

UNIVERSIDADE ESTADUAL DO OESTE DO PARANÁ
CÂMPUS CASCAVEL
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA AGRÍCOLA

**INFLUÊNCIA DE CONFIGURAÇÕES AMOSTRAIS NA QUALIDADE DE ESTIMAÇÃO
ESPACIAL SOB O USO DE MODELOS ESPACIAIS BIVARIADOS**

JACQUELINE GABRIELA CANTÚ

CASCAVEL - PARANÁ
FEVEREIRO – 2015

JACQUELINE GABRIELA CANTÚ

**INFLUÊNCIA DE CONFIGURAÇÕES AMOSTRAIS NA QUALIDADE DE ESTIMAÇÃO
ESPACIAL SOB O USO DE MODELOS ESPACIAIS BIVARIADOS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de Mestre em Engenharia Agrícola, área de concentração em Sistemas Biológicos e Agroindustriais, da Universidade Estadual do Oeste do Paraná, campus Cascavel.

Orientadora: Prof^a. Dr^a. Luciana Pagliosa Carvalho Guedes.

Co-orientador: Prof. Dr. Miguel Angel Uribe Opazo.

CASCADEL – PARANÁ
FEVEREIRO – 2015

JACQUELINE GABRIELA CANTÚ

"Influência de configurações amostrais na qualidade de estimação espacial sob o uso de modelos espaciais bivariados"

Dissertação apresentada ao Programa de Pós-Graduação "*Stricto Sensu*" em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de Mestre em Engenharia Agrícola, área de concentração Sistemas Biológicos e Agroindustriais, **aprovada** pela seguinte banca examinadora:



Orientadora: Prof.ª Dra. Luciana Pagliosa Carvalho Guedes
Universidade Estadual do Oeste do Paraná – Campus de Cascavel



Prof.ª Dra. Rosangela Aparecida Botinha Assumpção
Universidade Tecnológica Federal do Paraná



Prof. Dr. Jerry Adriani Johann
Universidade Estadual do Oeste do Paraná – Campus de Cascavel

Cascavel, 02 de fevereiro de 2015.

BIOGRAFIA

Jacqueline Gabriela Cantú nasceu no estado do Paraná – Brasil em 03 de dezembro de 1991. Concluiu a graduação em Licenciatura em Matemática pela Universidade Estadual do Oeste do Paraná, UNIOESTE, em 2012. Título da monografia apresentada como requisito parcial para aprovação na disciplina Introdução à Pesquisa: “Modelos não-paramétricos de análise de sobrevivência aplicados a casos de câncer de estômago”, escrita sob orientação da professora Sandra Maria Tieppo. Participou do Programa Institucional de Bolsas de Iniciação à Docência (PIBID), subprojeto de Matemática no período de maio de 2010 a dezembro de 2012. Ingressou como bolsista no Programa de Pós-Graduação em Engenharia Agrícola pela Universidade Estadual do Oeste do Paraná, UNIOESTE em 2013.

Aos meus pais, Ilse e Jacó.
Dedico com muito carinho, por todo o apoio e o incentivo.

AGRADECIMENTOS

Meus sinceros agradecimentos...

A Deus, por me conceder saúde, capacidade e força para concluir este trabalho e por ter me cercado de pessoas prestativas e atenciosas;

Aos meus pais, Ilse e Jacó, pelo incentivo e o apoio nos momentos em que o desânimo e as forças precisavam ser restabelecidas, e por compartilharem comigo a realização de mais um sonho;

Às minhas irmãs, Fernanda e Emanuele, pela amizade, o companheirismo e os momentos de distração;

Ao meu namorado, Rodolfo, por estar sempre ao meu lado dividindo os momentos bons e os difíceis;

À professora orientadora Luciana Pagliosa Carvalho Guedes pela paciência, a disponibilidade, a confiança e a condução desse trabalho;

Ao professor co-orientador Miguel Angel Uribe Opazo pelas sugestões cujo contributo foi essencial para a realização desse trabalho;

Aos colegas e professores do Laboratório de Estatística Aplicada (LEA) e Laboratório de Estatística Espacial (LEE) pelos momentos agradáveis e de trabalho árduo que dividimos ao longo desses dois anos, de modo especial às colegas e amigas Denise e Regiane;

A Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Fundação Araucária pelo apoio financeiro;

Enfim, a todos que de alguma maneira contribuíram para a conclusão desse trabalho, seja com um sorriso, seja com um incentivo ou simplesmente com a força positiva do pensamento.

Aprenda como se você fosse viver para sempre.
Viva como se você fosse morrer amanhã.
Mahatma Gandhi

INFLUÊNCIA DE CONFIGURAÇÕES AMOSTRAIS NA QUALIDADE DE ESTIMAÇÃO ESPACIAL SOB O USO DE MODELOS ESPACIAIS BIVARIADOS

RESUMO

Estudos da variabilidade espacial do solo estão baseados na geoestatística, que se apresenta como um método cujos dados provêm de fenômenos naturais e que consideram a localização geográfica do fenômeno. Se numa área em estudo o pesquisador tem interesse em pesquisar a variabilidade das variáveis e se há evidências que os passos que descrevem a estrutura espacial dessas variáveis não são independentes pode-se propor o estudo de um modelo geoestatístico bivariado. Este trabalho concentrou-se em avaliar a influência da variação dos parâmetros do modelo bivariado com componente de correlação parcialmente comum (*bivariate Gaussian common component model* – BGCCM) no cálculo do coeficiente de correlação linear de Pearson e analisar a influência que tamanhos e configurações amostrais podem apresentar na estimação do modelo BGCCM e na predição espacial de variáveis em localizações não amostradas. Além disso, para amostras colocadas, construíram-se os semivariogramas cruzados e comparou-se com o modelo univariado e BGCCM, em relação às estimativas do modelo e as medidas associadas à predição espacial. Para isso, essas metodologias foram aplicadas em conjuntos de dados simulados e dados experimentais, provenientes de uma propriedade agrícola. O estudo de simulação da análise da influência da variação dos parâmetros do modelo bivariado BGCCM no cálculo do coeficiente de correlação linear de Pearson entre as variáveis descritas do modelo bivariado BGCCM revelou que o coeficiente de correlação linear de Pearson não pode ser considerado na tomada de decisão quanto à presença de dependência espacial conjunta entre pares de variáveis. No estudo com simulações observou-se que as maiores diferenças das medidas de acurácia e da soma quadrada da diferença entre as predições espaciais ocorreram quando se comparou os modelos univariado e semivariograma cruzado com o BGCCM. Ainda no estudo de simulação observou-se que para os dados balanceados as malhas regular e irregular apresentaram melhor eficiência quanto à predição espacial. O estudo com dados reais mostrou que, sob a abordagem do modelo BGCCM, observou-se a presença de dependência espacial principalmente entre pH e Mn para dados co-locados e balanceados do ano agrícola 2010/2011; e entre as variáveis dentro dos seguintes pares: (Prod, RSP0-10), (Prod, RSP11-20), (Prod, RSP21-30) e (Prod, Mn) para dados co-locados e balanceados do ano agrícola 2013/2014; e (Prod, RSP11-20) e (Prod, RSP21-30) para dados co-locados e desbalanceados do ano agrícola 2013/2014. Ainda considerando o estudo com dados reais comparando os modelos univariado, semivariograma cruzado e BGCCM, mostraram diferenças na soma quadrada da diferença da predição e nas medidas acurácia, tanto para dados balanceados como para os desbalanceados. No entanto, considerando os dados reais e a amostra controle, a qualidade da predição espacial usando o modelo BGCCM se mostrou ineficiente quando comparada com a qualidade obtida na predição espacial usando o modelo univariado. Porém, esse resultado pode ter sido influenciado pela escolha da configuração amostral utilizada.

Palavras-Chave: geoestatística, modelo bivariado com componente de correlação parcialmente comum, simulação, semivariograma cruzado.

THE INFLUENCE OF SAMPLE CONFIGURATIONS IN QUALITY PET UNDER THE USE OF SPATIAL MODELS BIVARIATE

ABSTRACT

The soil spatial variability's studies are based in geostatistics which appears as a method whose data comes from natural phenomena and consider the geographical location of the phenomenon. If in an area under study the researcher has interest in searching the variability of variables and has evidence that the steps which describe the spatial structure of this variables aren't independent a geostatistical bivariate model study can be proposed. This work concentrates on evaluating the variation's influence of the bivariate Gaussian common component model (BGCCM) parameters in calculating the Pearson correlation coefficient and analyzing the influence that sizes and sampling settings may present at the BGCCM's estimation and at spatial prediction variables in non-sampled locations. Moreover, for co-placed samples, crossed semivariograms were built and compared with univariate model and BGCCM, in relation to estimates of the model and the sizes associated with the spatial prediction. In order to do it, these methodologies were applied in simulated data sets and experimental data, from an agricultural property. The simulation study of the parameters variation influence's analysis of the bivariate model BGCCM in calculating the Pearson correlation coefficient between described variables of bivariate model BGCCM revealed that the Pearson's linear correlation coefficient can't be considered in decision-making about the presence of joint spatial dependence between pairs of variables. In the study with simulations, it was observed that the biggest differences of accuracy measures and the square sum of the spatial prediction's difference occurred when the univariate models and crossed semivariogram were compared to the BGCCM. Moreover, the simulation's study observed that for balanced data the regular and irregular meshes showed better efficiency as the spatial prediction. The study with real data showed that under the BGCCM approach, spatial dependence was observed, mainly between pH and Mn for co-placed and balanced data of the agriculture year 2010/2011; and between the variable inside of the next pairs: (Prod, RSP0-10), (Prod, RSP11-20), (Prod, RSP21-30) and (Prod, Mn) for co-placed and balanced data of the agriculture year 2013/2014; and (Prod, RSP11-20) and (Prod, RSP21-30) for co-placed and unbalanced data of the agriculture year 2013/2014. Still considering the real data study comparing the univariate models, crossed semivariogram and BGCCM, differences could be observed in the square sum of prediction's difference and in the accuracy measures, both for balanced and unbalanced data. However, considering the real data and the control sample, the spatial prediction's quality using the BGCCM model was inefficient when compared to the quality resulted from the spatial prediction using the univariate model. Nevertheless, this result may have been influenced by the choice of sample configuration.

Key-words: bivariate Gaussian common component model, crossed semivariogram, geostatistics, simulation.

SUMÁRIO

LISTA DE TABELAS	xi
LISTA DE FIGURAS	xiv
1 INTRODUÇÃO	1
2 OBJETIVOS	3
2.1 Objetivo geral.....	3
2.2 Objetivos específicos	3
3 REVISÃO BIBLIOGRÁFICA	4
3.1 Geoestatística.....	4
3.2 Semivariograma experimental	4
3.3 Continuidade espacial.....	8
3.4 Modelos para a função semivariância	9
3.5 Modelagem espacial	12
3.6 Modelos geoestatísticos lineares gaussianos univariados	12
3.7 Modelos geoestatísticos bivariados gaussianos.....	13
3.8 Estimação dos parâmetros do modelo por máxima verossimilhança	17
3.9 Critérios de validação de modelo	22
3.10 Krigagem e Cokrigagem	23
3.11 Acurácia em mapas temáticos	28
3.12 Amostragem espacial	30
3.13 Tipos de amostragem	30
3.14 Tamanho amostral	35
3.15 Espaçamento amostral	36
4 MATERIAL E MÉTODOS	38
4.1 Análise da influência da variação dos parâmetros do modelo bivariado BGCCM no cálculo do coeficiente de correlação linear de Pearson.....	38
4.2 Comparação de modelos	40
4.2.1 Estudo com dados simulados	40
4.2.2 Estudo com dados experimentais	41
4.2.2.1 Análise para dados co-locados e balanceados	42
4.2.2.2 Análise para dados co-locados e desbalanceados.....	44
4.3 Comparação de malhas	45
5 RESULTADOS E DISCUSSÕES	48

5.1 Análise da influência da variação dos parâmetros do modelo bivariado BGCCM no cálculo do coeficiente de correlação linear de Pearson.....	48
5.2 Comparação de modelos com dados simulados.....	51
5.3 Comparação de modelos com dados experimentais co-locados e balanceados do ano agrícola 2010/2011	54
5.4 Comparação de modelos com dados experimentais co-locados e balanceados do ano agrícola 2013/2014	63
5.5 Comparação de modelos com dados experimentais co-locados e desbalanceados do ano agrícola 2013/2014	72
5.6 Comparação de malhas amostrais para dados co-locados e balanceados	82
5.7 Comparação de malhas amostrais para dados co-locados e desbalanceados	86
6 CONCLUSÕES.....	89
7 CONSIDERAÇÕES FINAIS	90
REFERÊNCIAS	91

LISTA DE TABELAS

Tabela 1 Matriz genérica dos erros.....	28
Tabela 2 Parâmetros que definem a estrutura de dependência espacial bivariada, considerando o modelo BGCCM	39
Tabela 3 Porcentagem de amostras segundo classificação de Dancey e Reidy (2007) para as amostras de a1 a a10	48
Tabela 4 Estatísticas descritivas das amostras variando ϕ_0	49
Tabela 5 Porcentagem de amostras segundo classificação de Dancey e Reidy (2007) para as amostras variando ϕ_0	50
Tabela 6 Estatísticas descritivas das amostras variando ϕ_0, ϕ_1 e ϕ_2 simultaneamente	50
Tabela 7 Porcentagem de amostras segundo classificação de Dancey e Reidy (2007) para as amostras variando ϕ_0, ϕ_1 e ϕ_2 simultaneamente.....	51
Tabela 8 Estatística descritiva da soma quadrada da diferença da predição espacial das variáveis Y_1 e Y_2 , comparando os modelos univariado (u), semivariograma cruzado (sc) e modelo BGCCM (mb)	52
Tabela 9 Estatística descritiva da exatidão global das variáveis Y_1 e Y_2 , comparando os modelos univariado (u), semivariograma cruzado (sc) e modelo BGCCM (mb)	52
Tabela 10 Estatística descritiva do índice Kappa das variáveis Y_1 e Y_2 , comparando os modelos univariado (u), semivariograma cruzado (sc) e modelo BGCCM (mb)	52
Tabela 11 Estatística descritiva do índice Tau das variáveis Y_1 e Y_2 , comparando os modelos univariado (u), semivariograma cruzado (sc) e modelo BGCCM (mb)	53
Tabela 12 Medidas de correlação linear de Pearson entre as variáveis pH e cálcio, pH e magnésio, pH e manganês	55
Tabela 13 Medidas descritivas das variáveis pH, cálcio, magnésio e manganês.....	55
Tabela 14 Critérios utilizados para a escolha do melhor modelo ajustado para o modelo univariado.....	58
Tabela 15 Critérios utilizados para a escolha do melhor modelo ajustado para o semivariograma cruzado.....	59
Tabela 16 Critérios utilizados para a escolha do melhor modelo ajustado para o modelo BGCCM	59
Tabela 17 Modelos espaciais ajustados para o modelo univariado e os parâmetros estimados pelo método de máxima verossimilhança	60
Tabela 18 Modelos espaciais ajustados para o semivariograma cruzado e os parâmetros estimados pelo método de mínimos quadrados	60

Tabela 19 Modelos espaciais ajustados para o modelo BGCCM e os parâmetros estimados pelo método de máxima verossimilhança	60
Tabela 20 Medidas de acurácia para o modelo univariado e semivariograma cruzado.....	62
Tabela 21 Medidas de acurácia para modelo univariado e BGCCM	63
Tabela 22 Medidas de acurácia para o semivariograma cruzado e modelo BGCCM.....	63
Tabela 23 Medidas de correlação linear de Pearson entre as variáveis.....	64
Tabela 24 Estatística descritiva das variáveis produtividade, resistência do solo a penetração e manganês.....	64
Tabela 25 Critérios utilizados para a escolha do melhor modelo ajustado para o modelo univariado.....	68
Tabela 26 Critérios utilizados para a escolha do melhor modelo ajustado para o semivariograma cruzado.....	68
Tabela 27 Critérios utilizados para a escolha do melhor modelo ajustado para o modelo BGCCM	69
Tabela 28 Modelos espaciais ajustados para o modelo univariado e os parâmetros estimados pelo método de máxima verossimilhança	70
Tabela 29 Modelos espaciais ajustados para o semivariograma cruzado e os parâmetros estimados pelo método de mínimos quadrados	70
Tabela 30 Modelos espaciais ajustados para o modelo BGCCM e os parâmetros estimados pelo método de máxima verossimilhança	70
Tabela 31 Medidas de acurácia para o modelo univariado e semivariograma cruzado.....	71
Tabela 32 Medidas de acurácia para o modelo univariado e BGCCM	71
Tabela 33 Medidas de acurácia para o semivariograma cruzado e modelo BGCCM.....	71
Tabela 34 Medidas descritivas das variáveis produtividade, resistência do solo a penetração e manganês.....	73
Tabela 35 Critérios utilizados para a escolha do melhor modelo ajustado para o modelo univariado.....	77
Tabela 36 Critérios utilizados para a escolha do melhor modelo ajustado para o modelo BGCCM	77
Tabela 37 Modelos espaciais ajustados para o modelo univariado e os parâmetros estimados pelo método de máxima verossimilhança	78
Tabela 38 Modelos espaciais ajustados para o modelo BGCCM e os parâmetros estimados pelo método de máxima verossimilhança	78
Tabela 39 Medidas de acurácia para o modelo univariado e BGCCM	78
Tabela 40 Estatística descritiva do erro de predição espacial para o grupo controle	81
Tabela 41 Estatística descritiva da soma quadrada da diferença da predição espacial das variáveis 1 e 2.....	82
Tabela 42 Estatística descritiva da exatidão global das variáveis 1 e 2.	82

Tabela 43 Estatística descritiva do índice Kappa das variáveis 1 e 2.....	83
Tabela 44 Estatística descritiva do índice Tau das variáveis 1 e 2.....	83
Tabela 45 Estatística descritiva das somas quadradas do erro da predição espacial para o grupo de controle para ambas as variáveis.....	85
Tabela 46 Estatística descritiva da soma quadrada da diferença da predição espacial das variáveis 1 e 2.....	86
Tabela 47 Estatística descritiva da exatidão global das variáveis 1 e 2	86
Tabela 48 Estatística descritiva do índice Kappa das variáveis 1 e 2.....	86
Tabela 49 Estatística descritiva do índice Tau das variáveis 1 e 2.....	87

LISTA DE FIGURAS

Figura 1 Representação gráfica do modelo teórico do semivariograma experimental.....	7
Figura 2 Malha amostral irregular.	31
Figura 3 Malha amostral estratificada.	32
Figura 4 Malha amostral sistemática regular.....	33
Figura 5 Malha amostral alinhada em uma direção (a) e alinhada em ambas as direções (b).	34
Figura 6 Malha amostral <i>lattice plus close pairs</i> ($8 \times 8, 3, 4 \times 4$).....	35
Figura 7 Malha amostral <i>lattice plus in-fill</i>	35
Figura 8 Exemplo de malha regular.	38
Figura 9 Exemplo de uma simulação.....	40
Figura 10 Área experimental e grade de amostragem.	42
Figura 11 Localizações restantes.....	44
Figura 12 Malhas amostrais: regular (a), irregular (b) e <i>lattice plus close pairs</i> (c).....	46
Figura 13 Gráfico <i>Boxplot</i> do coeficiente de correlação linear de Pearson para as simulações das amostras de a1 a a10.....	48
Figura 14 Gráfico <i>Boxplot</i> do coeficiente de correlação linear de Pearson para as simulações das amostras variando ϕ_0	49
Figura 15 Gráfico <i>Boxplot</i> do coeficiente de correlação linear de Pearson para as simulações das amostras variando ϕ_0, ϕ_1 e ϕ_2 simultaneamente.....	50
Figura 16 Gráficos <i>Boxplots</i> das somas quadradas da diferença de predição espacial para as variáveis Y_1 (a) e Y_2 (b), comparando os modelos univariado (u), semivariograma cruzado (sc) e modelo BGCCM (mb).....	53
Figura 17 Gráfico <i>Boxplot</i> para a exatidão global das variáveis Y_1 e Y_2 , comparando os modelos univariado (u), semivariograma cruzado (sc) e modelo BGCCM (mb).	53
Figura 18 Gráfico <i>Boxplot</i> para o índice Kappa das variáveis Y_1 e Y_2 , comparando os modelos univariado (u), semivariograma cruzado (sc) e modelo BGCCM (mb).	54
Figura 19 Gráfico <i>Boxplot</i> para o índice Tau das variáveis Y_1 e Y_2 , comparando os modelos univariado (u), semivariograma cruzado (sc) e modelo BGCCM (mb).	54
Figura 20 Gráfico <i>Boxplot</i> para as variáveis pH, cálcio, magnésio e manganês.....	55
Figura 21 Histograma para as variáveis pH, cálcio, magnésio e manganês.....	56
Figura 22 Gráfico <i>Postplot</i> para as variáveis pH (a), cálcio (b), magnésio (c) e manganês (d).	57
Figura 23 Semivariograma experimental univariado para as variáveis pH (a), cálcio (b), magnésio (c) e manganês (d).	58

Figura 24 Semivariograma cruzado para as variáveis pH e cálcio (a), pH e magnésio (b), pH e manganês (c), sendo que para todas as variáveis foi utilizado o pH na estrutura de espacial bivariada.	58
Figura 25 Mapas de variabilidade para o pH, pH/u (a), pH e Ca/sc (b), pH e Ca/mb (c), pH/u (d), pH e Mg/sc (e), pH e Mg/mb (f), pH/u (g), pH e Mn/sc (h) e pH e Mn/mb (i).	61
Figura 26 Mapas de variabilidade Ca/u (a), Ca e pH/sc (b), Ca e pH/mb (c), Mg/u (d), Mg e pH/sc (e), Mg e pH/mb (f), Mn/u (g), Mn e pH/sc (h), Mn e pH/mb (i).	62
Figura 27 Gráfico <i>Boxplot</i> para as variáveis Prod, RSP0-10, RSP11-20, RSP21-30 e Mn... 65	65
Figura 28 Histograma para as variáveis Prod, RSP0-10, RSP11-20, RSP21-30 e Mn.....	65
Figura 29 Gráfico <i>Postplot</i> para as variáveis Prod (a), RSP0-10 (b), RSP11-20 (c), RSP21-30 (d) e manganês (e), classificados em intervalos de mesma amplitude.....	66
Figura 30 Semivariograma experimental univariado para as variáveis Prod (a), RSP0-10 (b), RSP11-20 (c), RSP21-30 (d) e manganês (e).....	67
Figura 31 Semivariograma cruzado para as variáveis RSP0-10 (a), RSP11-20 (b), RSP21-30 (c) e manganês (d), sendo que para todas as variáveis foi considerada a produtividade na estrutura espacial bivariada.	68
Figura 32 Mapas de variabilidade RSP0-10/U (a), RSP0-10/SC (b), RSP0-10/BGCCM (c), RSP11-20/U (d), RSP11-20/SC (e), RSP11-20/BGCCM (f), RSP21-30/U (g), RSP21-30/SC (h), RSP21-30/BGCCM (i), Mn/U (j), Mn/SC (k) e Mn/BGCCM (l).	72
Figura 33 Gráfico <i>Boxplot</i> para as variáveis produtividade, RSP0-10, RSP11-20, RSP21-30 e manganês.....	73
Figura 34 Histograma para as variáveis produtividade, RSP0-10, RSP11-20, RSP21-30 e manganês.....	74
Figura 35 Gráfico <i>Postplot</i> para as variáveis RSP0-10 (a), RSP11-20 (b), RSP21-30 (c) e manganês (d).....	75
Figura 36 Semivariograma experimental univariado para as variáveis RSP0-10 (a), RSP11-20 (b), RSP21-30 (c) e manganês (d).	76
Figura 37 RSP0-10/u (a), RSP0-10/mb (b), RSP11-20/u (c), RSP11-20/mb (d), RSP21-30/u (e), RSP21-30/mb (f), Mn/u (g) e Mn/mb (h).	80
Figura 38 Mapas de erros RSP0-10/u (a), RSP0-10/mb (b), RSP11-20/u (c), RSP11-20/mb (d), RSP21-30/u (e), RSP21-30/mb (f), Mn/u (g) e Mn/mb (h), classificados em intervalos de mesma amplitude.	81
Figura 39 Gráfico <i>Boxplot</i> para as somas quadradas da diferença da predição para as variáveis 1 e 2.....	83
Figura 40 Gráfico <i>Boxplot</i> para a exatidão global das variáveis 1 e 2.	84
Figura 41 Gráfico <i>Boxplot</i> para o índice Kappa das variáveis 1 e 2.	84
Figura 42 Gráfico <i>Boxplot</i> para o índice Tau das variáveis 1 e 2.	84

Figura 43 Gráfico <i>Boxplot</i> das somas quadradas do erro de predição espacial para o grupo de controle para as variáveis 1 e 2.	85
Figura 44 Gráfico <i>Boxplot</i> para as somas quadradas da diferença da predição espacial das variáveis 1 e 2.....	87
Figura 45 Gráfico <i>Boxplot</i> para a exatidão global das variáveis 1 e 2.	87
Figura 46 Gráfico <i>Boxplot</i> para o índice Kappa das variáveis 1 e 2.....	88
Figura 47 Gráfico <i>Boxplot</i> para o índice Tau das variáveis 1 e 2.	88

1 INTRODUÇÃO

O uso irracional de insumos agrícolas vem causando impactos negativos ao meio ambiente, tal como, a utilização excessiva de produtos químicos ao solo como adubos, inseticidas, herbicidas e agrotóxicos que, uma vez aplicados nas áreas de cultivo, são levados pelas águas das chuvas aos rios, além de penetrarem no solo atingindo os lençóis freáticos. Devido a esses impactos, a produção agrícola passou a ser objeto de estudo de diversos pesquisadores, os quais buscam alternativas de modo a minimizar o uso de insumos agrícolas e maximizar a produção de alimentos, sem perda de qualidade.

Segundo Guedes et al. (2008), o recente processo de globalização da economia vem impulsionando o aprimoramento agrícola. Logo, pesquisadores e agricultores estão adotando um novo sistema de agricultura sustentável, chamado de agricultura de precisão.

A agricultura de precisão trata-se de uma prática integrada de informações e tecnologias, na qual objetiva-se estudar a área agrícola, por meio do mapeamento da variabilidade espacial dos componentes do solo.

Segundo Molin (2002), a agricultura de precisão é um sistema de gerenciamento da produção que leva em consideração a variabilidade espacial da produtividade e dos fatores de produção, sendo possível identificar e quantificar as necessidades locais do solo para aplicar insumos na dosagem correta e em locais onde há maior necessidade. Assim, para a implementação de um programa de agricultura de precisão são necessários estudos da variabilidade espacial do solo por meio da geoestatística.

De acordo com Vieira (2000), a geoestatística está fundamentada na Teoria das Variáveis Regionalizadas e surgiu na África do Sul a partir de estudos com dados de concentração de ouro do matemático Daniel Krige em 1951, e tem por objetivo principal estudar o grau de dependência espacial entre as amostras observadas dentro do campo experimental. A geoestatística apresenta-se como um método que se utiliza de procedimentos estatísticos aplicados a problemas cujos dados provêm de fenômenos naturais e que levam em consideração a localização geográfica do fenômeno em estudo.

Por meio de técnicas geoestatísticas é possível estimar um modelo espacial que tem como objetivo descrever o comportamento de algumas características da estrutura de dependência espacial da variável sob estudo que não podem ser descritas exclusivamente por modelos determinísticos.

Na literatura são apresentados diversos modelos geoestatísticos, os quais se dividem em dois subconjuntos: modelos univariados quando envolve o estudo de uma única variável georreferenciada e modelos multivariados, isto é, modelos que consideram mais de uma variável georreferenciada, desde que exista uma relação entre os atributos. No caso de

haver apenas duas variáveis envolvidas no estudo, os dados são denominados como bivariados.

Winter (1971) relata que uma desvantagem dos modelos univariados é não se levar em conta as correlações entre as variáveis medidas nas mesmas unidades experimentais. Ferreira e Duarte (1992) destacam ainda outro problema, que é o de conduzir a conclusões divergentes, quando das análises realizadas para cada variável (GODOY, 1985, apud SILVA et al., 1997, p. 57).

A grande vantagem dos modelos multivariados é oferecer um resultado único da análise, assegurando, assim, um poder muito maior de generalização dos resultados, em comparação com os modelos univariados (GODOY, 1985, apud SILVA et al., 1997, p. 58).

Especificamente para o estudo da variabilidade espacial conjunta de duas variáveis, a utilização de modelos geoestatísticos bivariados é vantajosa quando na área sob estudo tem-se que uma das variáveis do solo é mais dispendiosa para ser observada do que outra variável, sendo que ambas são correlacionadas. Os resultados fornecem mapas de variabilidade descrevendo o comportamento de ambos os atributos na área de estudo.

Bognola et al. (2008) argumentam que explorar metodologias multivariadas pode reduzir custos amostrais de variáveis cujas medidas sejam dispendiosas, mantendo-se a previsão das estimativas ao se explorarem as relações espaciais entre diferentes variáveis.

Além disso, Bognola et al. (2008), ao avaliarem a variabilidade espacial do rendimento de *Pinus taeda* L. em função de teores de argila do solo em uma área localizada no município de Rio Negrinho, estado de Santa Catarina, observaram que devido à presença de continuidade espacial no estudo, o uso do estimador geoestatístico multivariado, para a estimativa do incremento médio anual volumétrico, foi preferido ao estimador geoestatístico usual univariado, por duas razões, proporcionou estimativas mais detalhadas da variável de interesse e é de mais baixo custo quando comparado com o estimador clássico (krigagem ordinária), em vista de se obterem os teores de argila através do mapeamento detalhado dos solos de forma mais econômica do que os levantamentos dendrométricos de pré-corte, praticados pelas empresas florestais madeireiras.

Fonseca (2008) avaliou o comportamento de modelos geoestatísticos bivariados e verificou que os modelos bivariados com componente de correlação parcialmente comum e de co-regionalização possuem desempenhos semelhantes para estimação dos parâmetros e krigagem. No entanto, a análise de dados observacionais do solo mostrou que os modelos geoestatísticos bivariados são ferramentas úteis para a agricultura de precisão, pois sua utilização pode baratear os custos dos monitoramentos de solos e ao mesmo tempo gerar mapas preditivos com alta qualidade.

Na literatura, não há evidências de trabalhos, os quais têm por objetivo estudar a influência do tamanho e da configuração amostral de duas variáveis georreferenciadas sob

estudo na estimação de modelos geoestatísticos bivariados, bem como na predição espacial dessas variáveis em localizações não amostrais.

2 OBJETIVOS

2.1 Objetivo geral

Analisar a influência que tamanhos e configurações amostrais podem apresentar na estimação de modelos geoestatísticos bivariados e na predição espacial de variáveis em localizações não amostradas.

2.2 Objetivos específicos

- Estudar a influência da variação dos parâmetros do modelo bivariado BGCCM no cálculo do coeficiente de correlação linear de Pearson;
- Estudar a variabilidade de propriedades químicas e físicas do solo por meio dos modelos geoestatísticos univariados e bivariados;
- Construir mapas temáticos que representam a variabilidade espacial do conjunto de dados na área em estudo;
- Comparar os modelos geoestatísticos univariados e bivariados, por meio das simulações e para cada tamanho e configuração amostral, quanto à estimação do modelo e a qualidade de predição espacial;
- Para cada simulação obtida para cada modelo geoestatístico bivariado, comparar a qualidade de estimação do modelo e a qualidade da predição espacial, considerando os diferentes tamanhos e configurações amostrais;
- Avaliar a eficiência de modelos bivariados na estimação da produtividade da soja, propriedades químicas e físicas do solo, considerando uma estrutura de dependência espacial bivariada;
- Para amostras co-locadas, considerando os dados reais e simulados, comparar o modelo geoestatístico bivariado com o modelo obtido pelo semivariograma cruzado, em relação às estimativas do modelo e as medidas associadas à predição espacial.

3 REVISÃO BIBLIOGRÁFICA

3.1 Geoestatística

A geoestatística surgiu na África do Sul a partir de estudos com dados de concentração de ouro do matemático Daniel G. Krige em 1951 e está baseada na Teoria das Variáveis Regionalizadas, idealizada por Matheron nas décadas de 50 e 60 (VIEIRA, 2000).

De acordo com Faraco et al. (2008), a Teoria das Variáveis Regionalizadas, leva em consideração o princípio que os valores de uma variável distribuída no espaço são relacionados com a posição que ocupam, isto é, leva em consideração à localização geográfica do fenômeno em estudo.

Vieira (2000) descreve que Daniel G. Krige concluiu que a variância dos dados depende da distância entre as amostras. Assim a modelagem geoestatística procura obter uma função aleatória para explicar o comportamento de uma ou mais variáveis que possuem localizações espaciais conhecidas.

A abordagem geoestatística se diferencia dos modelos lineares nos pressupostos em que todas as observações não são assumidas independentes e que existe um efeito aleatório latente na parte explanatória do modelo (FONSECA, 2008).

Segundo Guedes (2008), as técnicas da geoestatística permitem, a partir de alguns pontos amostrais, conhecer a continuidade das variáveis de interesse, em toda a área de estudo, demonstrando a variação espacial do fenômeno por meio de mapas de variabilidade.

Para Ferraz et al. (2012), a geoestatística é uma ferramenta importante para análise de dados, no âmbito da agricultura de precisão. Vieira (2000) acrescenta ainda que o objetivo da geoestatística aplicada à agricultura de precisão é caracterizar a variabilidade espacial dos atributos do solo e das plantas e fazer estimativa, utilizando o princípio da variabilidade espacial a fim de se identificar inter-relações desses atributos no espaço e no tempo, além de permitir estudar padrões de amostragem adequados.

Outro objetivo da geoestatística é medir o grau de dependência espacial entre as amostras observadas dentro do campo experimental. De acordo com Landim (2006), um dos métodos de estudo de dependência espacial é por meio da construção de semivariograma.

3.2 Semivariograma experimental

O semivariograma experimental é um gráfico de dispersão que relaciona os valores das estimativas da semivariância $\hat{\gamma}(h)$ com distância euclidiana entre dois pontos quaisquer

do campo aleatório(h). Espera-se que a semivariância aumente com a distância h até atingir um valor máximo no qual se estabiliza (CRESSIE, 1993).

Segundo Guerra (1988), existem três tipos de semivariogramas: verdadeiro, experimental e teórico. O semivariograma verdadeiro representa a dependência real que é desconhecida. O semivariograma experimental é obtido por meio dos pontos amostrais e o semivariograma teórico é obtido ajustando um modelo matemático ao semivariograma experimental.

De acordo com Marins et al. (2008), a função semivariância $\gamma(h)$ é definida como sendo a metade da esperança matemática do quadrado da diferença do valor da variável regionalizada ou processo estocástico $Z(s_i)$ e $Z(s_i + h)$ separados por um espaçamento h entre dois pontos, descrito na Equação 1.

$$\gamma(h) = \frac{1}{2} E \left[(Z(s_i) - Z(s_i + h))^2 \right] \quad \text{Eq.(1)}$$

Na literatura apresentam-se vários estimadores para o cálculo da função semivariância para uma variável, os mais utilizados são de Matheron, descrito por Journel e Huijbregts (1978), apresentado na Equação 2 e o proposto por Cressie e Hawkins (1980), descrito na Equação 3.

$$\hat{\gamma}(h) = \frac{1}{2 N(h)} \sum_{i=1}^{N(h)} [Z(s_i) - Z(s_i + h)]^2 \quad \text{Eq.(2)}$$

$$\hat{\gamma}(h) = \frac{\frac{1}{2 N(h)} \left(\sum_{i=1}^{N(h)} |Z(s_i) - Z(s_i + h)|^{1/2} \right)^4}{0,457 + (0,494/N(h))} \quad \text{Eq.(3)}$$

Em que $N(h)$ é o número de pares de valores medidos da dissimilaridade entre $Z(s_i)$ e $Z(s_i + h)$ separados por h, sendo que $Z(s_i + h)$ é o valor da variável regionalizada na amostra localizada em $(s_i + h)$ e $Z(s_i)$ é o valor da variável regionalizada na amostra localizada em (s_i) .

Segundo Cressie (1993), o estimador clássico de Matheron é não viciado, mas é influenciado pela presença de *outliers*, isto é, é o mais estável desde que os dados apresentam distribuição normal de probabilidade. Entretanto, na presença de *outliers* o estimador de Cressie e Hawkins é o mais apropriado, uma vez que é mais robusto.

Teixeira (2013) comparou estimadores de semivariância, realizando simulações de diferentes tipos de bancos de dados, com o objetivo de verificar o comportamento de cada estimador diante de determinada característica. Observou que o estimador de Cressie e Hawkins demonstrou bom desempenho, de acordo com o aumento da quantidade de *outliers*.

Para descrever a continuidade espacial entre duas variáveis é utilizado o semivariograma cruzado (Equação 4), no qual é possível certificar a relação entre duas variáveis espacialmente correlacionadas, mostrando que a variabilidade espacial de uma variável é acompanhada pela variabilidade espacial da outra variável. Segundo Vieira (1998), o semivariograma cruzado é utilizado apenas quando existe o mesmo número de parcelas amostrais observadas para as duas variáveis em estudo e que essas parcelas amostrais tenham posições geográficas coincidentes.

$$\hat{\gamma}(h) = \frac{1}{2 N(h)} \sum_{i=1}^{N(h)} [Z_i(s_i) - Z_i(s_i + h)] \times [Z_j(s_i) - Z_j(s_i + h)] \quad \text{Eq.(4)}$$

onde i e j representam a i -ésima e a j -ésima variável.

Para a construção do semivariograma, além de definir um estimador define-se também uma distância de corte (*cutoff*). Clark (1979) recomenda utilizar 50% da distância máxima entre pontos.

Se a amostragem possui uma estrutura regular, uma escolha natural para os valores de h é a própria distância entre os pontos amostrais. Se a amostragem possui uma estrutura irregular define-se um *lag* e uma tolerância, que são medidas que correspondem a um intervalo de distância pré-definida e então calcula a semivariância para todos os pares de pontos compreendidos no *lag* (CÂMARA; MEDEIROS, 1988).

De acordo com Guedes et al. (2008), a construção do semivariograma pode ser feita por meio de um único semivariograma, chamado omnidirecional, o qual contém informações de todas as direções, ou por meio do semivariograma direcional que possui apenas informações da direção pré-definida.

Segundo Câmara e Medeiros (1998), as direções convencionais usadas para a construção dos semivariogramas direcionais são 0° , 45° , 90° e 135° , com um raio de tolerância para o *lag* igual a $22,5^\circ$, seguindo o sistema de medida de direção azimute, que considera o norte como sendo a direção 0° . No entanto, a construção de semivariograma nas demais direções não é impedida de ser realizada.

A construção do semivariograma direcional identifica se a estrutura de dependência espacial da variável em estudo depende da direção. Esse comportamento indica que o fenômeno em estudo tem característica chamada de anisotropia (GUEDES et al., 2008).

Silva et al. (2007) relatam que não há consenso, na bibliografia, a respeito do número mínimo de pares de pontos amostrais suficientes para a estimativa da semivariância experimental a cada *lag*. Landim (2006) afirma que se deve utilizar o mínimo de 30 pares, enquanto Webster e Oliver (2007) recomendam o mínimo de 100 pares de pontos amostrais.

Construído o semivariograma experimental, podem-se identificar alguns parâmetros, tais como alcance $a = g(\varphi_3)$, patamar $(\varphi_1 + \varphi_2)$, efeito pepita (φ_1) e contribuição (φ_2) (Figura 1).

