	0,0035	0,0001	11,007	1,004	833,4	784,8	795,3
	Gaussiano	0,0044	0,0002	11,046	1,008	835,3	784,7	795,2
	Matérn k=1,5	0,0030	0,0001	11,090	1,008	843,4	786,0	796,5

Tabela 15 Estimação dos parâmetros das variáveis cobre, zinco, manganês, carbono, cálcio, magnésio, alumínio e fósforo por MV

Modelos	Média (μ)		Efeito pepita (φ_1)	Contribuição (φ_2)	Patamar ($\varphi_1 + \varphi_2$)	Alcance (a)	Fa	EPR
	β_0	β_1						
Cu Exponencial	-20426,5	0,0028	0,804	0,2832	1,0872	367,47	1,0	73,95
Zn Exponencial	2,94		0	3,127	3,127	270,54	1,0	0,00
Mn Esférico	77,33		173,85	309,34	483,19	595,32	1,0	35,98
C Gaussiano	29,44		9,028	4,49	13,518	149,54	6,08	66,79
Ca Exponencial	5,3987		1,2636	0,5308	1,7944	230,52	3,94	70,42
Mg Esférico	1,8184		0,3705	0,1962	0,5667	193,14	1,0	65,38
Al Exponencial	0,287		0	0,1268	0,1268	121,86	1,0	0,00
P Esférico	19,28		0	120,7	120,7	109,11	1,0	0,00

