

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

**ANÁLISE GEOESTATÍSTICA DE MAPAS TEMÁTICOS DA  
PRODUTIVIDADE DA SOJA COM DIFERENTES GRADES  
AMOSTRAIS**

**FRANCIELE BUSS FRESCKI KESTRING**

CASCAVEL

2011

**FRANCIELE BUSS FRESCKI KESTRING**

**ANÁLISE GEOESTATÍSTICA DE MAPAS TEMÁTICOS DA PRODUTIVIDADE  
DA SOJA COM DIFERENTES GRADES AMOSTRAIS**

Dissertação apresentada ao Programa de Pós-Graduação *Stricto Sensu* em Engenharia Agrícola, da Universidade Estadual do Oeste do Paraná, em cumprimento parcial aos requisitos para a obtenção do título de mestre em Engenharia Agrícola, área de concentração Engenharia de Sistemas Agroindustriais.

Orientador: Prof<sup>o</sup>. Dr. Miguel Angel Uribe  
Opazo

Co-Orientadora: Prof<sup>a</sup>. Dra. Luciana Pagliosa  
Guedes Carvalho

CASCADEL

2011

Dados Internacionais de Catalogação-na-Publicação (CIP)  
Biblioteca Central do Campus de Cascavel – Unioeste  
Ficha catalográfica elaborada por Jeanine da Silva Barros CRB-9/1362

K52a      Kestring, Franciele Buss Frescki  
            Análise geoestatística de mapas temáticos da produtividade da soja  
            com diferentes grades amostrais / Franciele Buss Frescki Kestring—  
            Cascavel, PR: UNIOESTE, 2011.  
            30 cm.

            Orientador: Prof. Dr. Miguel Angel Uribe Opazo  
            Co-Orientadora: Profa. Dra. Luciana Pagliosa Guedes Carvalho  
            Dissertação (Mestrado) – Universidade Estadual do Oeste do  
Paraná.  
            Programa de Pós-Graduação Stricto Sensu em Engenharia Agrícola,  
            Centro de Ciências Exatas e Tecnológicas.  
            Bibliografia.

            1. Geoestatística. 2. Soja - Produtividade. 3. Mapas temáticos –  
            Grades amostrais. I. Universidade Estadual do Oeste do Paraná. II.  
            Título.

CDD 21ed. 631.4

---

<sup>i</sup> Correções de língua portuguesa, língua inglesa e normas do PGEAGRI realizadas por Noeli Teresinha Unser, revisora cadastrada no Programa.

## **FRANCIELE BUSS FRESCKI KESTRING**

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 Engenharia de Sistemas Agroindustriais. Banca examinadora:

Orientador: Prof. Dr. Miguel Angel Uribe Opazo  
Centro de Ciências Exatas e Tecnológicas, UNIOESTE - Cascavel

Prof. Dr. Marcio Antonio Villas Boas  
Centro de Ciências Exatas e Tecnológicas, UNIOESTE - Cascavel

Prof. Dr. Erivelto Mercante  
Centro de Ciências Exatas e Tecnológicas, UNIOESTE - Cascavel

Prof. Dr. Jorim Souza da Virgems Filho  
Departamento de Matemática e Estatística, UEPF - Ponta Grossa

## BIOGRAFIA RESUMIDA

Nome: Franciele Buss Frescki Kestring

Ano de nascimento: 1987

Naturalidade: Medianeira-PR

Graduada em Matemática pela Universidade Estadual do Oeste do Paraná - UNIOESTE (2008). Atualmente, aluna regular do Programa de Pós-Graduação *Stricto Sensu* em Engenharia Agrícola, nível de mestrado, área de concentração Engenharia de Sistemas Agroindustriais. Professora do quadro próprio do magistério da rede estadual paranaense e professora temporária da Universidade Estadual do Oeste do Paraná, campus Medianeira.

## EPÍGRAFE

*Deus não joga dados com o Universo.*

Albert Einstein

## DEDICATÓRIA

Aos meus pais, por tudo o que  
fizeram e me permitiram fazer.  
Também ao Fernando, que esteve  
ao meu lado em todos os momentos.

## AGRADECIMENTOS

A Deus, por permitir esta conquista.

Ao Prof<sup>o</sup>. Dr. Miguel Angel Uribe Opazo, pelo apoio, constante orientação, e por ser compreensivo nos momentos de crise e por acreditar em meu potencial.

À Prof<sup>a</sup> Dra. Luciana Pagliosa Guedes de Carvalho, pelo empenho, co-orientação e motivação para a realização deste trabalho.

Aos professores do Programa de Pós-Graduação *Stricto Sensu* em Engenharia Agrícola, pelos ensinamentos, pelas opiniões e pelo estímulo para prosseguir com os trabalhos.

À UNIOESTE e seus funcionários, um muito obrigada, pela oportunidade e todo o serviço prestado ao longo desse período.

Aos meus pais, Vergilino Frescki e Maria Lurdes Buss Frescki e meu irmão Fernando Buss Frescki, por todo o amor, compreensão, incentivo e carinho.

Especialmente, ao meu marido, Fernando Kestring, por ter compreendido e respeitado minha ausência quando necessário e por ter sempre dado o apoio que precisei.

Aos colegas de caminhada, Naimara, Fernanda, Rosangela, Denise, Everton, Priscila, Jaqueline, dentre outros, pelas correções, pelo auxílio com o R e com o  $\LaTeX$ .

Aos colegas de trabalho e alunos, por compreenderem as faltas e incentivarem este passo em minha carreira.

A todos que de forma direta ou indireta colaboraram para que se tornasse possível a realização desse trabalho.

## RESUMO

### ANÁLISE GEOESTATÍSTICA DE MAPAS TEMÁTICOS DA PRODUTIVIDADE DA SOJA COM DIFERENTES GRADES AMOSTRAIS

Com o aumento da produção agrícola mundial, o processo de produção agrícola tornou-se alvo do estudo de diversos pesquisadores. Estudos sobre a variabilidade espacial da produtividade da soja são de grande importância para o desenvolvimento de novas tecnologias, que beneficiam a agricultura. A análise geoestatística torna possível realizar previsões dos resultados, tendo como um de seus produtos os mapas temáticos. Este trabalho descreve algumas técnicas para a construção e comparação de mapas temáticos, utilizando a krigagem. A análise foi realizada com dados da produtividade de soja em  $t\ ha^{-1}$  do ano agrícola 2004/2005 numa área experimental com grades de amostragem com distâncias de 25x25 m, 50x50 m, 75x75 m, 100x100 m e monitor de colheita, comparando-se os mapas, utilizando a matriz de erros e a matriz de confusão. Além de uma melhor precisão dos mapas de variabilidade espacial gerados, a análise dos índices de acurácia possibilita um melhor planejamento das malhas amostrais para futuros estudos. As medidas de acurácia obtidas por meio da matriz de erros são opções significativas para realizar a comparação entre mapas temáticos, uma vez que fornecem índices globais e também por classes.

**Palavras-chave:** matriz de erros, matriz de confusão, índices de acurácia.

## ABSTRACT

### GEOSTATISTICAL ANALYSIS OF THEMATIC MAPS OF SOYBEAN YIELD WITH DIFFERENT SAMPLING GRIDS

Studies on spatial variability of soybeans yield are of great importance for the development of new technologies that improve the world agricultural production. One of methods that allows this study is geostatistics. The geostatistical analysis makes possible the predictions of results and one of its products are thematic maps. Thus, this trial describes some techniques to draw and compare thematic maps using kriging. The analysis was based on data from soybean yield in  $t\ ha^{-1}$  according to harvest year 2004/2005 in an experimental area with sampling grades whose distances were: 25x25 m, 50x50 m, 75x75 m and 100x100 m plus a harvest monitor. The maps were compared using error matrix and confusion matrix. In addition, there was a better accuracy of the spatial variability maps that were drawn, while the analysis of coefficients of accuracy allows a better planning of sampling mesh for future studies. The measures of accuracy that were obtained by error matrix are significant options to make comparison among thematic maps, once they provide global indices and also by classes.