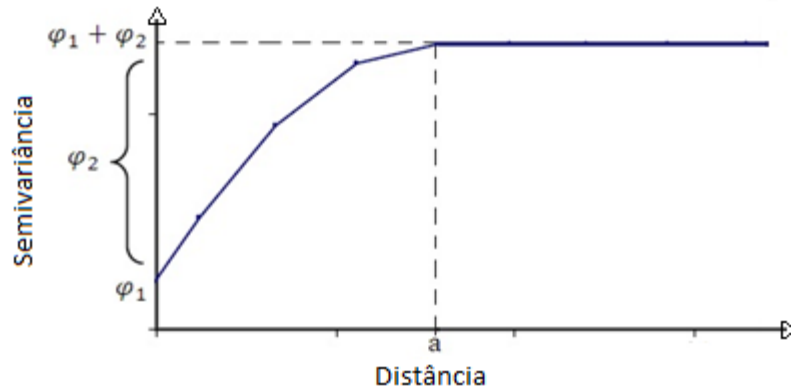


Figura 1 Representação gráfica do modelo teórico do semivariograma experimental.

Esses parâmetros descrevem a estrutura de dependência espacial da variável georreferenciadas e, segundo Vieira (2000), são definidos como:

Alcance: é o raio de influência, que representa a distância máxima onde há dependência espacial entre pontos, isto é, autocorrelação espacial entre as amostras. Para distâncias maiores que o alcance as amostras são consideradas independentes entre si.

Patamar: é o valor que o semivariograma assume quando a distância assume o valor do alcance. Teoricamente o valor do patamar é igual ao valor da variância do processo estocástico $Z(s_i)$. A partir deste valor considera-se que não existe mais dependência espacial entre as amostras.

Efeito pepita: representa a descontinuidade do semivariograma para distâncias menores que a menor distância entre as amostras, isto é, representa a variabilidade em pequenas escalas. Por definição, $\hat{\gamma}(0) = 0$ quando $h=0$. Porém na prática, à medida que h se aproxima de zero, a semivariância tende a um valor positivo, chamado de efeito pepita. No caso dos valores da semivariância serem constantes ou muito próximos para qualquer valor de h , tem-se o efeito pepita puro e, neste caso, as amostras não possuem dependência espacial.

Contribuição: é a diferença entre o patamar e o efeito pepita.

Comparando as Equações 2 e 4, observa-se que matematicamente o estimador dos valores das semivariâncias para a construção do semivariograma experimental é um caso particular do semivariograma cruzado, quando ambas as variáveis são coincidentes.

Segundo Silva et al. (2003), considerando os métodos de ajuste por mínimos quadrados, a estimação dos parâmetros do modelo que descrevem a estrutura de dependência espacial do semivariograma cruzado é similar ao semivariograma experimental para uma variável; no entanto, as interpretações para os parâmetros são diferentes. Ainda

segundo Silva et al. (2003), o alcance obtido para o semivariograma cruzado representa a distância máxima de dependência espacial entre as variáveis em estudo e o patamar, se existir, se aproxima do valor da covariância entre ambas as variáveis. Quando as variáveis forem inversamente correlacionadas o semivariograma cruzado poderá produzir resultados negativos.

Silva et al. (2003), ao estudarem a correlação espacial das variáveis fósforo e potássio com a produtividade de soja em uma área agrícola experimental, utilizando o semivariograma cruzado, constataram a existência de uma fraca correlação espacial do fósforo com a produtividade de soja; e do potássio com a produtividade de soja. Concluindo então, que para a área estudada as estimativas de produtividade podem ser realizadas sem considerar as variáveis fósforo e potássio.

Por meio do semivariograma, além de identificar os parâmetros: alcance, patamar, efeito pepita e contribuição; é possível identificar características relacionadas à estrutura e localização do fenômeno natural que representam. Dentre as características destaca-se a continuidade espacial (AMARAL et al., 2010).

3.3 Continuidade espacial

Quanto à continuidade, espera-se que elementos amostrais próximos apresentem maior dependência do que elementos amostrais distantes; assim pontos amostrais mais próximos tendem a ser mais semelhantes em relação a pontos mais espaçados. Dessa maneira, é esperado que a semivariância $\hat{\gamma}(h)$ aumente com a distância (h) (CRESSIE, 1993).

Uma forma de descrever a continuidade espacial de uma variável é por meio do gráfico h-scatterplot construído para várias distâncias h . Segundo Isaaks e Srivastava (1989), o h-scatterplot é um gráfico de dispersão de $Z(s)$ versus $Z(s + h)$. Assim, se a nuvem de pontos estiver situada em torno da reta $Z(s) = Z(s + h)$, na direção de 45° passando pela origem, indica que os valores amostrais separados pela distância h são semelhantes. Se conforme aumenta a distância de h a nuvem de pontos se afastarem da reta $Z(s) = Z(s + h)$, então a continuidade espacial na direção h considerada é maior, isto é, os valores são menos similares (KAVANAGH, 2001).

Além do gráfico h-scatterplot, outros métodos gráficos e quantitativos podem ser utilizados para descrever a continuidade espacial, como, por exemplo, função correlação, função covariância e variograma.

O efeito pepita relativo (EPR) (Equação 5) proposto por Cambardella et al. (1994) expressa à intensidade da dependência espacial.

$$EPR = \frac{\varphi_1}{\varphi_1 + \varphi_2} \times 100 \quad \text{Eq.(5)}$$

Cambardella et al. (1994) classificam como sendo a dependência espacial forte para valores $EPR \leq 25\%$; moderada quando o IDE estiver entre 25% e 75% e fraca dependência para valores $EPR \geq 75\%$. Quando $EPR = 100\%$ indica que o semivariograma apresenta efeito pepita puro, indicando que a variável possui independência espacial.

3.4 Modelos para a função semivariância

A partir do semivariograma é possível ajustar uma função contínua às estimativas da semivariância por meio dos modelos teóricos (CAMARGO et al.1998). O modelo ajustado deve ser o mais simples possível e representa uma tendência de $\hat{\gamma}(h)$ em relação aos valores de h .

Segundo Vieira (2000), os modelos podem ser denominados como modelos transitivos, que possuem patamar e modelos não transitivos, sem patamar. Os principais modelos transitivos utilizados na literatura são:

3.4.1 Modelo esférico

De acordo com Vieira (2000), o modelo esférico apresenta crescimento rápido na origem e atinge o patamar $\frac{2}{3}$ do alcance (φ_3). Observa-se que para pequenos valores de h esse modelo apresenta um comportamento linear. Segundo Diggle e Ribeiro Jr. (2007), as principais características do modelo esférico são: existência de alcance, patamar finito e próximo à origem, esse modelo é uma vez diferenciável. É importante destacar que este modelo não tem segunda derivada. A função semivariância é descrita na Equação 6.

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[\frac{3}{2} \left(\frac{h}{\varphi_3} \right) - \frac{1}{2} \left(\frac{h}{\varphi_3} \right)^3 \right], & 0 < h \leq \varphi_3 \\ \varphi_1 + \varphi_2, & h > \varphi_3 \end{cases} \quad \text{Eq.(6)}$$

A função de covariância (Equação 7) é definida como sendo $C(h) = C(0) - \gamma(h)$, onde $C(h)$ é a covariância espacial e $C(0) = \varphi_1 + \varphi_2$.

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & h = 0 \\ \varphi_2 \left[1 - \frac{3}{2} \left(\frac{h}{\varphi_3} \right) + \frac{1}{2} \left(\frac{h}{\varphi_3} \right)^3 \right], & 0 < h \leq \varphi_3 \\ 0, & h > \varphi_3 \end{cases} \quad \text{Eq.(7)}$$

A função de correlação espacial (Equação 8) é definida com sendo $\rho(h) = 1 - \frac{\gamma(h)}{C(h)}$.

$$\rho(h) = \begin{cases} 1, & h = 0 \\ 1 - \frac{3}{2} \left(\frac{h}{\varphi_3} \right) + \frac{1}{2} \left(\frac{h}{\varphi_3} \right)^3, & 0 < h \leq \varphi_3 \\ 0, & h > \varphi_3 \end{cases} \quad \text{Eq.(8)}$$

3.4.2 Modelo exponencial

Segundo Vieira (2000), a diferença fundamental entre o modelo exponencial e o modelo esférico é que o exponencial atinge o patamar apenas assintoticamente, enquanto o modelo esférico atinge o patamar no valor do alcance. O parâmetro φ_3 é determinado visualmente como a distância após a qual o semivariograma se estabiliza. Além disso, o alcance prático é definido como a distância na qual o valor do modelo é 95% do patamar (ISAAKS; SRIVASTAVA, 1989). Além disso, o modelo exponencial é contínuo, mas não diferenciável na origem. A função semivariância é descrita na Equação 9.

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[1 - \exp \left(-\frac{h}{\varphi_3} \right) \right], & 0 < h \leq \varphi_3 \\ \varphi_1 + \varphi_2, & h > \varphi_3 \end{cases} \quad \text{Eq.(9)}$$

A função de covariância (Equação 10) é definida como sendo $C(h) = C(0) - \gamma(h)$.

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & h = 0 \\ \varphi_2 \left[\exp \left(-\frac{h}{\varphi_3} \right) \right], & 0 < h \leq \varphi_3 \\ 0, & h > \varphi_3 \end{cases} \quad \text{Eq.(10)}$$

A função de correlação espacial (Equação 11) é definida com sendo $\rho(h) = 1 - \frac{\gamma(h)}{C(h)}$.

$$\rho(h) = \begin{cases} 1, & h = 0 \\ \exp \left(-\frac{h}{\varphi_3} \right), & 0 < h \leq \varphi_3 \\ 0, & h > \varphi_3 \end{cases} \quad \text{Eq.(11)}$$

3.4.3 Modelo gaussiano

Este modelo apresenta comportamento parabólico na origem e é muito utilizado para modelar fenômenos contínuos (SOARES, 2000). De acordo com Borssoi (2007), atinge o patamar assintoticamente e o alcance prático é expresso como sendo $\sqrt{3}\varphi_3$. A função semivariância é descrita na Equação 12.

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left\{ 1 - \exp \left[- \left(\frac{h}{\varphi_3} \right)^2 \right] \right\}, & 0 < h \leq \varphi_3 \\ \varphi_1 + \varphi_2, & h > \varphi_3 \end{cases} \quad \text{Eq.(12)}$$

A função de covariância (Equação13) é definida como sendo $C(h) = C(0) - \gamma(h)$.

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

A função de correlação espacial (Equação14) é definida com sendo $\rho(h) = 1 - \frac{\gamma(h)}{C(h)}$.

$$\rho(h) = \begin{cases} 1, & h = 0 \\ \exp \left[- \left(\frac{h}{\varphi_3} \right)^2 \right], & 0 < h \leq \varphi_3 \\ 0, & h > \varphi_3 \end{cases} \quad \text{Eq.(14)}$$

3.4.5 Modelo Família Matérn

A função semivariância é descrita na Equação 15.

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[\left(1 - 2^{k-1} \Gamma(k) \right)^{-1} \left(\frac{h}{\varphi_3} \right)^k K_k \left(\frac{h}{\varphi_3} \right) \right], & h > 0 \end{cases} \quad \text{Eq.(15)}$$

sendo k um parâmetro relacionado à suavidade do processo, de forma que quanto maior k , maior será suavidade, $\Gamma(\cdot)$ a função gama $K_k(\cdot)$ é a função de Bessel de ordem k (CORREIA, 2013).

A função de covariância (Equação 16) é definida como sendo $C(h) = C(0) - \gamma(h)$.

$$C(h) = \begin{cases} 0, & h = 0 \\ \varphi_2 \left[\left(2^{k-1} \Gamma(k) \right)^{-1} \left(\frac{h}{\varphi_3} \right)^k K_k \left(\frac{h}{\varphi_3} \right) \right], & h > 0 \end{cases} \quad \text{Eq.(16)}$$

A função de correlação espacial (Equação 17) é definida com sendo $\rho(h) = 1 - \frac{\gamma(h)}{C(h)}$.

$$\rho(h) = \begin{cases} \left(2^{k-1} \Gamma(k) \right)^{-1} \left(\frac{h}{\varphi_3} \right)^k K_k \left(\frac{h}{\varphi_3} \right), & h > 0 \end{cases} \quad \text{Eq.(17)}$$

3.5 Modelagem espacial

De acordo com Fonseca (2008), a modelagem espacial é um conjunto de técnicas utilizadas para tentar descrever o comportamento de um ou mais atributos que não podem ser descritos exclusivamente por meio de modelos determinísticos. Isto é, os modelos espaciais são representações de situações reais que tentam explicar a variabilidade dos processos estocásticos por meio de uma ou mais variáveis.

Em geoestatística é apresentada uma gama de modelos, os quais se dividem em dois subconjuntos: modelos univariados, quando envolvem o estudo de uma única variável de interesse e modelos multivariados, isto é, modelos que consideram mais de uma variável, desde que exista uma relação entre as variáveis.

3.6 Modelos geoestatísticos lineares gaussianos univariados

Os primeiros modelos estatísticos propostos foram os lineares univariados, que assumem erros aleatórios independentes e identicamente distribuídos com distribuição de probabilidade gaussiana (FONSECA, 2008).

Dos modelos geoestatísticos o mais simples é o gaussiano estacionário, cujas suposições são definidas por Diggle e Ribeiro Jr. (2007):

- i) $\{Z(s): s \in \mathbb{R}^2\}$ é um processo estocástico gaussiano com média μ , variância $\text{Var}\{Z(s)\}$ e função de correlação $\rho(h) = \text{Corr}\{Z(s), Z(s')\}$, em que $h = \|s - s'\|$ representa a distância entre as localizações s e s' ;
- ii) $\{Y(s): s \in \mathbb{R}^2\}$, y_i são as realizações mutuamente independentes da variável aleatória Y_i , com distribuição normal, média $E[Y_i|Z(\cdot)] = Z(s_i)$ e variância condicional τ^2 .

Consideremos que em alguma área A exista um campo aleatório gaussiano Z , na qual se deseja fazer uma amostragem de n localizações espaciais e observar os valores de atributos georreferenciados nas localizações amostradas. Assim, existe um vetor $\mathbf{Y}(s)$ de $n \times 1$ valores observados no conjunto de localizações $s = (s_1, s_2, \dots, s_n)$ que, segundo Diggle e Ribeiro Jr (2007), sob o pressuposto de gaussianidade, pode ser modelado como descrito na Equação 18.

$$Y(s) = \mu + Z(s) + \varepsilon(s) \quad \text{Eq.(18)}$$

sendo que μ é o termo determinístico do modelo e, segundo Martins (2010), pode ser escrito como na Equação 19.

$$\mu = \beta_0 + \sum_{j=1}^p \beta_j X_j \quad \text{Eq.(19)}$$

em que β_j , $j = 0, \dots, p$ são constantes a serem determinadas e X_j , $j = 1, \dots, p$ covariáveis espaciais. Considerando as n localizações amostrais, esse termo determinístico pode ser escrito matricialmente como $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, em que \mathbf{X} é chamada de matriz de delineamento, com dimensão $n \times q$ e $\boldsymbol{\beta}$ é um vetor $q \times 1$ que contém os $(q - 1)$ parâmetros associados às covariáveis X_1, X_2, \dots, X_j .

Ainda da Equação 18 tem-se que $Z(s)$ é um campo aleatório gaussiano que possui vetor de médias nulo de dimensão $n \times 1$ e matriz de covariância $\Sigma_{n \times n}$, e $\boldsymbol{\varepsilon}(s)$ que é o ruído branco ou erro estocástico para cada localização, sendo que os erros nas n localizações amostrais são considerados mutuamente independentes e identicamente distribuídos com distribuição de probabilidade normal com média zero, desvio padrão τ e podem ser escritos matricialmente pelo vetor $\boldsymbol{\varepsilon}$ de dimensão $n \times 1$.

A partir da Equação 18 tem-se que a distribuição de probabilidade de $\mathbf{Y}(s)$, com vetor de médias $\mathbf{X}\boldsymbol{\beta}$ e matriz de covariâncias dada por $\Sigma_Y = \Sigma + \tau^2 \mathbf{I}$, em que Σ é a matriz de dimensão $n \times n$, cujos elementos representam as correlações entre as observações realizadas em diferentes localizações, \mathbf{I} é uma matriz identidade $n \times n$. Tem-se o vetor $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{\varphi}^*, \tau^2)$, sendo $\boldsymbol{\varphi}^*$ um vetor de parâmetros associados à função de correlação utilizada.

Em alguns processos naturais o uso de modelos espaciais lineares gaussianos univariados não é válido devido a sua simplicidade, o que levou a necessidade de desenvolver técnicas mais sofisticadas para tentar modelar processo que possui estrutura mais complexa de variabilidade (FONSECA, 2008).

3.7 Modelos geoestatísticos bivariados gaussianos

Os modelos bivariados possuem estruturas mais complexas quando comparados com os modelos univariados, uma vez que levam em consideração duas variáveis de interesse, desde que exista uma correlação linear de Pearson entre os dois atributos, isto é, estuda-se o comportamento conjunto das duas variáveis (FONSECA, 2008).

Segundo Fonseca (2008), a utilização de modelos geoestatísticos bivariados se torna vantajosa quando na área sob estudo tem-se que uma das variáveis do solo é mais dispendiosa para ser observada do que outra variável, sendo que ambas são correlacionadas. Sendo assim, as variáveis podem ter sido obtidas por amostragens balanceada ou desbalanceada, co-locadas ou não e em diferentes configurações amostrais. No que segue, usam-se os termos 'amostragem balanceada' quando o número de

observações é o mesmo para ambas as variáveis. O termo 'co-locada' é utilizado quando as amostras de ambas as variáveis são tomadas nas mesmas localizações.

Segundo Righetto (2012), quando o interesse está em mais de um campo aleatório, supõe-se dois, acredita-se que o passo inicial é o de ajustar um modelo geoestatístico para cada um dos dois vetores das variáveis observadas; porém, nem sempre é o melhor caminho, pois há casos em que existe correlação entre as variáveis aleatórias estudadas e, nestes casos, é interessante utilizar modelos geoestatísticos bivariados. Desta forma, os modelos geoestatísticos gaussianos bivariados são utilizados quando envolvem o estudo da variabilidade espacial de duas variáveis georreferenciadas de interesse. Assim existindo dois campos aleatórios gaussianos pode ser modelado como descrito na Equação 20.

$$\mathbf{Y}_i = \mu_i + Z_i + \varepsilon(s_i), \quad i = 1,2 \quad \text{Eq.(20)}$$

sendo que μ_i é o termo determinístico do modelo associada a variável Y_i , $i = 1,2$ e, segundo Martins (2010), pode ser escrito como na Equação 19; Z_i é um campo aleatório gaussiano, e considerando n_i localizações amostradas para cada variável Y_i , tem-se que o campo aleatório Z_i possui vetor de médias nulo de dimensão $n_i \times 1$ e matriz de covariância Σ_i com dimensão $n_i \times n_i$. Segundo Diggle e Ribeiro Jr. (2007), o modelo descrito na Equação 20 está sendo considerado na sua forma mais simples, sem ruído branco. Logo, pela Equação 20 tem-se que $\mathbf{Y}_i = \mu_i + Z_i$, $i = 1,2$.

Segundo Fonseca (2008), $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$ possui distribuição gaussiana n-variada, sendo $n = n_1 + n_2$, com vetor de médias $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ e matriz de covariância positiva definida Σ_Y e que possui o comportamento empírico de correlações. Além disso, a matriz Σ_Y pode ser particionada como na Equação 21.

$$\Sigma_Y = \begin{bmatrix} \Sigma_1 & \Sigma_{1,2} \\ \Sigma_{1,2} & \Sigma_2 \end{bmatrix} \quad \text{Eq.(21)}$$

em que Σ_i é a matriz de dimensão $n_i \times n_i$ das covariáveis da variável Y_i , $i = 1,2$ e $\Sigma_{1,2}$ é a matriz com covariâncias cruzadas entre Y_1 e Y_2 , de dimensão $n_1 \times n_2$.

Na literatura, existem vários trabalhos publicados relacionados ao estudo de modelos geoestatísticos multivariados. Silva (2007) produziu mapas em um contexto multivariado, analisando os componentes principais para a redução do conjunto de variáveis suportes à sua primeira componente, incorporando sua informação de variabilidade espacial à outra variável de interesse principal, em uma estrutura de modelo bivariada. Para isso analisou dados de dois problemas distintos em sua natureza. Em um dos casos os dados reportavam variáveis de produtividade de soja associados a dados físicos e químicos, distribuídos em uma área com sistema de plantio direto. No outro, as variáveis

representavam o incremento médio anual – IMA de *Pinus taeda* L. Silva (2007) comenta, ainda, que a análise revelou a capacidade do modelo bivariado em explorar a informação espacial contida nos dados quando as localizações das variáveis não eram as mesmas e a capacidade de identificar zonas de regionalização onde o modelo univariado não o fez, principalmente em se tratando de amostras pequenas da variável de interesse.

Fonseca (2008), ao estudar dados provenientes da análise de solo de uma propriedade agrícola com 51,8 ha de área, onde foram amostradas 67 localizações georeferenciadas, utilizou os valores mensurados de pH e da saturação por bases do solo, que foram submetidas à análise descritiva espacial, modelagens geoestatísticas univariadas, bivariadas e predições espaciais. Para verificar vantagens quanto à adoção de modelos univariados ou bivariados, a amostra da saturação por bases, que possui coleta mais dispendiosa, foi dividida em uma subamostra de modelagem e uma subamostra de controle. A primeira foi utilizada para fazer a modelagem geoestatística e a segunda foi utilizada para comparar as precisões das predições espaciais nas localizações omitidas no processo de modelagem. Além disso, realizaram implementações computacionais e comparações entre dois modelos bivariados. Os modelos geoestatísticos univariados e bivariados em conjunto com estimadores por máxima verossimilhança se mostraram eficientes para gerar mapas preditivos do pH e da saturação por bases da fazenda Tupã. A análise dos dados mostrou que os modelos geoestatísticos bivariados são uma ferramenta útil para a agricultura de precisão, pois sua utilização pode baratear os custos dos monitoramentos de solos e ao mesmo tempo gerar mapas preditivos com alta qualidade.

3.8.1 Modelos espaciais gaussianos bivariados com componente de correlação parcialmente comum

Diggle e Ribeiro Jr. (2007) propuseram uma abordagem para problemas bivariados denominada Modelo gaussiano bivariado com componente de correlação parcialmente comum, abreviada como BGCCM. Esse modelo utiliza as decomposições dos campos aleatórios latentes da Equação 20 e é descrito como na Equação 22.

$$Z_i = \sigma_{0i}S_0 + \sigma_iS_i, i = 1,2 \quad \text{Eq.(22)}$$

Sendo $\sigma = (\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2)$ um vetor de parâmetros de dispersão associados à Equação 18 e S_0, S_1 e S_2 campos aleatórios gaussianos mutuamente independentes, com vetores de médias nulos, variâncias unitárias e correlações determinadas por funções de correlação válidas. Segundo Fonseca (2008), as correlações, quando combinadas, vão gerar correlações cruzadas válidas entre as variáveis observáveis Y_1 e Y_2 , uma vez que o campo aleatório S_0, S_1 e S_2 são usados para gerar a matriz Σ_Y . Logo, a Equação 20 pode ser determinada pela Equação 23.

$$\begin{cases} Y_1 = \mu_1 + \sigma_{01}S_0 + \sigma_1S_1 \\ Y_2 = \mu_2 + \sigma_{02}S_0 + \sigma_2S_2 \end{cases} \quad \text{Eq.(23)}$$

Sejam $Y_i(s_l)$ e $Y_i(s_k)$ observações da variável Y_i em duas quaisquer localizações de s_l e s_k , separados por uma distância $h = h_{lk}$, $l, k = 1, 2, \dots, n_i$ e $i = 1, 2$. Assim o elemento de matriz $\Sigma_{i,(l,k)}$ é dado por $\text{Cov}(h) = \sigma_{0i}^2\rho_0(h) + \sigma_i^2\rho_i(h)$, sendo ρ_0 e ρ_i as funções de correlações adotadas para S_i , $i = 1, 2$, respectivamente. Pelas propriedades de covariância pode-se encontrar $\Sigma_{1,2}$, que é igual a $\sigma_{01}\sigma_{02}\mathbf{R}_0$, sendo \mathbf{R}_0 uma matriz $n_1 \times n_2$ das correlações cruzadas entre as respostas e depende da função de correlação adotada para S_0 .

Considerando duas localizações amostradas s_l e s_k , sendo $h = h_{lk}$ a distância entre elas. As covariâncias para um vetor $(Y_i(s_l), Y_i(s_k), Y_j(s_l), Y_j(s_k))^T$ são dadas por:

$$\Sigma_Y = \begin{bmatrix} \sigma_{01}^2 + \sigma_1^2 & \sigma_{01}\rho_0(h) + \sigma_1^2\rho_1(h) & \sigma_{01}\sigma_{02} & \sigma_{01}\sigma_{02}\rho_0(h) \\ \sigma_{01}\rho_0(h) + \sigma_1^2\rho_1(h) & \sigma_{01}^2 + \sigma_1^2 & \sigma_{01}\sigma_{02}\rho_0(h) & \sigma_{01}\sigma_{02} \\ \sigma_{01}\sigma_{02} & \sigma_{01}\sigma_{02}\rho_0(h) & \sigma_{02}^2 + \sigma_2^2 & \sigma_{01}\rho_0(h) + \sigma_2^2\rho_2(h) \\ \sigma_{01}\sigma_{02}\rho_0(h) & \sigma_{01}\sigma_{02} & \sigma_{01}\rho_0(h) + \sigma_2^2\rho_2(h) & \sigma_{02}^2 + \sigma_2^2 \end{bmatrix}$$

Dessa forma, a distribuição de probabilidade do vetor $\mathbf{Y} = (Y_1, Y_2)$ está estabelecida e depende do vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}^*, \boldsymbol{\sigma}^{2*}, \phi^*_0, \phi^*_1, \phi^*_2)$, sendo $\boldsymbol{\beta}^*$ um vetor de parâmetros associados μ e ϕ^*_j é um parâmetro associado à escolha da função ρ_j , $j = 0, 1, 2$. Segundo Fonseca (2008), para a estimação dos parâmetros e a predição espacial pode-se utilizar as mesmas técnicas envolvidas nos modelos geoestatísticos univariados. Além disso, é necessário utilizar a reparametrização adicionais em $\boldsymbol{\theta}$ para encontrar a função de verossimilhança concentrada.

Considerando a Equação 23 a covariância entre Y_1 e Y_2 pode ser expressa como na Equação 24.

$$\begin{aligned} \text{Cov}(Y_1(x_i), Y_2(x_j)) &= \text{Cov}(\mu_1 + S_0(x_i) + S_1(x_i), \mu_2 + S_0(x_j) + S_1(x_j)) \Rightarrow \\ \text{Cov}(Y_1(x_i), Y_2(x_j)) &= \text{Cov}(S_0(x_i), S_0(x_j)) \end{aligned} \quad \text{Eq.(24)}$$

A Equação 24 pode ser escrita também como:

$$\text{Cov}(Y_1(x_i), Y_2(x_j)) = \sigma_{01}\sigma_{02}\rho(h)$$

3.8 Estimação dos parâmetros do modelo por máxima verossimilhança

O ajuste de modelo é feito inicialmente de forma visual, selecionando valores para o efeito pepita, patamar e alcance e em seguida consideram-se critérios estatísticos.

Considerando o caso estacionário do modelo geoestatístico univariado dado pela Equação 18, onde o processo $Z(s_i)$ pode ser escrito como um conjunto de observações de Y com distribuição de probabilidades de acordo com a Equação 25.

$$\mathbf{Y} \sim N(\mathbf{D}\boldsymbol{\beta}, \sigma^2\mathbf{R} + \tau^2\mathbf{I}) \quad \text{Eq.(25)}$$

Os parâmetros gerais do modelo a serem estimados são $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \phi, \tau^2)$ onde ϕ é um parâmetros da função de correlação.

Os dados $y = \{y_1, y_2, \dots, y_n\}$ que representam uma realização do processo estocástico espacial em n coordenadas, possuem distribuição gaussiana n -variada, isto é, $\mathbf{Y} \sim N_n(\boldsymbol{\mu}; \boldsymbol{\Sigma})$, onde $\boldsymbol{\mu}$ é o vetor de médias e $\boldsymbol{\Sigma}$ é a matriz de variâncias e covariâncias de dimensão $n \times n$. Segundo Dudewicz e Mishra (1988), a distribuição conjunta de \mathbf{Y} é descrita na Equação 26.

$$f_Y(y) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{Y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu})\right\} \quad \text{Eq.(26)}$$

Sendo \mathbf{Y} um vetor gaussiano correlacionado, sua função de verossimilhança é descrita na Equação 27.

$$L(\boldsymbol{\theta}) = f(y|\boldsymbol{\theta}) = \frac{(|\sigma^2\mathbf{R} + \tau^2\mathbf{I}|)^{-1/2}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}(\mathbf{Y} - \mathbf{D}\boldsymbol{\beta})' (|\sigma^2\mathbf{R} + \tau^2\mathbf{I}|)^{-1} (\mathbf{Y} - \mathbf{D}\boldsymbol{\beta})\right\} \quad \text{Eq.(27)}$$

O método de máxima verossimilhança (MARDIA; MASHALL, 1984) consiste em maximizar a função verossimilhança, isto é, obter o vetor de parâmetro $\boldsymbol{\theta}$ que maximize a função verossimilhança. Por simplicidade matemática, utiliza-se o logaritmo natural da função de verossimilhança dado na Equação 28.

$$l(\boldsymbol{\theta}) = -\frac{1}{2} \log(2\pi)^n - \frac{1}{2} \log(|\sigma^2\mathbf{R} + \tau^2\mathbf{I}|) - \frac{1}{2} (\mathbf{Y} - \mathbf{D}\boldsymbol{\beta})' (|\sigma^2\mathbf{R} + \tau^2\mathbf{I}|)^{-1} (\mathbf{Y} - \mathbf{D}\boldsymbol{\beta}) \quad \text{Eq.(28)}$$

Utilizando a reparametrização $\frac{\tau^2}{\sigma^2} = v^2$, proposta por Diggle e Ribeiro Jr. (2007), então $\text{Var}(Y) = \boldsymbol{\Sigma} = |\sigma^2\mathbf{R} + \tau^2\mathbf{I}| = \sigma^2 \left(\mathbf{R} + \frac{\tau^2}{\sigma^2}\mathbf{I}\right) = \sigma^2\mathbf{V}$, onde \mathbf{V} é uma matriz de dimensão $n \times n$.

Substituindo $\sigma^2\mathbf{R} + \tau^2\mathbf{I}$ por $\sigma^2\mathbf{V}$ na Equação 28, obtém-se a Equação 29.

$$l(\theta) = -\frac{1}{2} \left[n \log(2\pi) + \log(|\sigma^2\mathbf{V}|) + (\mathbf{Y} - \mathbf{D}\boldsymbol{\beta})' (\sigma^2\mathbf{V})^{-1} (\mathbf{Y} - \mathbf{D}\boldsymbol{\beta}) \right] \quad \text{Eq.(29)}$$

Agora substituindo $\sigma^2\mathbf{V}$ por Σ na Equação 29, obtém-se a Equação 30.

$$l(\theta) = -\frac{1}{2} \left[n \log(2\pi) + \log(|\Sigma|) + (\mathbf{Y} - \mathbf{D}\boldsymbol{\beta})' (\Sigma)^{-1} (\mathbf{Y} - \mathbf{D}\boldsymbol{\beta}) \right] \quad \text{Eq.(30)}$$

Desenvolvendo-se os produtos matriciais da Equação 30, tem-se que:

$$(\mathbf{Y} - \mathbf{D}\boldsymbol{\beta})' (\Sigma)^{-1} (\mathbf{Y} - \mathbf{D}\boldsymbol{\beta}) = \mathbf{Y}' (\Sigma)^{-1} \mathbf{Y} - \mathbf{Y}' (\Sigma)^{-1} \mathbf{D}\boldsymbol{\beta} - \boldsymbol{\beta}' \mathbf{D}' (\Sigma)^{-1} \mathbf{Y} + \boldsymbol{\beta}' \mathbf{D}' (\Sigma)^{-1} \mathbf{D}\boldsymbol{\beta}$$

Como $\mathbf{Y}' (\Sigma)^{-1} \mathbf{D}\boldsymbol{\beta} = \boldsymbol{\beta}' \mathbf{D}' (\Sigma)^{-1} \mathbf{Y}$ tem-se a Equação 31.

$$(\mathbf{Y} - \mathbf{D}\boldsymbol{\beta})' (\Sigma)^{-1} (\mathbf{Y} - \mathbf{D}\boldsymbol{\beta}) = \mathbf{Y}' (\Sigma)^{-1} \mathbf{Y} - 2\mathbf{Y}' (\Sigma)^{-1} \mathbf{D}\boldsymbol{\beta} + \boldsymbol{\beta}' \mathbf{D}' (\Sigma)^{-1} \mathbf{D}\boldsymbol{\beta} \quad \text{Eq.(31)}$$

Substituindo a Equação 31 na Equação 30, obtém-se a Equação 32.

$$l(\theta) = -\frac{1}{2} \left[n \log(2\pi) + \log(|\Sigma|) + \mathbf{Y}' (\Sigma)^{-1} \mathbf{Y} - 2\mathbf{Y}' (\Sigma)^{-1} \mathbf{D}\boldsymbol{\beta} + \boldsymbol{\beta}' \mathbf{D}' (\Sigma)^{-1} \mathbf{D}\boldsymbol{\beta} \right] \quad \text{Eq.(32)}$$

Em que $\mathbf{Y}' (\Sigma)^{-1} \mathbf{D}\boldsymbol{\beta}$ é um escalar com dimensão $\mathbf{Y}_{n \times 1}, \Sigma_{n \times n}, \mathbf{D}_{n \times n}, \boldsymbol{\beta}_{n \times 1}$.

Segundo Fonseca (2008), para fazer a estimação dos parâmetros e a predição espacial nos modelos geoestatísticos bivariado pode-se utilizar as mesmas técnicas envolvidas nos modelos geoestatísticos univariados com reparametrizações adequadas em $\boldsymbol{\theta}$ para encontrar a função de verossimilhança concentrada.

Segundo Kolman (1997), se \mathbf{A} é uma matriz quadrada simétrica definida positiva e $\mathbf{X} = [X_1, X_2, \dots, X_n]'$ um vetor, então:

$$\begin{aligned} \text{I)} \quad \frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} &= \mathbf{A}' \text{ (transposta)} \\ \text{II)} \quad \frac{\partial \mathbf{x}' \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} &= 2\mathbf{A}\mathbf{x} \text{ (forma quadrática)} \end{aligned}$$

Calculando a derivada parcial do logaritmo natural da função de verossimilhança de θ com relação a $\boldsymbol{\beta}$, e utilizando dos resultados de Kolman (1997), obtém-se a Equação 33.

$$\frac{\partial l(\theta)}{\partial \boldsymbol{\beta}} = -\frac{1}{2} \left[-2(\mathbf{Y}' \Sigma^{-1} \mathbf{D})' + 2(\mathbf{D}' \Sigma^{-1} \mathbf{D})\boldsymbol{\beta} \right] \Rightarrow \quad \text{Eq.(33)}$$

$$\frac{\partial l(\theta)}{\partial \beta} = \mathbf{D}'\Sigma^{-1}\mathbf{Y} - \mathbf{D}'\Sigma^{-1}\mathbf{D}\beta$$

Se $\frac{\partial l(\theta)}{\partial \beta} = 0$, então da Equação 33 obtém-se o estimador de máxima verossimilhança para o parâmetro β que é descrito na Equação 34.

$$\begin{aligned} \mathbf{D}'\Sigma^{-1}\mathbf{Y} - \mathbf{D}'\Sigma^{-1}\mathbf{D}\hat{\beta} &= \mathbf{0} \Rightarrow \\ \mathbf{D}'\Sigma^{-1}\mathbf{Y} &= \mathbf{D}'\Sigma^{-1}\mathbf{D}\hat{\beta} \\ \hat{\beta} &= \mathbf{D}'\Sigma^{-1}\mathbf{Y}(\mathbf{D}'\Sigma^{-1}\mathbf{D})^{-1} \end{aligned} \quad \text{Eq.(34)}$$

Considerando-se também que $\Sigma = \sigma^2\mathbf{V}$ e que $|\Sigma| = \sigma^2|\mathbf{V}|$, e desenvolvendo-se os produtos matriciais da Equação 34, tem-se a Equação 35.

$$l(\theta) = -\frac{1}{2} \left[n \log(2\pi) + \log(|\sigma^2\mathbf{V}|) + \mathbf{Y}'(\sigma^2\mathbf{V})^{-1}\mathbf{Y} - \beta'\mathbf{D}'(\sigma^2\mathbf{V})^{-1}\mathbf{D}\beta \right] \quad \text{Eq.(35)}$$

logo:

$$\begin{aligned} l(\theta) &= -\frac{1}{2} \left[n \log(2\pi) + \log[(\sigma^2)^n|\mathbf{V}|] + \frac{\mathbf{Y}'\mathbf{V}^{-1}\mathbf{Y}}{\sigma^2} - 2\frac{\mathbf{Y}'\mathbf{V}^{-1}\mathbf{D}\beta}{\sigma^2} + \frac{\beta'\mathbf{D}'\mathbf{V}^{-1}\mathbf{D}\beta}{\sigma^2} \right] \Rightarrow \\ l(\theta) &= -\frac{1}{2} \left[n \log(2\pi) + \log(\sigma^2) + \log|\mathbf{V}| + \frac{(\mathbf{Y} - \mathbf{D}\beta)'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{D}\beta)}{\sigma^2} \right] \end{aligned}$$

onde $\left[\frac{(\mathbf{Y} - \mathbf{D}\beta)'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{D}\beta)}{\sigma^2} \right]$ é uma soma de quadrados ponderada pela matriz de covariâncias.