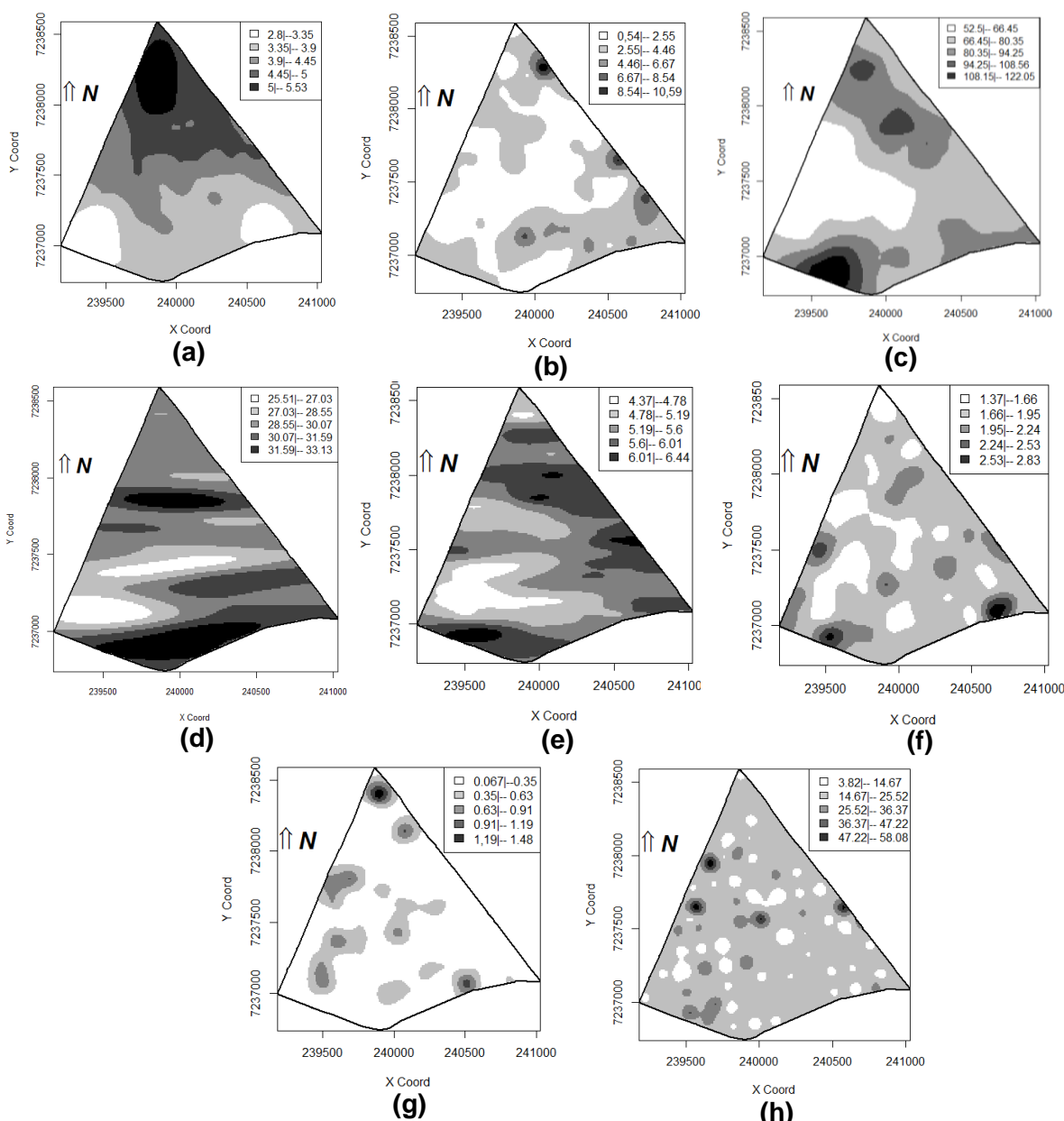


Figura 47 Mapa temático das variáveis: a) Cobre Cu (mg/dm³), (b) Zinco Zn (mg/dm³), (c) Manganês Mn (mg/dm³), (d) Carbono C (g/dm³). (e) Cálcio Ca (cmolc/dm³), (f) Magnésio Mg (cmolc/dm³), (g) Alumínio Al (cmolc/dm³) e (h) Fósforo P (mg/dm³)

Das amostragens simuladas, as que mais se aproximam da utilizada no experimento agrícola (1 amostra/2ha adicionada de 19 pontos próximos) são as *lattice plus close pairs* (9x9,19,3) e (9x9,19,5), no qual se utiliza a mesma quantidade de pontos próximos e com uma distribuição semelhante dos pontos próximos. Porém, nas simulações, as amostragens não apresentaram os melhores resultados quanto à qualidade da estimativa dos parâmetros do modelo geoestatístico e da predição espacial feita pela krigagem. Os resultados apontaram a *lattice plus close pairs* na versão 7x7,51,1 como a amostragem mais eficiente tanto para dados isotrópicos e estacionários como não estacionários ou anisotrópicos.

Neste experimento o sistema de amostragem adotado mostrou-se eficiente na caracterização da estrutura de dependência espacial das variáveis estudadas. Porém, em muitos experimentos, nem sempre o mesmo sistema de amostragem consegue detectar a estrutura de dependência espacial. Pois, as diferentes escalas de variação dos atributos do solo induzem grande dificuldade no desenvolvimento de um plano de amostragem, o qual utilize uma malha amostral com espaçamento único, quando vários atributos do solo estão envolvidos (MONTANARI et al., 2012).

Tais afirmações concordam com os resultados obtidos por Kerry et al. (2010), Cherubin et al. (2014) e Pias et al. (2014), os quais observaram que, apenas um plano amostral com único espaçamento não é eficiente na caracterização da variabilidade espacial quando vários atributos do solo são analisados. Ainda verificaram que, com o aumento da dimensão da malha amostral, há uma tendência do aumento da ocorrência de distribuições aleatórias. Isso ocorre devido aos erros de medidas, de amostragem ou microvariações não detectadas, considerando-se que o espaçamento de amostragem utilizado é maior que o necessário para detectar dependência espacial (CAMBARDELLA et al., 1994).

Portanto, de acordo com os resultados e argumentações apresentados, este estudo constitui-se de um importante referencial que poderá ser considerado no planejamento de futuras estratégias de amostragem de solo a serem adotadas nas áreas de AP. Pois, a *lattice plus close pairs* permite trabalhar com uma amostragem regular não muito densa e adiciona de pontos próximos (com distâncias menores) que podem facilitar a detecção de dependência espacial e trazer melhores resultados quanto à qualidade da estimação dos parâmetros do modelo geoestatístico e da estimação de valores da variável georreferenciada em localizações não amostradas.