**Keywords:** error matrix, confusion matrix, coefficients of accuracy.

## SUMÁRIO

<b>LISTA DE TABELAS .....</b>	<b>x</b>
<b>LISTA DE FIGURAS .....</b>	<b>xiii</b>
<b>1 INTRODUÇÃO .....</b>	<b>1</b>
<b>2 REVISÃO BIBLIOGRÁFICA .....</b>	<b>3</b>
2.1 Introdução à Geoestatística .....	3
2.2 Modelos teóricos da função semivariância .....	7
2.2.1 Modelo esférico .....	7
2.2.2 Modelo exponencial .....	8
2.2.3 Modelo gaussiano .....	9
2.2.4 Família Matérn .....	10
2.2.5 Família exponencial potência .....	11
2.3 Métodos de estimação de parâmetros .....	12
2.3.1 Máxima verossimilhança - MV .....	12
2.3.2 Máxima verossimilhança restrita - MVR .....	12
2.4 Critérios de validação dos ajustes de modelos geoestatísticos .....	13
2.4.1 Validação cruzada .....	13
2.4.2 Critério de informação de Akaike .....	14
2.5 Construção de mapas por krigagem .....	15
2.6 Comparação de mapas temáticos .....	17
2.6.1 Matriz de erros .....	18
2.6.2 Índice Kappa .....	19
2.6.3 A estatística Tau .....	20
2.6.4 Exatidão global, acurácia do produtor e do usuário .....	20
2.6.5 Outras medidas de acurácia .....	21
2.6.6 Matriz de confusão .....	21

<b>3</b>	<b>MATERIAL E MÉTODOS</b>	<b>24</b>
3.1	Caracterização da área	24
3.2	Análise exploratória de dados	25
3.3	Análise estatística espacial	25
3.4	Critérios de seleção de modelos	26
3.5	Comparação dos mapas temáticos	26
<b>4</b>	<b>RESULTADOS E DISCUSSÕES</b>	<b>28</b>
4.1	Estatísticas descritivas	28
4.2	Análise de tendência e estacionariedade	30
4.3	Identificação da continuidade espacial	30
4.4	Identificação da anisotropia	31
4.5	Modelagem espacial	33
4.5.1	Modelagem espacial para os dados da grade 100x100 m	33
4.5.2	Modelagem espacial para os dados da grade 75x75 m	38
4.5.3	Modelagem espacial para os dados da grade 50x50 m	42
4.5.4	Modelagem espacial para os dados da grade 25x25 m	46
4.6	Comparação dos mapas gerados	49
<b>5</b>	<b>CONCLUSÕES</b>	<b>58</b>
<b>6</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>59</b>
<b>7</b>	<b>REFERÊNCIAS</b>	<b>60</b>
<b>8</b>	<b>ANEXOS</b>	<b>66</b>

## LISTA DE TABELAS

Tabela 1	Matriz genérica dos erros .....	18
Tabela 2	Relação entre o valor do índice Kappa e o desempenho de classificação obtido (KRIPPENDORFF, 1980) .....	19
Tabela 3	Índices de acurácia por classe .....	21
Tabela 4	Índices de acurácia global .....	22
Tabela 5	Matriz genérica de confusão da $i$ -ésima classe (JENSEN, 1986) .....	22
Tabela 6	Índices obtidos a partir da matriz de confusão por classe .....	23
Tabela 7	Matriz genérica de confusão total .....	23
Tabela 8	Resumo das estatísticas para as quatro grades analisadas e monitor de colheita .....	28
Tabela 9	Determinação de parâmetros para os métodos de estimação e modelos - grade 100x100 m .....	34
Tabela 10	Seleção de melhor modelo para os dados de produtividade da soja na grade 100x100 m .....	36
Tabela 11	Classificação dos pixels do mapa para a grade amostral 100x100 m .....	37
Tabela 12	Estatísticas descritivas para os valores estimados pelo modelo Matérn, com $\kappa = 2,0$ , para a grade 100x100 m, em $t\ ha^{-1}$ .....	38
Tabela 13	Determinação de parâmetros para os métodos de estimação e modelos - grade 75x75 m .....	38
Tabela 14	Validação cruzada, critério de informação de Akaike, $IAC$ , e máximo do logaritmo da função verossimilhança para os dados de produtividade da soja para a grade 75x75 m .....	40
Tabela 15	Classificação dos pixels do mapa para a grade amostral 75x75 m .....	41
Tabela 16	Estatísticas descritivas para os valores preditos pelo modelo Matérn, com $\kappa = 0,7$ , para a grade 75x75 m, em $t\ ha^{-1}$ .....	41
Tabela 17	Determinação de parâmetros para os métodos de estimação e modelos - grade 50x50 m .....	42
Tabela 18	Seleção do melhor modelo para os dados de produtividade da soja para a grade 50x50 m .....	44
Tabela 19	Classificação dos pixels do mapa para a grade amostral 50x50 m .....	45
Tabela 20	Estatísticas descritivas para os valores preditos pelo modelo Matérn, com $\kappa = 0,7$ , ajustado por máxima verossimilhança, para a grade 50x50 m, em $t\ ha^{-1}$ .....	45
Tabela 21	Determinação de parâmetros para os métodos de estimação e modelos - grade 25x25 m .....	46

Tabela 22	Validação cruzada para a grade 25x25 m .....	48
Tabela 23	Classificação dos pixels do mapa para a grade amostral 25x25 m .....	49
Tabela 24	Estatísticas descritivas para os valores preditos pelo modelo Matérn, com $\kappa = 0,7$ , ajustado por máxima verossimilhança, para a grade 25x25 m, em $t\ ha^{-1}$ .....	49
Tabela 25	Quantidade de pixels por classe para os mapas temáticos da produtividade da soja e porcentagem do total de pixels .....	51
Tabela 26	Matriz de erros para comparação dos mapas da grade 25x25 m (referência) e 50x50 m (modelo) .....	51
Tabela 27	Índice de Acurácia para as classes na produtividade da soja na comparação entre os mapas temáticos nas grades 25x25 m (referência) e 50x50 m (modelo), em percentual .....	52
Tabela 28	Matriz de erros para comparação dos mapas da grade 25x25 m (referência) e 75x75 m (modelo) .....	52
Tabela 29	Índice de Acurácia para as classes na produtividade da soja na comparação entre os mapas temáticos nas grades 25x25 m (referência) e 75x75 m (modelo) .....	53
Tabela 30	Matriz de erros para comparação dos mapas da grade 25x25 m (referência) e 100x100 m (modelo) .....	53
Tabela 31	Índices de Acurácia para as classes na produtividade da soja na comparação entre os mapas temáticos nas grades 25x25 m (referência) e 100x100 m (modelo) .....	54
Tabela 32	Índices de acurácia na produtividade da soja na comparação entre os mapas temáticos nas grades 25x25 m (referência) e os mapas 100x100, 75x75 e 50x50 m (modelo), valores em percentual .....	54
Tabela 33	Índices obtidos a partir da matriz de confusão para a grade 100x100 m (%) ..	55
Tabela 34	Índices obtidos a partir da matriz de confusão para a grade 75x75 m (%) ...	55
Tabela 35	Índices obtidos a partir da matriz de confusão para a grade 50x50 m (%) ...	55
Tabela 36	Matriz de confusão total para a grade 100x100 m .....	56
Tabela 37	Matriz de confusão total para a grade 75x75 m .....	56
Tabela 38	Matriz de confusão total para a grade 50x50 m .....	56
Tabela 39	Produção (t) estimada para cada modelo, por classes .....	56
Tabela 40	Matriz de confusão da primeira classe 100x100 m .....	66
Tabela 41	Matriz de confusão da segunda classe 100x100 m .....	66
Tabela 42	Matriz de confusão da terceira classe 100x100 m .....	66
Tabela 43	Matriz de confusão da quarta classe 100x100 m .....	66
Tabela 44	Matriz de confusão da quinta classe 100x100 m .....	66
Tabela 45	Matriz de confusão da primeira classe 75x75 m .....	66

Tabela 46	Matriz de confusão da segunda classe 75x75 m	67
Tabela 47	Matriz de confusão da terceira classe 75x75 m	67
Tabela 48	Matriz de confusão da quarta classe 75x75 m	67
Tabela 49	Matriz de confusão da quinta classe 75x75 m	67
Tabela 50	Matriz de confusão da primeira classe 50x50 m	67
Tabela 51	Matriz de confusão da segunda classe 50x50 m	67
Tabela 52	Matriz de confusão da terceira classe 50x50 m	67
Tabela 53	Matriz de confusão da quarta classe 50x50 m	68
Tabela 54	Matriz de confusão da quinta classe 50x50 m	68