Calculando-se a derivada de $l(\theta)$ com relação a σ^2 obtém-se a Equação 36.

$$\frac{\partial l(\theta)}{\partial \sigma^2} = -\frac{1}{2} \left[\frac{n}{\sigma^2} - \frac{(\mathbf{Y} - \mathbf{D}\beta)'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{D}\beta)}{(\sigma^2)^2} \right] \quad \text{Eq.(36)}$$

Se $\frac{\partial l(\theta)}{\partial \sigma^2} = 0$ e considerando o vetor de parâmetros $(\beta, \sigma^2, \phi, v^2)'$, tem-se a Equação

37.

$$\begin{aligned} \frac{\partial l(\theta)}{\partial \sigma^2} &= \left[\frac{n}{\sigma^2} - \frac{(\mathbf{Y} - \mathbf{D}\beta)'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{D}\beta)}{(\sigma^2)^2} \right] = 0 \Rightarrow \\ \hat{\sigma}^2 &= \left[\frac{(\mathbf{Y} - \mathbf{D}\beta)'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{D}\beta)}{n} \right] \end{aligned} \quad \text{Eq.(37)}$$

Retornando-se a Equação 34 e substituindo Σ por $\sigma^2\mathbf{V}$, obtém-se a Equação 38.

$$\begin{aligned}\hat{\beta} &= \left(\frac{\mathbf{D}'\mathbf{V}^{-1}\mathbf{D}}{\sigma^2} \right)^{-1} \frac{\mathbf{D}\mathbf{V}^{-1}\mathbf{Y}}{\sigma^2} \Rightarrow \\ \hat{\beta} &= (\mathbf{D}'\mathbf{V}^{-1}\mathbf{D})^{-1} \sigma^2 \frac{\mathbf{D}\mathbf{V}^{-1}\mathbf{Y}}{\sigma^2} \Rightarrow \\ \hat{\beta} &= (\mathbf{D}'\mathbf{V}^{-1}\mathbf{D})^{-1} \mathbf{D}\mathbf{V}^{-1}\mathbf{Y}\end{aligned}\quad \text{Eq.(38)}$$

Que depende dos parâmetros \emptyset e \mathbf{v}^2 . Neste caso, a matriz de correlação é dada na Equação 39.

$$\mathbf{V} = \begin{pmatrix} 1 + \mathbf{v}^2 & \rho(u_{12}) & \dots & \rho(u_{1n}) \\ \rho(u_{21}) & 1 + \mathbf{v}^2 & \dots & \rho(u_{n2}) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(u_{n1}) & \rho(u_{n2}) & \dots & 1 + \mathbf{v}^2 \end{pmatrix}\quad \text{Eq.(39)}$$

O logaritmo natural da função verossimilhança concentrada é dada pela Equação 40.

$$\begin{aligned}l(\theta) &= -\frac{1}{2} \left[n \log(2\pi) + n \log \left(\frac{(\mathbf{Y} - \mathbf{D}\beta)' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{D}\beta)}{n} \right) + \log|\mathbf{V}| \right. \\ &\quad \left. + \frac{(\mathbf{Y} - \mathbf{D}\beta)' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{D}\beta)}{\frac{(\mathbf{Y} - \mathbf{D}\beta)' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{D}\beta)}{n}} \right]\end{aligned}\quad \text{Eq.(40)}$$

Assim,

$$\begin{aligned}l(\emptyset, \mathbf{v}^2) &= -\frac{1}{2} \left[n \log(2\pi) + n \log \left(\frac{(\mathbf{Y} - \mathbf{D}\beta)' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{D}\beta)}{n} \right) + \log|\mathbf{V}| + n \right] \Rightarrow \\ l(\emptyset, \mathbf{v}^2) &= -\frac{1}{2} \left[n \log(2\pi) + n \log \left((\mathbf{Y} - \mathbf{D}\beta)' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{D}\beta) \right) - n \log n + \log|\mathbf{V}| + n \right]\end{aligned}$$

Agora, considerando o caso estacionário do modelo geoestatístico bivariado dado pela Equação 20, onde o processo $Z(s_i)$ pode ser escrito como um conjunto de observações de Y com distribuição de probabilidades gaussiana n -variada, ou seja, $\mathbf{Y} \sim N_n(\boldsymbol{\mu}; \boldsymbol{\Sigma})$. A distribuição conjunta de \mathbf{Y} é semelhante ao caso univariado (Equação 25). Sendo Y um vetor gaussiano correlacionado, sua função de verossimilhança é composta pela distribuição conjunta de probabilidades e por simplicidade matemática, utiliza-se o logaritmo natural da função de verossimilhança descrita na Equação 41.

$$l(\theta) = -\frac{1}{2} \left[n \log(2\pi) + \log(|\Sigma_Y|) + (\mathbf{Y} - \boldsymbol{\mu}_Y)' (\Sigma_Y)^{-1} (\mathbf{Y} - \boldsymbol{\mu}_Y) \right] \quad \text{Eq.(41)}$$

Considerando duas localizações amostradas para o modelo geoestatístico bivariado com componente de correlação parcialmente comum a matriz de covariância Σ_Y , é dada pelas Equações 42, 43 e 44.

$$\Sigma_1 = \sigma_{01}^2 \mathbf{R}_0 + \sigma_1^2 \mathbf{R}_1 \quad \text{Eq.(42)}$$

$$\Sigma_{1,2} = \sigma_{01} \sigma_{02} \mathbf{R}_0 \quad \text{Eq.(43)}$$

$$\Sigma_2 = \sigma_{02}^2 \mathbf{R}_0 + \sigma_2^2 \mathbf{R}_2 \quad \text{Eq.(44)}$$

Utilizando as seguintes reparametrizações $\sigma = \sigma_{01}, \eta = \frac{\sigma_{02}}{\sigma_{01}}, t_1 = \frac{\sigma_1}{\sigma_{01}}$ e $t_2 = \frac{\sigma_2}{\sigma_{02}}$, tem-se as Equações 45, 46 e 47.

$$\begin{aligned} \Sigma_1 &= \sigma^2 (\mathbf{R}_0 + t_1^2 \mathbf{R}_1) \Rightarrow \\ \Sigma_1 &= \sigma^2 \mathbf{V}_1 \end{aligned} \quad \text{Eq.(45)}$$

Em que \mathbf{V}_1 é uma matriz $n_1 \times n_1$, em que cada elemento dessa matriz pode ser escrito como $v_1 = \rho_0(h) + t_1^2 \rho_1(h)$.

$$\begin{aligned} \Sigma_{1,2} &= \sigma^2 \eta \mathbf{R}_0 \Rightarrow \\ \Sigma_{1,2} &= \sigma^2 \mathbf{V}_{1,2} \end{aligned} \quad \text{Eq.(46)}$$

Em que $\mathbf{V}_{1,2}$ é uma matriz $n_1 \times n_2$, em que cada elemento dessa matriz pode ser escrito como $v_{1,2} = \rho_0(h)$.

$$\begin{aligned} \Sigma_2 &= \sigma^2 \eta^2 (\mathbf{R}_0 + t_2^2 \mathbf{R}_2) \Rightarrow \\ \Sigma_2 &= \sigma^2 \eta^2 \mathbf{V}_2 \end{aligned} \quad \text{Eq.(47)}$$

Em que \mathbf{V}_2 é uma matriz $n_2 \times n_2$, em que cada elemento dessa matriz pode ser escrito como $v_2 = \rho_0(h) + t_2^2 \rho_2(h)$.

Assim a equação do modelo geoestatístico bivariado com componente de correlação parcialmente comum é dado na Equação 48.

$$\Sigma_Y = \sigma^2 V_{BGCCM} = \sigma^2 \begin{bmatrix} V_1 & \eta V_{1,2} \\ \eta V_{1,2}' & \eta^2 V_2 \end{bmatrix} \quad \text{Eq.(48)}$$

Assim o logaritmo natural da função de verossimilhança é descrito como sendo que $\Sigma_Y = \sigma^2 V_{BGCCM}$.

Os parâmetros $\widehat{\sigma}^2$ e $\widehat{\beta}$ do modelo bivariado são obtidos de forma semelhante ao caso univariado. Além disso, a função de verossimilhança concentrada também tem expressão semelhante ao caso univariado, mudando apenas a matriz V e os parâmetros na qual a função é concentrada.

3.9 Critérios de validação de modelo

O ajuste de modelos não é uma tarefa simples, por isso na escolha do modelo além do conhecimento do comportamento do fenômeno e do critério visual, são utilizados critérios estatísticos, como validação cruzada e critério de informação de Akaike.

A validação cruzada é uma técnica de avaliação de erros de estimativa que permite comparar valores estimados e amostrados, tendo como informação a amostra de dados de que se dispõe. Isto é, consiste em retirar da amostra temporariamente uma localização do conjunto de dados amostrais e o valor da mesma localização é estimado por krigagem usando os valores das amostras restantes (ISAAKS; SRIVASTAVA, 1989). Calculada a estimativa, pode-se compará-la ao valor da amostra que foi temporariamente removida do conjunto de dados amostrais. Repete-se este procedimento para todas as demais amostras disponíveis

Conhecendo todos os valores amostrados e estimados, pode-se avaliar a qualidade do modelo ajustado uma vez que é possível identificar o erro médio (EM) – descrito na Equação 49, desvio padrão do erro médio (S_{EM}) – descrito na Equação 50, erro reduzido (ER) – descrito na Equação 51, desvio padrão do erro reduzido (S_{ER}) – descrito na Equação 52 e o erro absoluto (EA) – descrito na Equação 53 (FARACO et al., 2008). Segundo Mello et al. (2005), o melhor modelo é o que fornece o valor de EM e ER mais próximo de zero, o valor de S_{EM} menor e o valor de S_{ER} mais próximo de um. Também é conveniente avaliar o erro absoluto médio (EA), que deve ser o menor possível.

$$EM = \frac{1}{n} \sum_{i=1}^n (Z(s_i) - \widehat{Z}(s_i)) \quad \text{Eq.(49)}$$

$$S_{EM} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Z(s_i) - \widehat{Z}(s_i))^2} \quad \text{Eq.(50)}$$

$$ER = \frac{1}{n} \sum_{i=1}^n \frac{(Z(s_i) - \hat{Z}(s_i))}{\sigma(\hat{Z}(s_i))} \quad \text{Eq.(51)}$$

$$S_{ER} = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{|Z(s_i) - \hat{Z}(s_i)|}{\sigma(\hat{Z}(s_i))}} \quad \text{Eq.(52)}$$

$$EA = \sum_{i=1}^n |(s_i) - \hat{Z}(s_i)| \quad \text{Eq.(53)}$$

O critério de informação de Akaike (AIC), descrito na Equação 54, procura uma solução satisfatória e a simplicidade do modelo.

$$AIC = -2(\theta) + 2k \quad \text{Eq.(54)}$$

sendo que θ é o logaritmo natural da função máxima verossimilhança e k é o número de parâmetros do modelo ajustado. A escolha para o melhor ajuste é feita com base no modelo que apresentar menor valor de AIC.

O Critério de informação Bayesiana (BIC), descrito na Equação 55.

$$BIC = -2 \times (\theta) + k \log(n) \quad \text{Eq.(55)}$$

em que, θ é o logaritmo natural da função máxima verossimilhança, k é o número de parâmetros, n é o número de observações. Menor valor do AIC e BIC indica o melhor ajuste do modelo (OLINDA, 2010).

3.10 Krigagem e Cokrigagem

Uma etapa importante da geoestatística é a construção de mapas de variabilidade espacial, o que é possível por meio de métodos de interpolação.

Em geoestatística, o método de interpolação por krigagem tem sido utilizado para estimar valores de uma variável em locais não amostrados a partir de valores amostrais vizinhos (VIEIRA, 2000). Possibilitando a construção de mapas temáticos, os quais são possíveis fazer o mapeamento do solo, auxiliando desta forma na otimização na produção, de forma a minimizar custos e maximizar lucros.

Ortiz (2002) ressalta que a identificação da estrutura da variabilidade espacial fornece um melhor conhecimento do comportamento da variável na área em estudo e possibilita um melhor planejamento das futuras amostragens.

De acordo com Guedes (2008), os mapas de produtividade permitem identificar as localizações de áreas críticas, em termos de rentabilidade, enquanto que mapas dos atributos do solo e do relevo podem auxiliar na identificação dos fatores que estão limitando a produtividade. Segundo Camargo (1998), a krigagem fornece estimadores exatos com propriedades de não tendenciosidade e eficiência.

Câmara e Medeiros (1998) afirmam que a krigagem ordinária (Equação 55) é baseada na ideia de regressão linear, onde a estimativa de um valor desconhecido $Z(s_0)$ é obtida pela combinação linear de n valores conhecidos $Z(s_i)$.

$$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i) \quad \text{Eq.(55)}$$

onde $\hat{Z}(s_0)$ é o valor estimado para a localização s_0 não amostrada, n é o número de pontos amostrais e λ_i são os pesos associados a cada valor conhecido da variável em locais amostrado $Z(s_i)$.

Os pesos de krigagem devem satisfazer a condição de estimador não tendencioso e com variâncias mínimas, isto é, a esperança do erro seja zero e a variância do erro seja mínima, assim tem-se que (Equação 56):

$$\begin{aligned} E[\hat{Z}(s_0) - Z(s_0)] &= 0 \Rightarrow \\ E\left[\sum_{i=1}^n \lambda_i Z(s_i) - Z(s_0)\right] &= 0 \Rightarrow \\ \mu \sum_{i=1}^n \lambda_i - \mu &= 0 \end{aligned} \quad \text{Eq.(56)}$$

A Equação 56 é satisfeita se e somente se:

$$\sum_{i=1}^n \lambda_i = 1 \quad \text{Eq.(57)}$$

Além da esperança do erro ser zero, a variância da estimativa deve ser mínima, ou seja,

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

Da Equação 57 tem-se que:

$$\begin{aligned} \text{Var}[\hat{Z}(s_0) - Z(s_0)] &= E[(\hat{Z}(s_0) - Z(s_0))^2] \Rightarrow \\ \text{Var}[\hat{Z}(s_0) - Z(s_0)] &= E\left[\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j Z(s_i) Z(s_j) + E[Z(s_0)] - 2 \sum_{i=1}^n E[\lambda_i Z(s_i) \cdot Z(s_0)]\right] \Rightarrow \\ \text{Var}[\hat{Z}(s_0) - Z(s_0)] &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \text{Cov}(Z(s_i) Z(s_j)) + \text{Var}\left[Z(s_0) - 2 \sum_{i=1}^n \lambda_i \text{Cov}[Z(s_i) \cdot Z(s_0)]\right] \end{aligned} \quad \text{Eq.(58)}$$

Minimizar $\text{Var}[\hat{Z}(s_0) - Z(s_0)] \Rightarrow \frac{\partial}{\partial \lambda_i} \text{Var}[\hat{Z}(s_0) - Z(s_0)]$. Assim os pesos λ_i são obtidos a partir do sistema de equações descritas na Equação 59.

$$\begin{cases} \sum_{i=1}^n \lambda_i \gamma(s_i, s_j) + \sigma = \gamma(s_i, s_0) \\ \sum_{i=1}^n \lambda_i = 1 \end{cases} \quad \text{Eq.(59)}$$

onde σ é um multiplicador de Lagrange, introduzido para minimizar a variância do erro. Esse sistema de equações pode ser escrito da forma matricial como na Equação 60.

$$[\boldsymbol{\gamma}] \times [\boldsymbol{\alpha}] = [\boldsymbol{\delta}] \quad \text{Eq.(60)}$$

onde $[\boldsymbol{\gamma}]$ é a matriz das semivariâncias entre os pares de pontos amostrados, $[\boldsymbol{\alpha}]$ é o vetor de pesos e $[\boldsymbol{\delta}]$ é o vetor das semivariâncias da amostra em relação ao ponto a ser estimado, isto é:

$$[\boldsymbol{\gamma}] = \begin{bmatrix} \gamma(s_1, s_1) & \gamma(s_1, s_2) & \dots & \gamma(s_1, s_n) & 1 \\ \gamma(s_2, s_1) & \gamma(s_2, s_2) & \dots & \gamma(s_2, s_n) & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ \gamma(s_n, s_1) & \gamma(s_n, s_2) & \dots & \gamma(s_n, s_n) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix}$$

$$[\boldsymbol{\alpha}] = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \sigma \end{bmatrix}$$

$$[\delta] = \begin{bmatrix} \gamma(s_1, s_0) \\ \gamma(s_2, s_0) \\ \vdots \\ \gamma(s_n, s_n) \\ 1 \end{bmatrix}$$

Em relação aos ponderadores do estimador de krigagem, Soares (2000) afirma que quanto mais próximas estiverem as amostras do ponto a estimar, maior será o seu peso no estimador. Além disso, quanto mais correlacionadas estiverem as amostras, maior o efeito de agrupamento ou redundância e menor será o seu peso individual na construção do estimador.

O método de interpolação considerando duas variáveis é a cokrigagem ordinária. Righetto (2012) afirma que este interpolador é útil quando duas ou mais variáveis são amostradas nos mesmos pontos amostrais, dentro de um mesmo domínio espacial e apresentam significativo grau de correlação entre elas.

A cokrigagem considera uma variável principal Z_2 e uma variável secundária Z_1 . Suponha que se deseja estimar o valor de Z_2 na posição s_0 , por uma combinação linear das variáveis Z_1 e Z_2 , descrita na Equação 61.

$$\hat{Z}_2(s_0) = \sum_{i=1}^{n1} \lambda_{1i} Z_1(s_i) + \sum_{i=1}^{n2} \lambda_{2j} Z_2(s_i) \quad \text{Eq.(61)}$$

onde λ_{1i} e λ_{2i} são pesos associados a cada valor de Z_1 e Z_2 , respectivamente. Da mesma forma da krigagem os pesos na cokrigagem devem satisfazer a condição de estimador não tendencioso e com variâncias mínimas. Assim, tem-se que:

$$\begin{aligned} E[\hat{Z}_2(s_0) - Z_2(s_0)] &= 0 \Rightarrow \\ E\left\{\sum_{i=1}^{n1} \lambda_{1i} Z_1(s_{1i}) + \sum_{i=1}^{n2} \lambda_{2j} Z_2(s_{2j}) - \hat{Z}_2(s_0)\right\} &= 0 \Rightarrow \\ \sum_{i=1}^{n1} \lambda_{1i} E[Z_1(s_{1i})] + \sum_{i=1}^{n2} \lambda_{2j} E[Z_2(s_{2j})] - E[\hat{Z}_2(s_0)] &= 0 \Rightarrow \\ m_1 \sum_{i=1}^{n1} \lambda_{1i} + m_2 \sum_{i=1}^{n2} \lambda_{2j} - m_2 &= 0 \end{aligned}$$

Ou,

$$m_1 \sum_{i=1}^{n_1} \lambda_{1i} + m_2 \left(\sum_{i=1}^{n_2} \lambda_{2j} - 1 \right) = 0 \quad \text{Eq.(62)}$$

A Equação 62 é satisfeita se e somente se:

$$\sum_{i=1}^{n_1} \lambda_{1i} = 0 \quad \text{e} \quad \sum_{i=1}^{n_2} \lambda_{2j} = 1 \quad \text{Eq.(63)}$$

A variância de estimação:

$$\begin{aligned} \text{Var}[\hat{Z}_2(s_0) - Z_2(s_0)] &= E \left[(\hat{Z}_2(s_0) - Z_2(s_0))^2 - \{E[\hat{Z}_2(s_0) - Z_2(s_0)]\}^2 \right] = \\ &= \text{Var} \left\{ \sum_{i=1}^{n_1} \lambda_{1i} Z_1(s_i) + \sum_{j=1}^{n_2} \lambda_{2j} Z_2(s_j) - Z_2(s_0) \right\} = \\ &= \left\{ \sum_{i=1}^{n_1} \sum_{i=1}^{n_1} \lambda_{1i} \lambda_{1j} \text{Cov}(Z(s_i)Z(s_j)) + \sum_{j=1}^{n_2} \sum_{j=1}^{n_2} \lambda_{2i} \lambda_{2j} \text{Cov}(Z(s_i)Z(s_j)) \right. \\ &\quad + 2 \sum_{i=1}^{n_2} \sum_{j=1}^{n_2} \lambda_{1i} \lambda_{2j} \text{Cov}(Z(s_i)Z(s_j)) \\ &\quad - 2 \sum_{i=1}^{n_2} \lambda_{1i} \text{Cov}(Z(s_i)Z(s_0)) \\ &\quad \left. - 2 \sum_{j=1}^{n_2} \lambda_{2j} \text{Cov}(Z(s_j)Z(s_0)) + \text{Cov}(Z(s_0)Z(s_0)) \right\} \end{aligned} \quad \text{Eq.(64)}$$

Da mesma forma que na krigagem, na cokrigagem os pesos λ_{1i} e λ_{2j} são obtidos a partir do sistema de equações descrito na Equação 65.

$$\left\{ \begin{array}{l} \sum_{i=1}^{n_1} \lambda_{1i} \gamma_{12}(s_{1i}, s_{1k}) + \sum_{i=1}^{n_2} \lambda_{2j} \gamma_{12}(s_{1k}, s_{2j}) - \sigma_1 = \gamma_{12}(s_{1k}, s_0) \\ \sum_{i=1}^{n_1} \lambda_{1i} \gamma_{12}(s_{1i}, s_{21}) + \sum_{i=1}^{n_2} \lambda_{2j} \gamma_{12}(s_{1k}, s_{2j}) - \sigma_2 = \gamma_{22}(s_{21}, s_0) \\ \sum_{i=1}^{n_1} \lambda_{1i} = 0 \\ \sum_{j=1}^{n_2} \lambda_{2j} = 1 \end{array} \right. \quad \text{Eq.(65)}$$

onde σ_1 e σ_2 são os multiplicadores de Lagrange, onde $k = 1, \dots, n_1$ e $l = 1, \dots, n_2$.

Este sistema de equações também pode ser escrito na forma matricial, como no caso de krigagem, e resolve-se de forma similar.

Carvalho e Assad (2003) incorporaram a altitude como variável auxiliar na determinação do mapa de variabilidade espacial na precipitação pluvial para o Estado de São Paulo por meio da comparação de dois interpoladores geoestatísticos multivariados (cokrigagem ordinária e cokrigagem co-locada). Os pesquisadores estudaram 1027 observações de precipitação pluvial anual média proveniente de estações climáticas, representando uma área de aproximadamente 248808,8 km (2,91% do território nacional), no período de 1957 a 1997, e concluíram que o uso de altitude como variável auxiliar beneficia os dois interpoladores estudados; entretanto, na cokrigagem ordinária ela auxilia numa maior uniformidade da distribuição espacial.

3.11 Acurácia em mapas temáticos

A partir do ajuste de modelo e da krigagem é possível construir mapas temáticos, os quais descrevem a variabilidade espacial da variável em estudo para toda a área de interesse. Construídos os mapas, é possível compará-los a fim detectar semelhanças e diferenças. Uma maneira de comparar os mapas é por meio da matriz de erros (KESTRING et al., 2011).

A matriz de erros apresentada na Tabela 1 é uma matriz $m \times m$, em que i representa as linhas, j as colunas e m número total de classes. Essa matriz descreve a precisão da classificação e caracteriza os erros, isto é, quantificam os pixels do mapa de referência e os do mapa modelo. Cada elemento da matriz representa a quantidade de pixels pertencente à classe i do mapa modelo e à classe j do mapa referência. A diagonal principal, isto é, quando $i = j$ representa os pixels que foram classificados por ambos os mapas, indicando o nível de concordância (FIGUEIREDO, 2007), enquanto que os elementos fora da diagonal principal representam as classificações distintas dos mapas. Assim, se os mapas forem idênticos, na classificação dos pixels, os elementos fora da diagonal principal da matriz de erros serão nulos.

Tabela 1 Matriz genérica dos erros

Mapa modelo	Mapa de referência					Total	
	1	2	.	.	.		m
1	n_{11}	n_{12}	.	.	.	n_{1m}	$n_{1\cdot}$
2	n_{21}	n_{22}	.	.	.	n_{2m}	$n_{2\cdot}$
.
.
.
m	n_{m1}	n_{m2}	.	.	.	n_{mm}	$n_{m\cdot}$
Total	$n_{\cdot 1}$	$n_{\cdot 2}$.	.	.	$n_{\cdot m}$	n

Uma medida que pode ser extraída da matriz de erros é a exatidão global (Equação 66).

$$EG = \frac{\sum_i^m n_{ii}}{n} \quad \text{Eq.(66)}$$

sendo m o número de classes, n o número total de pixels e n_{ii} número de pixels da diagonal principal. A exatidão global quantifica a similaridade entre os dois mapas; além disso, $EG > 0,85$ indica precisão entre os mapas.

Outra medida que pode ser extraída da matriz de erros é o índice Kappa (k_p) (Equação 67).

$$k_p = \frac{n \sum_i^m n_{ii} - \sum_i^m n_{i,j} \cdot n_{j,i}}{n^2 - \sum_i^m n_{i,j} \cdot n_{j,i}} \quad \text{Eq.(67)}$$

Sendo n_i os elementos da linha i e n_j os elementos da coluna j . Segundo Kripendorff (1980), $k_p < 0,67$ indica baixa exatidão, $0,67 \leq k_p < 0,8$ média exatidão e $k_p \geq 0,8$ alta exatidão entre os mapas temáticos construídos.

Outra maneira de avaliar a acurácia é por meio do índice Tau (T) (Equação 68).

$$T = \frac{EG - \frac{1}{m}}{1 - \frac{1}{m}} \quad \text{Eq.(68)}$$

Caon e Genú (2013) estabeleceram uma densidade amostral que ofereça precisão nos mapas de atributos químicos gerados sobre diferentes densidades amostrais, em duas profundidades, e mensurar sua implicação prática na adubação e calagem para a cultura da soja, na região de Guarapuava. Compararam mapas para pH, H+Al, P, K, Ca, Mg, V% e a necessidade de calagem, nas camadas 0-10 e 0-20 cm, nas densidades amostrais de $\frac{1}{4}$, $\frac{1}{2}$, 1, 2 e 4 ha, sendo a primeira densidade tomada como referência de comparação. Os modos de comparação foram a exatidão global e o índice Kappa. Concluíram que a diminuição da densidade amostral depreciou a qualidade dos mapas gerados. Todas as densidades amostrais diferiram entre si quanto à exatidão, porém foram encontrados bons níveis de exatidão até a densidade amostral de uma amostra por hectare.

Figueiredo e Vieira (2007) analisaram o comportamento dos índices de exatidão global, Kappa e Tau, e para isso criaram várias matrizes de confusão (erros) hipotéticas, variando primeiramente o número de classes para informacionais de cada classificação, em seguida variando o total de amostras utilizado para gerar a matriz de confusão e, por fim,

variando o total de pixels e discrepando os valores amostrados, para validação, entre as classes informacionais para verificar o comportamento dos três índices dentro de cinco níveis de desempenho de classificação. Concluíram que as diferenças entre os três índices não são constantes ao longo dos níveis de desempenho da classificação. No geral, a diferença entre os índices Kappa e Tau foi muito pequena. Embora a exatidão global apresente um valor mais alto, os coeficientes de concordância Kappa e Tau são mais consistentes por envolver no valor final todas as células da matriz de confusão.

3.12 Amostragem espacial

Tendo em vista as dificuldades para observar todos os elementos da população, de modo especial custo e tempo, torna-se viável e praticável o uso de apenas uma parte da população denominada amostra. Os elementos que compõem a amostra são chamados de unidades amostrais.

Unidade amostral ou unidade experimental é a menor proporção da área experimental. Segundo Banzatto e Kronka (1989), a escolha da unidade experimental ou parcela deve ser orientada de forma a minimizar o erro experimental.

Amostragem é o método de seleção das unidades amostrais. No entanto, na seleção das unidades amostrais é necessário garantir que a amostra seja representativa da população. A menos de certas discrepâncias inerentes à aleatoriedade, a amostra deve possuir as mesmas características básicas da população no que diz respeito à(s) variável(eis) que deseja pesquisar (NETO, 2002).

Dungan et al. (2002) descrevem três componentes amostrais como essenciais para a percepção da estrutura espacial: tamanho da unidade amostral, que é a superfície ou volume em que a variável de interesse será mensurada; a forma da unidade amostral e a distância entre unidades amostrais. Os dois primeiros componentes, tamanho e forma, têm influência direta sobre a variável mensurada, pois mudanças no tamanho e na forma da unidade amostral criam uma nova variável, induzindo a alterações na variância e na estrutura de continuidade espacial. O último componente, distância entre as unidades amostrais, tem influência sobre a vizinhança (ODA-SOUZA, 2010).

3.13 Tipos de amostragem

Uma ampla variedade de metodologias de amostragem tem sido desenvolvida a fim de propiciar amostras de menor custo em diversos contextos de pesquisa. A escolha do método de amostragem depende do propósito do experimento, das características da população em estudo, características da variável, tempo e recursos econômicos (NETO, 2002).

Alguns métodos de amostragem espacial são: irregular ou aleatória simples, estratificada, regular ou sistemática, *lattice plus close pairs* e *lattice plus in-fill*.

Malha amostral irregular é um método de amostragem na qual os elementos são selecionados por acaso (Figura 2). Os pontos x_i são selecionados de forma independentemente da distribuição uniforme da área.

Segundo Webster e Oliver (2007), a amostragem irregular é a forma mais simples de amostragem. Cada unidade (elemento) da amostra é escolhida sem relação a qualquer outra e, além disso, todas as unidades têm a mesma chance de seleção.

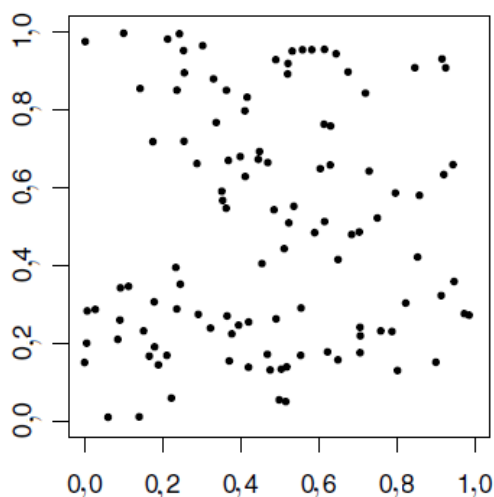


Figura 2 Malha amostral irregular.

Na malha amostral estratificada os elementos são selecionados em grupos, sendo que em cada grupo os elementos tendem a ter características similares (Figura 3). Além disso, os grupos entre si são heterogêneos.

As subdivisões (estratos) são tipicamente pequenos quadrados, mas podem ter outras formas, de igual área. Pelo menos dois pontos de amostragem são escolhidos de forma aleatória dentro de cada estrato. Uma forma de estratificar a região para melhorar a precisão das estimativas é dividi-la em função de determinados atributos. Essa prática é muito comum em estudos do solo (WEBSTER; OLIVER, 2007).

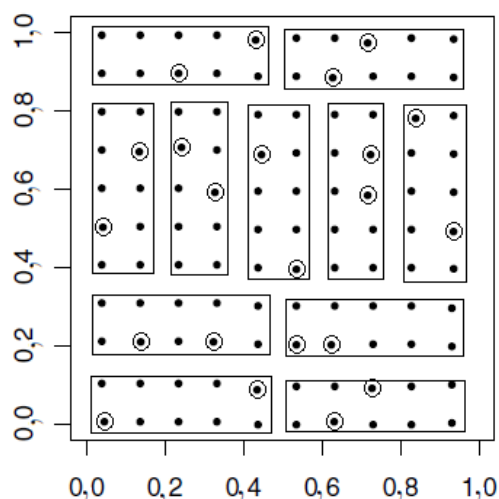


Figura 3 Malha amostral estratificada.

Na amostragem sistemática ou regular, escolhe-se um ponto amostral inicial aleatoriamente e, em seguida, especificam-se segundo um padrão regular as $n - 1$ localizações dos pontos amostrais. Exemplos de amostragem sistemática regular são: malha quadrada, malha retangular, malha triangular e malha hexagonal.

Na malha quadrada os pontos amostrais x_i selecionados são equidistantes, de forma que as linhas e colunas imaginárias sejam ortogonais (Figura 4-a). A equidistância entre os pontos é determinada pelo tamanho da área, objetivo da pesquisa e a experiência do pesquisador.

Guedes (2008) e Milne (1959) apontam a malha amostral quadrada como amostragem sistemática centrada, uma vez que a área experimental é dividida em quadrados iguais e a unidade amostral estará no centro do quadrado.

Milne (1959), ao comparar as amostragens aleatória e estratificada com a amostragem sistemática centrada produz uma melhor estimativa para a média, menor erro padrão, intervalo de confiança para a média populacional mais eficiente.

Na malha amostral retangular os pontos amostrais x_i selecionados são regularmente espaçados, sendo que as linhas e colunas formam retângulos imaginários (Figura 4-b).

Em experimentação agrícola, Banzatto e Kronka (1989) afirmam que as parcelas retangulares são mais eficientes na superação da heterogeneidade do solo quanto seu eixo maior está na direção de maior variação do solo.

Webster e Oliver (2007) acrescentam que as malhas retangulares são mais práticas e a perda de precisão, em comparação com malhas triangulares, é normalmente tão pequena que elas são preferidas.

Na malha amostral triangular os pontos amostrais x_i selecionados são os vértices dos triângulos equiláteros formados por linhas imaginárias (Figura 4-c).

Na malha amostral hexagonal, os pontos amostrais x_i selecionados são os vértices dos hexágonos formados por linhas imaginárias (Figura 4-d). Segundo Cressie (1993), a malha hexagonal regular pode ser obtida a partir da malha triangular equilátera excluindo os pontos apropriados. Por outro lado, a malha triangular equilátera é obtida a partir da malha hexagonal regular pela soma dos pontos do centro dos hexágonos.

Yfantis et al. (1987) acrescentam que a malha hexagonal contém apenas dois terços da quantidade de dados da malha triangular. Assim, a confiabilidade do semivariograma estimada a partir de uma malha hexagonal não é tão boa como a triangular.

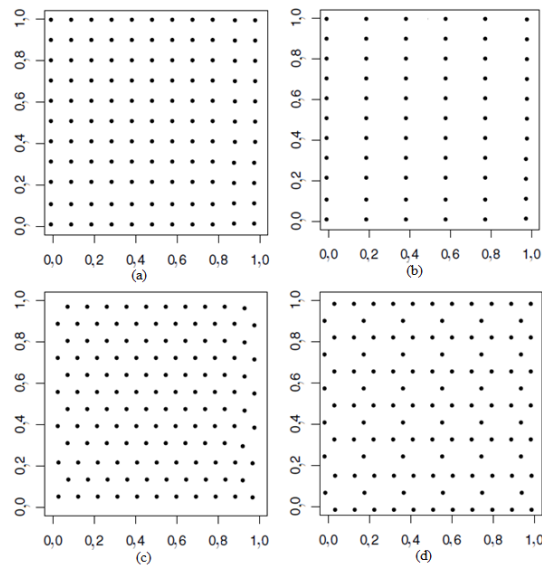


Figura 4 Malha amostral sistemática regular.

Quenouille (1949) generaliza e amplia as formas de se definir a configuração amostral, empregando combinações nas duas direções da área de interesse das amostragens: irregular, regular e estratificada, surgindo, assim, diversos tipos de amostragem. Assim pode-se obter, por exemplo, amostragem aleatória em uma direção e sistemática em outra. Considera, ainda, se as amostras são alinhadas ou desalinhadas em uma ou ambas as direções.

Na Figura 5-a é apresentado um exemplo de amostragem irregular na direção x e na direção y amostragem regular. Já na Figura 5-b é apresentado um exemplo de amostragem alinhada em ambas as direções, isto é, amostragem regular nas direções x e y .

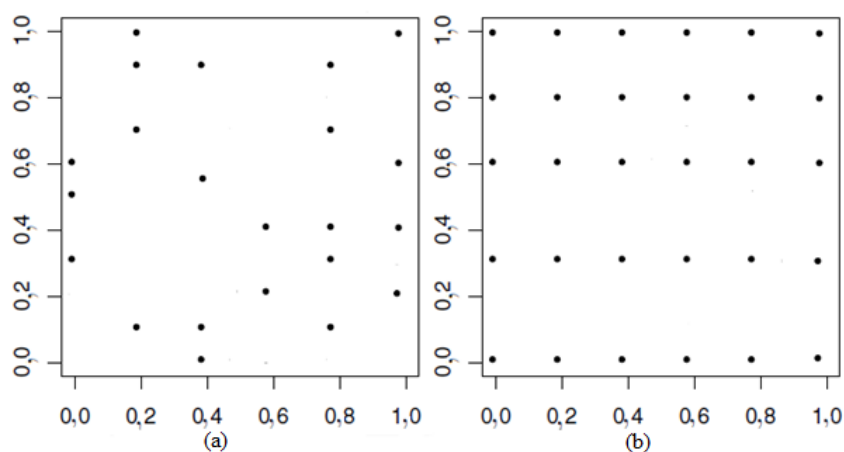


Figura 5 Malha amostral alinhada em uma direção (a) e alinhada em ambas as direções (b).

A malha amostral *lattice plus close pairs* é um método de amostragem sistemática centrada com pares de pontos próximos (Figura 6). Segundo Diggle e Ribeiro Jr. (2007), essa amostragem consiste em localizações regulares $k \times k$ com espaçamento Δ acrescentando m pontos, os quais estão localizados de forma uniformemente aleatória dentro de uma circunferência de raio $\delta = \alpha\Delta$, cujo centro será um ponto da amostragem regular, escolhido aleatoriamente.

De Bastiani et al. (2012) realizaram um estudo sobre a variabilidade espacial de dados de resistência do solo à penetração (RSP) nas camadas 0,0-0,1 m, 0,1-0,2 m e 0,2-0,3 m de profundidade, utilizando métodos estatísticos em forma univariada. Os dados referem-se à amostragem sistemática centrada com pares de pontos próximos (*lattice plus close pairs*), com distância máxima de 141 m entre pontos e em alguns locais aleatórios; a amostragem foi realizada com distâncias a 75 e 50 m entre pontos. Neste estudo, verificaram que pelos métodos geoestatísticos foi possível detectar que a RSP nas três camadas analisadas apresentaram graus de compactação mais elevado na região norte da área estudada. Por meio dos mapas temáticos construídos foi possível diagnosticar as áreas onde é necessário aplicar técnicas para amenizar o problema de compactação do solo. A construção de mapas temáticos utilizando modelo linear espacial (LSM) foi utilizado para realizar estimativas da RSP de 0,2- 0,3 m de profundidade em função das camadas anteriores. As medidas de acurácia permitiram realizar a comparação entre o mapa temático univariado utilizando krigagem ordinária e o mapa temático utilizando o LSM com krigagem universal. Os valores dos índices mostraram existir discordância entre os mapas univariado e o mapa que utiliza o LSM tendo como covariáveis a RSP de camadas anteriores.

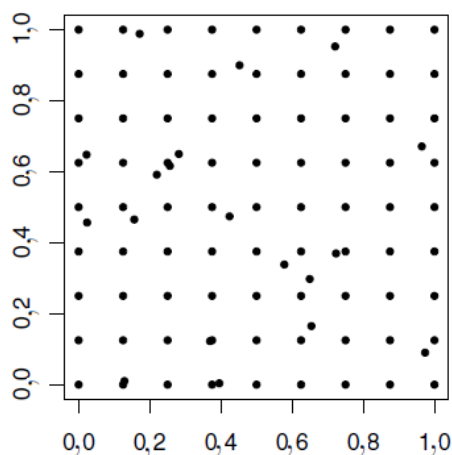


Figura 6 Malha amostral *lattice plus close pairs*($8 \times 8, 3, 4 \times 4$).

A malha *lattice plus in-fill* é um método de amostragem sistemática centrada com amostragens sistemáticas centradas menores (Figura 7). Segundo Diggle e Ribeiro Jr. (2007), essa amostragem consiste em localizações regulares $k \times k$ com espaçamento Δ acrescentando m malhas amostrais regulares menores $r \times r$, adicionando assim $r^2 - 4$ localizações adicionais. Para esta malha é usada a seguinte notação: $(k \times k, m, r \times r)$. Na Figura 7 é apresentado um exemplo de amostragem *lattice plus in-fill* ($8 \times 8, 3, 4 \times 4$).

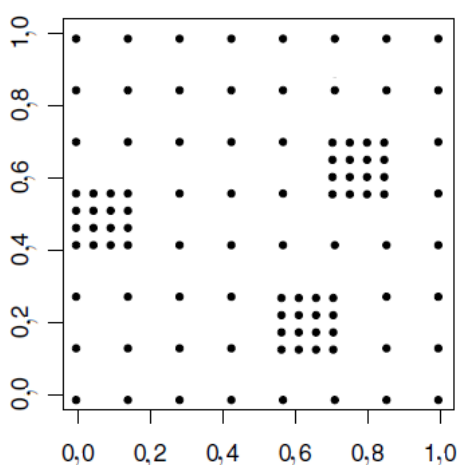


Figura 7 Malha amostral *lattice plus in-fill*.