13 CONCLUSÕES

Os resultados apresentados mostraram que a configuração amostral utilizada na análise da dependência espacial de variáveis georreferenciadas podem afetar a qualidade da estimativa dos parâmetros do modelo geoestatístico e as estimativas espaciais de valores não amostrados. Pois, os diferentes sistemas de amostragem aqui pesquisados apresentaram comportamentos distintos quanto à qualidade da estimativa dos parâmetros do modelo ajustado à função semivariância e da qualidade da predição espacial feita pela krigagem. O processo de encontrar a amostragem ideal é uma das grandes dificuldades da geoestatística, pois nem sempre as configurações amostrais eficientes para a estimativa dos parâmetros são necessariamente eficientes para predição espacial.

A *lattice plus close pairs 7x7,51,1* em todas as versões analisadas nas simulações, tanto para dados isotrópicos e estacionários, como não estacionários ou anisotrópicos, foi a que apresentou os melhores resultados na qualidade da estimativa dos parâmetros do modelo e da predição espacial. Também, observou-se que ela apresentou o melhor desempenho quanto à qualidade da estimação do efeito pepita que, por sua vez, é o parâmetro que mais exerce influência na qualidade da predição espacial feita pela krigagem ordinária.

Além disso, a amostragem sistemática apresentou os piores resultados para as características destacadas (estimação do modelo e predição espacial), porém, é a mais utilizada nas áreas manejadas com agricultura de precisão no Brasil. Já os sistemas de amostragens *lattice plus in-fill* e aleatória conseguiram apresentar bons resultados apenas para uma destas características. Portanto, com base nos resultados apresentados nas simulações e a amostragem utilizada no experimento, recomenda-se que em posteriores experimentos nessa área, se considere o aumento da quantidade de pontos próximos e um raio menor dos pontos próximos, e ainda para variáveis com tendência direcional e anisotropia à concentração dos pontos próximos na direção da tendência e da anisotropia e na direção ortogonal a estes fenômenos.

14 REFERÊNCIAS

AGUIRRE, G. A.; STEDINGER, J. R.; TESTER, J. W. **Geothermal resource assessment: a case study of spatial variability and uncertainty analysis for the state of New York and Pennsylvania**. Stanford University, Stanford, California, February, p.11-13, 2013.

AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In: PETROV, B. N.; CSAKI, F. (eds.). INTERNATIONAL SYMPOSIUM ON INFORMATION THEORY, 2. 1973, Budapest. **Proceedings...** Budapest: Akademiai Kiado, p. 267-281, 1973.

ANCHIETA, L. **Amostragem de solo em agricultura de precisão: particularidades e recomendações**. 2012. Dissertação (Mestrado em ciências: solos e nutrição de plantas) - Escola superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2012.

ANDRIOTTI, J. L. S. Notas de geoestatística. **Acta geológica leopoldencia** XXV. (55): 3-14), 2002. Disponível em: http://www.cprm.gov.br/publique/media/Art_Andriotti_notas_geoestatistica.pdf, acesso em 09/2015.

AZEVEDO, C. (2011). **Introdução à Amostragem Estratificada (AE)**. Disponível em: http://www.ime.unicamp.br/~cnaber/aula_AE.pdf . Acesso em 10/2015.

BARNES, R. J. The variogram sill and the sample variance. **Mathematical geology**, v. 23, p. 673-697, 1991.

BERNARDI, A. C. C.; RABELLO, L. M.; INAMASU, R. Y.; GREGO, C. R.; ANDRADE, R. G. Variabilidade espacial de parâmetros físico-químicos do solo e biofísicos de superfície em cultivo de sorgo. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 18, n.6, 2014.

BEAL, A.; CLAEYS-BRUNO, M.; SERGENT, M. Constructing space-filling designs using an adaptive WSP algorithm for spaces with constraints. **Chemometrics and Intelligent Laboratory Systems**, v. 133, p. 84-91, 2013.