## LISTA DE FIGURAS

Figura 1	Relação entre a função semivariância $\gamma(h)$ e a função covariância $C(h)$ . ....	7
Figura 2	Representação gráfica do modelo esférico. ....	8
Figura 3	Representação gráfica do modelo exponencial. ....	9
Figura 4	Representação gráfica do modelo gaussiano. ....	10
Figura 5	Representação gráfica do modelo Família Matérn para $k=0,1$ , $k=0,2$ , $k=0,5$ , $k=1$ e $k=2$ , $c_0 = 0$ , $c_1 = 1$ e range $r = 20$ . Adaptado de Minasny & McBratney (2005). ....	11
Figura 6	Mapa de localização da área experimental. ....	24
Figura 7	Área em estudo - esquema amostral. Fonte: Faraco (2006). ....	25
Figura 8	Localização dos pontos nas grades: (a) Grade 100x100 m, (b) Grade 75x75 m, (c) Grade 50x50 m, (d) Grade 25x25 m. ....	26
Figura 9	Gráficos box-plot: (a) 100x100 m; (b) 75x75 m; (c) 50x50 m; (d) 25x25 m; (e) monitor de colheita ....	29
Figura 10	Gráficos post-plot: (a) 100x100 m; (b) 75x75 m; (c) 50x50 m; (d) 25x25 m; (e) Monitor de colheita para a variável produtividade da soja em $t\ ha^{-1}$ . Md: mediana; Q1: 1º quartil, Q3: 3º quartil; min: valor mínimo; max: valor máximo ....	30
Figura 11	Semivariogramas experimentais para a variável regionalizada produtividade da soja nas grades (a) 100x100 m, (b) 75x75 m, (c) 50x50 m, (d) 25x25 m no ano agrícola 2004/2005 com envelopes simulados utilizando 99 simulações de Monte Carlo. ....	31
Figura 12	Semivariogramas direcionais para a variável produtividade da soja em $t\ ha^{-1}$ para as grades: (a) 100x100 m, (b) 75x75 m, (c) 50x50 m, (d) 25x25 m, nas direções $0^\circ$ , $45^\circ$ , $90^\circ$ e $135^\circ$ ....	32
Figura 13	Semivariogramas omnidirecionais para as grades: (a) 100x100 m, (b) 75x75 m, (c) 50x50 m, (d) 25x25 m, (e) Monitor de colheita ....	33
Figura 14	Semivariogramas ajustados à grade 100x100 m para: (a) máxima verossimilhança com modelo esférico; (b) máxima verossimilhança restrita com modelo esférico; (c) máxima verossimilhança com modelo exponencial; (d) máxima verossimilhança restrita com modelo exponencial; (e) máxima verossimilhança com modelo gaussiano; (f) máxima verossimilhança restrita com modelo gaussiano; (g) máxima verossimilhança com Família Matérn $\kappa = 0, 7$ ; (h) máxima verossimilhança restrita com Família Matérn $\kappa = 0, 7$ ; (i) máxima verossimilhança com Família Matérn $\kappa = 1, 0$ ; (j) máxima verossimilhança restrita com Família Matérn $\kappa = 1, 0$ ; (k) máxima verossimilhança com Família Matérn $\kappa = 1, 5$ ; (l)	

	máxima verossimilhança restrita com Família Matérn $\kappa = 1,5$ ; (m) máxima verossimilhança com Família Matérn $\kappa = 2,0$ ; (n) máxima verossimilhança restrita com Família Matérn $\kappa = 2,0$ . . . . .	35
Figura 15	Mapa construído para a grade amostral 100x100 m . . . . .	37
Figura 16	Semivariogramas ajustados à grade 75x75 m para (a)máxima verossimilhança com modelo esférico; (b) máxima verossimilhança restrita com modelo esférico; (c) máxima verossimilhança com modelo exponencial; (d) máxima verossimilhança restrita com modelo exponencial; (e) máxima verossimilhança com modelo gaussiano; (f) máxima verossimilhança restrita com modelo gaussiano; (g) máxima verossimilhança com modelo Família Matérn $k=0,7$ ; (h) máxima verossimilhança restrita com modelo Família Matérn $k=0,7$ ; (i) máxima verossimilhança com Família Matérn $k=1,0$ ; (j) máxima verossimilhança restrita com Família Matérn $k=1,0$ ; (k) máxima verossimilhança com Família Matérn $k=1,5$ ; (l) máxima verossimilhança restrita com Família Matérn $k=1,5$ ; (m) máxima verossimilhança com Família Matérn $k=2,0$ ; (n) máxima verossimilhança restrita com Família Matérn $k=2,0$ . . . . .	39
Figura 17	Mapa construído para a grade 75x75 m . . . . .	41
Figura 18	Semivariogramas ajustados à grade 50x50 m para (a)máxima verossimilhança com modelo esférico; (b) máxima verossimilhança restrita com modelo esférico; (c) máxima verossimilhança com modelo exponencial; (d) máxima verossimilhança restrita com modelo exponencial; (e)máxima verossimilhança com modelo Família Matérn $\kappa=0,7$ ; (f)máxima verossimilhança restrita com modelo Família Matérn $\kappa=0,7$ ; (g) máxima verossimilhança com Família Matérn $\kappa=1,0$ ; (h) máxima verossimilhança restrita com Família Matérn $\kappa=1,0$ ; (i)máxima verossimilhança com Família Matérn $\kappa=1,5$ ; (j)máxima verossimilhança restrita com Família Matérn $k=1,5$ ; (k) máxima verossimilhança com Família Matérn $\kappa=2,0$ ; (l) máxima verossimilhança restrita com Família Matérn $\kappa=2,0$ . . . . .	43
Figura 19	Mapa temático para a grade 50x50 m. . . . .	45
Figura 20	Semivariogramas ajustados à grade 25x25 m para (a) máxima verossimilhança com modelo esférico; (b) máxima verossimilhança restrita com modelo esférico; (c) máxima verossimilhança com modelo exponencial; (d) máxima verossimilhança restrita com modelo exponencial; (e) máxima verossimilhança com modelo Família Matérn $\kappa=0,7$ ; (f) máxima verossimilhança restrita com modelo Família Matérn $\kappa=0,7$ ; (g) máxima verossimilhança com Família Matérn $\kappa=1,0$ ; (h) máxima verossimilhança restrita com Família Matérn $\kappa=1,0$ ; (i) máxima verossimilhança com Família Matérn $\kappa=1,5$ ; (j) máxima verossimilhança restrita com Família Matérn $\kappa=1,5$ ; (k) máxima verossimilhança com Família Matérn $\kappa=2,0$ ; (l) máxima verossimilhança restrita com Família Matérn $\kappa=2,0$ . . . . .	47
Figura 21	Mapa temático da produtividade da soja para a grade 25x25 m . . . . .	49
Figura 22	Mapas temáticos da produtividade da soja para as grades (a) 100x100 m; (b) 75x75 m; (c) 50x50 m e (d) 25x25 m . . . . .	50

## 1 INTRODUÇÃO

Com as tecnologias para a mecanização e o aumento da produção agrícola mundial, este processo de produção tornou-se alvo do estudo de diversos pesquisadores, buscando uma agricultura sustentável, que possa otimizar a produção agrícola de forma racional, reduzindo o uso de insumos. Em particular, a produção de soja é de extrema importância econômica para o Brasil. Na safra 2010/2011 a produção mundial de soja foi de 263,7 milhões de toneladas, com uma área plantada de 103,5 milhões de hectares, sendo que o Brasil é o segundo maior produtor mundial de soja (USDA, 2011). Na safra 2010/2011, o país alcançou uma produtividade média de  $3125 \text{ kg h}^{-1}$ , ocupando uma área de mais de 24 milhões de hectares, o que totalizou uma produção de 75 milhões de toneladas (CONAB, 2011).

Para a tomada de decisão do agronegócio, o pesquisador ou o produtor precisam de um objeto de análise que possa ser interpretado. O mapa temático de uma variável pode ser tal objeto, segundo o qual, o produtor pode realizar um estudo de sua área de produção. Constrói-se um mapa temático da variável objeto de estudo (produtividade das culturas, atributos físico-químicos do solo, propriedades da planta) e, posteriormente, faz-se uma realimentação de informações. É de grande importância a construção de mapas que sejam confiáveis, que revelem a variabilidade espacial do atributo em questão. Esta informação tem grande utilidade para o agricultor, pois pode-se identificar locais com determinada característica e investigar os motivos que levam essas áreas a apresentarem essas possíveis variações de produtividade (ou qual seja o atributo em estudo). Assim, o mapa temático gerado pode ser utilizado como base, na determinação de áreas de manejo localizadas.

Os processos de construção e classificação de mapas temáticos, construídos por meio da análise de dados georreferenciados são cada vez mais numerosos. Entretanto, nem todos são confiáveis, sendo necessário obter um método seguro, que permita determinar a eficiência da classificação realizada.