3.14 Tamanho amostral

O tamanho amostral leva em consideração as variáveis que se deseja pesquisar, o grau de precisão desejado pelo pesquisador, custos e o tamanho populacional.

Segundo Banzatto e Kronka (1989), o objetivo do trabalho experimental também influencia no tamanho da parcela (unidades amostrais), por exemplo, estudar o efeito da profundidade de semeadura do sorgo granífero sobre o desenvolvimento inicial das plantas, não necessita trabalhar com parcelas tão grandes quanto as que seriam necessárias para um estudo de produção da cultura.

É natural pensar que grandes amostras oferecem maior exatidão das características da área em estudo, isto é, quanto maior o tamanho amostral maior será a aproximação alcançada. Por outro lado, amostras grandes elevam os custos da pesquisa, além de demandar mais tempo. No entanto, o tamanho amostral deve ser baseado de modo que os desvios padrão das amostras se aproximem do desvio correspondente de toda área.

Storck et al. (2006) afirmam que ao considerar amostras independentes, recomenda-se que o número de unidades experimentais seja, no mínimo, igual a 20, para se obter uma precisão razoável em experimentos de campos.

Coelho et al. (2009) avaliaram a influência da densidade amostral de três grades com: 128, 64 e 32 parcelas e do tipo de interpolador krigagem na elaboração de mapas temáticos da produtividade da soja. Os autores calcularam o coeficiente de desvio relativo (CDR) para expressar a diferença média dos valores interpolados em cada grade em relação à grade considerada como referência (128 parcelas). O CDR mostrou-se eficiente na comparação de mapas e possibilitou concluir que, para utilizar todo o potencial da krigagem e esse método ser recomendado, é necessário ter muitos pontos para que um bom semivariograma seja construído. Além disso, a diminuição do número de amostras correspondeu a aumento no CDR e, portanto, a mapas mais dissimilares em relação à referência. A amostragem que obteve o melhor resultado utilizou o interpolador inverso da distância.

3.15 Espaçamento amostral

O espaçamento ideal entre amostras é o que fornece boa precisão da variabilidade espacial das variáveis pesquisadas. Assim é natural pensar que em espaçamentos menores tem-se maior representatividade da variabilidade espacial. Entretanto, o custo da coleta e a análise das amostras, poderá inviabilizar o trabalho.

Uribe-Opazo et al. (2007) pesquisaram a estrutura de dependência espacial de um conjunto de dados reais, referentes à produtividade da soja, ajustando modelos teóricos a semivariogramas experimentais construídos com diferentes grades amostrais de 50 x 50, 75 x 75 e 100 x 100 m. De acordo com os dados, concluíram que a maior dificuldade no ajuste é devida a estimativas incorretas dos valores da função semivariância, relacionadas com poucas amostras ou baixa densidade amostral, visto que a configuração amostral de 100 x 100 m por hectare foi insuficiente para que os semivariogramas experimentais evidenciassem a estrutura de dependência espacial. Vale ressaltar que os autores consideraram as grades amostrais $n \times n$ como malha regular quadrada, na quais duas localizações consecutivas possuem distância n nas direções 0 e 90°.

Assim é natural pensar que o tamanho e o tipo de amostragem têm influência na avaliação da correlação espacial entre as variáveis. Desta forma, a escolha do tamanho e das configurações amostrais para o estudo da variabilidade espacial é de suma importância,

uma vez que a disposição espacial pode ser considerada significativa ou insignificante podendo induzir relações errôneas entre as variáveis.

Diferentes escolhas na estrutura espacial, tais como: espaçamento da grade amostral, tamanho e forma da unidade amostral e valores de *cutoff* podem induzir a diferentes resultados, que certamente irão influenciar no ajuste de modelo e na construção do mapa temático.

Estudo relacionado à escolha da configuração amostral mais eficiente pode ser definida como um problema de otimização, que consiste em buscar o menor tamanho amostral para essa configuração, sem prejudicar a eficiência da predição espacial, e consequentemente minimizar os custos do processo de coleta das amostras. Ruiz-Cárdenas et al. (2009) utilizaram essa metodologia para problemas de racionalização de estações de rede de monitoramento ambiental.

Guedes, et al. (2011) relatam que quando o tamanho amostral é grande, métodos que buscam a melhor configuração amostral pelo estudo de todas as possíveis combinações de soluções são inviáveis computacionalmente e, portanto, uma alternativa de busca por melhores configurações amostrais são as estratégias metaheurísticas de busca de soluções ótimas denominadas algoritmo de recozimento simulado e algoritmo genético. Os pesquisadores propuseram um algoritmo genético híbrido para redimensionar um conjunto amostral de 256 pontos da variável teor de potássio, reduzindo o tamanho amostral em 50%, mas minimizando as perdas que possam ser geradas na acurácia dos resultados relativos à predição espacial. Eles concluíram que a amostragem otimizada pelo algoritmo genético híbrido apresentou as menores estimativas para a média da variância da predição, inclusive em relação ao valor dessa estimativa obtida pelo conjunto de dados inicial. Além disso, as estimativas do número e da porcentagem de valores preditos pertencentes a cada intervalo da classificação para determinação da quantidade de produto a ser aplicado em cada localização amostral foram mais próximas aos valores obtidos no conjunto de dados inicial.

4 MATERIAL E MÉTODOS

4.1 Análise da influência da variação dos parâmetros do modelo bivariado BGCCM no cálculo do coeficiente de correlação linear de Pearson

Para avaliar a influência da variação dos parâmetros do modelo bivariado BGCCM no cálculo do coeficiente de correlação linear de Pearson foram simuladas 10 amostras para três configurações de parâmetros, sendo que em cada amostra foi realizado 100 simulações. Ainda em cada amostra foi gerado 100 dados do vetor $\mathbf{Y} = (Y_1, Y_2)^T$ e suas respectivas localizações espaciais, sendo que os dados gerados pertencem a uma área regular com limites dos eixos x e y variando de 0 a 100 u.m. e uma malha regular de localizações (Figura 8). Para realizar as simulações foi adotada a função Matérn com k igual a 2. Uma vez que essa função apresentou melhor erro médio, desvio padrão do erro médio, erro relativo, desvio padrão do erro relativo e erro absoluto no estudo com dados reais.

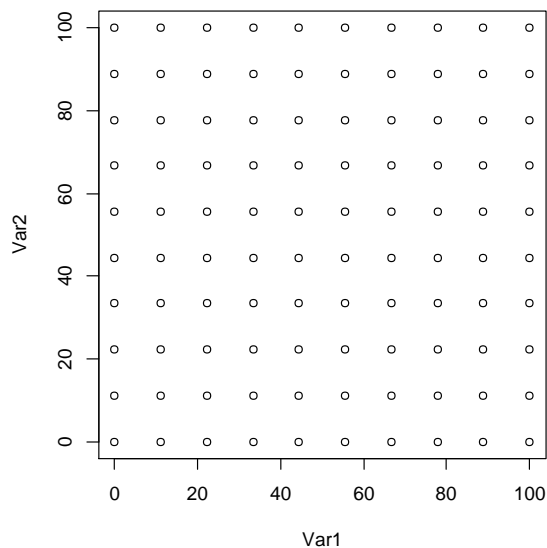


Figura 8 Exemplo de malha regular.

O primeiro ensaio foi realizado considerando os parâmetros de média e alcance fixos em $\mu_1 = 10, \mu_2 = 50, \phi_0 = \phi_1 = \phi_2 = 20$. Logo os alcances práticos são aproximadamente iguais a 108 u.m. e variado os parâmetros de dispersão (σ). Sendo que no primeiro momento os parâmetros σ_{01}, σ_1 e σ_{02} foram considerados fixos, em 16 e variado σ_2 em 100%, 75%, 50%, 25% e 0% do valor de $\sigma_{01} = \sigma_1 = \sigma_{02}$, totalizando 5 amostras. No segundo momento os parâmetros σ_{01}, σ_1 e σ_{02} foram considerados fixos, em 4 e variado σ_2 também em 100%, 75%, 50%, 25% e 0% do valor de $\sigma_{01} = \sigma_1 = \sigma_{02}$, totalizando 5 amostras (Tabela 3).

Tabela 2 Parâmetros que definem a estrutura de dependência espacial bivariada, considerando o modelo BGCCM

Amostras	σ_{01}	σ_1	σ_{02}	σ_2
a1	16	16	16	16
a2	16	16	16	12
a3	16	16	16	8
a4	16	16	16	4
a5	16	16	16	0
a6	4	4	4	4
a7	4	4	4	3
a8	4	4	4	2
a9	4	4	4	1
a10	4	4	4	0

Nas duas situações do primeiro ensaio, tem-se que: (a) em uma das variáveis (Y_1) existe a mesma influência dos dois componentes que definem a sua estrutura de dependência espacial; (b) ambas as variáveis tem mesma estrutura de dependência espacial, quanto ao seu componente comum e (c) o decréscimo de σ_2 foi gerado com o intuito de diminuir a influência do componente independente que define a estrutura de dependência espacial em uma das variáveis (Y_2), o que gera como consequência um aumento da influência do componente espacial comum às duas variáveis.

O segundo ensaio foi realizado considerando os parâmetros de média, dispersão e alcance de Y_1 e Y_2 fixos em $\mu_1 = 10$, $\mu_2 = 50$, $\phi_1 = 20$, $\phi_2 = 20$, $\sigma_{01} = 5$, $\sigma_1 = 2$, $\sigma_{02} = 5$ e $\sigma_2 = 2$, logo $\text{Var}(Y_1) = \sigma_{01}^2 + \sigma_1^2 = 29$ e $\text{Var}(Y_2) = \sigma_{02}^2 + \sigma_2^2 = 29$. E foi variado o alcance (ϕ_0), isto é, o raio de dependência espacial da estrutura comum das duas variáveis em 30, 27, 25, 20, 18, 15, 12, 10, 5 e 3, totalizando 10 amostras.

O terceiro ensaio foi realizado considerando os parâmetros de média e dispersão fixos em $u_1 = 10$, $u_2 = 50$, $\sigma_{01} = 5$, $\sigma_1 = 2$, $\sigma_{02} = 5$ e $\sigma_2 = 2$, logo $\text{Var}(Y_1) = \sigma_{01}^2 + \sigma_1^2 = 29$ e $\text{Var}(Y_2) = \sigma_{02}^2 + \sigma_2^2 = 29$. E variado o alcance ϕ_0 , ϕ_1 e ϕ_2 , simultaneamente, em 30, 25, 20, 18, 15, 12, 10, 8, 5 e 3, totalizando 10 amostras.

Para cada simulação, em cada configuração de parâmetros, calculou-se o coeficiente de correlação linear de Pearson entre Y_1 e Y_2 . Em seguida foi obtida a estatística descritiva dos coeficientes de correlação linear de Pearson de cada configuração e foram construídos os gráficos Boxplots. Além disso, para cada configuração, foram obtidas as porcentagens de simulações que pertencem à classificação de Dancey e Reidy (2007), em que $(r_{Y_1, Y_2}) \leq 0,4$ indica fraca associação linear, $0,4 < (r_{Y_1, Y_2}) < 0,7$ moderada e $(r_{Y_1, Y_2}) \geq 0,7$ forte associação linear.

4.2 Comparação de modelos

4.2.1 Estudo com dados simulados

Nessa fase do trabalho foi comparada a predição espacial obtida pelo modelo univariado, semivariograma cruzado e o modelo bivariado BGCCM. Para isso, foram simulados conjuntos de dados bivariados baseado em processos espaciais gaussianos e com estruturas de dependência espaciais conhecidas, considerando o modelo BGCCM. Além disso, foram pré-determinados que os conjuntos de dados possuíssem distribuição normal multivariada de probabilidade.

Em cada amostra foram feitas 100 simulações do vetor $\mathbf{Y} = (Y_1, Y_2)^T$, sendo que a determinação de cada amostra simulada foi realizada para uma área regular com limites dos eixos x e y variando de 0 a 100 u.m., com configuração amostral denominada *lattice plus close pairs*, sendo 81 pontos de localizações regulares (9x9) e 19 pontos acrescentados, os quais estão localizados de forma uniformemente aleatória dentro de uma circunferência de raio igual a 10 (Figura 9). Além disso, as variáveis Y_1 e Y_2 são co-locadas e balanceadas.

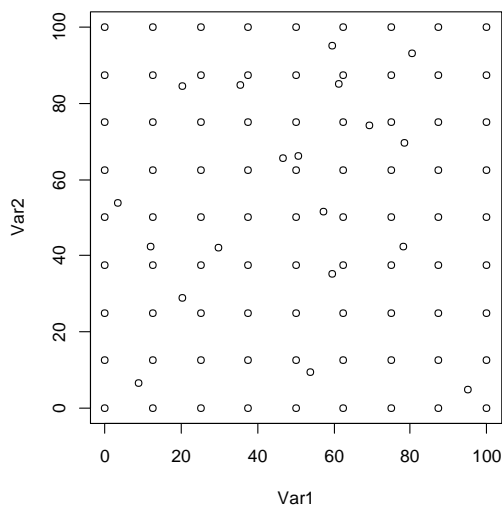


Figura 9 Exemplo de uma simulação.

Para realizar as simulações foi adotada a função Matérn com k igual a 2, uma vez que, essa função apresentou melhor erro médio, desvio padrão do erro, erro relativo, desvio padrão do erro relativo e erro absoluto no estudo com dados reais. Além disso, os parâmetros de média, alcance e dispersão foram considerados fixos em $\mu_1 = 10, \mu_2 = 50, \phi_0 = \phi_1 = \phi_2 = 18$ e $\sigma_{01} = \sigma_1 = \sigma_{02} = \sigma_2 = 3$, logo $\text{Var}(Y_1) = \sigma_{01}^2 + \sigma_1^2 = 18$ e $\text{Var}(Y_2) = \sigma_{02}^2 + \sigma_2^2 = 18$. Utilizando essa configuração paramétrica tem-se que os campos aleatórios possuem a mesma variabilidade e os parâmetros escolhidos para estruturar as correlações geram alcances práticos de aproximadamente 96 u.m., ou seja, 68% da área

total, indicando um valor alto do raio de dependência espacial na estrutura comum a Y_1 e Y_2 e também nas estruturas espaciais independentes de Y_1 e Y_2 .

Obtidas as estimativas dos parâmetros do modelo univariado e modelo bivariado BGCCM pelo método de máxima verossimilhança e semivariograma cruzado pelo método de mínimos quadrados ordinários foram realizadas as predições espaciais para Y_1 e Y_2 em localizações não amostradas para uma malha regular com limites dos eixos x e y variando de 0 a 100 u.m., totalizando 10201 pontos preditos. Para a predição espacial considerando o modelo univariado foi utilizado a krigagem ordinária e para as predições espaciais feitas considerando o semivariograma cruzado e modelo bivariado BGCCM foi utilizada a cokrigagem ordinária.

Para comparar os modelos geoestatísticos univariados e bivariados, quanto à predição espacial, foram calculadas as somas quadradas da diferença da predição espacial, medidas de acurácia: índices de exatidão global e índices de concordância: Kappa e Tau. Os índices de acurácia foram baseados na matriz de erros considerando 10 classes de intervalos de mesma amplitude.

4.2.2 Estudo com dados experimentais

Nessa etapa do desenvolvimento deste trabalho foram aplicadas as técnicas geoestatísticas a dados experimentais, provenientes de uma propriedade agrícola.

Os dados são provenientes de uma pesquisa realizada em uma área agrícola, no município de Cascavel, cuja localização geográfica é, aproximadamente, 24,95° Sul de latitude, 53,37° Oeste de longitude e altitude média de 650 m. Os dados são referentes ao ano agrícola de 2010/2011 e 2013/2014, com área total de 167,35 ha, com experimentos conduzidos por pesquisadores do grupo de pesquisa do Laboratório de Estatística Espacial (LEE) da Universidade Estadual do Oeste do Paraná – UNIOESTE, *campus* de Cascavel.

O banco de dados foi realizado sob uma amostragem sistemática centrada com pares de pontos próximos (*lattice plus close pairs*), com distância máxima de 141 m entre pontos. Em alguns locais, escolhidos de forma aleatória, a amostragem foi realizada com distâncias menores: 75 e 50 m entre pontos, obtendo-se 102 pontos amostrais (Figura 10). Todas as amostras foram georreferenciadas e localizadas com auxílio de um aparelho receptor de sinal com o sistema de posicionamento global (GPS) GEOEXPLORE 3, num sistema espacial de coordenadas UTM.

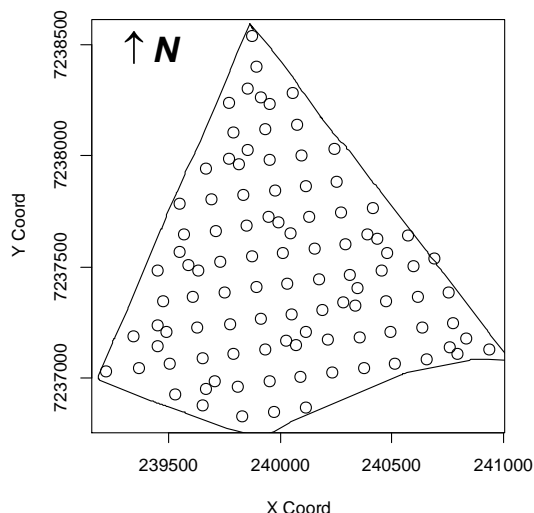


Figura 10 Área experimental e grade de amostragem.

Em cada ponto da área experimental coletaram-se quatro subamostra de solo que após misturada deu origem a amostra de cada local.

As amostras foram encaminhadas a COODETEC onde se realizaram as análises químicas do solo, sendo cobre (Cu), zinco (Zn), ferro (Fe), manganês (Mn), fósforo (P), carbono (C), pH, alumínio trocável ($H+Al_3$), cálcio (Ca), magnésio (Mg), alumínio (Al) e potássio (K). No Laboratório de Controle de Qualidade de Produtos Agrícolas (LACOM) da Universidade Estadual do Oeste do Paraná – UNIOESTE, *campus* de Cascavel, realizaram-se as análises físicas do solo, sendo, umidade (%), resistência do solo a penetração (Kpa) e densidade global (g/cm^3) nas profundidades de 0-10 cm, 10-20 cm e 20-30 cm. A resistência do solo a penetração foi medida com o auxílio de um medidor automatizado de compactação do solo, modelo PLG5500 da marca FALKER.

Para a seleção das variáveis no estudo de modelos geoestatísticos univariados e bivariados, foram levadas em consideração as variáveis que apresentaram maiores valores do coeficiente de correlação linear de Pearson.

Além disso, coletou-se uma parcela da soja para determinar a produtividade ($t\ ha^{-1}$). A produtividade foi estimada considerando a quantidade de grãos de soja colhida em cada parcela, sendo que cada parcela possui $1\ m^2$. Após a triagem, realizaram-se a pesagem de grãos para cada parcela e verificando o teor de água para posterior correção para 13%. O valor obtido foi convertido para $t\ ha^{-1}$.

4.2.2.1 Análise para dados co-locados e balanceados

Os dados experimentais do ano agrícola 2010/2011 possuem 102 localizações e o ano agrícola 2013/2014 em 100 localizações, pois foram perdidas duas amostras a respeito

da produtividade. As amostras, para ambos os anos agrícolas, são balanceadas e colocadas.

As variáveis selecionadas para o ano agrícola 2010/2011 foram Cálcio (Ca), Magnésio (Mg) e Manganês (Mn), por apresentarem os maiores valores do coeficiente de correlação linear de Pearson com o pH.

Para o ano agrícola 2013/2014 foram selecionadas as variáveis resistência do solo a penetração nas camadas de 0-10 cm (RSP0-10), 11-20 cm (RSP11-20), 21-30 cm (RSP21-30) e o Manganês (Mn), por apresentarem os maiores valores do coeficiente de correlação linear de Pearson com a produtividade (Prod).

Primeiramente realizou-se a análise estatística univariada dos dados co-locados e balanceados. Para a análise estatística exploratória dos dados experimentais foram utilizadas as medidas da estatística descritiva, teste de normalidade, gráfico *Boxplot* e Histograma, a fim de avaliar o comportamento geral das variáveis e detectar possíveis pontos discrepantes. Para a análise espacial dos dados, por meio das técnicas geoestatísticas, foram construídos os gráficos *Postplot* a fim de identificar a estrutura de dependência espacial e detectar a existência de padrão espacial das variáveis em estudo.

Foram também elaborados os semivariogramas experimentais para cada uma das variáveis e estimado os parâmetros do modelo teórico que descreve a estrutura de dependência espacial. Os modelos teóricos utilizados foram: esférico, exponencial, gaussiano e Matérn com k igual 1; 1,5 e 2, sendo que os parâmetros foram estimados pelo método de máxima verossimilhança. Para a escolha do melhor modelo teórico utilizou-se os critérios de validação cruzada de modelos, sendo eles EM, S_{EM} , ER, S_{ER} , EA, AIC e BIC (FARACO et al., 2008).

Obtidas as estimativas dos parâmetros do modelo geoestatístico univariado foi realizada a predição espacial por meio da krigagem ordinária, para cada variável pesquisada, em cada ano agrícola, em localizações não amostradas para uma malha regular com limites dos eixos x e y variando de 0 a 100, totalizando 66935 pontos preditos.

Para a análise geoestatística bivariada dos dados co-locados e balanceados diversos modelos teóricos foram estimados para o vetor $\mathbf{Y} = (Y_1, Y_2)^T$, sendo eles: esférico, exponencial, gaussiano e Matérn com k igual 1; 1,5 e 2. A análise geoestatística bivariada foi feita seguindo duas metodologias. Na primeira metodologia construiu-se um semivariograma cruzado para cada par de variáveis e os parâmetros da função semivariância desse semivariograma cruzado foram estimados pelo método de mínimos quadrados ordinários. Na segunda metodologia, estimaram-se os parâmetros do modelo geoestatístico bivariado BGCCM, pelo método de máxima verossimilhança. Para avaliar a qualidade das estimativas do modelo foram utilizados os critérios de validação cruzada citados anteriormente.

Obtidas as estimativas dos parâmetros dos modelos bivariados foi realizada a predição espacial por meio da cokrigagem ordinária para cada variável, em localizações não amostradas para uma malha regular com limites dos eixos x e y variando de 0 a 100, totalizando 66935 pontos preditos, considerando a mesma malha utilizada na análise univariada dos dados experimentais dos anos agrícolas 2010/2011 e 2013/2014.

A fim de comparar os modelos geoestatísticos univariados e bivariados dos dados co-locados e balanceados, em relação à predição espacial, foram obtidas as somas quadradas da diferenças entre as predições, medidas de acurácia: índices de exatidão global e índices de concordância: Kappa e Tau. Os índices de acurácia foram baseados na matriz de erros considerando 10 classes de intervalos de mesma amplitude.

4.2.2.2 Análise para dados co-locados e desbalanceados

Para o estudo de amostras co-locadas e desbalanceadas foram omitidas 50 localizações da variável Y_2 (RSP0-10, RSP11-20, RSP21-30 e Mn) dos dados experimentais do ano agrícola 2013/2014, de forma aleatória (Figura 11). No entanto, a variável Y_1 (Prod) permaneceu com as 100 localizações.

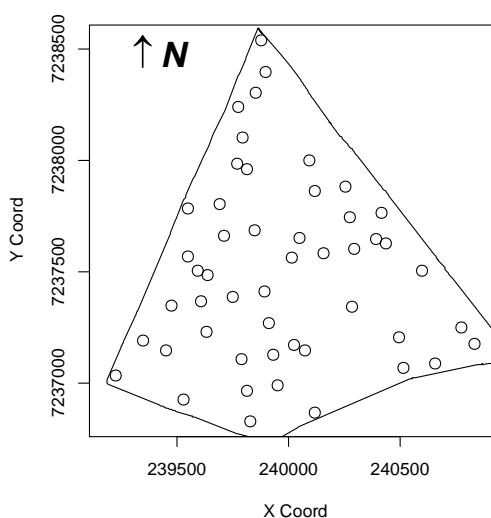


Figura 11 Localizações restantes.

Primeiramente realizou-se a análise univariada dos dados co-locados e desbalanceados. Para a análise estatística exploratória dos dados experimentais foram utilizadas as medidas das estatísticas descritivas, teste de normalidade, gráfico *Boxplot* e Histograma a fim de avaliar o comportamento geral das duas variáveis e detectar possíveis pontos discrepantes. Para a análise espacial dos dados por meio das técnicas geoestatísticas, foram construídos os gráficos *Postplot* a fim de identificar a estrutura de dependência espacial e detectar a existência de padrão espacial das variáveis em estudo.

Foram também elaborados os semivariogramas experimentais para cada uma das variáveis e estimado os parâmetros do modelo teórico que descreve a estrutura de dependência espacial. Os modelos teóricos utilizados foram: esférico, exponencial, gaussiano e Matérn com k igual 1; 1,5 e 2, sendo que os parâmetros foram estimados pelo método de máxima verossimilhança. Para a escolha do melhor modelo teórico, utilizou-se os critérios de validação cruzada de modelos, sendo eles EM, S_{EM} , ER, S_{ER} , EA, AIC e BIC (FARACO et al., 2008).

Obtidas as estimativas dos parâmetros do modelo geoestatístico univariado foi realizada a predição espacial por meio da krigagem ordinária para cada variável pesquisada, em localizações não amostradas para uma malha regular com limites dos eixos x e y variando de 0 a 100, totalizando 66935 pontos preditos.

Para a análise geoestatística bivariada dos dados co-locados e desbalanceados diversos modelos teóricos foram estimados para o vetor $\mathbf{Y} = (Y_1, Y_2)^T$, sendo eles: esférico, exponencial, gaussiano e Matérn com k igual 1; 1,5 e 2. Estimaram-se os parâmetros do modelo geoestatístico bivariado BGCCM, pelo método de máxima verossimilhança. Para avaliar a qualidade das estimativas do modelo foram utilizados os critérios de validação cruzada citados anteriormente.

Obtidas as estimativas dos parâmetros dos modelos bivariados foi realizada a predição espacial por meio da cokrigagem ordinária para cada variável, em localizações não amostradas para uma malha regular com limites dos eixos x e y variando de 0 a 100, totalizando 66935 pontos preditos, considerando a mesma malha utilizada na análise univariada dos dados experimentais co-locados e balanceados dos anos agrícolas 2010/2011 e 2013/2014.

A fim de comparar os modelos geoestatísticos univariados e bivariados dos dados co-locados e desbalanceados, em relação à predição espacial, foram obtidas as somas quadradas das diferenças entre as predições, medidas de acurácia: índices de exatidão global e índices de concordância: Kappa e Tau. Os índices de acurácia foram baseados na matriz de erros considerando 10 classes de intervalos de mesma amplitude.

Além disso, para avaliar a qualidade da predição espacial considerando os modelos geoestatísticos estimados pelas abordagens univariada e bivariada, foi realizada a predição espacial para as 50 localizações omitidas inicialmente, e calculadas as somas quadradas do erro de predição espacial.

4.3 Comparação de malhas

Nessa fase do trabalho foi observado o comportamento de três malhas amostrais, sendo elas: regular (r) (Figura 12-a), irregular (i) (Figura 12-b) e *lattice plus close pairs* (L) (Figura 12-c) sob a abordagem do modelo geoestatístico bivariado BGCCM, em relação à predição espacial, considerando conjuntos de dados simulados.

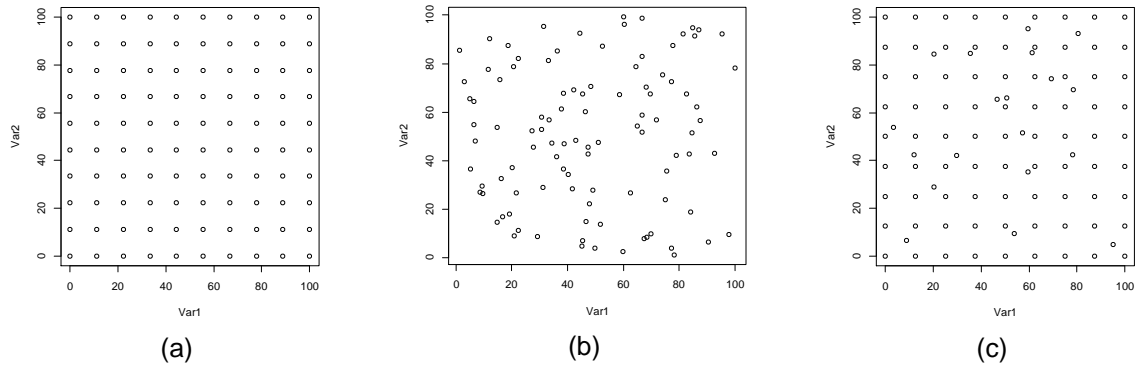


Figura 12 Malhas amostrais: regular (a), irregular (b) e *lattice plus close pairs* (c).

As simulações dos conjuntos de dados bivariados foram baseadas em processos espaciais gaussianos e com estruturas de dependência espaciais conhecidas, considerando o modelo BGCCM. Para isso foram pré-determinados que os conjuntos de dados possuíssem distribuição normal de probabilidade. Além disso, foi adotada a função Matérn com k igual a 2, uma vez que, essa função apresentou melhor erro médio, desvio padrão do erro médio, erro relativo, desvio padrão do erro relativo e erro absoluto no estudo com dados reais.

A determinação dos conjuntos de dados simulados foi realizada para três malhas amostrais sendo elas: regular, irregular e *lattice plus close pairs* com tamanhos amostrais iguais a 100. Para cada configuração amostral foram feitas 100 simulações do vetor $\mathbf{Y} = (Y_1, Y_2)^T$ e suas respectivas localizações. Os dados foram obtidos dentro de uma área regular com limites dos eixos x e y variando de 0 a 100 u.m.

Num primeiro momento cada conjunto amostral de Y_1 e Y_2 foi simulado balanceado e co-locado, ou seja, possuem o mesmo número de observações e foram tomadas nas mesmas localizações. No segundo momento cada conjunto amostral possui 100 dados para a variável Y_1 e 50 para a variável Y_2 , isto é, as amostras são desbalanceadas, no entanto, co-locadas.

Para cada simulação e configuração amostral foram considerados os parâmetros média, dispersão e alcance sempre fixos em $u_1 = 10, u_2 = 50$, $\varphi_0 = \varphi_1 = \varphi_2 = 18$ e $\sigma_{01} = \sigma_1 = \sigma_{02} = \sigma_2 = 3$, logo $\text{Var}(Y_1) = \sigma_{01}^2 + \sigma_1^2 = 18$ e $\text{Var}(Y_2) = \sigma_{02}^2 + \sigma_2^2 = 18$. Utilizando essa configuração paramétrica tem-se que os campos aleatórios possuem a mesma variabilidade e os parâmetros escolhidos para estruturar as correlações geram alcances práticos de aproximadamente 96 u.m., ou seja, 68% da área total, indicando um valor alto do raio de dependência espacial na estrutura comum a Y_1 e Y_2 e também nas estruturas espaciais independentes de Y_1 e Y_2 .

Além disso, para cada simulação e cada configuração amostral, os parâmetros do modelo BGCCM foram estimados pelo método de máxima verossimilhança. Em seguida, foram obtidas as predições espaciais para localizações não amostradas, utilizando a

cokrigagem ordinária. Sendo que a predição espacial foi realizada para uma malha regular com limites dos eixos x e y variando de 0 a 100 u.m, totalizando 10201 pontos preditos.

Com o intuito de comparar a predição espacial das três configurações amostrais, considerando todas as simulações, foram obtidas as estatísticas descritivas para as somas quadradas da diferença da predição espacial, medidas de acurácia: índices de exatidão global, Kappa e Tau e construídos os gráficos *Boxplots* para essas medidas.

Ainda com o intuito de avaliar a qualidade da predição espacial das três configurações amostrais foram simulados 144 dados do vetor $\mathbf{Y} = (Y_1, Y_2)^T$ e suas respectivas localizações sob os mesmos parâmetros iniciais. Cada amostra foi dividida em uma subamostra de modelagem e outra subamostra de controle. O grupo de modelagem foi utilizado para fazer a estimação dos parâmetros do modelo BGCCM e a predição dos valores das duas variáveis nas 44 localizações da subamostra de controle pela cokrigagem ordinária. Dessa forma o grupo de controle foi utilizado para avaliar a qualidade da predição espacial, usando como medidas as somas quadradas do erro da predição espacial e as medidas de acurácia: índices de exatidão global e índices de concordância: Kappa e Tau. Os índices de acurácia foram baseados na matriz de erros considerando 10 classes de intervalos de mesma amplitude.

Para a realização das simulações, análise estatística das medidas geradas e análise geoestatística de variáveis provenientes de modelos geoestatísticos univariados e bivariados foi utilizado o software R (R Development Core Team, 2006), utilizando o pacote *geoR* (RIBEIRO JR.; DIGGLE, 2001).

5 RESULTADOS E DISCUSSÕES

5.1 Análise da influência da variação dos parâmetros do modelo bivariado BGCCM no cálculo do coeficiente de correlação linear de Pearson

Para avaliar a influência da variação dos parâmetros no cálculo do coeficiente de correlação linear de Pearson foram obtidas as estatísticas descritivas (Tabelas 6 e 8) e construídos os gráficos *Boxplot* (s 13, 14 e 15), para cada configuração.

Para o primeiro ensaio observou-se que mesmo alterando a estrutura de dependência espacial bivariada, num primeiro momento considerando alta dispersão e num segundo baixa dispersão para σ_{01} , σ_1 e σ_2 , tem-se um comportamento similar entre as amostras, observando alta variabilidade do coeficiente de correlação linear de Pearson (Figura 13). Além disso, observa-se comportamento similar entre as porcentagens de amostras segundo classificação de Dancey e Reidy (2007) (Tabela 5).

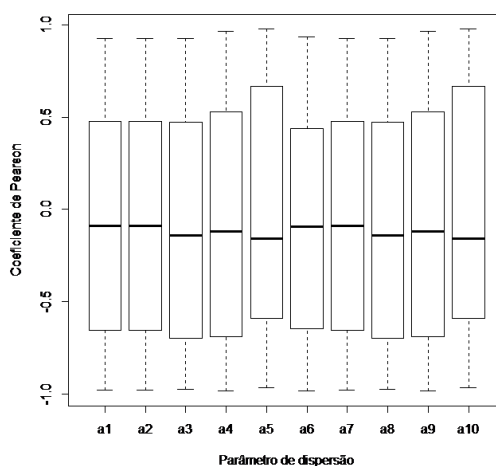


Figura 13 Gráfico *Boxplot* do coeficiente de correlação linear de Pearson para as simulações das amostras de a1 a a10.

Tabela 3 Porcentagem de amostras segundo classificação de Dancey e Reidy (2007) para as amostras de a1 a a10

Amostras	Fraca (%)	Moderada (%)	Forte (%)
a1	38	24	38
a2	34	25	41
a3	33	22	45
a4	31	28	41
a5	30	31	39
a6	38	24	38
a7	34	25	41
a8	33	22	45
a9	31	28	41
a10	30	31	39

No segundo ensaio notou-se que ao diminuir o raio de dependência espacial (ϕ_0) do componente comum de Y_1 e Y_2 , a variabilidade e a amplitude dos valores de correlação linear de Pearson (r_{Y_1, Y_2}) apresentam comportamento similar. Mesmo com ϕ_0 baixo ainda tem-se uma alta amplitude dos valores de correlação linear, com valores variando de -0,83 e 0,76 (Tabela 6 e Figura 14). No entanto, para os menores alcances observa-se que existe uma maior porcentagem de simulações em que o coeficiente de correlação de Pearson entre as variáveis Y_1 e Y_2 é considerado como fraca associação linear, segundo classificação de Dancey e Reidy (2007) (Tabela 7). Esses resultados evidenciam que em geral o decréscimo do raio de dependência espacial (ϕ_0) do componente comum de Y_1 e Y_2 não influencia nos valores do coeficiente de correlação linear de Pearson entre Y_1 e Y_2 .

Tabela 4 Estatísticas descritivas das amostras variando ϕ_0

Estatísticas	30	27	25	20	18	15	12	10	5	3
Mínimo	-0,99	-0,99	-0,99	-0,98	-0,98	-0,98	-0,98	-0,99	-0,96	-0,83
1° quartil	-0,67	-0,72	-0,72	-0,75	-0,73	-0,73	-0,72	-0,68	-0,28	-0,25
Mediana	-0,12	-0,15	-0,17	-0,09	-0,10	-0,06	-0,02	-0,03	0,00	0,07
Média	-0,07	-0,07	-0,07	-0,08	-0,08	-0,08	-0,07	-0,07	0,004	0,03
3° quartil	0,51	0,49	0,51	0,49	0,47	0,51	0,47	0,46	0,28	0,30
Máximo	0,96	0,95	0,94	0,98	0,99	0,98	0,84	0,92	0,89	0,76
Variância	0,41	0,42	0,43	0,43	0,43	0,42	0,40	0,36	0,16	0,15
DP	0,64	0,65	0,66	0,66	0,65	0,65	0,63	0,61	0,40	0,39

DP = desvio padrão.

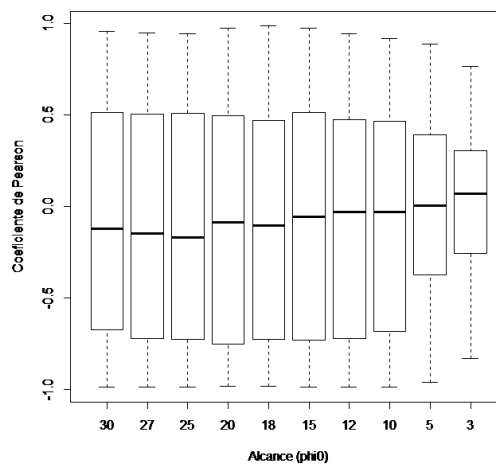


Figura 14 Gráfico *Boxplot* do coeficiente de correlação linear de Pearson para as simulações das amostras variando ϕ_0 .

Tabela 5 Porcentagem de amostras segundo classificação de Dancey e Reidy (2007) para as amostras variando ϕ_0

ϕ_0	Fraca (%)	Moderada (%)	Forte (%)
30	33	27	40
27	36	20	44
25	35	17	48
20	32	20	48
18	33	21	46
15	28	28	44
12	30	30	40
10	34	33	33
5	65	26	9
3	66	28	6

Foram observados que os resultados do terceiro ensaio são similares ao ensaio anterior, mesmo com o decréscimo de ϕ_1 e ϕ_2 simultâneo ao decréscimo de ϕ_0 (Tabela 8 e Figura 15), o que evidencia um comportamento similar, isto é, os valores de (r_{Y_1, Y_2}) apresentam alta amplitude. Como no caso anterior, observa-se para os menores alcances simultâneos de ϕ_0, ϕ_1 e ϕ_2 uma maior porcentagem de simulações em que pelo coeficiente de correlação linear de Pearson a associação linear entre as variáveis Y_1 e Y_2 é considerada fraca (DANCEY; REIDY, 2007) (Tabela 9).

Tabela 6 Estatísticas descritivas das amostras variando ϕ_0, ϕ_1 e ϕ_2 simultaneamente

Estatísticas	30	25	20	18	15	12	10	8	5	3
Mínimo	-0,99	-0,98	-0,98	-0,98	-0,97	-0,95	-0,94	-0,93	-0,89	-0,78
1° quartil	-0,79	-0,79	-0,75	-0,72	-0,69	-0,68	-0,69	0,00	-0,45	-0,37
Mediana	-0,14	-0,12	-0,08	-0,09	-0,08	-0,06	-0,07	0,00	-0,05	-0,04
Média	-0,08	-0,08	-0,08	-0,08	-0,08	-0,08	-0,08	-0,03	-0,06	-0,02
3° quartil	0,55	0,54	0,50	0,47	0,47	0,47	0,46	0,00	0,33	0,25
Máximo	0,97	0,96	0,98	0,98	0,98	0,98	0,97	0,94	0,87	0,85
Variância	0,46	0,45	0,44	0,42	0,40	0,38	0,35	0,12	0,24	0,17
DP	0,68	0,67	0,66	0,65	0,63	0,61	0,56	0,35	0,49	0,41

DP = desvio padrão.