BETTINI, C. Conceitos básicos de geoestatística. In: MEIRELLES, M. S. P.; CÂMARA, G.; ALMEIDA, C. M. (Ed.). **Geomática: modelos e aplicações ambientais**. cap. 4. Brasília: Embrapa, 2007.

BORSSOI, J. A. **Técnicas de diagnósticos em modelos espaciais lineares gaussianos**. 2007. 168 f. Dissertação (Mestrado em Engenharia Agrícola) – Universidade Estadual do Oeste do Paraná, Cascavel, 2007.

BORSSOI, J. A.; URIBE-OPAZO, M. A.; GALEA, M. Técnicas de diagnóstico de influência local na análise espacial da produtividade da soja. **Engenharia Agrícola**, v. 31, n. 2, p. 376-387, Jaboticabal, 2011.

CÂMARA, G.; MONTEIRO, A. M. V.; CARVALHO, M. S.; DRUCK, S (2002). **Análise Espacial de dados Geográficos**, 2ª edição (online), disponível em: <http://www.dpi.inpe.br/gilberto/livro/analise/>, acesso em 05/2015.

CAMARGO, E. C. G. **Geoestatística: fundamentos e aplicações**. Disponível em: http://www.dpi.inpe.br/gilberto/tutoriais/gis_ambiente/5geoest.pdf, acesso em 05/2015.

- CAMBARDELLA, C. A. et al. Fieldscale variability of soil properties in central Iowa soils. **Soil Science Society of America Journal**, v. 8, n. 6, p. 1501-1511, 1994.
- CARDOSO, G. G. G.; WANDERLEY, R. C.; SOUZA, M. L. C. Physical attributes of a pasture soil in southeast Goiás determined by geostatistics. **Revista Engenharia Agrícola**, v. 36, n. 1, p. 143-151, 2016.
- CHERUBIN, M. R.; SANTI, A. L.; EITELWEIN, M. T.; MENEGOL, D. R.; ROS, C. O. PIAS, O. H. C.; BERGHETTI, J. Eficiência de malhas amostrais utilizadas na caracterização da variabilidade espacial de fósforo e potássio. **Ciência Rural**, Santa Maria, v. 44, n. 03, 2014.
- COELHO, C. C.; SOUZA, E. G.; URIBE-OPAZO, M. A.; PINHEIRO NETO, R. Influência da densidade amostral e do tipo de interpolador na elaboração de mapas temáticos. **Acta Scientiarum Agronomy**, Maringá, v. 31, n. 1, p.165-174, 2009.
- CRESSIE, N. A. C. **Statistics for spatial data**. Reviewed Edition. Wiley, New York, 1993.
- DEBASTIANI, F. **Influência local em modelos espaciais lineares com distribuição da família de contornos elípticos**. Dissertação (Mestrado em estatística) – Universidade Federal de Pernambuco, 2012.
- DIGGLE, P.; LOPHAVEN, S. Bayesian geostatistical design. **Journal of Statistics**, Scandinavian, v. 33, p. 53-64, 2006.
- DIGGLE, P. J.; RIBEIRO JÚNIOR, P. J. **Model-based geostatistics**. New York: Springer, 228 p., 2007.
- EMILIANO, P. C. **Crerios de informao**: como eles se comportam em diferentes modelos? 2013. Tese (Doutorado em estatística e experimentação agropecuária) – Universidade Federal de Lavras, Lavras, 2013.
- FARACO, M. A.; URIBE-OPAZO, M. A. SILVA, E. A.; JOHANN, J. A.; BORSSOI, J. A. Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 32, n. 2, p. 463-476, 2008.
- FAZIO, V. S. **Interpolação espacial**: uma comparação analítica entre redes RBF e krigagem. 2013. Dissertação (Mestrado em ciências da computação) – Universidade Federal de Santa Catarina, Florianópolis, 2013.
- FERREIRA, I. O.; SANTOS, G. R.; RODRIGUES, D. D. Estudo sobre a utilização adequada da krigagem na representação computacional de superfícies batimétricas. **Revista brasileira de cartografia**, nº 65/5, p. 831-842, 2013.
- FLORIANO, E. P.; MULLER, I.; FINGER, C. A. G.; SCHNEIDER, P. R. Ajuste e seleção de modelos tradicionais para séries temporais de dados de altura de árvores. **Ciência Florestal**, Santa Maria, v. 16, n. 2, p.177-199, 2006.
- GUEDES, L. P. C. **Otimização de amostragem espacial**. 143f. Dissertação (Mestrado em Entomologia) – Escola superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2008.
- GUEDES, L. P. C.; RIBEIRO JUNIOR, P. J.; URIBE-OPAZO, M. A.; DE BASTIANI, F. Soybean yield maps using regular and optimized sample with different configurations by simulated annealing. **Revista Engenharia Agrícola**, v. 36, n. 1, p.114-125, 2016.