Nessa perspectiva, para construir mapas e realizar as avaliações pertinentes, o pesquisador necessitará de alguns métodos científicos, entre eles, a geoestatística. Quando se trabalha com dados distribuídos espacialmente, cada valor carrega consigo uma informação sobre sua vizinhança, uma vez que fenômenos naturais não ocorrem pontualmente. Quando dois dados são tomados em duas posições diferentes no espaço, existe um terceiro elemento essencial, associado a eles: a sua relação, que pode ser descrita por uma função do vetor de separação entre eles (BOURGAULT et al., 1996).

A geoestatística tem como princípio a Teoria das Variáveis Regionalizadas, que pode ser entendida como aplicação do formalismo de funções aleatórias ao reconhecimento e estimação de fenômenos naturais. Tem por objetivo estudar a estrutura de dependência espacial

de variáveis regionalizadas. Assim, a geoestatística pode ser entendida como o uso de um grupo de procedimentos estatísticos aplicáveis a alguns problemas, nos quais os dados são espacialmente referenciados. A geoestatística é construída a partir de conceitos básicos de estatística, em particular processos estocásticos. Em suma, a geoestatística é uma reunião de conceitos e métodos estatísticos para o estudo da variabilidade espacial de variáveis aleatórias contínuas, quando as informações coletadas são georreferenciáveis e apresentam dependência espacial.

Muitos dos atributos do solo apresentam estrutura de dependência espacial, tais como os físicos de resistência do solo à penetração, densidade do solo, umidade do solo e os químicos, como fósforo, potássio e cálcio. Trabalhos demonstram que os atributos do solo são espacialmente dependentes (MATHERON, 1963; VIEIRA et al., 1981; SOUZA et al., 1999; MERCANTE et al., 2003; GUEDES et al., 2008; FARACO et al., 2008, BAZZI et al., 2008; BORSSOI et al., 2009). A detecção dessa estrutura de dependência espacial em áreas experimentais pode permitir estimação mais eficiente dos contrastes entre tratamentos. Pontos próximos no espaço tendem a ter valores mais parecidos do que pontos mais afastados. Logo, o valor em um ponto pode ser estimado pelos valores de seus vizinhos, quando tratar-se de variáveis não independentes.

O estudo, discussão e aplicação da variabilidade espacial de fenômenos naturais e o sistema de agricultura de precisão vêm crescendo muito nos últimos anos, tendo como um dos motivadores, o estudo dos fatores que interferem na produtividade, para desenvolver novas técnicas que beneficiem a produção agrícola. Os objetivos da agricultura de precisão são o aumento da rentabilidade e a redução do impacto ambiental causado pelo uso de insumos.

O objetivo geral deste trabalho foi realizar a comparação de mapas temáticos de produtividade da soja com diferentes grades amostrais, construídos por técnicas geoestatísticas utilizando os métodos de estimação de máxima verossimilhança e máxima verossimilhança restrita. A comparação dos mapas temáticos é realizada baseando-se nas estatísticas de acurácia da matriz de erros e matriz de confusão.

## 2 REVISÃO BIBLIOGRÁFICA

A agricultura de precisão foi vista, inicialmente, como um conjunto de procedimentos para o manejo localizado de lavouras. Porém, com o desenvolvimento de pesquisas acerca da variabilidade das áreas agrícolas, passou a ser entendida como um sistema de gestão ou gerenciamento da produção agrícola (MOLIN, 2008). A prática de agricultura de precisão é um processo contínuo, constituindo um ciclo, sendo as etapas: amostragem aleatória sistêmica, análise do conjunto de dados, geração e interpretação das informações contidas nos mapas, medidas de correção da variabilidade, acompanhamento da lavoura, geração de mapas de colheita.

Os mapas temáticos são ferramentas importantes, utilizadas na área agrícola para visualizar a resposta de uma determinada tarefa realizada, como por exemplo, a produtividade de uma determinada cultura nos campos de produção, ou mesmo dados referentes a uma amostragem de solo. Este procedimento é necessário para que o agricultor possa intervir com insumos de forma adequada, econômica e contribuindo, assim, com o meio ambiente e com seu próprio lucro.

### 2.1 Introdução à Geoestatística

Uma variável regionalizada é uma variável distribuída no espaço e no tempo, para representar um fenômeno natural. A Teoria das Variáveis Regionalizadas foi desenvolvida pelo matemático e geólogo francês Georges François Paul Marie Matheron na década de 1960. Com esta teoria, Matheron afirmava que algumas variáveis estão intrinsecamente relacionadas com a sua localização geográfica, ou seja, estavam distribuídas no espaço (ou tempo). “Um valor medido carrega consigo uma informação sobre a sua vizinhança, uma vez que os fenômenos naturais não ocorrem pontualmente”. Nesta teoria, embasa-se a geoestatística.

Na geoestatística, os dados possuem estrutura espacial, considera-se um processo estocástico gaussiano  $\{Z(s) : s \in S \subset R^d\}$  (MARDIA & MARSHALL, 1984), em que:

- $Z$  representa o conjunto de variáveis aleatórias;
- $R^d$  representa o conjunto  $d$ -dimensional,  $d \geq 1$ . Observe que, quando  $d = 1$ , os dados estão em uma transeção; quando  $d = 2$ , os dados estão em um plano; e quando  $d = 3$ , os dados estão no espaço tridimensional (profundidade).
- $S$  é o espaço paramétrico, que é um subconjunto do espaço  $R^d$ ,  $d \geq 1$ .
- $\mathbf{Z}(s)$  é um vetor aleatório na posição  $s \in S$ .

Supõe-se que os dados  $Z(s_1), \dots, Z(s_n)$  deste processo são registrados em localizações espaciais conhecidas,  $s_i (i = 1, 2, \dots, n)$ , e gerados pelo seguinte modelo espacial linear, conforme (2.1).

$$Z(s_i) = \mu(s_i) + \varepsilon(s_i). \quad (2.1)$$

no qual, os termos determinísticos  $\mu(s_i)$  e estocástico  $\varepsilon(s_i)$ , dependem da localização espacial em que  $Z(s_i)$  foi obtida.

Considera-se que  $\varepsilon(s_i)$  tem média zero, isto é  $E[\varepsilon(s_i)] = 0$ , e a variação entre os pontos no espaço  $\varepsilon(s_i), \varepsilon(s_u)$  é determinada por alguma função de covariância  $C(s_i, s_u) = COV[\varepsilon(s_i), \varepsilon(s_u)]$ . Para algumas funções conhecidas de  $s$ , como  $X_1(s), \dots, X_p(s)$ , a média do processo estocástico é definida por (2.2).

$$\mu(s) = \mathbf{X}\boldsymbol{\beta} \quad (2.2)$$

em que:

$\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$  é o vetor de parâmetros desconhecidos a serem estimados;

$\mathbf{X} = [X_1(s), X_2(s), \dots, X_p(s)]$  é uma matriz de covariáveis  $n \times p$ , na qual  $\mathbf{X}_1(s)=1$ 's é o vetor de uns;

$\mu(s) = (\mu(s_1), \mu(s_2), \dots, \mu(s_n))^T$  é o vetor de médias.

Para a análise geoestatística utiliza-se a função semivariância ( $\gamma(u)$ ) para descrever a estrutura de dependência espacial em função da distância  $h$ , pois em algumas situações não se conhece a covariância, mas a semivariância pode existir mesmo assim. O semivariograma é um gráfico que expressa quantitativamente a variação de uma variável qualquer, de acordo com a distância ( $h$ ).