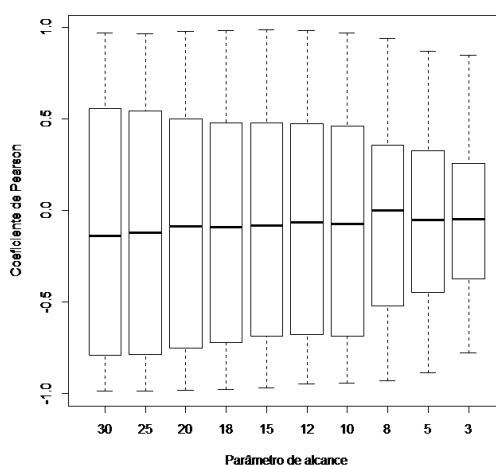


Figura 15 Gráfico *Boxplot* do coeficiente de correlação linear de Pearson para as simulações das amostras variando ϕ_0, ϕ_1 e ϕ_2 simultaneamente.

Tabela 7 Porcentagem de amostras segundo classificação de Dancey e Reidy (2007) para as amostras variando ϕ_0, ϕ_1 e ϕ_2 simultaneamente

ϕ_0, ϕ_1 e ϕ_2	Fraca (%)	Moderada (%)	Forte (%)
30	27	23	50
25	30	21	49
20	32	20	48
18	33	20	47
15	32	28	40
12	28	36	36
10	28	41	31
8	77	12	9
5	50	30	20
3	61	31	8

Nota-se que nos três ensaios houve pouca variação das estatísticas descritivas do coeficiente de correlação linear de Pearson. Além disso, é visível que a variabilidade dos dados não se altera a medida que variam os parâmetros de dispersão e alcance.

5.2 Comparação de modelos com dados simulados

Observou-se que a menor variabilidade entre as simulações e os menores valores para a soma quadrada da diferença da predição espacial para as variáveis Y_1 e Y_2 foram obtidas quando se comparou as predições feitas usando o modelo univariado e o semivariograma cruzado (Tabela 10 e Figura 16). Além disso, observou-se, por meio dessa medida, que a predição espacial obtida com o uso do modelo BGCCM foi a que apresentou maior diferença, quando comparado com as predições espaciais realizadas considerando o modelo geoestatístico univariado e o semivariograma cruzado.

Pelas medidas de acurácia (Tabelas 11 a 13 e Figuras 17 a 19) notou-se uma diferença relevante entre as predições espaciais utilizando os três modelos, pois na maioria das simulações obteve índice de exatidão global inferior a 85% (Tabela 11 e Figura 17). Apenas em 36% das simulações observou-se alta similaridade na predição espacial de Y_1 feita usando os modelos univariado e BGCCM ($EG \geq 85\%$, $Kp \geq 80\%$ e $T \geq 80\%$). Além disso, para a variável Y_1 o índice Kappa foi superior a 67% ao comparar o modelo BGCCM com o modelo univariado e semivariograma cruzado em 2% das simulações, já para o índice Tau foi respectivamente 5 e 3% das simulações.

No entanto, existe uma alta variabilidade das medidas de acurácia quando se comparou a predição espacial feita usando os modelos univariados e semivariograma cruzado com o modelo BGCCM. Sendo que os menores valores dessas medidas ocorrem na maioria das simulações quando se compara a predição espacial do modelo BGCCM com as demais.

Tabela 8 Estatística descritiva da soma quadrada da diferença da predição espacial das variáveis Y_1 e Y_2 , comparando os modelos univariado (u), semivariograma cruzado (sc) e modelo BGCCM (mb)

Estatísticas	Variável Y_1			Variável Y_2		
	u_sc	u_mb	sc_mb	u_sc	u_mb	sc_mb
Mínimo	913,6	23,6	900	1749	42906	76869
1° quartil	2127,2	62,2	3312	59700	162008	268826
Mediana	2781,5	28549,7	28503	92418	467856	569775
Média	2934,3	76464,1	75789	123607	2100224	2195521
3° quartil	3591,3	72578,8	67348	143215	1273665	1363989
Máximo	5722,3	1267266,3	1245121	488719	43607063	43226536
DP	991,71	163557,1	161739,1	102763,3	6076553	6025968
CV	33,79	213,90	213,41	83,14	289,33	274,47

DP = desvio padrão e CV = coeficiente de variação.

Tabela 9 Estatística descritiva da exatidão global das variáveis Y_1 e Y_2 , comparando os modelos univariado (u), semivariograma cruzado (sc) e modelo BGCCM (mb)

Estatísticas	Variável Y_1			Variável Y_2		
	u_sc	u_mb	sc_mb	u_sc	u_mb	sc_mb
Mínimo	0,3212	0,0292	0,0264	0,0	0,0047	0,0033
1° quartil	0,4858	0,1197	0,1071	0,0743	0,0320	0,0284
Mediana	0,5577	0,2069	0,1999	0,1035	0,0619	0,0461
Média	0,5579	0,4374	0,2878	0,1192	0,0714	0,0526
3° quartil	0,6229	0,9359	0,4896	0,1488	0,1047	0,0725
Máximo	0,7430	0,9771	0,7388	0,6560	0,1755	0,1377
DP	0,0934	0,3877	0,2101	0,0842	0,04568	0,0325
CV	16,74	88,64	72,96	70,68	63,98	61,74

DP = desvio padrão e CV = coeficiente de variação.

Tabela 10 Estatística descritiva do índice Kappa das variáveis Y_1 e Y_2 , comparando os modelos univariado (u), semivariograma cruzado (sc) e modelo BGCCM (mb)

Estatísticas	Variável Y_1			Variável Y_2		
	u_sc	u_mb	sc_mb	u_sc	u_mb	sc_mb
Mínimo	0,1712	0,0117	0,0168	-0,0991	0,0045	-0,0099
1° quartil	0,3898	0,0471	0,0409	-0,0176	0,0213	0,0155
Mediana	0,4723	0,1031	0,1051	0,0016	0,0297	0,0211
Média	0,4690	0,3824	0,2110	0,0165	0,0353	0,0214
3° quartil	0,5590	0,9256	0,3958	0,0309	0,0439	0,0259
Máximo	0,7002	0,9732	0,6971	0,5963	0,0988	0,0522
DP	0,1203	0,4193	0,2065	0,0762	0,0202	0,0109
CV	25,66	109,66	97,85	460,28	57,23	51,32

DP = desvio padrão e CV = coeficiente de variação.

Tabela 11 Estatística descritiva do índice Tau das variáveis Y_1 e Y_2 , comparando os modelos univariado (u), semivariograma cruzado (sc) e modelo BGCCM (mb)

Estatísticas	Variável Y_1			Variável Y_2		
	u_sc	u_mb	sc_mb	u_sc	u_mb	sc_mb
Mínimo	0,2458	-0,0787	-0,0818	-0,1111	-0,1059	-0,1074
1° quartil	0,4287	0,0218	0,0079	-0,0286	-0,0755	-0,0795
Mediana	0,5086	0,1188	0,1109	0,0039	-0,0423	-0,0599
Média	0,5088	0,3749	0,2087	0,0213	-0,0318	-0,0526
3° quartil	0,5809	0,9288	0,4329	0,0543	0,0052	-0,0305
Máximo	0,7144	0,9745	0,7098	0,6178	0,0839	0,0419
DP	0,1038	0,4308	0,2333	0,0936	0,0507	0,0361
CV	20,39	114,90	111,80	439,02	-159,66	-68,59

DP = desvio padrão e CV = coeficiente de variação.

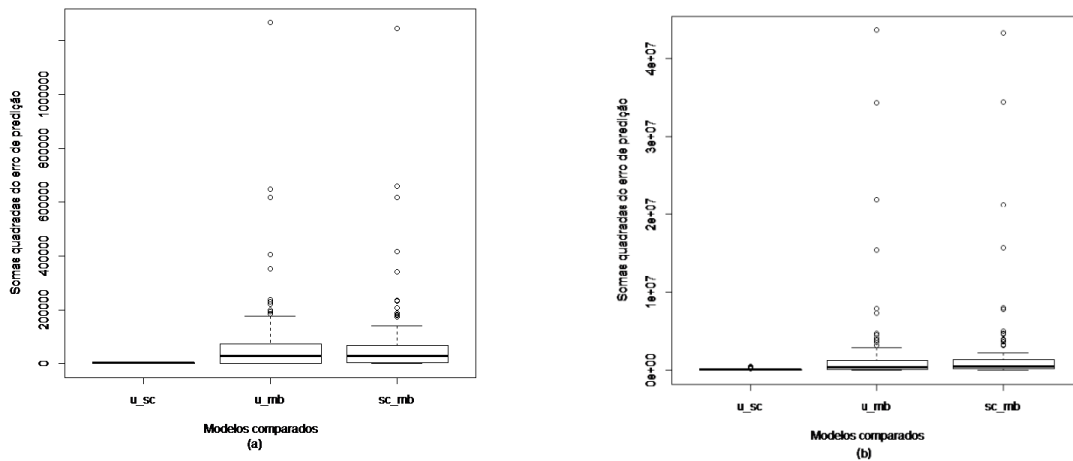


Figura 16 Gráficos *Boxplots* das somas quadradas da diferença de predição espacial para as variáveis Y_1 (a) e Y_2 (b), comparando os modelos univariado (u), semivariograma cruzado (sc) e modelo BGCCM (mb).

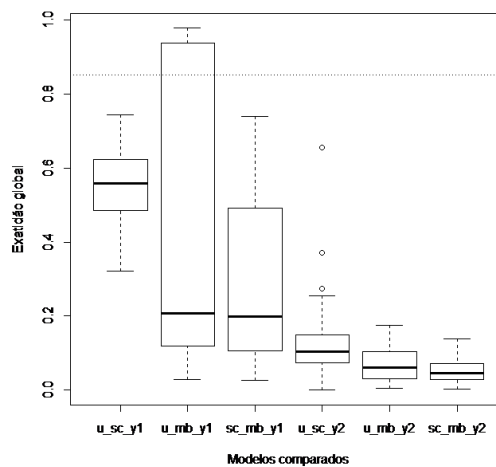


Figura 17 Gráfico *Boxplot* para a exatidão global das variáveis Y_1 e Y_2 , comparando os modelos univariado (u), semivariograma cruzado (sc) e modelo BGCCM (mb).

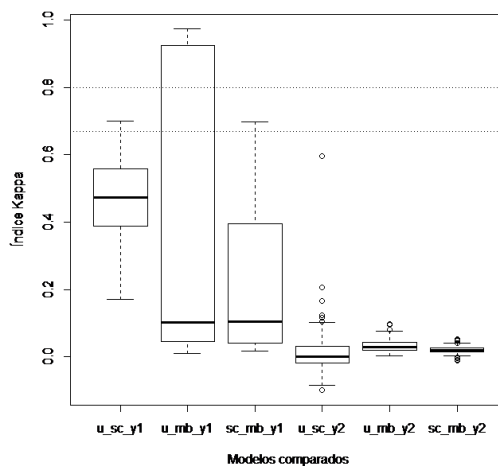


Figura 18 Gráfico *Boxplot* para o índice Kappa das variáveis Y_1 e Y_2 , comparando os modelos univariado (u), semivariograma cruzado (sc) e modelo BGCCM (mb).

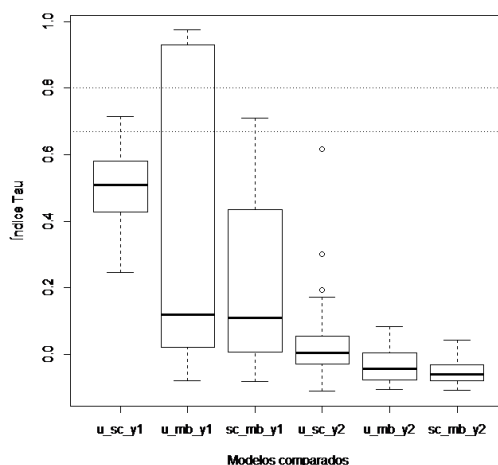


Figura 19 Gráfico *Boxplot* para o índice Tau das variáveis Y_1 e Y_2 , comparando os modelos univariado (u), semivariograma cruzado (sc) e modelo BGCCM (mb).

5.3 Comparação de modelos com dados experimentais co-locados e balanceados do ano agrícola 2010/2011

Diversas propriedades químicas e físicas do solo do ano agrícola 2010/2011 foram avaliadas e, entre eles selecionaram-se cálcio (Ca), magnésio (Mg), manganês (Mn) e pH, por apresentarem associação entre elas, verificada pelo coeficiente linear de Pearson (Tabela 14), observou-se que as propriedades pH e Ca, pH e Mg apresentaram forte associação linear ($r \geq 0,7$, DANCEY; REIDY, 2007), e pH e Mn apresentaram uma moderada correlação linear ($0,4 < r < 0,7$), DANCEY; REIDY, 2007), sendo que todas elas apresentaram correlação significativa a 5% de probabilidade. No entanto, como observado nas simulações, o coeficiente de correlação linear de Pearson não apresenta informação a respeito da associação bivariada.

Tabela 12 Medidas de correlação linear de Pearson entre as variáveis pH e cálcio, pH e magnésio, pH e manganês

Variáveis	Correlação (r)	p-valor*
pH e Ca	0,896	<0,01
pH e Mg	0,779	<0,01
pH e Mn	0,506	<0,01

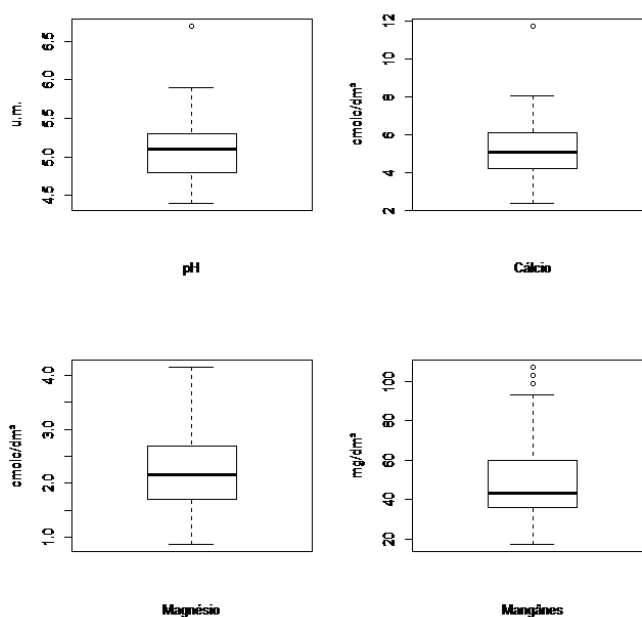
*p-valor ao nível de 5% de probabilidade

Analisando as medidas descritivas (Tabela 15) das variáveis verificou-se que o Mg e Mn apresentaram alta dispersão segundo classificação de Pimentel-Gomes (2000), e que o pH apresentou a maior homogeneidade. Além disso, observa-se para as variáveis pH, Ca e Mn a presença de pontos discrepantes (Figura 20).

Tabela 13 Medidas descritivas das variáveis pH, cálcio, magnésio e manganês

Estatísticas	Variáveis			
	pH	Ca(mg/dm ³)	Mg(mg/dm ³)	Mn(mg/dm ³)
n	102	102	102	102
Mínimo	4,40	2,37	0,86	17,00
Máximo	6,70	11,76	4,15	107,00
Média	5,095	5,201	2,252	49,320
Mediana	5,100	5,075	2,150	43,000
Variância	0,149	1,887	0,472	379,508
Desvio padrão	0,386	1,374	0,687	19,481
Coefficiente de variação	7,576	26,411	30,499	39,496
Curtose	1,460	3,565	-0,507	0,318
Assimetria	0,680	0,948	0,326	0,945
p-valor	<0,01*	<0,01*	<0,01*	0,216 ^{ns}

n = número de parcelas observadas, *significativo a 5% de probabilidade, ns = não significativo a 5% de probabilidade.

Figura 20 Gráfico *Boxplot* para as variáveis pH, cálcio, magnésio e manganês.

Para a avaliação da normalidade dos dados utilizou-se os coeficientes de assimetria, curtose, Histograma (Figura 21) e o teste de Shapiro Wilk (Tabela 15), sendo que as pH, Ca e Mn indicam que os dados não têm distribuição normal.

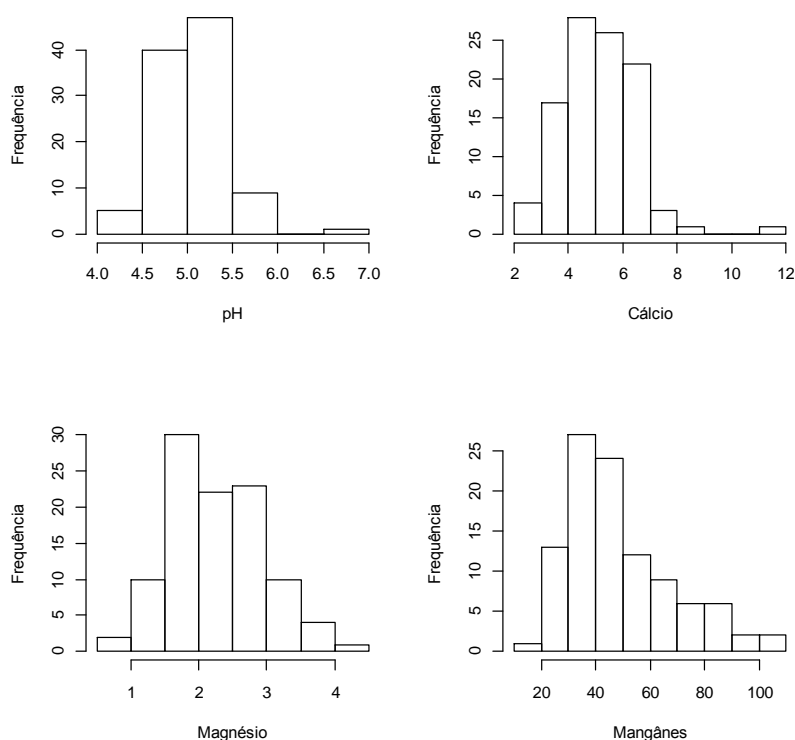


Figura 21 Histograma para as variáveis pH, cálcio, magnésio e manganês.

Ainda com o intuito de avaliar o comportamento das variáveis construíram-se os gráficos *Postplot* (Figura 22). Observa-se que na região superior da área agrícola locais com concentração dos menores valores para as variáveis.

Para as quatro variáveis em estudo verificou-se que não houve a presença de anisotropia, pois os gráficos *Postplot* não indicam visualmente qualquer tendência direcional em relação à continuidade espacial. Assim, por meio do cálculo das funções semivariâncias univariada e cruzada, construíram-se os semivariogramas omnidirecionais univariado e cruzado (Figuras 20 e 21) e ajustaram-se os modelos: esférico, exponencial, gaussiano e família Matérn com K igual a 1, 1,5 e 2, sendo que o melhor modelo para descrever cada função semivariância foi escolhido por meio das técnicas de validação cruzada. Os semivariogramas univariado e cruzado (Figuras 23 e 24) indicam que a semivariância aumenta conforme se aumenta os valores da distância.

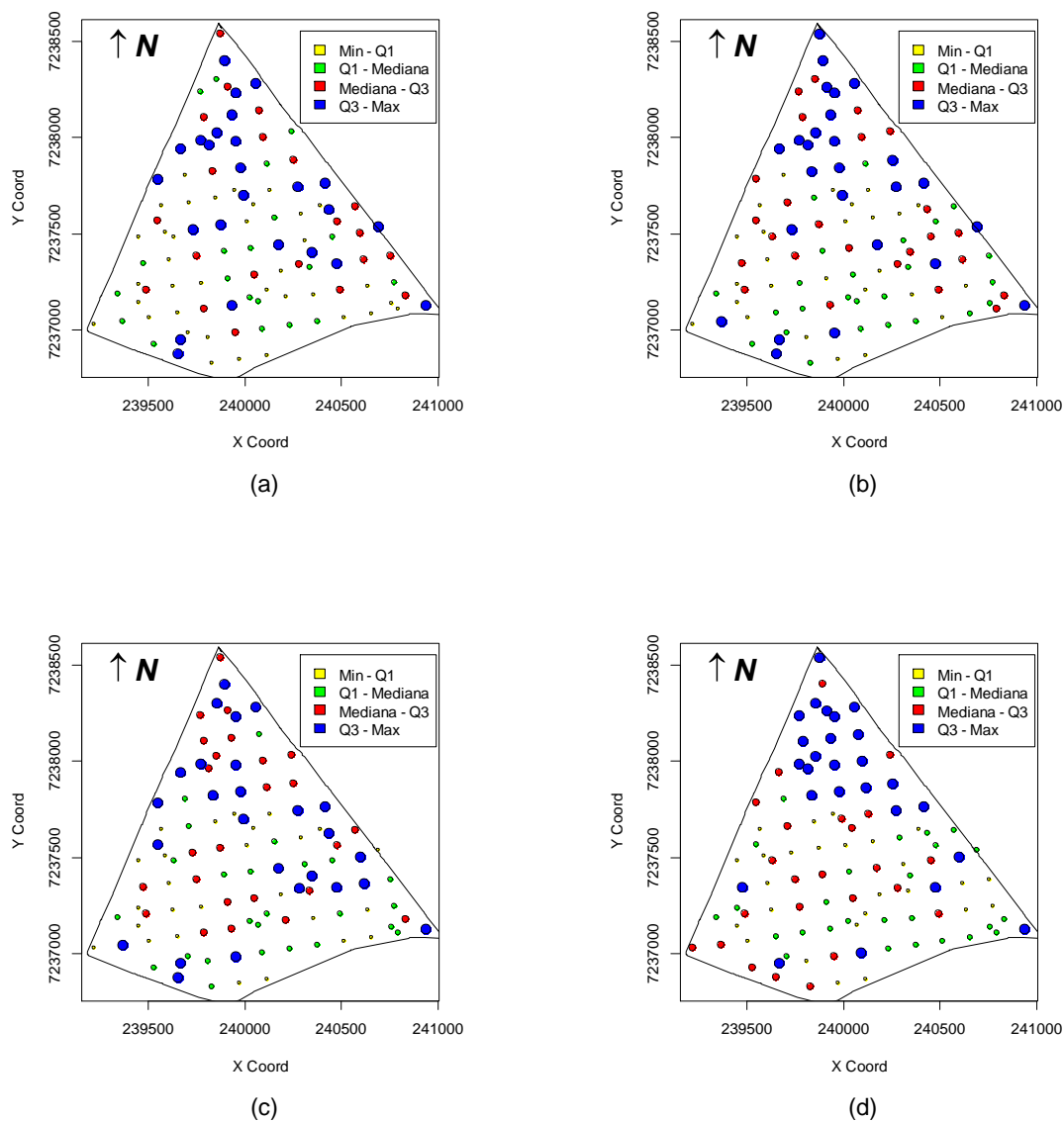


Figura 22 Gráfico *Postplot* para as variáveis pH (a), cálcio (b), magnésio (c) e manganês (d).

O modelo que melhor se ajustou para as variáveis pH, Ca, Mg e Mn do modelo univariado, respectivamente, foram exponencial, gaussiano, esférico e exponencial com parâmetros estimados por máxima verossimilhança (Tabela 16). Para a função semivariância cruzada os modelos ajustados foram Matérn com $k=2$, parâmetros estimados pelo método de mínimos quadrados (Tabela 17). Para o modelo BGCCM os modelos ajustados foram, respectivamente, Matérn com $k=1,5$, Matérn com $k=2$ e gaussiano com parâmetros estimados por máxima verossimilhança (Tabela 18).

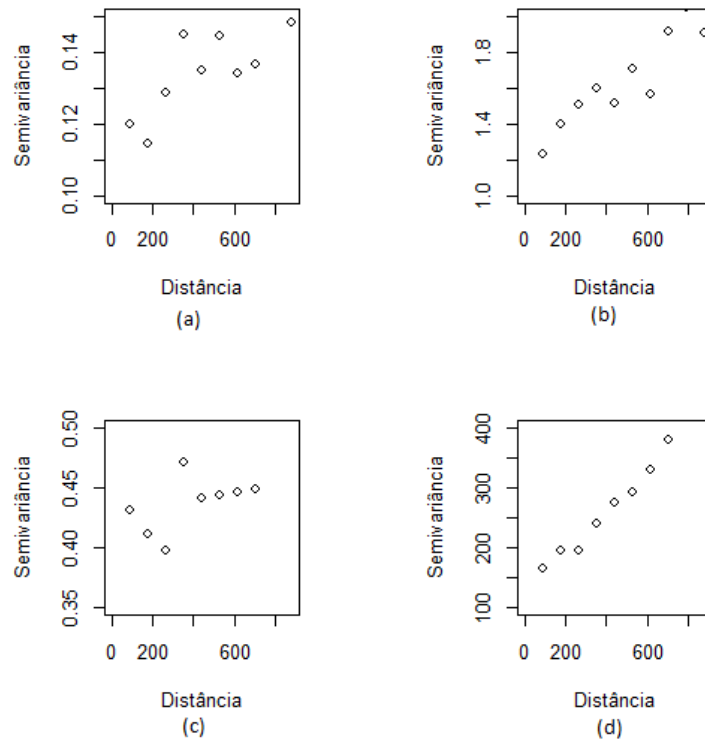


Figura 23 Semivariograma experimental univariado para as variáveis pH (a), cálcio (b), magnésio (c) e manganês (d).

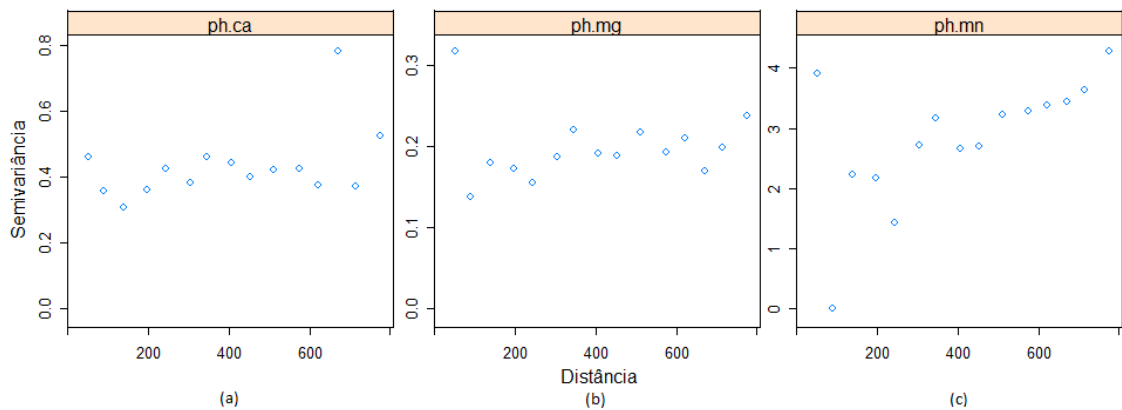


Figura 24 Semivariograma cruzado para as variáveis pH e cálcio (a), ph e magnésio (b), pH e manganês (c), sendo que para todas as variáveis foi utilizado o pH na estrutura de espacial bivariada.

Para a escolha dos melhores modelos foram levados em consideração o erro médio (EM), desvio padrão do erro (σ_{EM}), erro relativo (ER), erro médio relativo (σ_{ER}), erro absoluto (EA), Critério de Informação de Akaike (AIC) e critério de Informação Bayesiano (BIC).

Tabela 14 Critérios utilizados para a escolha do melhor modelo ajustado para o modelo univariado

Variável	Modelo	EM	σ_{EM}	ER	σ_{ER}	EA	AIC	BIC
pH	Exponencial	$3,68 \cdot 10^{-4}$	$3,61 \cdot 10^{-1}$	$5,12 \cdot 10^{-4}$	$1,00 \cdot 10^0$	$2,95 \cdot 10^1$	$9,34 \cdot 10^1$	$1,03 \cdot 10^2$
Ca	Gaussiano	$2,49 \cdot 10^{-3}$	$1,25 \cdot 10^0$	$9,90 \cdot 10^{-4}$	$1,00 \cdot 10^0$	$9,34 \cdot 10^1$	$3,47 \cdot 10^2$	$3,57 \cdot 10^2$
Mg	Esférico	$7,52 \cdot 10^{-5}$	$6,52 \cdot 10^{-1}$	$5,49 \cdot 10^{-5}$	$1,00 \cdot 10^0$	$5,30 \cdot 10^1$	$2,12 \cdot 10^2$	$2,23 \cdot 10^2$
Mn	Exponencial	$9,90 \cdot 10^{-3}$	$1,39 \cdot 10^1$	$2,02 \cdot 10^{-4}$	$1,00 \cdot 10^0$	$1,09 \cdot 10^3$	$8,48 \cdot 10^2$	$8,58 \cdot 10^2$

Tabela 15 Critérios utilizados para a escolha do melhor modelo ajustado para o semivariograma cruzado

Variável	Modelo	EM	σ_{EM}	ER	σ_{ER}	EA
pH	Matérn k=2	-0,00023	0,373446	0,021382	2,979441	30,2539
Ca	Matérn k=2	0,001854	1,284594	0,009936	0,968187	95,0631
pH	Matérn k=2	-0,00169	0,383284	0,004203	3,080826	30,9903
Mg	Matérn k=2	-0,00148	0,702608	0,005585	1,585671	59,7755
pH	Esférico	0,015454	0,335233	0,127282	2,642195	22,4134
Mn	Esférico	-0,43604	14,82593	-0,00181	0,079802	948,7706

Tabela 16 Critérios utilizados para a escolha do melhor modelo ajustado para o modelo BGCCM

Variável	Modelo	EM	σ_{EM}	ER	σ_{ER}	EA
pH	Matérn k=1,5	$-5,20 \cdot 10^{-5}$	$4,69 \cdot 10^0$	$3,44 \cdot 10^{-2}$	$8,46 \cdot 10^{-1}$	$3,65 \cdot 10^2$
Ca	Matérn k=1,5	$4,56 \cdot 10^{-1}$	$5,22 \cdot 10^0$	$-1,63 \cdot 10^{-2}$	$1,93 \cdot 10^{-1}$	$3,05 \cdot 10^2$
pH	Matérn k=2	$1,16 \cdot 10^{-1}$	$2,28 \cdot 10^0$	$6,93 \cdot 10^{-3}$	$9,66 \cdot 10^{-2}$	$1,63 \cdot 10^2$
Mg	Matérn k=2	$3,52 \cdot 10^{-2}$	$2,11 \cdot 10^0$	$-2,66 \cdot 10^{-3}$	$3,69 \cdot 10^{-2}$	$1,43 \cdot 10^2$
pH	Gaussiano	$-1,90 \cdot 10^{-4}$	$4,04 \cdot 10^{-1}$	$7,30 \cdot 10^{-1}$	$6,60 \cdot 10^0$	$3,22 \cdot 10^1$
Mn	Gaussiano	$-1,31 \cdot 10^0$	$4,29 \cdot 10^1$	$3,22 \cdot 10^{-2}$	$1,37 \cdot 10^0$	$2,73 \cdot 10^3$

Notou-se que o efeito pepita relativo para Mg no caso do modelo univariado e o par de variáveis pH e Mg no caso do semivariograma cruzado apresentaram fraca dependência espacial e as demais variáveis apresentaram moderada dependência espacial (Tabela 20).

Os alcances obtidos para o modelo univariado representam a distância em que os pontos amostrais estão correlacionados entre si. Assim, para as variáveis pH, Ca, Mg e Mn tem-se que o raio máximo de dependência espacial entre amostras é igual a 875,47 m, 1107,62 m, 1206,30 m e 1497,87 m, respectivamente (Tabela 19). Para o semivariograma cruzado o alcance representa a distância máxima de correlação espacial entre os pares de propriedades químicas pH e Ca, pH e Mg e pH e Mn e que foram de respectivamente 300, 200 e 450 m (Tabela 20).

Nota-se nas Tabelas 20 e 21 que os parâmetros de dispersão do modelo BGCCM e ajustado pelo semivariograma cruzado são valores pequenos, exceto, a dispersão de Mn no par de variáveis pH e Mn, sendo que tais valores indicam que o componente que define a estrutura de dependência espacial comum para as variáveis Y_1 e Y_2 é baixo. No entanto, valores pequenos de efeito pepita e patamar também foram observados nos parâmetros estimados dos modelos univariados para pH, Ca e Mg. Esses resultados podem ter sido influenciados pela escala de valores dessas variáveis. No entanto, o alcance ϕ_0 que define o raio de dependência espacial entre os pares de variáveis pH e Ca e pH e Mg é de respectivamente 899,38 e 917,84 m.

Para o par de variáveis pH e Mn observam-se os maiores valores dos parâmetros de dispersão, também observa-se que para o pH que o raio de dependência espacial do componente independente que define a estrutura de dependência espacial é igual a 0,077

m, sendo que esse indica um valor pequeno de raio de dependência espacial entre localizações distintas para pH.

Além disso, o raio de dependência espacial entre pH e Mn considerando localizações distintas é igual a 79,83 m. Por meio desses resultados e observando a pequena escala de valores de pH, Ca e Mg, considera-se que existe dependência espacial comum as variáveis em estudo em cada par de variável analisado.

Tabela 17 Modelos espaciais ajustados para o modelo univariado e os parâmetros estimados pelo método de máxima verossimilhança

Variáveis	Modelo	a	φ_1	φ_2	$\varphi_1 + \varphi_2$	EPR (%)
pH	Exponencial	875,4702	0,0991	0,0539	0,153	64,771
Ca	Gaussiano	1107,616	1,432	0,586	2,018	70,961
Mg	Esférico	1206,301	0,387	0,0838	0,4708	82,201
Mn	Exponencial	1497,865	122,7	245,1	367,8	33,361

a = alcance, φ_1 = efeito pepita, φ_2 = contribuição, $\varphi_1 + \varphi_2$ = patamar e EPR = efeito pepita relativo.

Tabela 18 Modelos espaciais ajustados para o semivariograma cruzado e os parâmetros estimados pelo método de mínimos quadrados

Variáveis	Modelo	a	φ_1	φ_2	$\varphi_1 + \varphi_2$	EPR (%)
pH e Ca	Matérn k=2	300	0,35412529	0,17128545	0,52541074	67,40
pH e Mg	Matérn k=2	200	0,19654325	-0,00845922	0,20500247	95,87
pH e Mn	Esférico	450	1,715608	1,118774	2,834382	41,48

a = alcance e φ_1 = efeito pepita, φ_2 = contribuição, $\varphi_1 + \varphi_2$ = patamar e EPR = efeito pepita relativo.

Tabela 19 Modelos espaciais ajustados para o modelo BGCCM e os parâmetros estimados pelo método de máxima verossimilhança

	pH e Ca	pH e Mg	pH e Mn
Modelo	Matérn k=1,5	Matérn k=2	Gaussiano
μ_1	3,794	3,413	5,094
μ_2	1,829	5,909	50,313
σ_{01}	$2,2 \cdot 10^{-14}$	$6,6 \cdot 10^{-14}$	$7,6 \cdot 10^{-3}$
σ_1	$4,5 \cdot 10^{-12}$	$5,4 \cdot 10^{-12}$	$1,2 \cdot 10^{-1}$
σ_{02}	$5,3 \cdot 10^{-11}$	$5,2 \cdot 10^{-12}$	$1,5 \cdot 10^2$
σ_2	$6,2 \cdot 10^{-12}$	$1,1 \cdot 10^{-11}$	$2,0 \cdot 10^2$
\emptyset_0	899,384	917,843	79,830
\emptyset_1	1318,384	1528,404	0,077
\emptyset_2	1319,444	1438,672	482,176

Por meio das estimativas dos parâmetros do modelo univariado, semivariograma cruzado e modelo BGCCM realizou-se a estimação espacial para localizações não amostradas das variáveis em estudo; no caso do modelo univariado, utilizando a krigagem ordinária, e nos casos bivariados usa-se a cokrigagem ordinária. Os resultados apresentados pela estimação espacial mostraram valores dissimilares quanto às estimações obtidas usando o modelo univariado e semivariograma cruzado ao comparar com o modelo

BGCCM para as variáveis pH, Ca e Mg, pois o modelo BGCCM apresentou valores superestimados em áreas que possuem alta propriedade química estudada e subestimados em áreas de baixa, conforme observa-se pelos intervalos das estimativas espaciais (Tabelas 22 a 24) e pelos mapas de variabilidade (Figuras 25 e 26). Os mapas foram construídos utilizando 10 classes de divisão, para buscar uma melhor representação da distribuição dos valores na área em estudo e comparados dois a dois.

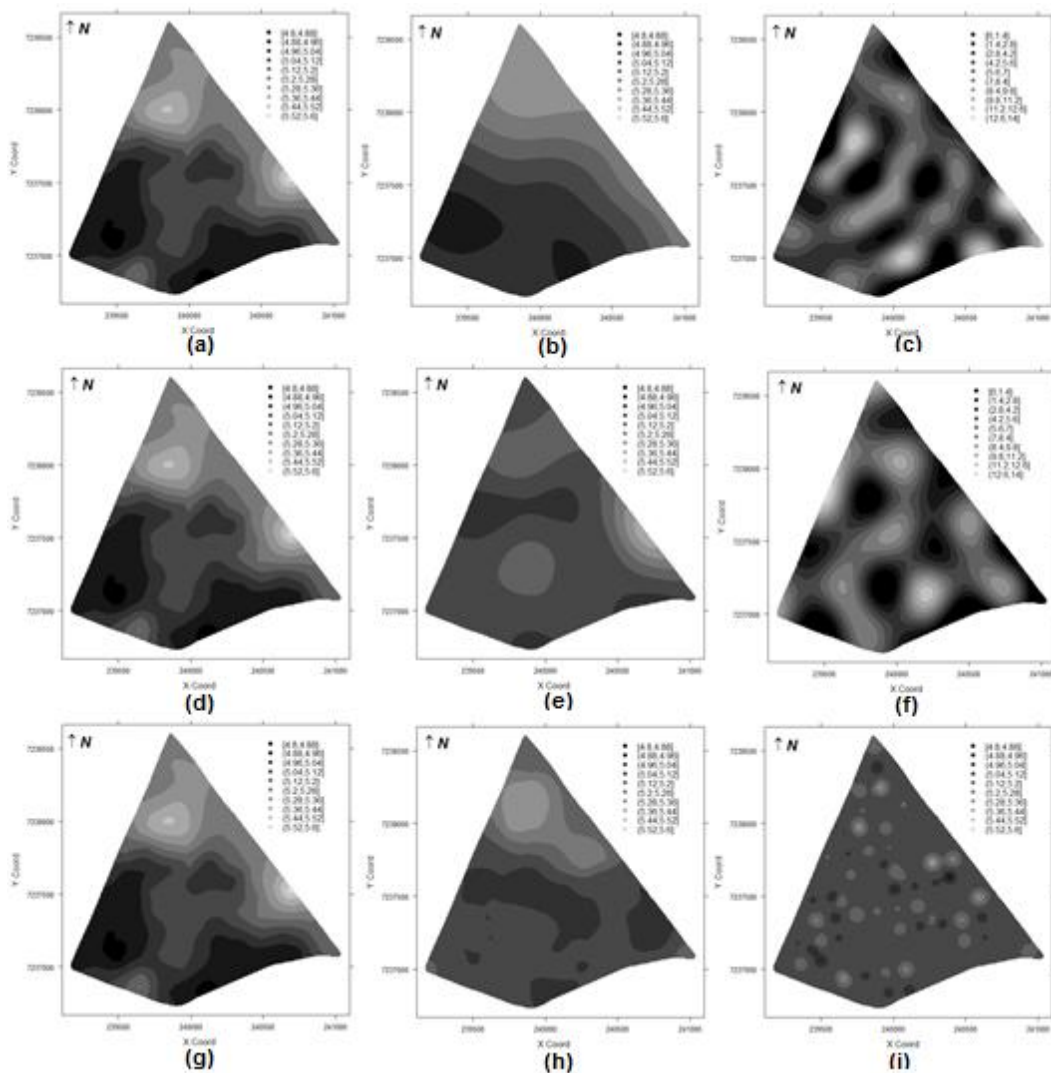


Figura 25 Mapas de variabilidade para o pH, pH/u (a), pH e Ca/sc (b), pH e Ca/mb (c), pH/u (d), pH e Mg/sc (e), pH e Mg/mb (f), pH/u (g), pH e Mn/sc (h) e pH e Mn/mb (i).