GUEDES, L. P. C.; URIBE-OPAZO, M. A. RIBEIRO JUNIOR, P. J. Influence of incorporating geometric anisotropy on the construction of thematic maps of simulated data and chemical attributes of soil. **Chilean Journal of Agricultural Research**. v. 73, n. 4, 2013.

GUEDES, L. P. C.; URIBE-OPAZO, M. A.; JOHANN, J. A.; SOUZA, E. G. Anisotropia no estudo da variabilidade espacial de algumas variáveis químicas do solo. **Revista Brasileira de Ciência do Solo**, v. 7, p. 2217-2226, 2008.

GUEDES, L. P. C.; URIBE-OPAZO, M. A.; RIBEIRO JÚNIOR, P. J. **Configurações amostrais otimizadas pela t mpera simulada em dados simulados com depend ncia espacial**. In: Simp sio de m todos num ricos computacionais da Universidade Federal do Paran  e XII Semana do Programa de P s-Gradua o em M todos Num ricos em Engenharia, Curitiba, 2015.

GUEDES, L. P. C.; RIBEIRO J NIOR, P. J.; PIEDADE, S. M. S.; URIBE-OPAZO, M. A. Optimization of spatial sample configurations using hybrid genetic algorithm and simulated annealing. **Chilean Journal of Statistics**, v. 2, n. 2, p. 39-50, 2011.

GRZEGOZEWSKI, D. M. **Influ ncia local para modelos geoestat sticos utilizando a produtividade da soja e atributos qu micos do solo**. 2012. Disserta o (Mestrado em Engenharia Agr cola) – Universidade Estadual do Oeste do Paran , Cascavel, 2012.

HALLAK, R.; PEREIRA FILHO, A. J. Metodologia para an lise de desempenho de simula es de sistemas convectivos na regi o metropolitana de S o Paulo com modelo ARPS: sensibilidade a varia es com os esquemas de advec o e assimila o de dados. **Revista Brasileira de meteorologia**, v. 26, n. 4, p. 591-608, 2011.

HELLE, K. B.; PEBESMA E. Optimising sampling designs for the maximum coverage problem of plume detection. **Spatial Statistics**, v. 13, p. 21-44, 2015.

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

JOHNSON, M. E.; MOORE, L. M.; YLVISAKER, D. Minimax and maximin distance design. **Journal of Statistical Planning and Inference**. North-Holland, p.131-148, 1990.

KERRY, R. et al. Sampling in precision agriculture. In: OLIVER, M. A. (Org.). **Geostatistical applications for precision agriculture**. Heidelberg: Springer-Verlag, p. 35-63, 2010.

KESTRING, F. B. F.; GUEDES, L. P. G.; DE BASTINI, F.; URIBE-OPAZO, M. A. Compara o de mapas tem ticos de diferentes grades amostrais para a produtividade da soja. **Revista Engenharia Agr cola**, Jaboticabal, v. 35, n. 4, p.733-743, 2015.

LANDIM, P. M. B. STURARO, J. R. **Krigagem indicativa aplicada a elabora o de mapas probabil stico de riscos**. Rio Claro: DGA, IGCE, UNESP, 2002. (Geoestat stica, Texto Did tico 6). Dispon vel em: <<http://www.rc.unesp.br/igce/aplicada/textodi.html>>. Acesso em 13 ago. 2015.