Os parâmetros de um semivariograma são:

- Alcance ( $a = g(\varphi_3)$ ): é a distância a partir da qual as amostras passam a ser não correlacionadas. Ou seja, o alcance reflete o grau de homogeneização entre as amostras. Dessa forma, quanto maior for o alcance, maior será a homogeneidade entre as amostras. O alcance é a distância que separa o campo estruturado (amostras correlacionadas) do campo aleatório (amostras não correlacionadas). O alcance é de fundamental importância para a interpretação dos semivariogramas. Ele indica a distância até onde os pontos amostrais estão correlacionados entre si (VIEIRA et al., 1983; SOUZA et al., 1997; VIEIRA, 1997), ou seja, os pontos localizados em uma área cujo raio seja o alcance, são mais semelhantes entre si, do que os separados por distâncias maiores;
- Efeito pepita ( $\varphi_1$ ): é o valor da função semivariância na origem. Teoricamente, esse valor deveria ser zero, pois duas amostras tomadas no mesmo ponto ( $h = 0$ ), deveriam ter os mesmos valores. Entretanto, quando não ocorre isto, atribui-se esta diferença a erros, como de amostragem e/ou análise;

- Contribuição ( $\varphi_2$ ): é denominada variância de dispersão e representa as diferenças espaciais entre os valores de uma variável tomada em dois pontos separados por distâncias cada vez maiores. Também representa a diferença entre o patamar ( $C = \varphi_1 + \varphi_2$ ) e o Efeito pepita ( $\varphi_1$ ). O parâmetro  $\varphi_2$  é conhecido como sill;

- Patamar ( $\varphi_1 + \varphi_2$ ): é o valor no qual o semivariograma estabiliza-se (no campo aleatório). Em outras palavras, é o ponto a partir do qual as amostras tornam-se não correlacionadas devido à grande distância que as separa. É a estimativa da variância dos dados.

É necessário que seja definido de antemão pelo pesquisador o cutoff, ou ponto de corte. Para a construção do semivariograma experimental se calculará até  $\frac{1}{4}$ ,  $\frac{1}{3}$  ou  $\frac{1}{2}$  do campo geométrico, pois depois desse valor o semivariograma perde o significado segundo Clark (1979) e Guerra (1988). Contudo, pode ser utilizado algum valor entre os mencionados para o cutoff, desde que se tenha bom senso nessa escolha. Segundo Diggle & Ribeiro Jr (2007), os parâmetros estimados podem ser afetados pelo valor do cutoff.

Considera-se a forma paramétrica da matriz de covariância  $\Sigma = [(\sigma_{ij})]$ , com  $\sigma_{ij} = COV(Z_{s_i}, Z_{s_j})$ , definida na equação (2.3):

$$\Sigma = \varphi_1 I_n + \varphi_2 R \quad (2.3)$$

em que:

- $\varphi_1$  é o efeito pepita ou erro de variância;
- $\varphi_2$  é a contribuição ou variância de dispersão;
- $R$  é uma matriz que é função de  $\varphi_3$ ,  $R = R(\varphi_3) = [(r_{iu})]$ , matriz  $n \times n$  simétrica com seus elementos da diagonal  $r_{ii} = 1, i = 1, \dots, n$ .  $r_{ij} = \frac{1}{2}C(h)$ , sendo  $h = ||i - j||$ , em que  $\varphi_3$  é função do alcance  $a$  do modelo;
- $I_n$  é a matriz identidade  $n \times n$ .

Existe uma relação entre a função semivariância  $\gamma(h)$  e a função covariância  $C(h)$ , por meio da equação (2.4):

$$\gamma(h) = C(0) - C(h) \quad (2.4)$$

sendo  $C(0)$  a variância dos dados.

É importante notar que, ao contrário da covariância ( $C(h)$ ) e da correlação ( $\rho(h) = 1 - \frac{\gamma(h)}{C(0)}$ ), a semivariância é uma medida de dissimilaridade, ou seja, é maior à medida em que as variáveis estão menos associadas. Esta medida exige uma hipótese de estacionariedade menos restritiva em relação às outras medidas possíveis, como por exemplo, a covariância, que exige estacionariedade de segunda ordem (DIGGLE & RIBEIRO JR., 2007). Espera-se que a semivariância aumente com a distância até ocorrer a estabilização em um patamar, correspondendo à distância limite de dependência espacial, definindo o alcance. Assim, quanto mais próximas entre si estiverem as medições, mais semelhantes elas tendem a ser. Então, medições situadas a distâncias maiores que o alcance são não correlacionadas entre si.

A definição da semivariância é dada pela hipótese intrínseca e é definida por meio da equação (2.5):

$$\gamma(h) = \frac{1}{2} E\{[Z(s_1) - Z(s_2)]^2\} \quad (2.5)$$

em que:  $s_1$  e  $s_2$  indicam a localização de dois pontos na região, separados por uma distância  $h$ , sendo  $h = \|s_1 - s_2\|$ .

Como a função semivariância é uma função que descreve, em relação a  $h$ , o valor da distância de  $Z(s+h) - Z(s)$ , conhecendo-se a distância que separa dois pontos estima-se o valor da variância da diferença das variáveis aleatórias em estudo nos pontos.

O estimador mais conhecido da função semivariância, definida na Equação (2.6) é o Estimador Clássico de Matheron (MATHERON, 1963). Seja  $\hat{\gamma}(h)$  a função de semivariância estimada de  $\gamma(h)$ , então, o Estimador Clássico de Matheron, denominado semivariância experimental é definido por:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum^{N(h)} [Z(s) - Z(s+h)]^2 \quad (2.6)$$

no qual:

- $N(h)$  é o número de pares de dados que apresentam um certa distância  $h$ , que deve ser considerado para este estudo como  $N(h) \geq 30$  (JOURNEL & HUIJBREGTS, 1978; GUERRA, 1998; WOLLENHAUPT et al., 1997);
- $Z$  é a variável regionalizada ou função aleatória;
- $s \in S \subset R^2$ .

Na Figura 1 apresenta-se um exemplo da relação expressa pela equação (2.4) entre a função semivariância  $\gamma(h)$  e a função covariância  $C(h)$ , na qual são apresentados os parâmetros efeito pepita  $\varphi_1$ , patamar  $\varphi_1 + \varphi_2$  e alcance  $a = g(\varphi_3)$ .

Quando a variável regionalizada apresenta estrutura de dependência espacial com patamar, o grau de dependência espacial pode ser determinado pelo coeficiente de efeito pepita relativo, EPR, apresentado na equação (2.7):

$$EPR = \frac{\varphi_1}{\varphi_1 + \varphi_2}. \quad (2.7)$$

A dependência espacial é considerada forte, quando o coeficiente de efeito pepita é menor que 0,25. Quando o coeficiente de efeito pepita está entre 0,25 e 0,75, a dependência é considerada moderada e, acima de 0,75, é considerada fraca (CAMBARDELLA et al., 1994).

Caso os semivariogramas direcionais não apresentem os mesmos padrões de comportamento, diz-se que o fenômeno é anisotrópico e estuda-se a semivariância em diferentes direções (GUEDES et al., 2008).

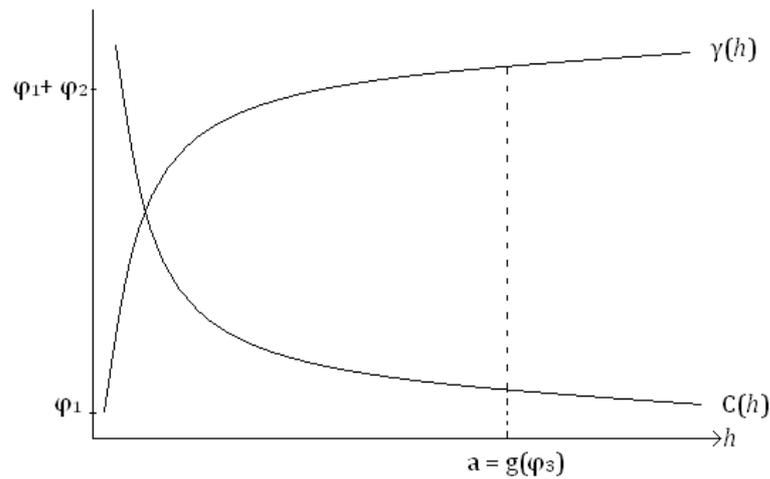


Figura 1 Relação entre a função semivariância  $\gamma(h)$  e a função covariância  $C(h)$ .

## 2.2 Modelos teóricos da função semivariância

O ajuste de um modelo teórico ao semivariograma experimental e a estimativa de seus parâmetros são alguns dos aspectos mais importantes das aplicações da Teoria das Variáveis Regionalizadas e podem ser as maiores fontes de ambiguidade nestas aplicações (VIEIRA et al., 1981). Todos os cálculos de geoestatística dependem do valor do modelo do semivariograma para cada distância especificada.

O gráfico do semivariograma experimental  $\hat{\gamma}(h)$  é formado por uma série de valores sobre os quais se pretende ajustar uma função. É imprescindível que o modelo ajustado represente a tendência  $\hat{\gamma}(h)$  em relação aos valores de  $h$ . Assim, as estimativas obtidas a partir da krigagem<sup>1</sup> serão mais exatas e, portanto, mais confiáveis.