De acordo com Foody (2002), é desejável que a exatidão global alcance índices superiores a 85% para indicar exatidão entre os mapas. Além disso, todos os índices Kappa e Tau, comparando-se as estimativas espaciais entre todos os modelos (univariado e bivariados) foram considerados de baixa exatidão, uma vez que os valores de Kp e T apresentaram valores menores que 67%, indicando uma relevante dissimilaridade entre os mapas, ou seja, pouca semelhança na classificação dos mapas, segundo classificação de Krippendorff (1980). Esses resultados corroboram com as conclusões observadas nos conjuntos de dados simuladas apresentadas em 5.2.

Além disso, na comparação do modelo univariado e o modelo ajustado a função semivariância cruzada para Mg e pH o índice Kappa e Tau resultou em valores negativos, fato justificado pela divisão da soma dos elementos da diagonal principal da matriz dos erros pelo total de elementos, ser menor que p_i , onde $p_i = \frac{1}{r}$, r é igual ao número de classes.

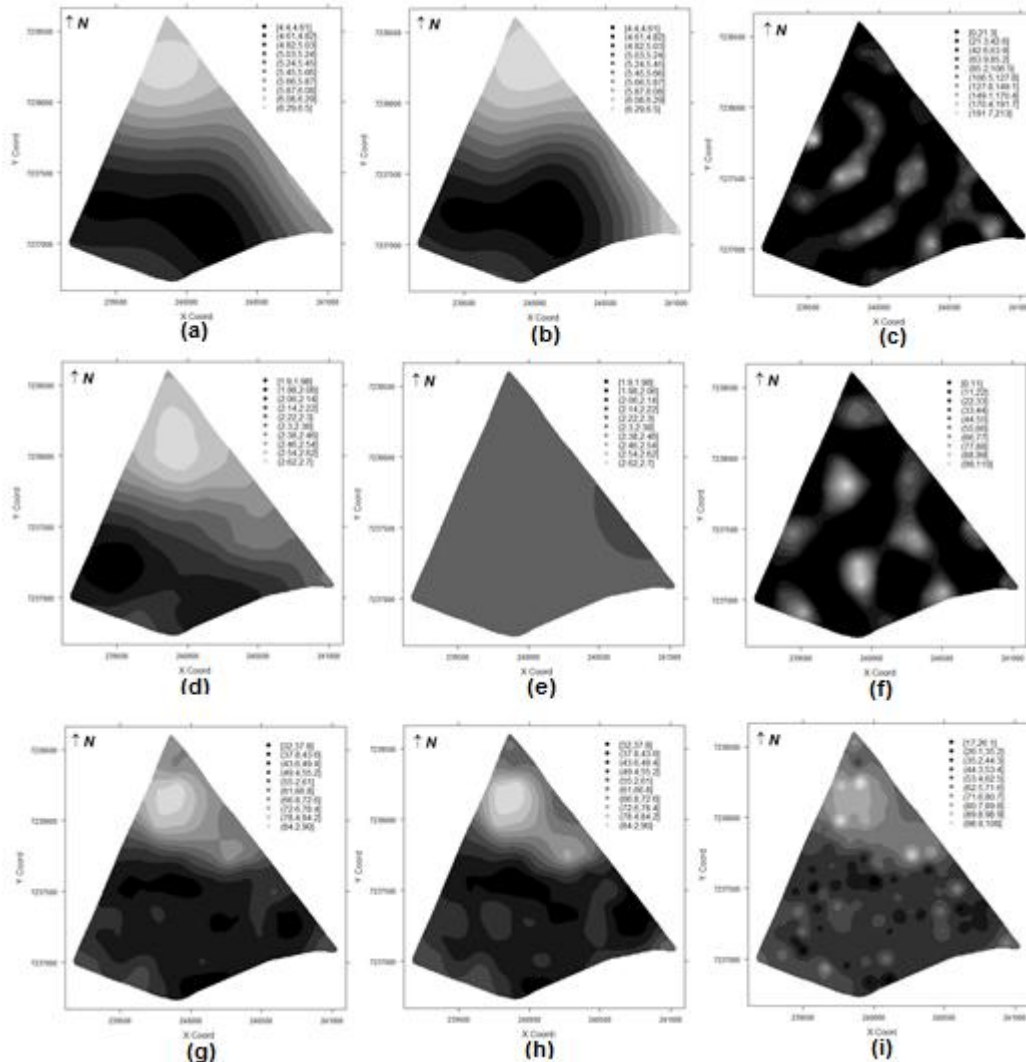


Figura 26 Mapas de variabilidade Ca/u (a), Ca e pH/sc (b), Ca e pH/mb (c), Mg/u (d), Mg e pH/sc (e), Mg e pH/mb (f), Mn/u (g), Mn e pH/sc (h), Mn e pH/mb (i).

Tabela 20 Medidas de acurácia para o modelo univariado e semivarograma cruzado

Variáveis	Intervalos	EG(%)	Kp(%)	T(%)
pH e Ca	4,8-5,6	41,05	29,51	34,50
pH e Mg	4,8-5,6	19,52	1,74	10,58
pH e Mn	4,8-5,6	24,69	8,24	16,33
Ca	4,4-6,5	67,71	62,98	64,12
Mg	1,9-2,7	9,25	-1,25	-0,83
Mn	32-90	68,38	59,44	64,87

EG = exatidão global, Kp = índice Kappa, T = índice Tau.

Tabela 21 Medidas de acurácia para modelo univariado e BGCCM

Variáveis	Intervalos	EG(%)	Kp(%)	T(%)
pH e Ca	0-14	20,27	0,93	11,41
pH e Mg	0-14	18,29	0,93	9,22
pH e Mn	4,8-5,6	20,79	1,41	11,99
Ca	0-213	21,14	10,42	12,38
Mg	0-110	11,85	5,79	2,06
Mn	17-108	68,39	55,01	64,89

EG = exatidão global, Kp = índice Kappa, T = índice Tau.

Tabela 22 Medidas de acurácia para o semivariograma cruzado e modelo BGCCM

Variáveis	Intervalos	EG(%)	Kp(%)	T(%)
pH e Ca	0-14	20,27	0,93	11,41
pH e Mg	0-14	18,29	0,93	9,22
pH e Mn	4,8-5,6	46,46	3,84	40,51
Ca	0-213	21,14	10,42	12,38
Mg	0-110	11,85	5,79	2,06
Mn	17-108	0,0	0,0	-11,11

EG = exatidão global, Kp = índice Kappa, T = índice Tau.

Neste caso, o uso do modelo bivariado BGCCM superestimou e subestimou valores para os pares de variáveis pH e Ca e pH e Mg e para as variáveis Ca, Mg e Mn, conforme pode ser observado nos intervalos de predição espacial usando esse modelo (Tabelas 21 e 22). Além disso, o intervalo de predição espacial usando os modelos estimados para o caso univariado e semivariograma cruzado são menores e se aproximam da amplitude, dessas variáveis, dos valores observados.

5.4 Comparação de modelos com dados experimentais co-locados e balanceados do ano agrícola 2013/2014

Diversas propriedades químicas e físicas do solo do ano agrícola 2013/2014 foram avaliadas e, entre elas selecionaram-se a produtividade (Prod), resistência do solo a penetração nas camadas de 0-10 cm (RSP0-10), 11-20 cm (RSP11-20), 21-30 cm (RSP21-30) e o Manganês (Mn), por apresentarem associação entre elas, verificada pelo coeficiente linear de Pearson (Tabela 25), observou-se que todas as propriedades apresentaram fraca associação linear. No entanto, como observado nas simulações o coeficiente de correlação linear de Pearson não apresenta informação a respeito da associação bivariada.

A amostra é co-locada e balanceada, ou seja, as variáveis possuem o mesmo número de observações e foram amostradas nas mesmas localizações.

Tabela 23 Medidas de correlação linear de Pearson entre as variáveis

Variáveis	Pearson	p-valor*
Prod e RSP0-10	0,1302	0,3623
Prod e RSP11-20	0,1332	0,3512
Prod e RSP21-30	0,1035	0,4698
Prod e Mn	0,1379	0,3343

*p-valor ao nível de 5% de probabilidade.

Além disso, foi levado em consideração que a variável Prod é menos dispendiosa para ser observada do que as variáveis RSP0-10, RSP11-20, RSP21-30 e Mn, uma vez que essa variável pode ser obtida por meio de um sensor de produtividade instalada nas máquinas colhedoras. A resistência do solo à penetração é necessária em estudos de qualidade física e estrutural do solo, para crescimento e desenvolvimento das raízes das plantas (SILVA et al., 2008), a qual possui relação com a produtividade. Dalchiavon et al. (2011) argumentam ainda que RSP pode influenciar negativamente o crescimento radicular da soja, afetando, sobremaneira, sua produtividade.

Analisando as medidas descritivas (Tabela 26) dos atributos do solo verificou-se que a RSP0-10 e Mn apresentaram alta dispersão, segundo classificação de Pimentel-Gomes (2000), enquanto que Prod, RSP11-20 e RSP21-30 apresentam menor dispersão, conforme observado na Figura 28. Além disso observa-se para as variáveis Prod, RSP11-20, RSP21-30 e Mn a presença de pontos discrepantes (Figura 27).

Tabela 24 Estatística descritiva das variáveis produtividade, resistência do solo a penetração e manganês

Estatísticas	Variáveis				
	Prod	RSP0-10	RSP11-20	RSP21-30	Mn
n	100	100	100	100	100
Mínimo	2,590	2428	5744	4763	26
Máximo	5,765	15130	17660	15670	143
Média	4,212	8195	10460	7057	61,190
Mediana	4,242	8253	10120	6748	58
Variância	4,582	7331545	6343141	2317882	423,611
Desvio padrão	0,584	2707,683	2518,559	1522,459	20,582
Coeficiente de variação	13,873	33,041	24,077	21,575	33,636
Curtose	0,273	-0,227	-0,083	9,118	1,495
Assimetria	-0,172	0,433	0,552	2,322	0,959
p-valor	0,790 ^{ns}	0,082 ^{ns}	0,040*	<0,01*	<0,01*

n = número de parcelas observadas, *significativo a 5% de probabilidade, ns = não significativo a 5% de probabilidade.

Ainda com o intuito de avaliar o comportamento das variáveis construíram-se os gráficos *Postplot*. Observou-se na Figura (29-e), na região superior da área agrícola, locais de concentração dos maiores valores para o Mn. No entanto, para as quatro propriedades em estudo verificou-se que não houve a presença de anisotropia, pois os gráficos *Postplot*

(Figura 29) não indicam visualmente qualquer tendência direcional em relação à continuidade espacial.

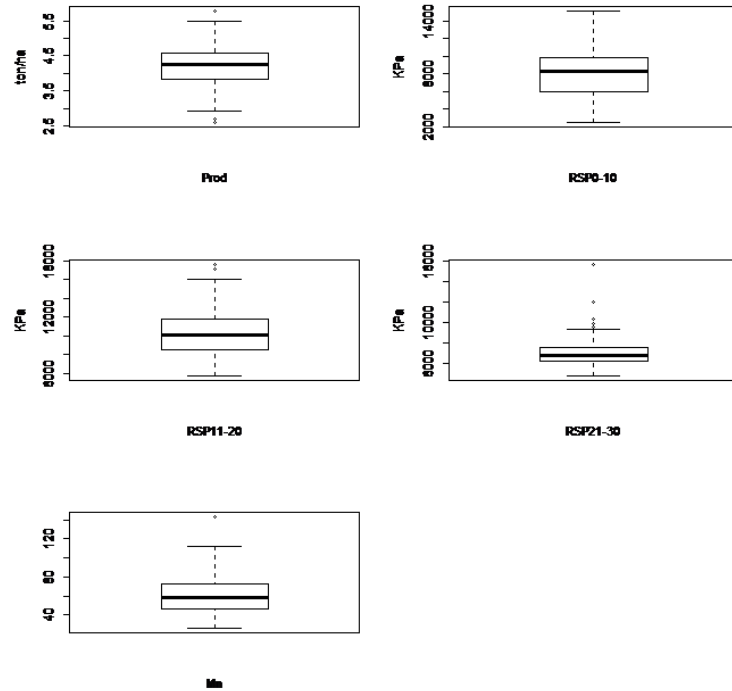


Figura 27 Gráfico *Boxplot* para as variáveis Prod, RSP0-10, RSP11-20, RSP21-30 e Mn.

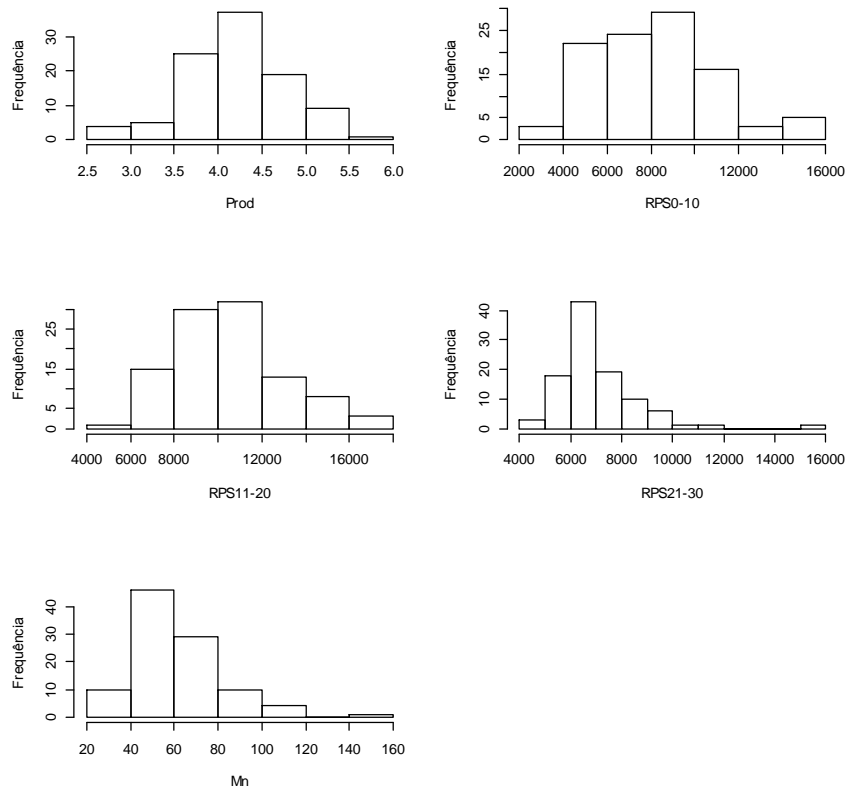


Figura 28 Histograma para as variáveis Prod, RSP0-10, RSP11-20, RSP21-30 e Mn.

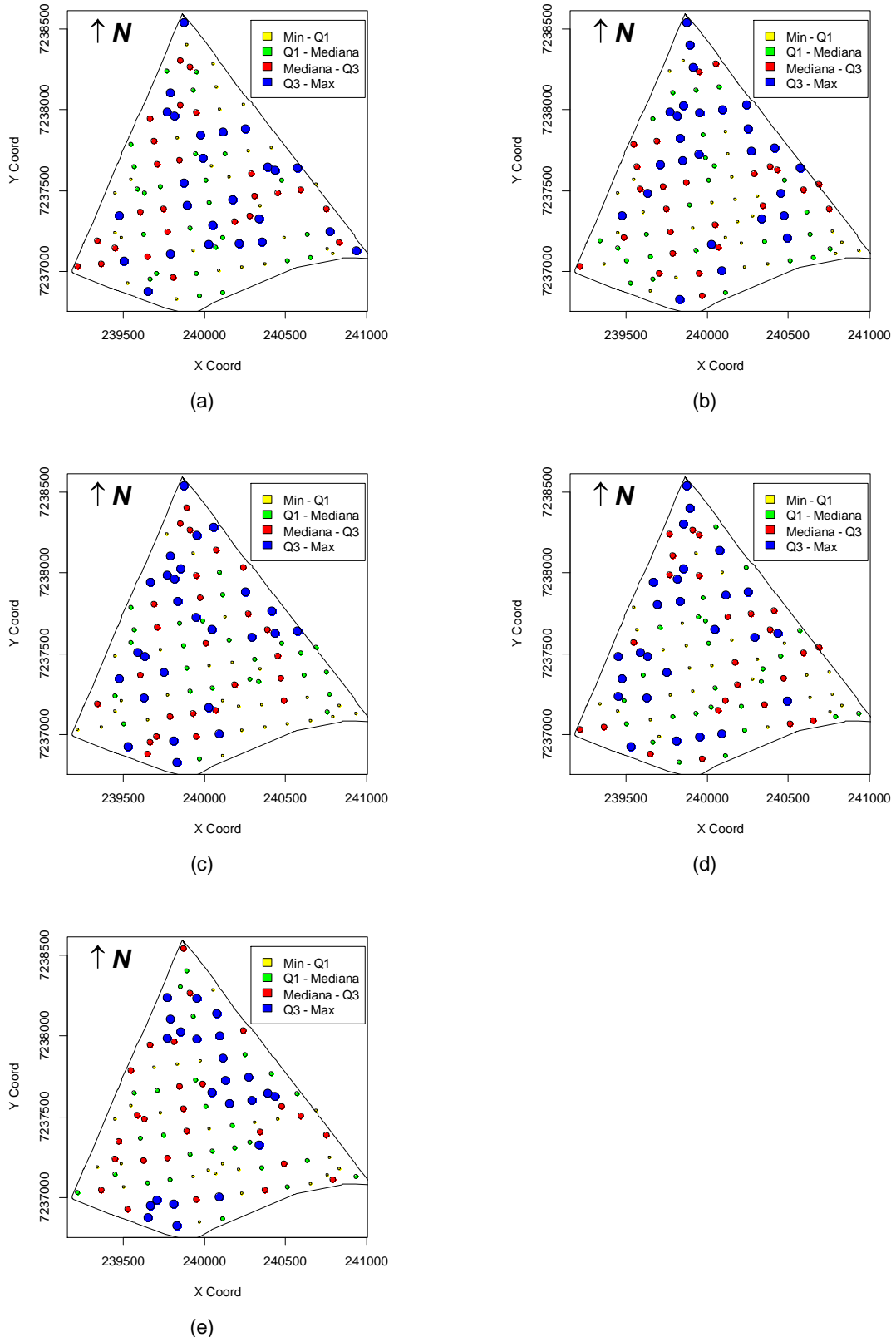


Figura 29 Gráfico *Postplot* para as variáveis Prod (a), RSP0-10 (b), RSP11-20 (c), RSP21-30 (d) e manganês (e), classificados em intervalos de mesma amplitude.

Por meio do cálculo dos valores das semivariâncias univariada e cruzada, em função da distância, construíram-se os semivariogramas omnidirecionais univariado e

cruzado (Figuras 30 e 31) e ajustaram-se os modelos: esférico, exponencial, gaussiano e família Matérn com K igual a 1, 1,5 e 2, sendo que o melhor modelo para descrever cada função semivariância foi escolhido por meio das técnicas de validação cruzada.

Os semivariogramas univariados (Figura 30) indicam que a semivariância aumenta conforme se aumenta os valores da distância. Já os semivariogramas cruzados (Figura 31), visualmente sugerem disposição aleatória dos pontos. No entanto, o efeito pepita relativo indica moderada dependência espacial, exceto para o par de variáveis Prod e Mn (Tabela 27).

O modelo que melhor se ajustou para o caso univariado a RSP0-10, RSP11-20, RSP21-30 e Mn respectivamente foram gaussiano, gaussiano, Matérn com K=2 e esférico com parâmetros estimados por máxima verossimilhança. Para o semivariograma cruzado foram respectivamente gaussiano, Matérn com K=2, esférico e Matérn com K=2, estimados pelo método de mínimos quadrados. Para o modelo BGCCM para as quatro propriedades foi o modelo Matérn com K=2 com parâmetros estimados por máxima verossimilhança.

Para a escolha do melhor modelo foi levado em consideração o erro médio (EM), desvio padrão do erro (σ_{EM}), erro relativo (ER), erro médio relativo (σ_{ER}), erro absoluto (EA), Critério de Informação de Akaike (AIC) e critério de Informação Bayesiano (BIC) (Tabelas 28 a 29).

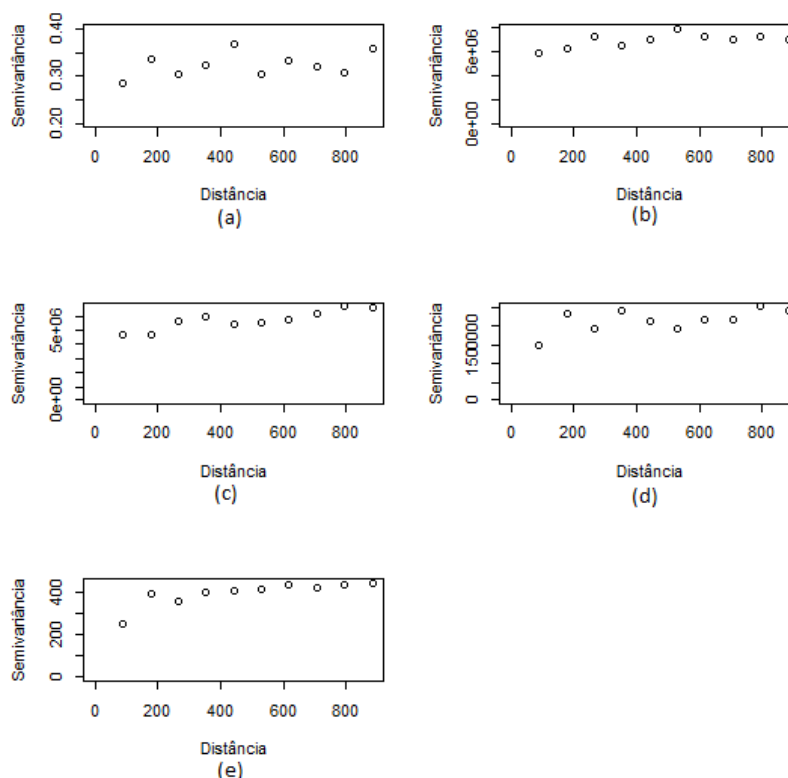


Figura 30 Semivariograma experimental univariado para as variáveis Prod (a), RSP0-10 (b), RSP11-20 (c), RSP21-30 (d) e manganês (e).

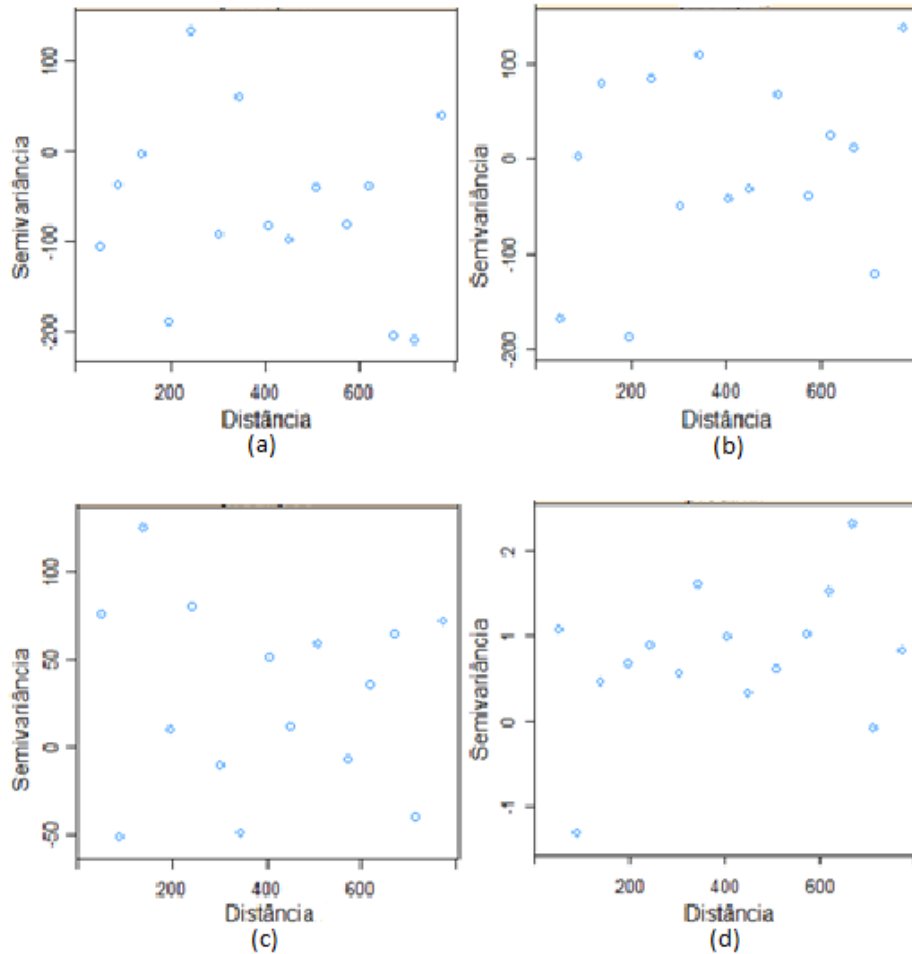


Figura 31 Semivariograma cruzado para as variáveis RSP0-10 (a), RSP11-20 (b), RSP21-30 (c) e manganês (d), sendo que para todas as variáveis foi considerada a produtividade na estrutura espacial bivariada.

Tabela 25 Critérios utilizados para a escolha do melhor modelo ajustado para o modelo univariado

Variáveis	Modelo	EM	σ_{EM}	ER	σ_{ER}	EA	AIC	BIC
Prod	Esférico	$-9,80 \cdot 10^{-4}$	$5,86 \cdot 10^{-1}$	$-8,50 \cdot 10^{-4}$	$1,01 \cdot 10^0$	$4,52 \cdot 10^1$	$1,82 \cdot 10^2$	$1,92 \cdot 10^2$
RSP0-10	Gaussiano	$-2,12 \cdot 10^0$	$2,61 \cdot 10^3$	$-4,01 \cdot 10^{-4}$	$1,01 \cdot 10^0$	$2,06 \cdot 10^5$	$1,90 \cdot 10^3$	$1,91 \cdot 10^3$
RSP11-20	Gaussiano	$-2,81 \cdot 10^0$	$2,27 \cdot 10^3$	$-6,01 \cdot 10^{-4}$	$1,01 \cdot 10^0$	$1,87 \cdot 10^5$	$1,88 \cdot 10^3$	$1,89 \cdot 10^3$
RSP21-30	Matérn k=2	$1,19 \cdot 10^{-1}$	$1,45 \cdot 10^3$	$-6,8 \cdot 10^{-5}$	$1,01 \cdot 10^0$	$9,81 \cdot 10^4$	$1,78 \cdot 10^3$	$1,79 \cdot 10^3$
Mn	Esférico	$-2,553 \cdot 10^{-1}$	$1,79 \cdot 10^1$	$-7,20 \cdot 10^{-3}$	$9,88 \cdot 10^{-1}$	$1,4 \cdot 10^3$	$8,81 \cdot 10^2$	$8,91 \cdot 10^2$

Tabela 26 Critérios utilizados para a escolha do melhor modelo ajustado para o semivariograma cruzado

Variáveis	Modelo	EM	σ_{EM}	ER	σ_{ER}	EA
Prod	Gaussiano	$1,90 \cdot 10^{-3}$	$6,02 \cdot 10^{-1}$	$7,11 \cdot 10^{-3}$	$1,93 \cdot 10^0$	$4,60 \cdot 10^1$
RSP0-10	Gaussiano	$1,50 \cdot 10^0$	$2,66 \cdot 10^3$	$2,08 \cdot 10^{-6}$	$4,01 \cdot 10^{-4}$	$2,05 \cdot 10^5$
Prod	Matérn k=2	$1,70 \cdot 10^{-3}$	$6,03 \cdot 10^{-1}$	$-6,61 \cdot 10^{-3}$	$1,93 \cdot 10^0$	$4,62 \cdot 10^1$
RSP11-20	Matérn k=2	$-2,68 \cdot 10^{-2}$	$2,34 \cdot 10^3$	$1,74 \cdot 10^{-6}$	$4,01 \cdot 10^{-4}$	$1,85 \cdot 10^5$
Prod	Esférico	$6,01 \cdot 10^{-4}$	$6,17 \cdot 10^{-1}$	$-1,71 \cdot 10^{-2}$	$2,09 \cdot 10^0$	$4,77 \cdot 10^1$
RSP21-30	Esférico	$-9,24 \cdot 10^0$	$1,49 \cdot 10^3$	$2,01 \cdot 10^{-5}$	$8,01 \cdot 10^{-4}$	$9,85 \cdot 10^4$
Prod	Matérn k=2	$3,01 \cdot 10^{-4}$	$6,04 \cdot 10^{-1}$	$7,58 \cdot 10^{-5}$	$1,94 \cdot 10^0$	$4,62 \cdot 10^1$
Mn	Matérn k=2	$-7,33 \cdot 10^{-2}$	$1,92 \cdot 10^{-1}$	$4,01 \cdot 10^{-4}$	$5,69 \cdot 10^{-2}$	$1,49 \cdot 10^3$

Tabela 27 Critérios utilizados para a escolha do melhor modelo ajustado para o modelo BGCCM

Variáveis	Modelo	EM	σ_{EM}	ER	σ_{ER}	EA
Prod	Matérn k=2	$9,01 \cdot 10^{-4}$	$5,97 \cdot 10^{-1}$	$7,91 \cdot 10^1$	$2,87 \cdot 10^0$	$4,64 \cdot 10^1$
RSP0-10	Matérn k=2	$-7,13 \cdot 10^1$	$9,67 \cdot 10^3$	$-4,82 \cdot 10^6$	$9,95 \cdot 10^5$	$3,58 \cdot 10^5$
Prod	Matérn k=2	$5,40 \cdot 10^{-3}$	$6,09 \cdot 10^{-1}$	$-2,08 \cdot 10^{-1}$	$3,21 \cdot 10^0$	$4,76 \cdot 10^1$
RSP11-20	Matérn k=2	$2,89 \cdot 10^2$	$9,10 \cdot 10^3$	$-1,82 \cdot 10^3$	$1,30 \cdot 10^4$	$3,76 \cdot 10^5$
Prod	Matérn k=2	$1,40 \cdot 10^{-3}$	$5,94 \cdot 10^{-1}$	$-1,12 \cdot 10^{-1}$	$2,15 \cdot 10^0$	$4,60 \cdot 10^1$
RSP21-30	Matérn k=2	$-3,55 \cdot 10^1$	$5,76 \cdot 10^3$	$3,06 \cdot 10^5$	$3,01 \cdot 10^{-4}$	$1,88 \cdot 10^5$
Prod	Matérn k=2	$-3,94 \cdot 10^{-2}$	$6,83 \cdot 10^{-1}$	$-1,46 \cdot 10^0$	$8,51 \cdot 10^0$	$4,93 \cdot 10^1$
Mn	Matérn k=2	$-5,72 \cdot 10^0$	$3,43 \cdot 10^1$	$-5,46 \cdot 10^0$	$5,59 \cdot 10^0$	$1,26 \cdot 10^3$

Considerando as estimativas dos parâmetros do modelo geoestatístico univariado (Tabela 30) e as estimativas dos parâmetros do modelo ajustado para o semivariograma cruzado (Tabela 31) notou-se que o efeito pepita relativo para as variáveis foi considerado como moderada dependência espacial, exceto para a Prod no ajuste de modelo univariado e o par de variáveis Mn e Prod no ajuste de modelo ao semivariograma cruzado, que apresenta forte dependência espacial segundo classificação de Cambardella et al. (1994).

Os alcances obtidos para o modelo univariado representam a distância em que os pontos amostrais estão correlacionados entre si. Assim, para as variáveis Prod, RSP0-10, RSP11-20, RSP21-30 e Mn tem-se que o raio máximo de dependência espacial entre amostras é igual a 87, 75, 259,17, 402,08, 216,68 e 222,87 m, respectivamente.

No caso do semivariograma cruzado o alcance representa a distância máxima de dependência espacial entre as variáveis, neste caso, a correlação entre a variável Prod e as propriedades químicas, RSP0-10, RSP11-20, RSP21-30 e Mn, foram respectivamente 40, 400, 200 e 200 m.

Além disso, para o modelo BGCCM o raio de dependência espacial referente ao componente que define a estrutura espacial comum as duas variáveis (ϕ_0) para o par de variáveis Prod e Mn foi inferior ao raio de dependência espacial obtido no semivariograma cruzado.

Nota-se na Tabela 30 que os parâmetros de dispersão do modelo BGCCM para a variável 1 são valores pequenos, sendo que tais valores indicam que o componente que define a estrutura de dependência espacial comum para a variável 1 é baixo. No entanto, valores pequenos de efeito pepita e patamar também foram observados nos parâmetros estimados do modelo univariado para a variável Prod. Esses resultados podem ter sido influenciados pela escala de valores desta variável.

O alcance ϕ_0 que define o raio de dependência espacial entre os pares de variáveis Prod e RSP0-10, Prod e RSP11-20, Prod e RSP21-30 e Prod e Mn é de, respectivamente, 652,67, 2937,36, 211,98 e 86,09 m. No entanto, o raio de dependência espacial do

componente independente que define a estrutura de dependência espacial das variáveis 1 e 2 são menores que o raio de dependência espacial entre os pares de variáveis, indicando pequeno raio de dependência espacial de cada variável entre localizações distintas.

Tabela 28 Modelos espaciais ajustados para o modelo univariado e os parâmetros estimados pelo método de máxima verossimilhança

Variáveis	Modelo	a	φ_1	φ_2	$\varphi_1 + \varphi_2$	EPR (%)
Prod	Esférico	$8,77 \cdot 10^1$	$0,00 \cdot 10^0$	$3,37 \cdot 10^{-1}$	$3,37 \cdot 10^{-1}$	$0,00 \cdot 10^0$
RSP0-10	Gaussiano	$2,59 \cdot 10^2$	$4,80 \cdot 10^6$	$2,54 \cdot 10^6$	$7,34 \cdot 10^6$	$6,53 \cdot 10^1$
RSP11-20	Gaussiano	$4,02 \cdot 10^2$	$4,06 \cdot 10^6$	$2,23 \cdot 10^6$	$6,30 \cdot 10^6$	$6,45 \cdot 10^1$
RSP21-30	Matérn k=2	$2,16 \cdot 10^2$	$7,73 \cdot 10^5$	$1,57 \cdot 10^6$	$2,34 \cdot 10^6$	$3,29 \cdot 10^1$
Mn	Esférico	$2,22 \cdot 10^2$	$1,03 \cdot 10^2$	$2,92 \cdot 10^2$	$3,96 \cdot 10^2$	$2,61 \cdot 10^1$

a = alcance, φ_1 = efeito pepita, φ_2 = contribuição, $\varphi_1 + \varphi_2$ = patamar e EPR = efeito pepita relativo.

Tabela 29 Modelos espaciais ajustados para o semivariograma cruzado e os parâmetros estimados pelo método de mínimos quadrados

Variáveis	Modelo	a	φ_1	φ_2	$\varphi_1 + \varphi_2$	EPR (%)
Prod e RSP0-10	Gaussiano	400	-54,2167	-35,3139	89,5306	60,56
Prod e RSP11-20	Matérn k=2	400	-35,2704	82,1389	117,4093	30,04
Prod e RSP21-30	Esférico	200	60,7739	-22,4052	83,1791	73,06
Prod e Mn	Matérn k=2	200	0,1925	1,0936	1,2861	14,97

a = alcance, φ_1 = efeito pepita, φ_2 = contribuição, $\varphi_1 + \varphi_2$ = patamar e EPR = efeito pepita relativo.

Tabela 30 Modelos espaciais ajustados para o modelo BGCCM e os parâmetros estimados pelo método de máxima verossimilhança

	Prod e RSP0-10	Prod e RSP11-20	Prod e RSP21-30	Prod e Mn
Modelo	Gaussiano	Matérn k=2	Matérn k=2	Matérn k=2
μ_1	4,1556	4,2013	4,1667	4,2066
μ_2	8144,7138	10421,4091	7024,8416	60,4894
σ_{01}	$4,11 \cdot 10^{-2}$	$5,54 \cdot 10^{-2}$	$2,61 \cdot 10^{-1}$	$2,21 \cdot 10^{-3}$
σ_1	$3,38 \cdot 10^{-1}$	$3,41 \cdot 10^{-1}$	$3,47 \cdot 10^{-1}$	$3,36 \cdot 10^{-1}$
σ_{02}	$1,74 \cdot 10^4$	$9,31 \cdot 10^2$	$7,33 \cdot 10^5$	$1,49 \cdot 10^1$
σ_2	$7,25 \cdot 10^6$	$6,13 \cdot 10^6$	$2,31 \cdot 10^6$	$3,86 \cdot 10^2$
ϕ_0	652,6689	2937,3578	2111,9816	86,0952
ϕ_1	40,4370	14,8575	15,5323	13,5531
ϕ_2	56,6405	29,6861	28,9621	28,0050

Por meio das estimativas dos parâmetros do modelo univariado, semivariograma cruzado e modelo BGCCM realizaram-se a estimação espacial para localizações não amostradas das variáveis em estudo e construídos os mapas de variabilidade (Figura 32). Para comparar a predição espacial obtida por esses três modelos foram utilizadas as medidas de acurácia (Tabelas 33 a 35).

Observam-se maiores valores para a exatidão global na estimação espacial para RSP21-30, ao comparar os três modelos; no entanto, somente ao comparar o modelo univariado com o semivariograma cruzado à exatidão global foi acima de 85% e os índices

de Kappa e Tau (Tabela 33) foram acima de 67%, indicando similaridade entre os mapas, segundo classificação de Krippendorff (1980).

Tabela 31 Medidas de acurácia para o modelo univariado e semivariograma cruzado

Variáveis	Intervalo	EG(%)	Kp(%)	T(%)
RSP0-10	2425-15125	63,44	45,40	59,38
RSP11-20	6230-18030	60,81	45,16	56,46
RSP21-30	4290-15690	92,54	87,42	91,71
Mn	19-139	51,32	27,22	45,91

EG = exatidão global, Kp = índice Kappa, T = índice Tau.

Tabela 32 Medidas de acurácia para o modelo univariado e BGCCM

Variáveis	Intervalo	EG(%)	Kp(%)	T(%)
RSP0-10	2425-15125	51,44	31,27	46,05
RSP11-20	6230-18030	40,38	9,84	33,75
RSP21-30	4290-15690	61,82	35,22	57,58
Mn	19-139	54,41	20,26	49,35

EG = exatidão global, Kp = índice Kappa, T = índice Tau.