LANDIM, P.M.B. **Sobre geoestat stica e mapas**. Terra e Did tica, v. 2 n. 1, p.19-33, 2006. Dispon vel em: <http://ocs.ige.unicamp.br/ojs/terraedidatica/article/viewFile/1008/442>, acesso em 05/2015.

LANDIM, P. M. B. **Introdu o aos m todos de estima o espacial para confec o de mapas**. DGA, IGCE, UNESP/ Rio Claro, Lab. Geomatem tica, Texto Did tico 02, 20 pp.

2000. Disponível em: <http://www.rc.unesp.br/igce/aplicada/DIDATICOS/LANDIM/interpo.pdf>, acesso em 05/2015.

LOURENÇO, R. W.; LANDIM, P. M. B. Mapeamento de áreas de risco a saúde pública por meio de métodos geoestatísticos. **Caderno de Saúde Pública**, 21(1), p. 150-160, Rio de Janeiro, 2005.

MACHADO, A. A.; DEMÉTRIO, C. G. B.; FERREIRA, D. F.; SILVA, J. G. C. Estatística Experimental: uma abordagem no planejamento e no uso de recursos computacionais. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 50: SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 11, 2005, Londrina. **Minicurso...** Londrina: UEL, p. 25-50, 2005.

MARQUES, A. P. S.; MARCATO JUNIOR, IMAI, N. N.; TACHIBANA, V. M. **Aplicação da Krigagem ordinária na inferência espacial de plantas aquáticas submersas**. In: VI Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação. Recife, Pernambuco, 2012.

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

MCBRATNEY A. B., WEBSTER R. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. **J. soil sci.** v. 37, p. 617-639, 1986.

MELLO, J. M.; BATISTA, J. L. F.; RIBEIRO JÚNIOR, P. J.; OLIVEIRA, M.S. Ajuste e seleção de modelos espaciais de semivariograma visando à estimativa volumétrica de *Eucalyptus Grandis*. **Scientia Forestalis**, Piracicaba, n. 69, p. 25-37, 2005.

DEL MONEGO, M.; RIBEIRO JÚNIOR, P. RAMOS, P. Comparing the performance of geostatistical models with additional information from covariates for sewage plume characterization. **Springer-Verlag Berlin Heidelberg**, n. 22 , p. 5850–5863, 2014.

MONTANARI, R.; MARQUES JÚNIOR, J.; PEREIRA, G. T.; SOUZA, Z. M. Forma da paisagem como critério para otimização amostral de latossolos sob cultivo de cana-de-açúcar. **Pesquisa Agropecuária Brasileira**, Brasília, v. 40, n. 1, p. 69-77, 2005.

MONTANARI, R. The use of scaled semivariograms to plan soil sampling in sugarcane fields. **Precision Agriculture**, v. 13, n. 5, p. 542-552, 2012.

MONDO, V. H. V. et al. Spatial variability of soil fertility and its relationship with seed physiological potential in a soybean production area. **Revista Brasileira de Sementes**, Londrina, v. 34, n. 2, p. 193-201, 2012.

MORAL, F. J.; TERRÓN, J. M.; MARQUES DA SILVA, J.R. Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. **Soil & Tillage Research**. p. 335-343. 2010.

MULLER, A. **Simulação estocástica: o método de monte carlo**. Trabalho de conclusão de curso (Laboratório de estatística II), Curitiba, 2008.

MUKAKA, M. **A guide to appropriate use of Correlation coefficient**. In: medical research. 2012.

NOGUEIRA, C. H. **Análise de variância com dependência espacial sob uma abordagem geoestatística**. 2013. Dissertação (Mestrado em estatística e experimentação agrícola) – Universidade Federal de Lavras, Lavras, 2013.

OLIVEIRA, I. R.; TEIXEIRA, D. B.; PANOSSO, A. R.; CAMARGO, L. A.; MARQUES JÚNIOR, J.; PEREIRA, G. T. Modelagem geoestatística das incertezas da distribuição espacial do fósforo disponível no solo, em área de cana-de-açúcar. **Revista Brasileira Ciência do Solo**, v. 37, n. 06, Viçosa, 2013.