Existem diversos modelos na literatura, tais como esférico, exponencial, gaussiano, Família Matérn, Família exponencial potência, entre outros, que são utilizados para o ajuste de um modelo teórico ao semivariograma experimental (DIGGLE & RIBEIRO JR, 2007). Neste texto serão citados esses modelos.

### 2.2.1 Modelo esférico

O modelo esférico é muito utilizado na representação de fenômenos que apresentam dependência espacial. A expressão para o modelo esférico é apresentado na equação (2.8).

$$\gamma(h) = \begin{cases} 0, & \text{se } h = 0 \\ \varphi_1 + \varphi_2 \left[ \frac{3}{2} \left( \frac{h}{\varphi_3} \right) - \frac{1}{2} \left( \frac{h}{\varphi_3} \right)^3 \right], & \text{se } 0 < h \leq \varphi_3 \\ \varphi_1 + \varphi_2, & \text{se } h > \varphi_3 \end{cases} \quad (2.8)$$

<sup>1</sup>Krigagem é um método de regressão usado em geoestatística para aproximar ou interpolar dados. A krigagem usa informações a partir do variograma para encontrar os pesos ótimos a serem associados às amostras com valores conhecidos que irão estimar pontos desconhecidos

Uma característica do modelo esférico é que apresenta um crescimento rápido na origem, atingindo o patamar numa distância  $h$ , finita, igual a  $\frac{2}{3}$  do alcance. A partir dessa distância, a semivariância torna-se constante. O modelo esférico pode ser representado graficamente conforme a Figura 2:

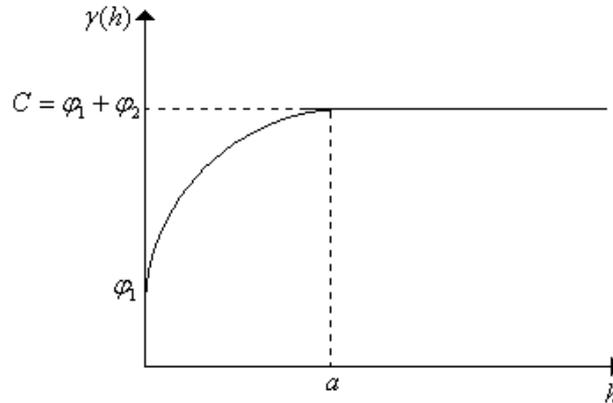


Figura 2 Representação gráfica do modelo esférico.

A função covariância para o modelo esférico é descrita pela equação (2.9):

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & \text{se } h = 0 \\ \varphi_2 \left[ 1 - \frac{3}{2} \left( \frac{h}{\varphi_3} \right) + \frac{1}{2} \left( \frac{h}{\varphi_3} \right)^3 \right], & \text{se } 0 < h \leq \varphi_3 \\ 0, & \text{se } h > \varphi_3 \end{cases} \quad (2.9)$$

No modelo esférico tem-se a função de correlação espacial, expressa pela equação (2.10):

$$\rho(h) = \begin{cases} 1, & \text{se } h = 0 \\ 1 - \frac{3}{2} \left( \frac{h}{\varphi_3} \right) + \frac{1}{2} \left( \frac{h}{\varphi_3} \right)^3, & \text{se } h > 0 \end{cases} \quad (2.10)$$

Vieira (2000) explica que, no modelo esférico, o alcance é obtido selecionando-se os valores iniciais do efeito pepita e do patamar e traçando-se uma reta que intercepte o eixo  $Oy$  em  $\varphi_1$  e seja tangente aos primeiros pontos mais próximos de  $h = 0$ . A reta cruzará o patamar à distância  $a' = \frac{2}{3}a$ , sendo  $a = \varphi_3$ . Esse modelo é utilizado em  $R$ ,  $R^2$  e  $R^3$ . Destaca-se que o modelo esférico não possui segunda derivada.

## 2.2.2 Modelo exponencial

Este modelo caracteriza-se por uma tendência linear na origem, alcançando um patamar apenas assintoticamente.

A expressão para o modelo exponencial é definida pela equação (2.11).

$$\gamma(h) = \begin{cases} 0, & \text{se } h = 0 \\ \varphi_1 + \varphi_2 \left[ 1 - \exp\left(\frac{-h}{\varphi_3}\right) \right], & \text{se } h > 0 \end{cases} \quad (2.11)$$

O modelo exponencial apresenta a seguinte forma, de acordo com a Figura 3:

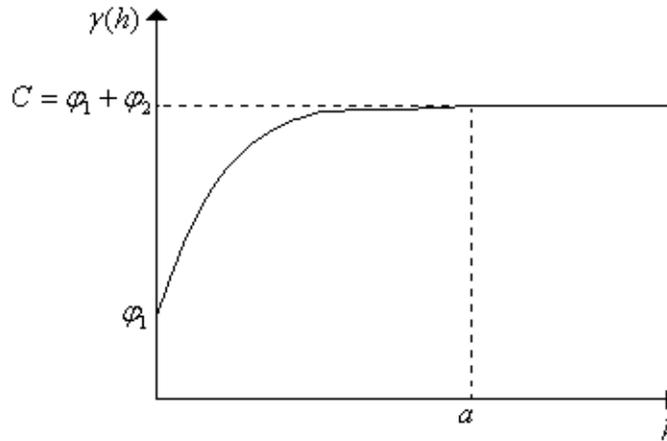


Figura 3 Representação gráfica do modelo exponencial.

A função covariância para o modelo exponencial é da forma:

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & \text{se } h = 0 \\ \varphi_2 \left[ \exp\left(\frac{-h}{\varphi_3}\right) \right], & \text{se } h > 0. \end{cases} \quad (2.12)$$

Para o modelo exponencial, a função de correlação espacial é definida como:

$$\rho(h) = \begin{cases} 1, & \text{se } h = 0 \\ \exp\left(-\frac{h}{\varphi_3}\right), & \text{se } h > 0. \end{cases} \quad (2.13)$$

No modelo exponencial, inicialmente  $a = 3\varphi_3$  é determinado visualmente e  $\varphi_1$  e  $\varphi_2$  são determinados como no modelo esférico. O modelo exponencial pode ser utilizado em  $R$ ,  $R^2$  e  $R^3$  (URIBE-OPAZO & KAVANAGH, 2006).

### 2.2.3 Modelo gaussiano

A expressão para o modelo gaussiano é:

$$\gamma(h) = \begin{cases} 0, & \text{se } h = 0 \\ \varphi_1 + \varphi_2 \left\{ 1 - \exp\left[-\left(\frac{h}{\varphi_3}\right)^2\right] \right\}, & \text{se } h > 0. \end{cases} \quad (2.14)$$

O modelo gaussiano tem a representação gráfica conforme a Figura 4:

A função covariância para o modelo gaussiano é da forma:

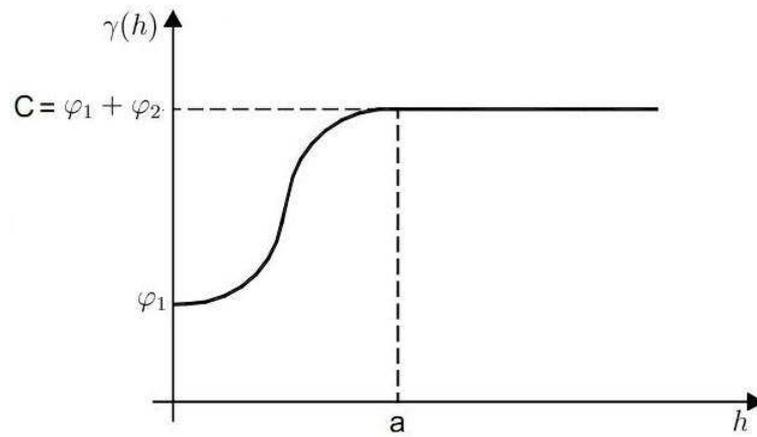


Figura 4 Representação gráfica do modelo gaussiano.

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & \text{se } h = 0 \\ \varphi_2 \left\{ \exp \left[ - \left( \frac{h}{\varphi_3} \right)^2 \right] \right\}, & \text{se } h > 0. \end{cases} \quad (2.15)$$

Para o modelo gaussiano, a correlação espacial é dada por:

$$\rho(h) = \begin{cases} 1, & \text{se } h = 0 \\ \exp \left[ - \left( \frac{h}{\varphi_3} \right)^2 \right], & \text{se } h > 0 \end{cases} . \quad (2.16)$$

O modelo gaussiano apresenta alcance extenso e seu patamar é semelhante ao modelo exponencial. O que os distingue é que o modelo gaussiano demonstra um comportamento em forma de parábola próximo à origem. O alcance é definido como  $a = \sqrt{3}\varphi_3$ . Também pode ser utilizado em  $R$ ,  $R^2$  e  $R^3$  (URIBE-OPAZO & KAVANAGH, 2006).