Tabela 33 Medidas de acurácia para o semivariograma cruzado e modelo BGCCM

Variáveis	Intervalo	EG(%)	Kp(%)	T(%)
RSP0-10	2425-15125	44,13	19,64	37,93
RSP11-20	6230-18030	39,43	5,52	32,69
RSP21-30	4290-15690	59,50	30,23	55,00
Mn	19-139	49,76	12,09	44,18

EG = exatidão global, Kp = índice Kappa, T = índice Tau.

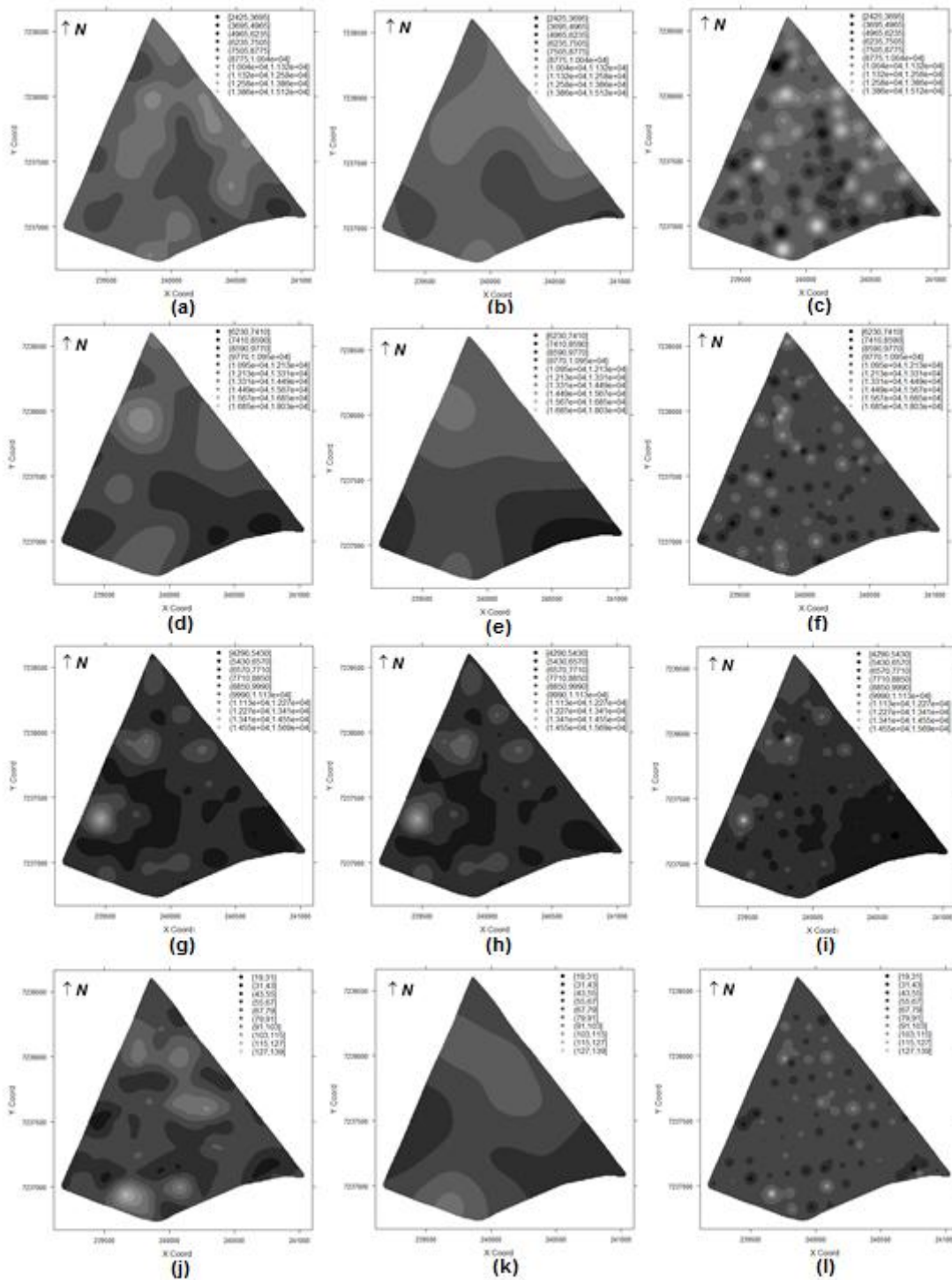


Figura 32 Mapas de variabilidade RSP0-10/U (a), RSP0-10/SC (b), RSP0-10/BGCCM (c), RSP11-20/U (d), RSP11-20/SC (e), RSP11-20/BGCCM (f), RSP21-30/U (g), RSP21-30/SC (h), RSP21-30/BGCCM (i), Mn/U (j), Mn/SC (k) e Mn/BGCCM (l).

5.5 Comparação de modelos com dados experimentais co-locados e desbalanceados do ano agrícola 2013/2014

Com o intuito de avaliar a predição espacial obtida pelo modelo univariado e o modelo BGCCM para amostras co-locadas e desbalanceadas, foram omitidos 50% dos dados do ano agrícola 2013/2014, assim a RSP0-10, RSP11-20, RSP21-30 e Mn possuem

um total de 50 pontos amostrais e a produtividade, variável menos dispendiosa, possui 100 pontos.

Analisando as medidas descritivas (Tabela 36) das variáveis verificou-se que a RSP0-10 e Mn apresentaram a maior dispersão, considerada alta segundo classificação de Pimentel-Gomes (2000), enquanto que as variáveis Prod, RSP11-20 e RSP21-30 obtiveram menor dispersão. Além disso, observa-se na Figura 33 a presença de pontos discrepantes para a variável RSP21-30.

Tabela 34 Medidas descritivas das variáveis produtividade, resistência do solo a penetração e manganês

Estatísticas	Variáveis				
	Prod	RSP0-10	RSP11-20	RSP21-30	Mn
n	100	50	50	50	50
Mínimo	2,590	2428	5744	4809	29,00
Máximo	5,765	15130	17660	15670	106,00
Média	4,212	8424	11070	7380	62,94
Mediana	4,242	8374	11030	6979	60,00
Variância	4,582	7238140	6418456	3104278	374,711
Desvio padrão	0,584	2690,379	2533,467	1761,896	19,357
Coefficiente de variação	13,873	31,937	22,889	23,874	30,755
Curtose	0,273	-0,194	-0,251	8,115	-0,617
Assimetria	-0,172	0,197	0,151	2,379	0,394
p-valor	0,7904 ^{ns}	0,9091 ^{ns}	0,9944 ^{ns}	<0,01*	0,2521

n = número de parcelas observadas, *significativo a 5% de probabilidade, ns = não significativo a 5% de probabilidade.

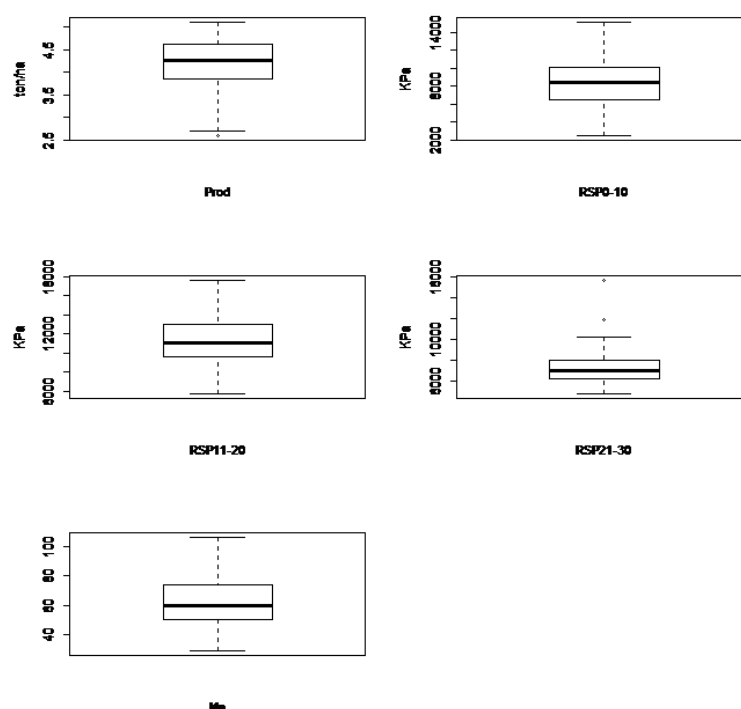


Figura 33 Gráfico *Boxplot* para as variáveis produtividade, RSP0-10, RSP11-20, RSP21-30 e manganês.

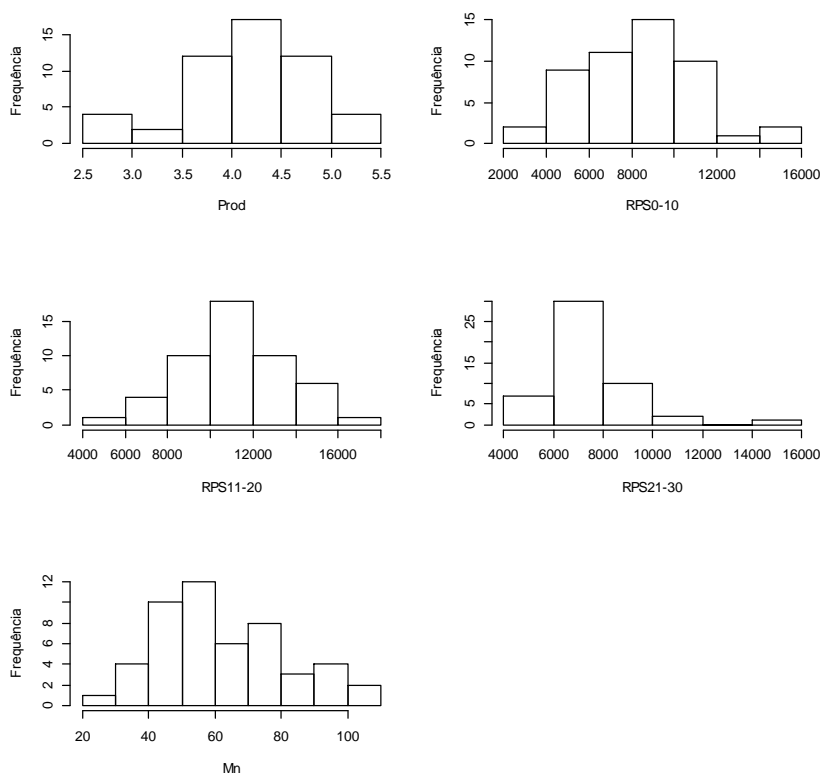


Figura 34 Histograma para as variáveis produtividade, RSP0-10, RSP11-20, RSP21-30 e manganês.

Ainda com o intuito de avaliar o comportamento das propriedades químicas construíram-se os gráficos *Postplot* (Figura 35). Observou-se na Figura (35-d) na região superior do mapa locais de concentração de maiores valores para o Mn. No entanto, para as quatro propriedades em estudo verificou-se que não houve a presença de anisotropia, pois os gráficos *Postplot* não indicam visualmente qualquer tendência direcional em relação à continuidade espacial.

Por meio do cálculo das funções semivariâncias univariadas, em função da distância, construíram-se os semivariogramas omnidirecionais univariado (Figura 36), ajustaram-se os modelos: esférico, exponencial, gaussiano e família Matérn com K igual a 1, 1,5 e 2, sendo que o melhor modelo para descrever cada função semivariância foi escolhido por meio das técnicas de validação cruzada.

As maiores semivariâncias foram obtidas nos semivariogramas univariado para a RSP0-10, RSP11-20 e RSP21-30 (Figura 36-b, 36-c e 36-d) devido à escala de valores das RSP ser maior que a escala da Prod e Mn.

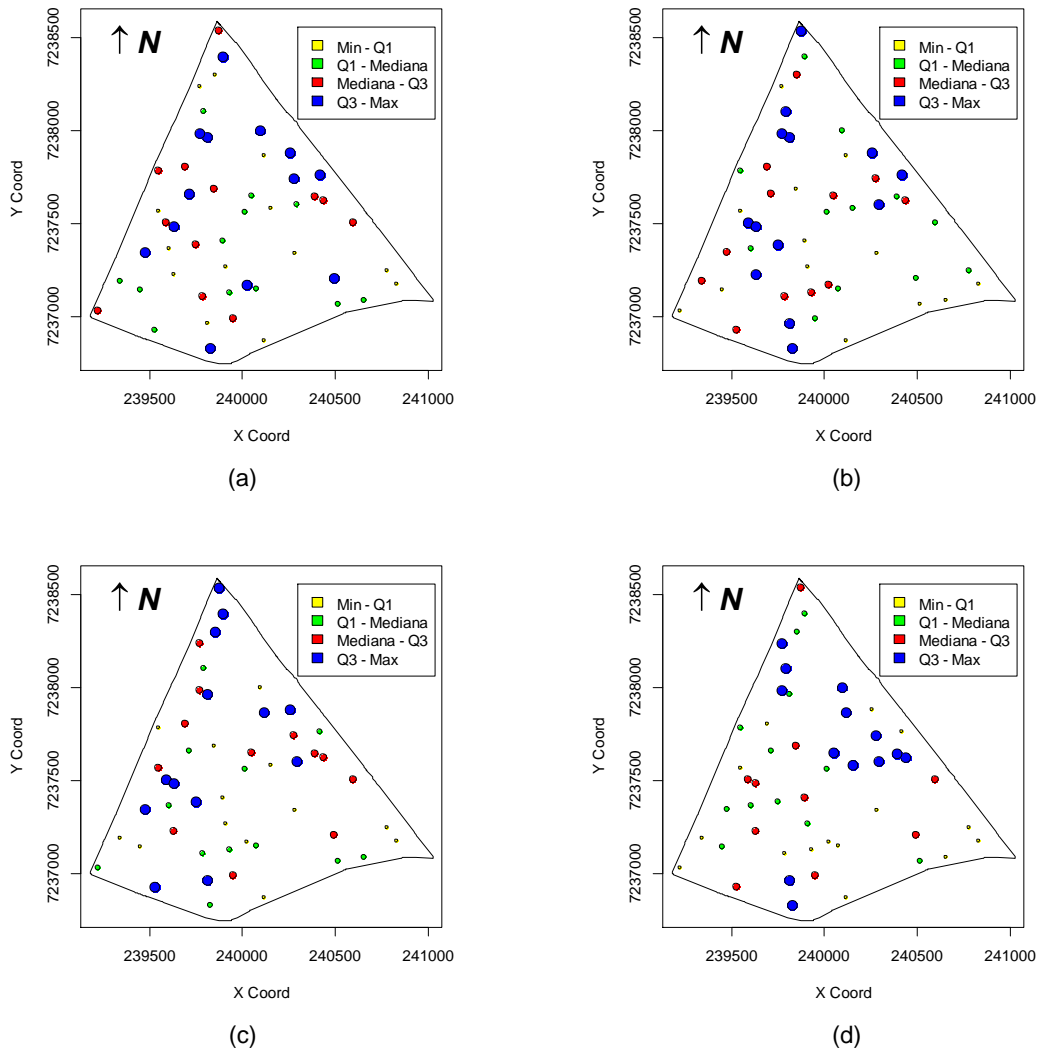


Figura 35 Gráfico *Postplot* para as variáveis RSP0-10 (a), RSP11-20 (b), RSP21-30 (c) e manganês (d).

O modelo que melhor se ajustou para a RSP0-10, RSP11-20, RSP21-30 e Mn respectivamente foram esférico, gaussiano, esférico e gaussiano com parâmetros estimados por máxima verossimilhança (ML), para a escolha do melhor modelo foi levado em consideração o erro médio (EM), desvio padrão do erro médio (σ_{EM}), erro relativo (ER), erro médio relativo (σ_{ER}), erro absoluto (EA), Critério de Informação de Akaike (AIC) e critério de Informação Bayesiano (BIC) (Tabelas 37 e 38).

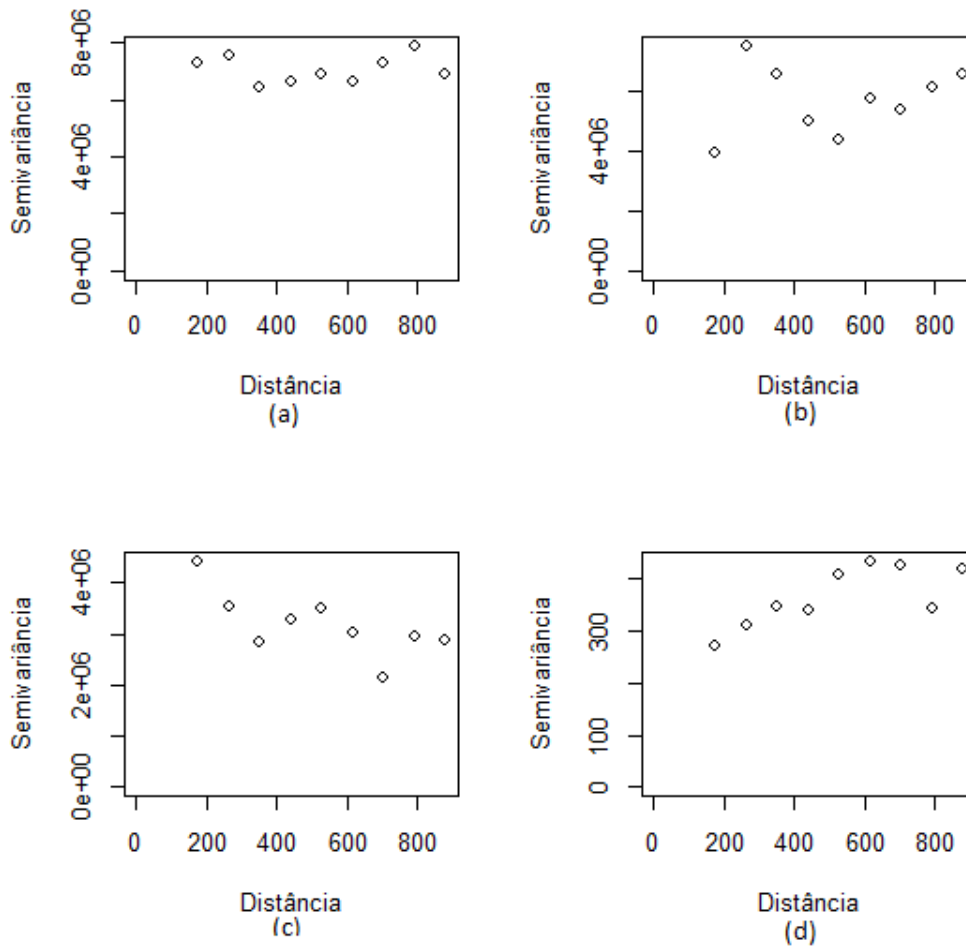


Figura 36 Semivariograma experimental univariado para as variáveis RSP0-10 (a), RSP11-20 (b), RSP21-30 (c) e manganês (d).

Nota-se na Tabela 37 que os parâmetros de dispersão do modelo BGCCM para a variável 1 são valores pequenos, sendo que tais valores indicam que o componente que define a estrutura de dependência espacial comum para a variável 1 é baixo. No entanto, valores pequenos de efeito pepita e patamar também foram observados nos parâmetros estimados do modelo univariado para a variável Prod. Esses resultados podem ter sido influenciados pela escala de valores desta variável.

O alcance ϕ_0 que define o raio de dependência espacial entre os pares de variáveis Prod e RSP0-10 e Prod e Mn são de respectivamente 0,006 m e 0,02 m, indicando pequeno raio de dependência espacial entre esses pares de variáveis, enquanto que para as variáveis Prod e RSP11-20 e Prod e RSP21-30 são, respectivamente, 367,09 e 15,87 m. Esses resultados indicam que houve dependência espacial entre Prod e RSP11-20 e entre Prod e RSP21-30.

Tabela 35 Critérios utilizados para a escolha do melhor modelo ajustado para o modelo univariado

Variável	Modelo	EM	σ_{EM}	ER	σ_{ER}	EA	AIC	BIC
Prod	Esférico	$-6,90 \cdot 10^{-3}$	$6,15 \cdot 10^{-1}$	$-5,95 \cdot 10^{-3}$	$1,01 \cdot 10^0$	$2,41 \cdot 10^1$	$1,00 \cdot 10^2$	$1,08 \cdot 10^2$
RSP0-10	Esférico	$-3,99 \cdot 10^2$	$2,64 \cdot 10^3$	$-8,09 \cdot 10^{-3}$	$1,01 \cdot 10^0$	$1,04 \cdot 10^5$	$9,36 \cdot 10^2$	$9,44 \cdot 10^2$
RSP11-20	Gaussiano	$-2,49 \cdot 10^2$	$2,50 \cdot 10^3$	$-5,10 \cdot 10^{-3}$	$1,01 \cdot 10^0$	$1,01 \cdot 10^5$	$9,31 \cdot 10^2$	$9,38 \cdot 10^2$
RSP21-30	Esférico	$-7,36 \cdot 10^0$	$1,78 \cdot 10^3$	$-2,14 \cdot 10^{-3}$	$1,01 \cdot 10^0$	$6,01 \cdot 10^4$	$8,96 \cdot 10^2$	$9,03 \cdot 10^2$
Mn	Gaussiano	$3,61 \cdot 10^{-2}$	$1,67 \cdot 10^1$	$-1,14 \cdot 10^{-3}$	$1,01 \cdot 10^0$	$6,87 \cdot 10^2$	$4,37 \cdot 10^2$	$4,45 \cdot 10^2$

Tabela 36 Critérios utilizados para a escolha do melhor modelo ajustado para o modelo BGCCM

Variável	Modelo	EM	σ_{EM}	ER	σ_{ER}	EA
Prod	Matérn k=2	$9,00 \cdot 10^4$	$5,97 \cdot 10^{-1}$	$7,91 \cdot 10^{-1}$	$2,87 \cdot 10^0$	$4,64 \cdot 10^1$
RSP0-10	Matérn k=2	$-7,13 \cdot 10^2$	$9,67 \cdot 10^3$	$-4,82 \cdot 10^6$	$9,95 \cdot 10^5$	$3,58 \cdot 10^5$
Prod	Matérn k=2	$5,40 \cdot 10^{-3}$	$6,09 \cdot 10^{-1}$	$-2,08 \cdot 10^{-1}$	$3,21 \cdot 10^0$	$4,76 \cdot 10^1$
RSP11-20	Matérn k=2	$2,89 \cdot 10^2$	$9,10 \cdot 10^3$	$-1,82 \cdot 10^3$	$1,30 \cdot 10^4$	$3,76 \cdot 10^5$
Prod	Matérn k=2	$1,40 \cdot 10^{-3}$	$5,94 \cdot 10^{-1}$	$-1,12 \cdot 10^{-1}$	$2,15 \cdot 10^0$	$4,60 \cdot 10^1$
RSP21-30	Matérn k=2	$-3,55 \cdot 10^1$	$5,76 \cdot 10^3$	$3,06 \cdot 10^5$	$3,01 \cdot 10^{-4}$	$1,88 \cdot 10^5$
Prod	Matérn k=2	$-3,94 \cdot 10^{-2}$	$6,83 \cdot 10^{-1}$	$-1,46 \cdot 10^0$	$8,51 \cdot$	$4,93 \cdot 10^1$
Mn	Matérn k=2	$-5,72 \cdot 10^0$	$3,43 \cdot 10^1$	$-0,54 \cdot 10^0$	$5,59 \cdot 10^0$	$1,26 \cdot 10^3$

Para as variáveis Prod e RSP0-10 o efeito pepita relativo apresentou uma forte dependência espacial, enquanto que RSP11-20, RSP21-30 e Mn moderada dependência espacial (Tabela 39). Os alcances obtidos para o modelo univariado representam a distância em que os pontos amostrais estão correlacionados entre si. Assim, para as variáveis Prod, RSP0-10, RSP11-20, RSP21-30 e Mn tem-se que o raio máximo de dependência espacial entre amostras é igual a 118,2, 139, 55, 287,07, 149,74 e 392,34 m, respectivamente.

Para o modelo BGCCM tem-se que os valores do raio de dependência espacial referente ao componente que define a estrutura espacial comum as duas variáveis (ϕ_0) para o par de variáveis Prod e RSP0-10 e Prod e Mn foram pequenos indicando baixo raio de dependência entre as variáveis (Tabela 40). Além disso, os valores obtidos para o parâmetro de dispersão para a variável 1 foram todos pequenos e tais parâmetros estão associados a S_0 e S_1 , indicando que o componente que define a estrutura de dependência espacial comum para a variável 1 é baixo e também que a estrutura de dependência espacial para 1 pode ser explicada pela média das respectivas variáveis.

Tabela 37 Modelos espaciais ajustados para o modelo univariado e os parâmetros estimados pelo método de máxima verossimilhança

Variáveis	Modelo	a	φ_1	φ_2	$\varphi_1 + \varphi_2$	EPR (%)
Prod	Esférico	118,1968	0,0	0,3658	0,3658	0,0
RSP0-10	Esférico	139,5496	0,0	7020217,6	7020217,6	0,0
RSP11-20	Gaussiano	287,0686	4402736,0	1901273,9	6304009	69,84
RSP21-30	Esférico	149,7422	1489640,9	1554618,2	3044259,1	48,93
Mn	Gaussiano	392,3396	182,07	187,40	369,47	49,28

a = alcance, φ_1 = efeito pepita, φ_2 = contribuição, $\varphi_1 + \varphi_2$ = patamar e EPR = efeito pepita relativo.

Tabela 38 Modelos espaciais ajustados para o modelo BGCCM e os parâmetros estimados pelo método de máxima verossimilhança

	Prod e RSP0-10	Prod e RSP11-20	Prod e RSP21-30	Prod e Mn
Modelo	Matérn k=2	Matérn k=2	Matérn k=2	Matérn k=2
μ_1	4,212	4,109	4,209	4,209
μ_2	8277,401	10702,501	7381,601	60,382
σ_{01}	$1,16 \cdot 10^{-1}$	$1,15 \cdot 10^{-1}$	$1,46 \cdot 10^{-1}$	$9,89 \cdot 10^{-3}$
σ_1	$3,04 \cdot 10^{-1}$	$3,44 \cdot 10^{-1}$	$2,09 \cdot 10^{-1}$	$3,41 \cdot 10^{-1}$
σ_{02}	$3,52 \cdot 10^6$	$1,15 \cdot 10^6$	$1,27 \cdot 10^6$	$3,05 \cdot 10^2$
σ_2	$3,05 \cdot 10^6$	$5,42 \cdot 10^6$	$1,92 \cdot 10^6$	$1,19 \cdot 10^2$
\emptyset_0	0,006	367,091	15,865	0,023
\emptyset_1	0,695	0,138	13,043	13,869
\emptyset_2	28,215	16,844	6,305	430,544

Por meio das estimativas dos parâmetros do modelo univariado e BGCCM realizaram-se as predições espaciais para localizações não amostradas das variáveis em estudo, utilizando, respectivamente, a krigagem e cokrigagem ordinária.

Observa-se que os índices de Kappa e Tau para a predição espacial da RSP21-30 foi acima de 67%, segundo classificação de Krippendorff (1980), indica similaridade entre os mapas quando comparando as estimações espaciais entre o modelo univariado e o modelo BGCCM para amostras co-localizadas e desbalanceadas. A predição espacial para o Mn apresentou maior dissimilaridade entre os mapas, conforme pode ser observado na Tabela 41 e Figura 37.

Tabela 39 Medidas de acurácia para o modelo univariado e BGCCM

Variáveis	Intervalo	EG (%)	Kp (%)	T (%)
RSP0-10	2460-15060	65,35	7,39	61,49
RSP11-20	6480-17330	52,98	17,82	47,76
RSP21-30	5430-13380	77,43	8,46	74,92
Mn	40-90	36,45	25,05	29,38

EG = exatidão global, Kp = índice Kappa, T = índice Tau.

A fim de comparar as abordagens univariada e bivariada quanto à predição espacial foi realizada a predição para as 50 localizações omitidas inicialmente das variáveis RSP e Mn. O modelo univariado se mostrou mais eficiente na predição espacial para as variáveis RSP, uma vez que, a soma quadrada do erro da predição espacial foi menor. Para a

variável Mn, observa-se uma similaridade quanto à eficiência da predição espacial usando os modelos univariado e bivariado (Tabela 42 e Figura 38).

A má qualidade da predição espacial pelo modelo bivariado BGCCM pode estar relacionada com a malha amostral adotada, pois a malha *lattice plus close pairs* foi a que apresentou os piores resultados no estudo com simulações, quando esta foi comparada com a configuração amostral regular e irregular. Além disso, a má qualidade pode estar relacionada também à diminuição de 50% das localizações de uma variável.

Fonseca (2008), trabalhando com dados de pH e saturação por base do solo provenientes de uma propriedade agrícola usando a malha regular, encontrou uma melhor eficiência da predição espacial usando o modelo BGCCM quando comparado com a predição espacial do modelo univariado.

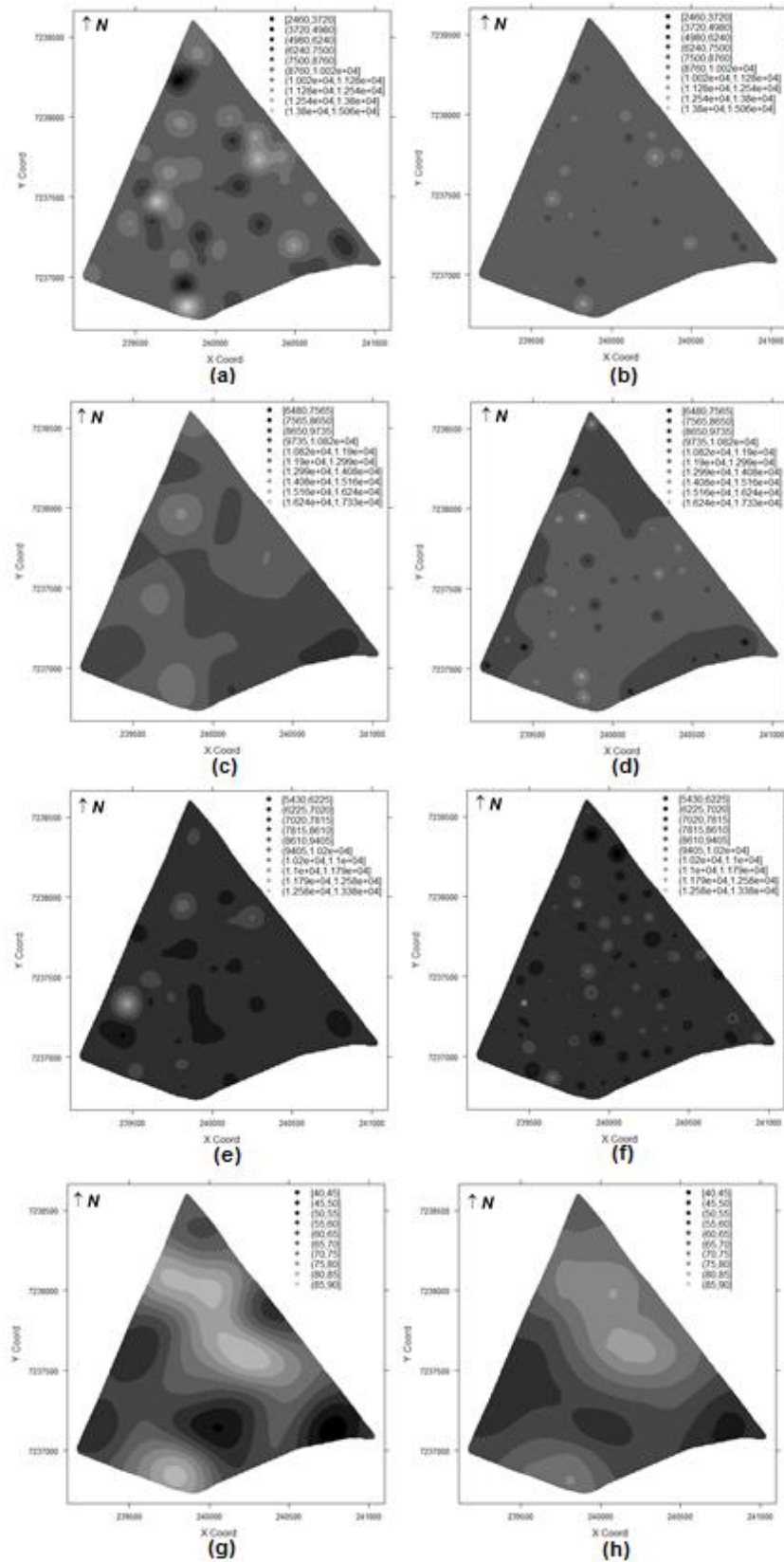


Figura 37 RSP0-10/u (a), RSP0-10/mb (b), RSP11-20/u (c), RSP11-20/mb (d), RSP21-30/u (e), RSP21-30/mb (f), Mn/u (g) e Mn/mb (h).

Tabela 40 Estatística descritiva do erro de predição espacial para o grupo controle

	Mínimo	Máximo	Média	Mediana	DP	SQ
RSP0-10 u	-4376,0	6573,0	-105,4	-419,6	2805,1	$3,86 \cdot 10^8$
RSP0-10 mb	-5151,0	7524,0	-168,3	-692,3	3054,9	$4,58 \cdot 10^8$
RSP11-20 u	-4439,0	4553,0	-1091,0	-1354,0	2132,9	$2,82 \cdot 10^8$
RSP11-20 mb	-4607,0	5873,0	-1175,0	1528,0	2366,4	$3,43 \cdot 10^8$
RSP21-30 u	-2688,0	2520,0	-624,5	-800,7	1135,5	$8,26 \cdot 10^7$
RSP21-30 mb	-2979,0	3211,0	-638,3	-992,0	1545,8	$1,37 \cdot 10^8$
Mn u	-36,5	79,9	-3,5	-8,4	21,5	$2,31 \cdot 10^4$
Mn mb	-37,0	77,7	-2,9	-5,2	21,1	$2,21 \cdot 10^4$

u = modelo univariado, mb = modelo bivariado, DP = desvio padrão, SQ = somas quadradas.

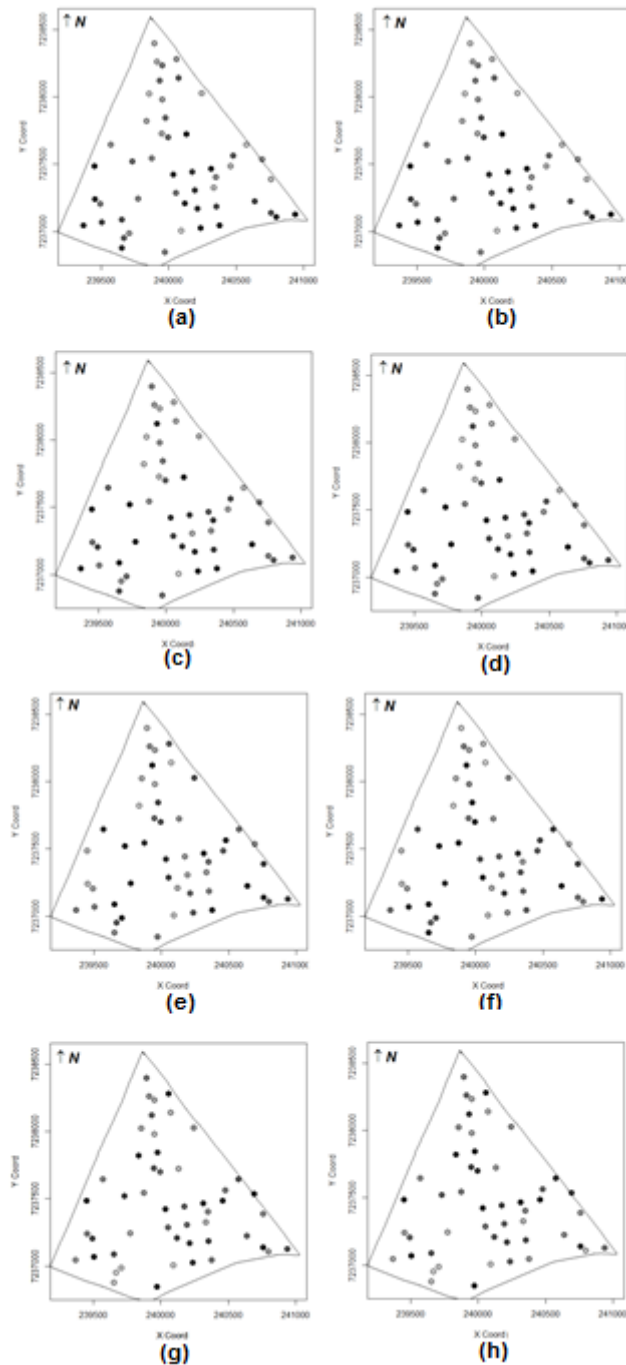


Figura 38 Mapas de erros RSP0-10/u (a), RSP0-10/mb (b), RSP11-20/u (c), RSP11-20/mb (d), RSP21-30/u (e), RSP21-30/mb (f), Mn/u (g) e Mn/mb (h), classificados em intervalos de mesma amplitude.

5.6 Comparação de malhas amostrais para dados co-locados e balanceados

Observou-se que a maioria dos menores valores para as estatísticas descritivas das somas quadrada da diferença da predição espacial foram obtida na comparação das malhas regular e *lattice plus close pairs*, para ambas as variáveis. Além disso, essa medida apresentou a menor variabilidade conforme pode ser observado pelo coeficiente de variação (Tabela 43 e Figura 39), indicando que na comparação entre as três configurações, por meio dessa medida, as que apresentam resultados mais similares quando a predição espacial foram as malhas regular e *lattice plus close pairs*.

Ao comparar as configurações tem-se que a maioria das medidas de acurácia são baixas indicando dissimilaridade entre as predições espaciais (Tabelas 44 a 46). Apenas na comparação das malhas irregular e *lattice plus close pairs* para a variável 1 observa-se 2% das simulações com valores de exatidão global e índice Tau acima de 85% e, também, para a variável 2, que apresentou 1%, conforme Figuras 40 a 42.

Tabela 41 Estatística descritiva da soma quadrada da diferença da predição espacial das variáveis 1 e 2

Estatísticas	Variável 1			Variável 2		
	r vs. i	r vs. L	i vs. L	r vs. i	r vs. L	i vs. L
Mínimo	$3,27 \cdot 10^4$	$1,39 \cdot 10^4$	$1,68 \cdot 10^4$	$1,03 \cdot 10^5$	$1,53 \cdot 10^5$	$6,55 \cdot 10^4$
1° quartil	$1,04 \cdot 10^5$	$5,98 \cdot 10^4$	$1,12 \cdot 10^5$	$5,16 \cdot 10^5$	$4,13 \cdot 10^5$	$5,16 \cdot 10^5$
Mediana	$1,73 \cdot 10^5$	$1,25 \cdot 10^5$	$1,70 \cdot 10^5$	$1,49 \cdot 10^6$	$1,19 \cdot 10^6$	$1,43 \cdot 10^6$
Média	$5,49 \cdot 10^5$	$2,60 \cdot 10^5$	$4,77 \cdot 10^5$	$2,58 \cdot 10^7$	$4,77 \cdot 10^6$	$2,50 \cdot 10^7$
3° quartil	$3,22 \cdot 10^5$	$2,64 \cdot 10^5$	$2,69 \cdot 10^5$	$3,26 \cdot 10^6$	$3,64 \cdot 10^6$	$3,26 \cdot 10^6$
Máximo	$1,59 \cdot 10^7$	$8,01 \cdot 10^6$	$1,59 \cdot 10^7$	$2,09 \cdot 10^9$	$6,21 \cdot 10^7$	$2,08 \cdot 10^9$
DP	$1,84 \cdot 10^6$	$8,03 \cdot 10^5$	$1,67 \cdot 10^6$	$2,09 \cdot 10^8$	$1,71 \cdot 10^7$	$2,08 \cdot 10^8$
CV	$3,35 \cdot 10^2$	$3,08 \cdot 10^2$	$3,50 \cdot 10^2$	$8,10 \cdot 10^2$	$3,58 \cdot 10^2$	$8,32 \cdot 10^2$

r = regular, i = irregular, L = *lattice plus close pairs*, DP = desvio padrão e CV = coeficiente de variação.