OLIVEIRA, R. B.; SILVA, A. F.; QUARTEZANI, W. Z.; LIMA, J. S. S.; ZIMBACK, C. R. Levantamento do tipo amostral, tamanho da área e número de pontos utilizados em análise geoestatística. In: **II Simpósio de Geoestatística Aplicada em Ciências Agrárias**, Botucatu, São Paulo, 2011.

OLIVEIRA JÚNIOR, J. C.; SOUZA, L. C. P.; MELO, V. F.; ROCHA, H. O. Variabilidade espacial de atributos mineralógicos de solo da formação guabirota, Curitiba (PR). **Revista Brasileira Ciência do solo**, v. 35, p. 1481-1490, 2011.

OPROMOLLA, P. A.; DALBEN, I.; CARDIM, M. Análise geoestatística de casos de hanseníase no Estado de São Paulo, 1991-2002. **Revista Saúde Pública**, v. 40, n. 5, p. 907-13, 2006.

OSÓRIO, M. P. L. **Redução de dimensão para modelos espaciais não gaussianos**. 2013. Dissertação (Mestrado em Estatística) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2013.

PAPANI, F. M. G. **Modelo especial birnbaum-saunders aplicado a dados agrícolas**. Tese (Doutorado em Engenharia Agrícola) - Universidade Estadual do Oeste do Paraná, Cascavel, 2016.

PAPANI, F. G.; URIBE-OPAZO, M. A.; LEIVA, V.; AYKROYD, R.G. Birnbaum–Saunders spatial modelling and diagnostics applied to agricultural engineering data. **Stochastic Environmental Research Risk Assessment**. Springer-Verlag Berlin Heidelberg, 2016.

PIAS, O. H. C.; SANTI, A. L.; CHERUBIN, M. R.; BERGHETTI, J.; OLIVEIRA, T. C. Caracterização da variabilidade espacial do índice relativo de clorofila na cultura do trigo. **Pesquisa Agropecuária trop**. Goiânia, v. 44, n. 4, p. 451-459, 2014.

PIMENTEL GOMES, F. **Curso de estatística experimental**. 13ª ed. São Paulo: Nobel, 2000, 479p.

PELLISSARI, A. L.; FIGUEIREDO FILHO, A.; CALDEIRA, S. F.; MACHADO, S. A. Geoestatística aplicada ao manejo de povoamentos florestais de teca, em períodos pré-desbaste seletivo, no estado do mato grosso. **Rev. Bras. Biom**. São Paulo, v.32, n.3, p. 430-444, 2014.

PONTES, J. M. **Geoestatística**: aplicações em experimentos de campo. 2002. 82f. Dissertação (Mestrado em agronomia – Estatística e experimentação agropecuária) Universidade Estadual de Lavras, Lavras, 2002.

QUENOUILLI, M. H. Approximate tests of correlation in time-series. **J. R. statist**, p. 68-84, 1949.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2015. Disponível em: <<http://www.r-project.org>>. Acesso em: 25 de Março, 2015.

RAGAGNIN, V. A.; SENA JUNIOR, D. G.; SILVEIRA NETO, A. N. Recomendação de calagem à taxa variada sob diferentes intensidades de amostragem. **Revista brasileira de Engenharia Agrícola e Ambiental**, v. 14, n. 6, Campina Grande, 2010.

REZENDE, F. C.; RIBEIRO, V. B.; ÁVILA, L. F.; FARIA, M. A.; SILVA, E. L. Variabilidade espacial do pH em área com cafeeiro fertirrigado e sistema tradicional. **Coffee Science**, Lavras, v. 7, n. 3, p. 198-207, 2012.

RIBEIRO JUNIOR, P. J.; DIGGLE, P. J. geoR: a package for geostatistical analysis. **R-NEWS**, London, v. 1, n. 2, p. 15-18, 2001.