#### 2.2.4 Família Matérn

Conforme Diggle & Ribeiro Jr (2007), essa família foi assim nomeada após Bertil Matérn, sueco, que a introduziu em 1960, em sua tese de doutorado, republicada, posteriormente, como Matérn (1986). É definida pela expressão (2.17).

$$\gamma(h) = \begin{cases} 0, & \text{se } h = 0 \\ \varphi_1 + \varphi_2 \left[ 1 - \left( 2^{k-1} \Gamma(k) \right)^{-1} \left( \frac{h}{\varphi_3} \right)^k K_k \left( \frac{h}{\varphi_3} \right) \right], & \text{se } h > 0 \end{cases} \quad (2.17)$$

em que  $\varphi_1, \varphi_2, \varphi_3$  e  $k$ , são parâmetros;  $K_k(\cdot)$  é a função de Bessel de terceiro tipo, de ordem  $k$ .

A Família Matérn é válida para  $\varphi_3, k > 0$  e corresponde à função de correlação exponencial quando  $k = 0,5$ ; e a função de correlação gaussiana quando  $k \rightarrow \infty$ . O gráfico da Figura 5 apresenta o modelo Família Matérn para  $k = 0,1; k = 0,2; k = 0,5; k = 1; k = 2$ :

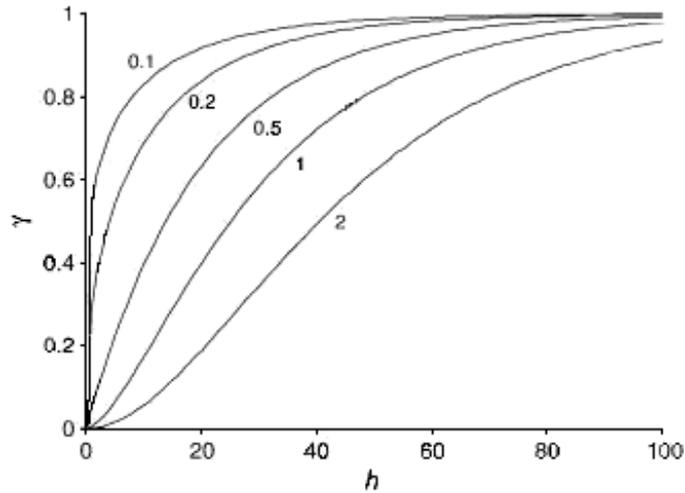


Figura 5 Representação gráfica do modelo Família Matérn para  $k=0,1$ ,  $k=0,2$ ,  $k=0,5$ ,  $k=1$  e  $k=2$ ,  $c_0 = 0$ ,  $c_1 = 1$  e range  $r = 20$ . Adaptado de Minasny & McBratney (2005).

A função de covariância para a Família Matérn é expressa pela equação (2.18):

$$C(h) = \begin{cases} 0, & \text{se } h = 0 \\ \varphi_2 \left[ 1 - \left( 2^{k-1} \Gamma(k) \right)^{-1} \left( \frac{h}{\varphi_3} \right)^k K_k \left( \frac{h}{\varphi_3} \right) \right], & \text{se } h > 0 \end{cases} \quad (2.18)$$

A função de correlação espacial é expressa pela equação (2.19):

$$\rho(h) = \{ 2^{k-1} \Gamma(k) \}^{-1} \left( \frac{h}{\varphi_3} \right)^k K_k \left( \frac{h}{\varphi_3} \right), h > 0 \quad (2.19)$$

### 2.2.5 Família exponencial potência

O modelo Família exponencial potência tem a equação (2.20) como expressão.

$$\gamma(h) = \left\{ \varphi_1 + \varphi_2 \left\{ 1 - \exp \left[ - \left| \frac{h}{\varphi_3} \right|^k \right] \right\} \right\}, \text{ se } h > 0 \quad (2.20)$$

Como a Família Matérn, tem um parâmetro de escala e gera funções de correlação monótonas e decrescentes em  $h$ . O modelo gaussiano é um caso particular do modelo exponencial potência, quando  $k = 2$ . (Quando  $k = 1$ , este modelo é equivalente ao modelo exponencial. O valor de  $k$  pode variar,  $0 \leq k \leq 2$ .)

A função covariância para a Família exponencial potência é dada pela equação (2.21):

$$C(h) = \{ 2^{k-1} \Gamma(k) \}^{-1} \left( \frac{h}{\varphi_3} \right)^k K_k \left( \frac{h}{\varphi_3} \right), h > 0. \quad (2.21)$$

A Família exponencial potência tem como função de correlação a equação:

$$\rho(h) = \exp \left[ - \left| \frac{h}{\varphi_3} \right|^k \right], \text{ se } 0 < k \leq 2. \quad (2.22)$$

## 2.3 Métodos de estimação de parâmetros

O ajuste de modelos teóricos a semivariogramas experimentais precisa de um método estatístico que estime o vetor de parâmetros  $\theta = (\varphi_1, \varphi_2 \text{ e } \varphi_3)^T$ , chamado  $\hat{\theta} = (\hat{\varphi}_1, \hat{\varphi}_2, \hat{\varphi}_3)^T$  de um semivariograma experimental. Na literatura, diversos métodos de estimação de parâmetros são utilizados, tais como: Mínimos quadrados ordinários (OLS), Mínimos quadrados ponderados, subdivididos em: WLS1, WLS2 e WLS3, IGF (Índice de Melhor Ajuste), Máxima Verossimilhança (MV) e Máxima Verossimilhança Restrita (MVR) (URIBE-OPAZO et al., 2007).

### 2.3.1 Máxima verossimilhança - MV

Seja uma variável em estudo, que apresenta distribuição gaussiana de probabilidades, utiliza-se o método de máxima verossimilhança para estimar os parâmetros da estrutura da matriz de covariância  $\Sigma$  definida na equação (2.3). O método de máxima verossimilhança consiste em maximizar o logaritmo da função verossimilhança, em relação aos parâmetros  $\varphi_1$ ,  $\varphi_2$  e  $\varphi_3$  e  $\beta$ 's do modelo. Esse método é iterativo e fornece sempre estimativas positivas de componentes de variância.

Para a estimação do vetor de parâmetros  $\delta = (\beta^T, \theta^T)^T$  desconhecido, escolhe-se um vetor que maximiza a função verossimilhança no domínio  $\Theta$ , sendo  $\Theta$  o espaço paramétrico.

Considere-se um processo gaussiano  $\{Z(s), s \in S\}$ , sendo que  $Z$  tem distribuição normal  $n$ -variada, com vetor de médias  $\mathbf{X}\beta$  e respectiva matriz de covariância  $\Sigma = \Sigma(\theta)$ , isto é,  $Z \sim N_n(\mathbf{X}\beta, \Sigma)$ .

O logaritmo da função de verossimilhança é uma função do vetor de parâmetros  $\delta = (\beta^T, \theta^T)^T \in \Theta$ , sendo expressa pela equação (2.23).

$$l(\delta) = -\frac{1}{2} \log(2\Pi) - \frac{1}{2} \log|\Sigma| - \frac{1}{2} (Z - \mathbf{X}\beta)^T \Sigma^{-1} (Z - \mathbf{X}\beta). \quad (2.23)$$

A estimativa de máxima verossimilhança consiste em maximizar o logaritmo da função verossimilhança, expresso pela equação (2.24).

$$l(\hat{\beta}, \hat{\theta}) = \max l(\beta, \theta). \quad (2.24)$$

### 2.3.2 Máxima verossimilhança restrita - MVR

Outro método utilizado para estimar os parâmetros das funções de covariância  $\Sigma$  é a Máxima Verossimilhança Restrita - MVR (ZIMMERMAN, 1989; CRESSIE, 1993; CHRISTENSEN, JOHNSON & PEARSON, 1993).

Segundo Schabenberger & Gotway (2005), um estimador de máxima verossimilhança restrita de  $\theta = (\varphi_1, \varphi_2, \varphi_3)^T$  é definido com o vetor  $\hat{\theta} \in \Theta$ , que maximiza a função de verossimilhança  $W = CZ$ , em que a matriz  $C$  é de ordem  $(n - 1) \times n$  de contrastes linearmente independentes (um contraste é qualquer combinação linear  $W = CZ$ , tal que  $E[CZ] = 0$ ).