Tabela 42 Estatística descritiva da exatidão global das variáveis 1 e 2.

Estatísticas	Variável 1			Variável 2		
	r vs. i	r vs. L	i vs. L	r vs. i	r vs. L	i vs. L
Mínimo	0,0159	0,0272	0,0020	0,0030	0,0147	0,0035
1° quartil	0,0889	0,1347	0,0874	0,1056	0,1253	0,0434
Mediana	0,1231	0,1658	0,1291	0,1613	0,1729	0,1404
Média	0,1269	0,1662	0,1518	0,1641	0,1763	0,1537
3° quartil	0,1653	0,2036	0,1776	0,1987	0,2344	0,2054
Máximo	0,2544	0,3152	0,9846	0,4204	0,3364	0,9450
DP	0,0503	0,0556	0,1291	0,0757	0,0761	0,1448
CV	39,62	33,44	85,03	46,09	43,15	94,19

r = regular, i = irregular, L = *lattice plus close pairs*, DP = desvio padrão e CV = coeficiente de variação.

Tabela 43 Estatística descritiva do índice Kappa das variáveis 1 e 2

Estatísticas	Variável 1			Variável 2		
	r vs. i	r vs. L	i vs. L	r vs. i	r vs. L	i vs. L
Mínimo	-0,0578	0,0576	-0,1464	-0,0726	-0,0446	-0,1504
1° quartil	-0,0136	0,0058	-0,0051	-0,0066	-0,0051	-0,0062
Mediana	0,0052	0,0259	0,0128	0,0000	0,0122	0,0095
Média	0,0047	0,0370	0,0103	0,0041	0,0166	0,0053
3° quartil	0,0213	0,0563	0,0317	0,0201	0,0262	0,0222
Máximo	0,1400	0,1973	0,0895	0,0497	0,2194	0,0661
DP	0,0294	0,0519	0,0363	0,02151	0,0394	0,0276
CV	619,15	140,24	351,83	521,52	236,77	517,08

r = regular, i = irregular, L = *lattice plus close pairs*, DP = desvio padrão e CV = coeficiente de variação.

Tabela 44 Estatística descritiva do índice Tau das variáveis 1 e 2

Estatísticas	Variável 1			Variável 2		
	r vs. i	r vs. L	i vs. L	r vs. i	r vs. L	i vs. L
Mínimo	-0,0935	-0,0809	-0,1089	-0,1077	-0,0948	-0,1072
1° quartil	-0,0124	0,0386	-0,0141	0,0063	0,0281	-0,0628
Mediana	0,0257	0,0731	0,0323	0,0681	0,0810	0,0448
Média	0,0299	0,0735	0,0576	0,0713	0,0847	0,0596
3° quartil	0,0725	0,1152	0,0862	0,1097	0,1493	0,1170
Máximo	0,1715	0,2391	0,9829	0,3561	0,2627	0,9389
DP	0,0559	0,0618	0,4348	0,0841	0,0845	0,1608
CV	186,39	83,95	248,95	117,98	99,76	269,68

r = regular, i = irregular, L = *lattice plus close pairs*, DP = desvio padrão e CV = coeficiente de variação.

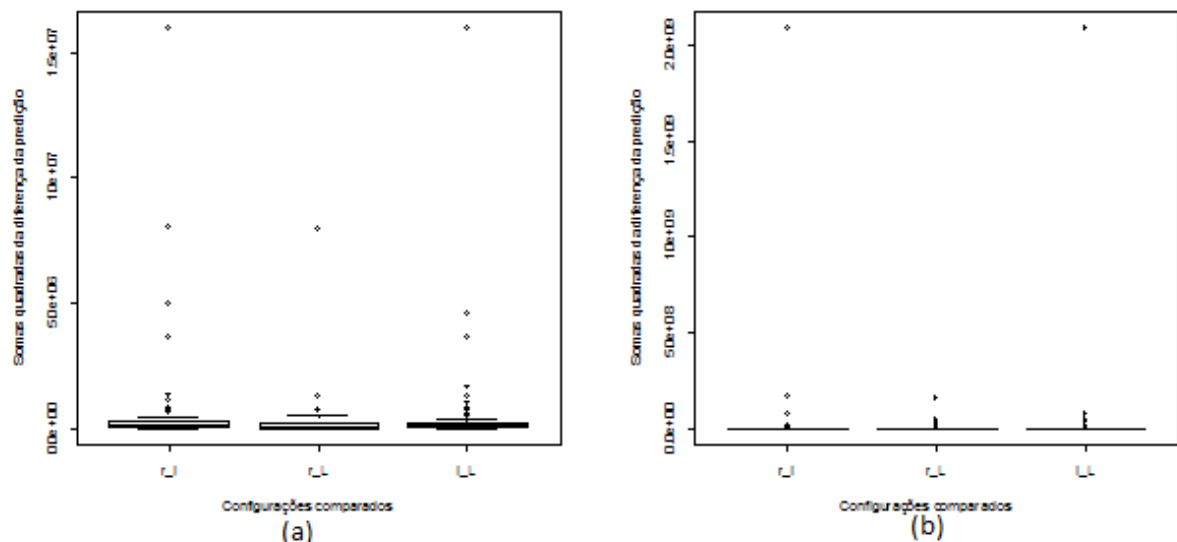


Figura 39 Gráfico *Boxplot* para as somas quadradas da diferença da predição para as variáveis 1 e 2.

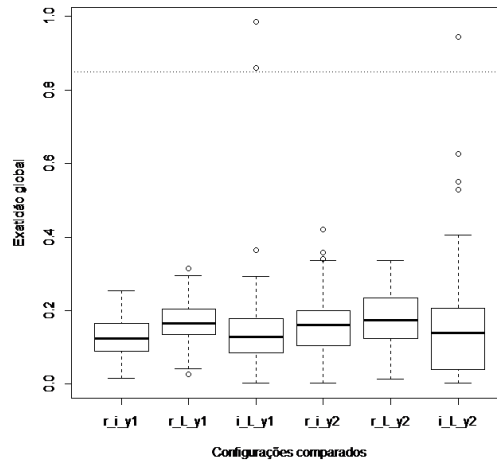


Figura 40 Gráfico *Boxplot* para a exatidão global das variáveis 1 e 2.

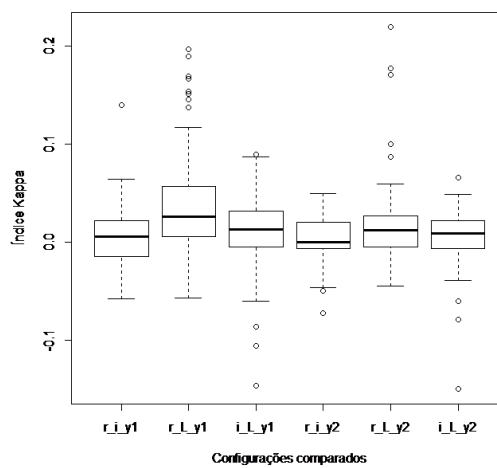


Figura 41 Gráfico *Boxplot* para o índice Kappa das variáveis 1 e 2.

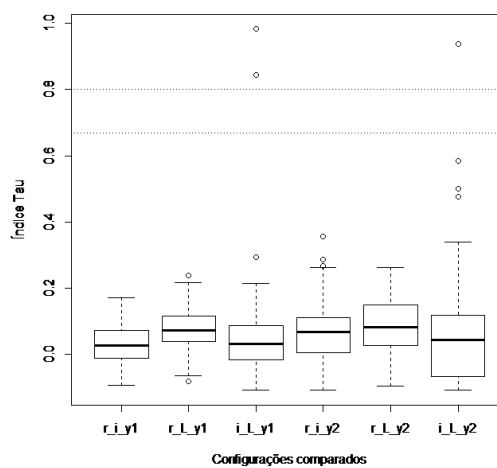


Figura 42 Gráfico *Boxplot* para o índice Tau das variáveis 1 e 2.

Utilizando a subamostra de controle observou-se que a malha regular e irregular apresentaram para a variável Y_1 os menores valores na maioria das medidas de estatística

descritiva das somas quadradas do erro da predição espacial e para Y_2 a malha irregular apresentou os menores valores das medidas descritivas (Tabela 47 e Figura 43)., evidenciando que na malha irregular ocorreu uma predição espacial mais eficiente em comparação com as malhas regular e *lattice plus close pairs*, considerando a análise de dependência espacial bivariada por meio do modelo BGCCM e conjuntos de dados colocados e balanceados.

Além disso, sob essa abordagem, observa-se que a malha amostral *lattice plus close pairs* para ambas as variáveis apresentou os maiores valores para todas as medidas das estatísticas descritivas das somas quadradas do erro da predição espacial (Tabela 47), indicando maior ineficiência quanto à predição espacial.

Tabela 45 Estatística descritiva das somas quadradas do erro da predição espacial para o grupo de controle para ambas as variáveis

Estatísticas	Variável 1			Variável 2		
	r	i	L	r	i	L
Mínimo	0,225	0,569	116,1	176,5	104,7	416,8
1° quartil	0,530	1,161	348,5	690,1	635,6	1142,0
Mediana	111,2	2,928	569,1	1545,0	1296,0	2041,0
Média	236,3	334,9	680,9	9615,0	3634,0	10070,0
3° quartil	256,3	252,2	878,1	4402,0	3531,0	4449,0
Máximo	4154,0	7813,0	4104,0	597600,0	28150,0	602100,0
DV	485,073	955,592	518,605	59668,8	5833,722	60072,5
CV	205,263	285,379	76,164	620,589	160,517	596,3996

r = regular, i = irregular, L = *lattice plus close pairs*, DV = desvio padrão, CV = coeficiente de variação.

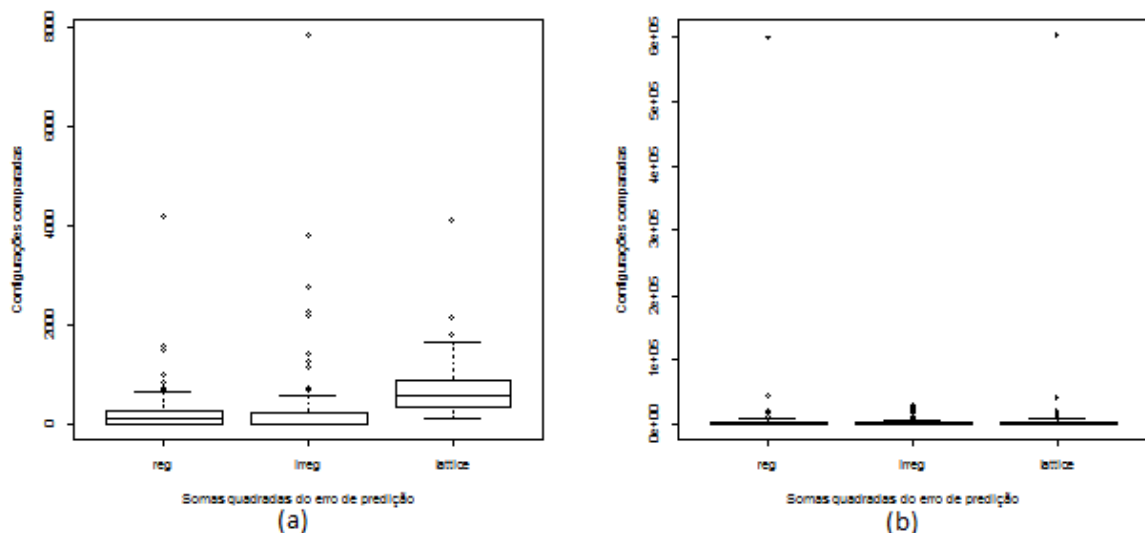


Figura 43 Gráfico *Boxplot* das somas quadradas do erro de predição espacial para o grupo de controle para as variáveis 1 e 2.

5.7 Comparação de malhas amostrais para dados co-locados e desbalanceados

Notou-se nas Tabelas 48 a 51 e Figuras 44 a 47 uma diferença relevante, entre as configurações amostrais, na predição espacial usando o modelo BGCCM, pois, em todas as simulações obtiveram-se baixos valores das medidas de acurácia, com exatidão global inferior a 85% e índices Kappa e Tau inferior a 67%. No entanto, de forma geral, a maioria das simulações apresentam comportamento similar.

Tabela 46 Estatística descritiva da soma quadrada da diferença da predição espacial das variáveis 1 e 2

Estatísticas	Variável 1			Variável 2		
	r vs. i	r vs. L	i vs. L	r vs. i	r vs. L	i vs. L
Mínimo	$3,44 \cdot 10^4$	$1,87 \cdot 10^4$	$2,99 \cdot 10^4$	$7,17 \cdot 10^4$	$6,92 \cdot 10^4$	$6,84 \cdot 10^4$
1° quartil	$1,08 \cdot 10^5$	$1,18 \cdot 10^5$	$9,56 \cdot 10^4$	$7,19 \cdot 10^5$	$7,60 \cdot 10^5$	$4,47 \cdot 10^5$
Mediana	$1,64 \cdot 10^5$	$1,77 \cdot 10^5$	$1,62 \cdot 10^5$	$2,14 \cdot 10^6$	$2,34 \cdot 10^6$	$1,60 \cdot 10^6$
Média	$2,41 \cdot 10^5$	$2,26 \cdot 10^5$	$2,30 \cdot 10^5$	$6,64 \cdot 10^6$	$4,53 \cdot 10^6$	$6,33 \cdot 10^6$
3° quartil	$2,52 \cdot 10^5$	$2,70 \cdot 10^5$	$2,42 \cdot 10^5$	5,24	$5,95 \cdot 10^6$	$4,32 \cdot 10^6$
Máximo	$2,32 \cdot 10^6$	$2,11 \cdot 10^6$	$2,41 \cdot 10^6$	$1,03 \cdot 10^8$	$2,59 \cdot 10^7$	$9,76 \cdot 10^7$
DP	$2,74 \cdot 10^5$	$2,26 \cdot 10^5$	$3,44 \cdot 10^5$	$1,58 \cdot 10^7$	$5,54 \cdot 10^6$	$1,54 \cdot 10^7$
CV	$1,14 \cdot 10^2$	$9,99 \cdot 10^1$	$1,49 \cdot 10^2$	$2,38 \cdot 10^2$	$1,22 \cdot 10^2$	$2,44 \cdot 10^2$

r = regular, i = irregular, L = *lattice plus close pairs*, DP = desvio padrão e CV = coeficiente de variação.

Tabela 47 Estatística descritiva da exatidão global das variáveis 1 e 2

Estatísticas	Variável 1			Variável 2		
	r vs. i	r vs. L	i vs. L	r vs. i	r vs. L	i vs. L
Mínimo	0,0050	0,0446	0,0021	0,0072	0,0326	0,0128
1° quartil	0,0874	0,0924	0,0866	0,1123	0,1075	0,1244
Mediana	0,1309	0,1269	0,1153	0,1731	0,1628	0,1918
Média	0,1449	0,1390	0,1462	0,1833	0,1831	0,2185
3° quartil	0,1832	0,1696	0,1898	0,2592	0,2469	0,2949
Máximo	0,4242	0,3183	0,5407	0,3972	0,4608	0,5972
DP	0,0849	0,0603	0,1008	0,0908	0,0984	0,1254
CV	58,57	43,39	68,98	49,55	53,74	5738

r = regular, i = irregular, L = *lattice plus close pairs*, DP = desvio padrão e CV = coeficiente de variação.

Tabela 48 Estatística descritiva do índice Kappa das variáveis 1 e 2

Estatísticas	Variável 1			Variável 2		
	r vs. i	r vs. L	i vs. L	r vs. i	r vs. L	i vs. L
Mínimo	-0,1060	-0,0945	-0,0812	-0,1009	0,0678	-0,1303
1° quartil	-0,0099	-0,0207	-0,0032	-0,0090	-0,0096	-0,0171
Mediana	0,0089	0,0078	0,0179	0,0070	0,0115	0,0018
Média	0,0112	0,0065	0,0167	0,0095	0,0142	0,0025
3° quartil	0,0321	0,0291	0,0363	0,0230	0,0371	0,0272
Máximo	0,1681	0,1386	0,1111	0,1467	0,1156	0,0967
DP	0,0434	0,0383	0,0376	0,0319	0,0359	0,0381
CV	388,45	587,39	225,11	336,77	252,39	1526,39

r = regular, i = irregular, L = *lattice plus close pairs*, DP = desvio padrão e CV = coeficiente de variação.

Tabela 49 Estatística descritiva do índice Tau das variáveis 1 e 2

Estatísticas	Variável 1			Variável 2		
	r vs. i	r vs. L	i vs. L	r vs. i	r vs. L	i vs. L
Mínimo	-0,1056	-0,0616	-0,10880	-0,1032	-0,0748	-0,0968
1° quartil	-0,0140	-0,0084	-0,0149	0,0136	0,0084	0,0271
Mediana	0,0342	0,0299	0,0170	0,0812	0,0698	0,1021
Média	0,0499	0,0434	0,0513	0,0926	0,0923	0,1317
3° quartil	0,0924	0,0773	0,0999	0,1768	0,1632	0,2165
Máximo	0,3602	0,2426	0,4897	0,3302	0,4009	0,5524
DP	0,0943	0,0670	0,1120	0,1009	0,1093	0,1393
CV	188,99	154,55	218,42	109,01	118,40	105,81

DP = desvio padrão e CV = coeficiente de variação.

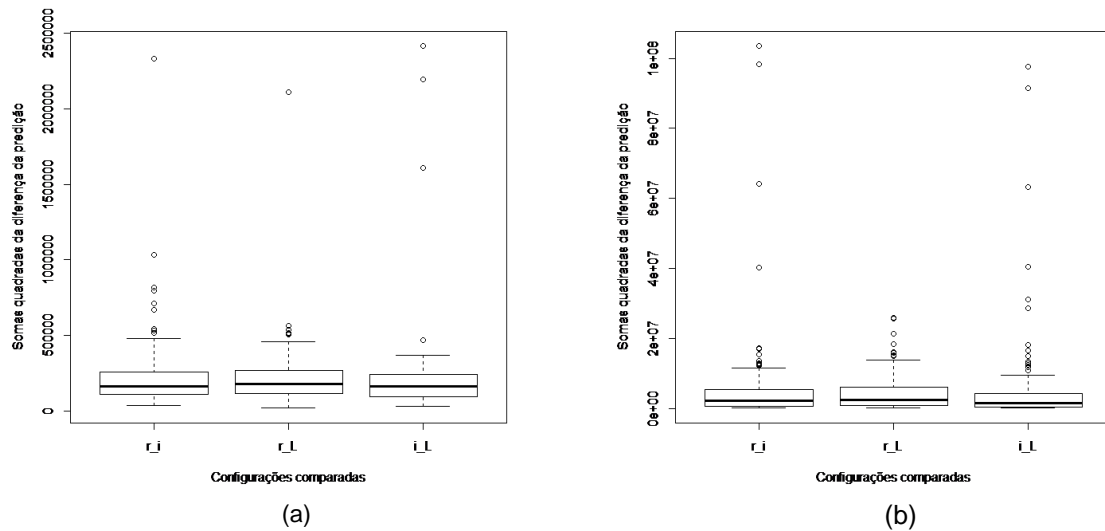


Figura 44 Gráfico *Boxplot* para as somas quadradas da diferença da predição espacial das variáveis 1 e 2.

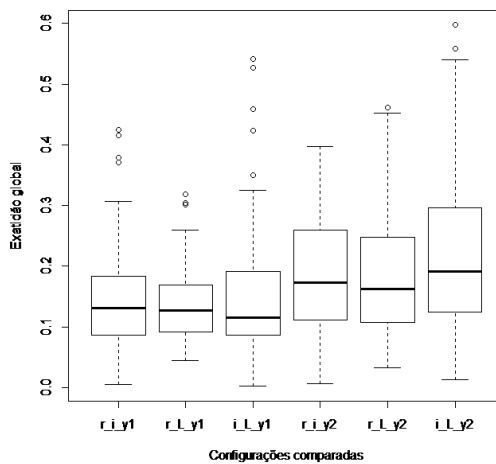


Figura 45 Gráfico *Boxplot* para a exatidão global das variáveis 1 e 2.

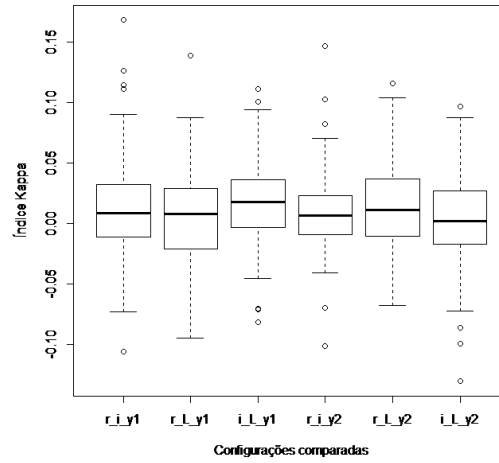


Figura 46 Gráfico *Boxplot* para o índice Kappa das variáveis 1 e 2.

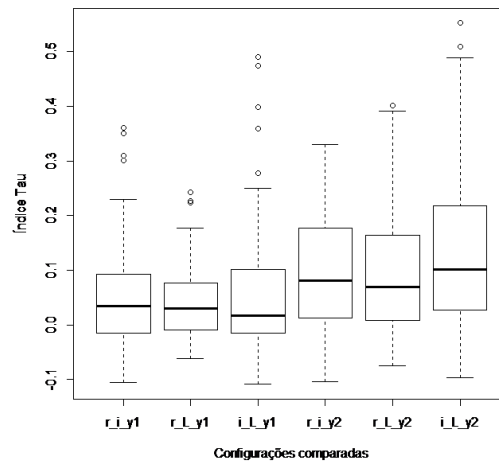


Figura 47 Gráfico *Boxplot* para o índice Tau das variáveis 1 e 2.

6 CONCLUSÕES

O estudo de simulação da análise da influência da variação dos parâmetros do modelo bivariado BGGCM no cálculo do coeficiente de correlação linear de Pearson entre as variáveis descritas do modelo bivariado BGCCM revelou que, para todas as configurações de parâmetros definidas, houve uma alta variabilidade do coeficiente de correlação linear de Pearson. Esse fato evidencia que o coeficiente de correlação linear de Pearson não pode ser considerado na tomada de decisão quanto à presença de dependência espacial conjunta entre pares de variáveis.

No estudo com simulações observou-se pelas medidas de acurácia uma diferença relevante entre as predições espaciais utilizando os modelos univariado, semivariograma cruzado e BGCCM. Além disso, essas medidas de acurácia e a soma quadrada da diferença entre as predições espaciais indicam que as maiores diferenças quanto à predição espacial ocorreram quando se comparou os modelos univariado e semivariograma cruzado com o BGCCM.

Ainda no estudo de simulação observou-se uma diferença relevante entre as predições espaciais utilizando as malhas regular, irregular e *lattice plus close pairs*, tanto para dados balanceados como para os desbalanceados, sendo que para os dados balanceados as malhas regular e irregular apresentaram melhor eficiência quanto à predição espacial.

O estudo com dados reais mostrou que sob a abordagem do modelo BGCCM observou-se a presença de dependência espacial, principalmente entre pH e Mn para dados co-locados e balanceados do ano agrícola 2010/2011; e entre as variáveis dentro dos seguintes pares: (Prod, RSP0-10), (Prod, RSP11-20), (Prod, RSP21-30) e (Prod, Mn) para dados co-locados e balanceados do ano agrícola 2013/2014; e (Prod, RSP11-20) e (Prod, RSP21-30) para dados co-locados e desbalanceados do ano agrícola 2013/2014.

O estudo com dados reais mostrou diferenças na predição espacial tanto para dados balanceados como para os desbalanceados. No entanto, para os dados desbalanceados para a variável Mn observou-se uma similaridade quanto à eficiência da predição espacial usando os modelos univariado e BGCCM.

No entanto, considerando os dados reais e a amostra controle, a qualidade da predição espacial usando o modelo BGCCM se mostrou ineficiente quando comparada com a qualidade obtida na predição espacial usando o modelo univariado. Porém, esse resultado pode ter sido influenciado pela escolha da configuração amostral utilizada, pois no conjunto de dados simulados a configuração *lattice plus close pairs* apresentou os piores resultados quanto à predição espacial quando comparada com os resultados obtidos para as configurações regular e irregular.

7 CONSIDERAÇÕES FINAIS

Em trabalhos futuros podem ser investigados outras configurações amostrais no estudo de simulação, verificando o comportamento dos modelos univariado e bivariados quanto à predição espacial. Além disso, podem ser estudadas medidas que descrevam a associação bivariada e que levem em consideração as localizações espaciais das variáveis em estudo.

REFERÊNCIAS

- AMARAL, L. de P.; FERREIRA, R. A.; WATZLAWICK, L. F. Análise da distribuição espacial de biomassa e carbono arbóreo acima do solo em floresta ombrófila mista. **Anais III Semana da Geomática/Colégio Politécnico da UFSM – Santa Maria**, 13-16 set, 2010. Disponível em: <http://200.132.36.199/3smgeo/downloads/Anais_3_Semana_da_Geomatica.pdf>. Acesso em: 20 abr. 2014.
- BANZATTO, D. A.; KRONKA, S. N. do. **Experimentação agrícola**. Jabotical, FUNEP, 247p., 1989.
- BOGNOLA, I. A.; RIBEIRO JR., P. J.; SILVA, E. A. A.; LINGNAU, C.; HIGA, A. R.; Modelagem uni e bivariada da variabilidade espacial de rendimento de Pinus taeda L. **Floresta**. Curitiba, PR, v.38, n.2, abr/jun, 2008.
- BORSSOI, J. A. **Técnicas de diagnóstico em modelos espaciais lineares gaussianos**. Dissertação de mestrado, Programa de Pós-Graduação em Engenharia Agrícola, Universidade de Estadual do Oeste do Paraná, 2007.
- CÂMARA, G.; MEDEIROS, J. S. GIS para meio ambiente. In: **Sistema de informações geográficas – GIS 98**, Curitiba, 1998, Curso D, Curitiba: Sagres, 185p., 1998.
- CAMARGO, E. C. G.; FUCKS, S. D.; CÂMARA, G. A análise espacial de superfícies. In: Câmara, G.; MEDEIROS, J.S. Eds. **Geoprocessamento para projetos ambientais**. São José dos Campos: INPE, 1998. Disponível em: <<http://www.dpi.inpe.br/gilberto/tutoriais/analise/cap3.pdf>>. Acesso em: 20 nov. 2014.
- CAMARGO, E. C. G. Geoestatística: fundamentos e aplicações. In: Câmara, G. & Medeiros, J.S. Eds. **Geoprocessamento para projetos ambientais**. São José dos Campos: INPE, 1998. Disponível em: <http://www.dpi.inpe.br/gilberto/tutoriais/gis_ambiente/5geoest.pdf>. Acesso em: 10 jul. 2012.
- CAMBARDELLA, C. A.; MOORMAN, T. B.; NOVAK, J. M.; PARKIN, T. B.; KARLEN, D. L.; TURCO, R. F.; KONOPKA, A. E. Field-scale variability of soil properties in Central Iowa Soils. **Soil Science Society of America Journal**, v.58, n.2, 1501-1511p., 1994.
- CAON, D.; GENUÍ, A. M.; Mapeamento de atributos químicos em diferentes densidades amostrais e influência na adubação e calagem. **Revista Brasileira de Engenharia Agrícola e Ambiental**. Campina Grande, PB. v.17, n.6, 629–639p., 2013.
- CARVALHO, J. R.P. de; ASSAD, E. D. Comparação de interpoladores espaciais multivariados para precipitação pluvial anual do Estado de São Paulo. **Embrapa Informática Agropecuária – Comunicado Técnico (INFOTECA-E)**. Campinas, SP, 2003. Disponível em: <<http://www.infoteca.cnptia.embrapa.br/handle/doc/8772>>. Acesso em: 08 abr. 2014.
- CLARK, I. **Practical Geostatistics**. London, Applied Science Publishers, 129p, 1979.
- COELHO, E. C.; SOUZA, E. G. de; URIBE-OPAZO, M. A.; NETO, R. P. Influência da densidade amostral e do tipo de interpolador na elaboração de mapas temáticos. **Acta Scientiarum Agronomy**. Maringá, PR, v.31, n.1, 165-174p., 2009.
- CORREIA, M. D. **Modelagem geoestatística da distribuição de carbono do solo e biomassa de herbáceas em sistema silvopastoril**. Dissertação de mestrado, Programa de Pós-Graduação em Biometria e Estatística Aplicada, Universidade de Federal Rural de Pernambuco, 2013.

CRESSIE, N. A. C. **Statistics for Spatial Data**. New York: A Wiley-Interscience, 900p.,1993.

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

DANCEY, C. P.; REIDY, J. **Statistics without for psychology**. Pearson/Prentice Hall, 619 p., 2007

DE BASTIANI, F.; URIBE-OPAZO, M. A.; DALPOSSO, G. H. Comparação de mapas de variabilidade espacial da resistência do solo à penetração construídos com e sem covariáveis usando um modelo espacial linear. **Engenharia Agrícola**. Jaboticabal, v.32, n. 2, 393-404p., 2012.

DIGGLE, P.J.; RIBEIRO JR., P.J. **Model-Based geoestatistic**. New York: Springer, 230p., 2007.

DUDEWICZ, E.; MISHRA, S. **Modern Mathematical Statistics**. Singapore: John Wiley & Sons, 1988.

DUGAN, J. L.; PERRY, J. N.; DALE, M. R. T.; LEGENDRE, P.; CITRON-POUSTY, S.; FORTIN, M. –J.; JAKOMULSKA, A.; MIRITI, M.; ROSENBERG, M. S. A balanced view of scale in spatial statistical analysis. **Ecography**, Copenhagen, v. 25, 626-640p., 2002.

FARACO, M. A.; URIBE-OPAZO, M. A.; SILVA, E. A. A. da; JOHANN, J.; 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, Minas Gerais, v. 32, n. 2, 463-476 p., 2008.

FERRAZ, G. A. S.; SILVA, F. M. de; COSTA, P. A. N. de; SILVA, A. C.; CARVALHO, F. de M. Agricultura de precisão no estudo de atributos químicos do solo e da produtividade de lavoura cafeeira. **Revista Coffee Science**, Lavras, v. 7, n. 1, 59-67p., jan./abr. 2012

FIGUEIREDO, G. C.; VIEIRA, C. A. O. Estudo do comportamento dos índices de Exatidão Global, Kappa e Tau, comumente usados para avaliar a classificação de imagens do sensoriamento remoto. **Anais XIII Simpósio Brasileiro de Sensoriamento Remoto**, Florianópolis, SC. INPE, p.5755-5762, 21-26 abril, 2007. Disponível em: <<http://marte.sid.inpe.br/col/dpi.inpe.br/sbsr@80/2006/11.13.17.35/doc/5755-5762.pdf>>. Acesso em: 08 abr. 2014.

FONSECA, B. H. F. **Um estudo sobre estimação e predição em modelos geoestatísticos bivariados**. Dissertação de mestrado, Programa de Pós-Graduação em Estatística e Experimentação Agronômica, Universidade de São Paulo, Escola Superior de Agricultura “Luiz de Queiroz”, São Paulo, 2008.

GUEDES, L. P. C.; RIBEIRO JR, 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 Agricultural Research**. v.2, n. 2, 39-50p., set, 2011.

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**, Viçosa, v.32, n.6, 2217-2226p., 2008.

GUEDES, L. P. C.; URIBE-OPAZO, M. A.; RIBEIRO JR, 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**, Chile, v.73 (4), out-dez, 2013.

GUEDES, L. P. C. Otimização da amostragem espacial. **Tese** (Doutorado em Estatística e Experimentação). Universidade de São Paulo, ESALQ/USP, Piracicaba, 143 p., 2008.

GUERRA, P. A. G. **Geoestatística operacional**. Brasília: Ministério das Minas e energia, Departamento Nacional de Produção Mineral, 150p., 1988.

ISAAKS, E. H.; SRISVASTAVA, R. M. **An introduction to applied geostatistic**. New York: Oxford University Press, 561p.,1998.

JOURNEL, A. G.; HUIJBREGTS, CH. J. **Mining Geostatistics**. London, Academic Press, 600p., 1978.

KAVANAGH, C. R. **Estimação de parâmetros no ajuste de modelos teóricos a semivariogramas experimentais**. Dissertação de mestrado, Programa de Pós-Graduação em Engenharia Agrícola, Universidade Estadual do Oeste do Paraná, Cascavel, PR, 2001.

KESTRING, F. B. F. **Análise geoestatística de mapas temáticos da produtividade da soja com diferentes grades amostrais**. Dissertação de mestrado, Programa de Pós-Graduação em Engenharia Agrícola, Universidade Estadual do Oeste do Paraná, Cascavel, 2008.

KOLMAN, B. **Introductory Linear Algebra with Applications**. USA: Prentice Hall, 1997.

KRIPENDORFF, K. **Content Analysis: an Introduction to its Methodology**. Beverly Hills: Sage Publications, 180p.,1980.

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

MARDIA, K. V.; MARSHALL, R. J. Maximum likelihood models for residual covariance in spatial regression. **Biometrika**, Oxford, v.71, n.1, 135-146p., 1984.

MARINS, A. C. de; URIBE-OPAZO, M. A.; JOHANN, J. A. Estimadores New1 e New2 no estudo da dependência espacial da produtividade da soja e atributos físicos do solo em uma área comercial. **Engenharia na Agricultura**, Viçosa, Minas Gerais, v. 16, n. 2, 133-143 p. abr./jun., 2008.

MARTINS, A. B. T. **Análise geoestatística de dados composicionais**. Tese de doutorado no Programa de Pós-Graduação em Métodos Numéricos em Engenharia. Universidade Federal do Paraná, Curitiba, PR, 2010.

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

MILANI, L.; SOUZA, E. G. de; URIBE-OPAZO, M. A.; FILHO, A. G.; JOHANN, J. A.; PEREIRA, J. O. Unidades de manejo a partir de dados de produtividade. **Acta Science Agronomic**. Maringá, v. 28, n. 4, 591-598 p., out./dez., 2006.

MILNE, A. The Centric systematic Area-Sample Treated as a Random Sample. **Biometrics**, Arlington, Newcastle, v.15, n.2, 270-297p., jun, 1959.

MOLIN, J. P. Definição de unidades de manejo a partir de mapas de produtividade. **Engenharia Agrícola**. Jaboticabal, v.22, n.1, 83-92p., 2002.

NETO, P. L. O. C. **Estatística**. Ed. Blucher Ltda, ed.2, 280p., 2002.

ODA-SOUZA, M.; BATISTA, J. L. F.; RIBEIRO JR, P. J.; RODRIGUES, R. R. Influência do tamanho e forma da unidade amostral sobre a estrutura de dependência espacial em quatro formações florestais do estado de São Paulo. **Floresta**, Curitiba, PR, v.40, n.4, 849-860p., out./dez, 2010.

OLINDA, R. A. de.; SCALON, J. D. Métodos de Monte Carlo para análise de processos Pontuais marcados. *Revista Brasileira de Biometria*. São Paulo, v.28, n.1, 39-56p., 2010.

ORTIZ, G. C. **Aplicação de métodos geoestatísticos para identificar a magnitude e a estrutura da variabilidade espacial de variáveis físicas do solo**. Dissertação de mestrado, Programa de Pós-Graduação em Agronomia, Escola Superior de agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, SP, 2002.

QUENOUILLE, M. H. Problems in plane sampling. **The Annals of Mathematical Statistics**, Harpenden, v.20, n.3, set, 1949.

R DEVELOPMENT CORE TEAM. R: **A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>. Acesso em: 20mar. 2014.

RIGHETTO, A. J. **Avaliação de modelos geoestatísticos multivariados**. Dissertação de mestrado, Programa de Pós-Graduação em Estatística e Experimentação Agrônômica, Universidade De São Paulo, Escola Superior de Agricultura ‘Luiz de Queiroz’, Piracicaba, SP, 2012.

RUIZ-CÁRDENAS, R.; FERREIRA, M. A. R.; SCHMIDT, A. M. Stochastic search algorithms for optimal design of monitoring networks. **Revista Environmetrics**, n. 21: 102–112p., mar, 2009.

SILVA, E. A. A.; URIBE-OPAZO, M. A.; SOUZA, E. G.; ROCHA, J. V. Um estimador robusto e os semivariograma cruzado na análise de variabilidade espacial de atributos de solo e planta. **Acta Scientiarum Agronomy**, Maringá, v.25, n.2, 365-371p., 2003.

SILVA, O. R. R. F.; MOREIRA, J. de A. N.; SANTOS, J. W. dos. Modelos univariados e multivariado na comparação de híbridos com o SISAI tradicional. **Revista Brasileira de Oleaginosas e Fibrosas**. Campinas, v.1, n.1, 57-62p., dez, 1997.

SILVA, R.; SOUZA, F. G.; CORTEZ, J. W.; FURLANI, C. E. A.; VIGNA, G. P. Variabilidade Espacial e Controle Estatístico do Processo de Perdas na Colheita Mecanizada do Algodoeiro. **Engenharia Agrícola**, Jaboticabal, v.27, n.3, 742-752p., set/dez, 2007.

SOARES, A. **Geoestatística para as ciências da terra e do ambiente**. Lisboa, IST Press, 206p., 2000.

STORCK, L.; GARCIA, D. C.; LOPES, S. J.; ESTEFANEL, V. **Experimentação Vegetal**. Santa Maria, Ed. UFSM, 198p., 2006.

TEIXEIRA, M. B. dos R. **Comparação entre estimadores de semivariância**. Dissertação de mestrado, Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, Universidade federal de Lavras, MG, 2013.

URIBE-OPAZO, M. A.; JOHANN, J. A.; VILAS BOAS, M. A.; BORSSOI, J. A.; LUNKES, C. **Fatores que influenciam na estimação de parâmetros no ajuste de modelos teóricos á semivariogramas experimentais no estudo da produtividade da soja**. 2007. Disponível em:<http://leg.ufpr.br/lib/exe/fetch.php/pessoais:joelmir:trabalho_apresentar_chile_miguel_2006_fim.doc>. Acesso em: 04 fev. 2014.

VIEIRA, S. R. Geoestatística Aplicada á Agricultura de Precisão. In: Sistemas de informações geográficas – **GIS 98**, Curitiba, 1998, Curso P, Curitiba: Sagres, 48p., 1998.

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

YFANTIS, E.A.; FLATMAN, G. T.; BEHAR, J. V. Efficiency of kriging estimation for square, triangular, and hexagonal grids. **Mathematical Geology**, v. 19, n. 3, 183-205 p.,1987.

WEBSTER, R.; OLIVER, M. A. **Geostatistics for Environmental Scientists**. 2nd Edition, John Wiley & Sons, Ltd, 315p.,2007.