ROBINSON, D. P.; LHOYD, C. D.; McKinley, J. M. Increasing the accuracy of nitrogen dioxide (NO₂) pollution mapping using geographically weighted regression (GWR) and geostatistics. **International Journal of Applied Earth Observation and Geoinformation**, p. 374–383, 2013.

ROSSITER, D.G. **Technical Note**: Co-kriging with the gstat package of the R environment for statistical computing. University of Twente, Faculty of Geo-Information Science & Earth, 2012.

ROSSONI, D. F. **Análise de Variância para experimentos com dependência espacial**. 2011. 109f. Dissertação (Mestrado em Estatística e Experimentação Agropecuária). Universidade Estadual de Lavras, Lavras, 2011.

SAGHAFIAN, B.; BONDARABADI, S. R. Validity of Regional Rainfall Spatial Distribution Methods in Mountainous Areas. **J. Hydrologic**, Iran, v. 14, n. 7, p. 771-771, 2009.

SANTOS, G. R.; OLIVEIRA, M. S.; LOUZADA, J. M.; SANTOS, M. R. T. Krigagem simples versus Krigagem Universal qual o preditor mais preciso? **Revista Energia na Agricultura**, v. 26, n. 2, p. 49-55, Botucatu, 2011.

SILVA, A. A.; NANNI, M. R.; SILVA JÚNIOR, C. A.; ROMAGNOLI, F.; CHICATI, M. L.; GASPAROTTO, A. C. Diferentes grades de amostragem para avaliação de atributos químicos do solo na distribuição espacial em áreas de cana-de-açúcar. **AGRARIAN ACADEMY, centro científico conhecer**, Goiânia, v. 1, n. 1, p. 159, 2014.

SILVA, A. M.; SILVA, R. M.; ALMEIDA, C. A. P.; CHAVES, J. J. S.; Modelagem geoestatística dos casos de dengue e da variação termopluiométrica em João Pessoa, Brasil. **Sociedade & Natureza**, Uberlândia, v. 27, n. 1, p. 157-169, 2015.

SOARES, A. **Coleção ensino da ciência e da tecnologia**: geoestatística para as ciências da terra e do ambiente. Lisboa. Técnico Lisboa, 3ª Edição, 2014.

SOBRAL, T. E. L.; BARRETO, G. **Análise dos critérios de informação para seleção de ordem em modelos auto regressivos**. In: 10ª Conferência Brasileira de Dinâmica, Controle e aplicações – DINCON, São Paulo, 2011.

SOUZA, Z. M.; SOUZA, G. S.; MARQUES JÚNIOR, J.; PEREIRA, G. T. Número de amostras na análise e na Krigagem de mapas de atributos do solo. **Ciência Rural**, v. 44, n. 2, p. 261-268, Santa Maria, 2014.

SILVA, A. F.; BARBOSA, A. P.; ZIMBACK, C. R. L.; LANDIM, P. M. B. Geostatistics and remote sensing methods in the classification of images of areas cultivated with citrus. **Engenharia Agrícola**, Jaboticabal, v. 33, n. 6, p. 1245-1256, 2013.

TEIXEIRA, M. B. R. **Comparação entre estimadores de semivariância**. 2013. Dissertação (Mestrado em estatística e experimentação agropecuária). Universidade Federal de Lavras, Lavras, 2013.

URIBE-OPAZO, M. A., BORSSOI, J. A., GALEA, M. Influence diagnostics in gaussian spatial linear models. **Journal of Applied Statistics**. v. 39, n. 3, 615-630, 2012.

WALVOORT, D. J. J.; BRUS, D. J.; de GRUIJTER, J. J. An r package for spatial coverage sampling and random sampling from compact geographical strata by k-means. **Comput. Geosci**. n. 36 , p. 1261-1267, 2010.

WILKS, D. S. **Statistical Methods in the Atmospheric Sciences**. International Geophysics Series. 2^a Edição, Estados Unidos da América, Academic Press, v. 91, 627 p., 2006.

YAMAMOTO, J. K.; LANDIM, P. M, B. (2013) **Geoestatística**: conceitos e aplicações. São Paulo: Oficina de Textos. 215p.