Se  $Z \sim N_n(\mathbf{X}\beta, \Sigma)$ , então,  $W = CZ \sim N(0, C\Sigma C^T)$ , assim, o estimador de máxima verossimilhança restrito de  $\theta$  consiste em maximizar o logaritmo da função verossimilhança restrita (BORSSOI et al., 2009; BORSSOI et al., 2011), definida na equação (2.25):

$$l_R(\theta) = -\frac{(n-1)}{2} \log(2\pi) - \frac{1}{2} \log|C\Sigma(\theta)C^T| - \frac{1}{2} W^T (C\Sigma(\theta)C^T)^{-1} W. \quad (2.25)$$

## 2.4 Critérios de validação dos ajustes de modelos geoestatísticos

Ao realizar-se o processo de modelagem geoestatística, encontram-se diversos modelos para um mesmo conjunto de dados. Pode haver dúvida se tal ajuste é o melhor possível. É preciso, então, definir qual o modelo que melhor se ajusta a eles. Dentre os métodos de validação de modelos, os que vêm sendo utilizados são aqueles nos quais as técnicas são baseadas na comparação entre valores teóricos de modelos geoestatísticos e valores empíricos.

A seguir apresentam-se os critérios de Validação Cruzada, Akaike e o Valor Máximo do Logaritmo da Função Verossimilhança.

### 2.4.1 Validação cruzada

A validação cruzada (cross validation) é uma técnica que permite a comparação das  $n$  observações com  $n$  valores estimados nos pontos onde foram coletadas as observações, ou seja, a estimação é avaliada nos locais onde foram retiradas as amostras (DAVIS, 1987).

Um valor  $Z(s_\alpha)$  amostrado é retirado temporariamente do conjunto de observações e é estimado com base nos demais valores,  $Z(s_i)$ ,  $i = 1, \dots, n$  e  $i \neq \alpha$ . Este procedimento é repetido para as  $n$  observações e obtém-se, assim,  $n$  pares ordenados do tipo  $(Z(s), \hat{Z}(s))$  (valor real, valor estimado).

Com esse conjunto de pares ordenados calculam-se os parâmetros estatísticos básicos das distribuições univariadas das estimativas e dos erros. É de esperar que numa boa estimativa os parâmetros de estatística descritiva univariada se assemelhem aos mesmos calculados para os valores reais, bem como é de desejar que os erros sejam mínimos e faz-se também a sua análise.

Seja  $U(S_{(i)}) = Z(s_i) - \hat{Z}(s_{(i)})$  para  $i = 1, 2, \dots, n$ . O Erro Médio  $EM$  por validação cruzada é obtido na equação (2.26):

$$EM = \frac{1}{n} \sum_{i=1}^n U(S_{(i)}) \quad (2.26)$$

em que  $n$  é o número de dados,  $Z(s_i)$  é o valor observado no ponto  $s_i$ ,  $\hat{Z}(s_{(i)})$  é o valor estimado por krigagem ordinária no ponto  $s_i$ , desconsiderando a observação  $Z(s_i)$ .

Webster & McBratney (1989), Cressie (1993), Vieira (2000), Mello et al. (2003) e Faraco et al. (2008) apresentaram o erro médio reduzido,  $ER$ , desvio padrão dos erros médios,  $DP_{EM}$ , desvio padrão dos erros reduzidos,  $DP_{ER}$  e erro absoluto,  $EA$  como medidas para avaliar modelos geoestatísticos.

O erro médio reduzido,  $ER$ , é dado pela equação (2.27):

$$ER = \frac{1}{n} \sum_{i=1}^n \frac{U(S_{(i)})}{\sigma(\hat{Z}(s_{(i)}))} \quad (2.27)$$

em que:  $\sigma(\hat{Z}(s_{(i)}))$  é o desvio padrão da krigagem no ponto  $s_i$ , sem considerar a observação  $Z(s_i)$ .

O desvio padrão do erro médio,  $DP_{ER}$ , é definido pela equação 2.28:

$$DP_{ER} = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{|U(S_{(i)})|}{\sigma(\hat{Z}(s_{(i)}))}} \quad (2.28)$$

O desvio padrão do erro reduzido,  $DP_{ER}$ , deve ser o mais próximo de um.

O erro absoluto  $EA$  é uma medida da magnitude dos erros na unidade variável. Se o conjunto de valores medidos e preditos por krigagem ordinária  $Z(s_i)$  e  $\hat{Z}(s_{(i)})$  for conhecido, é possível determinar o valor para o erro absoluto (na unidade da variável analisada), pela expressão (2.29):

$$EA = \sum_{i=1}^n |Z(s_i) - \hat{Z}(s_{(i)})|. \quad (2.29)$$

Segundo os autores citados acima, aplicando-se as condições de não tendenciosidade, os valores para  $EM$  e para  $ER$  devem ser o mais próximo de zero. Para a medida  $DP_{ER}$ , entende-se como melhor aquela que possui o menor valor.

Uma das limitações da Validação Cruzada é que ela gera pares de valores amostrados e estimados apenas no local da amostra.

#### 2.4.2 Critério de informação de Akaike

O Critério de informação de Akaike,  $IAC$  ou Akaike's information criterion, procura uma solução satisfatória entre o bom ajuste e o princípio da parcimônia, ou seja, ter preferência pela explicação mais simples para uma observação (AKAIKE, 1992). O Teorema de Akaike procura demonstrar que se dois modelos apresentam resultados igualmente satisfatórios, então, o modelo mais simples trará melhores desempenhos em novas predições. Logo, o  $IAC$  impõe uma penalidade para a complexidade dos modelos.

A estatística para o critério de informação de Akaike é definida pela expressão (2.30):

$$IAC = -2l(\delta) + 2k \quad (2.30)$$

em que:

$l(\delta)$ : é a função logaritmo de verossimilhança,  $(\delta = (\beta^T, \theta)^T)$ ;

$k$ : é o número de parâmetros do modelo ajustado.

O critério de informação de Akaike é uma ligação da função de verossimilhança com uma medida de informação perdida, quando um modelo representa de modo aproximado uma realidade.

## 2.5 Construção de mapas por krigagem

Na confecção de mapa temático, o pesquisador tem os pontos amostrais da área, de acordo com a variável regionalizada de interesse, que pode ser: produtividade da cultura, teor de água, potássio, nitrogênio, resistência do solo à penetração, entre outros atributos físicos e químicos do solo. Quando tem-se apenas estes pontos amostrais não se tem uma definição de como está se comportando toda a área em estudo, somente a parte amostrada. Para que se possa visualizar o comportamento da variável por toda a área em estudo, é necessário que se utilize algum método de interpolação para a criação de mapas temáticos. Esse método tornará possível a construção de um mapa de contorno, que permitirá que se realize uma análise de toda a área, inclusive das partes não amostradas.

Pesquisas têm sido realizadas para saber qual é o interpolador mais adequado para a construção dos mapas temáticos, e diversos pesquisadores como Landim (2006), Voltz & Webster (1990), Thompson (1992), Vieira (2000), Mello et al. (2003), Gomes et al. (2008), Faraco et al. (2008), concluem que o melhor método de interpolação é a krigagem. Frente a outros métodos de interpolação, a krigagem apresenta melhor desempenho, haja vista a predominância de menores erros produzidos por ele, como retratam os trabalhos supracitados. Mello et al. (2003), utilizando a estimativa de parâmetros da equação de chuvas intensas, constataram que a krigagem produziu menores erros médios, comparativamente ao método do inverso do quadrado da distância. Carvalho & Assad (2005) observaram que o interpolador geoestatístico apresentou melhor resultado que o inverso do quadrado da distância e que a curvatura mínima na espacialização da precipitação pluviométrica para o Estado de São Paulo, com menor quadrado médio do erro. Silva et al. (2008), estudando atributos químicos do solo, concluíram que a interpolação dos dados por krigagem apresentou uma melhor desempenho, visto que o inverso do quadrado da distância possuiu maior erro na estimação dos dados.

A krigagem ou krigeagem é uma técnica utilizada na geoestatística para estimar valores de propriedades, para locais onde valores dessa propriedade não foram medidos.

O termo krigagem é derivado do nome de Daniel G. Krige, que foi o pioneiro em introduzir o uso de médias móveis para evitar a superestimação sistemática de reservas em mineração (DELFINER & DELHOME, 1975). O que diferencia a krigagem de outros métodos de interpolação é a estimação de uma matriz de covariância espacial que determina os pesos atribuídos às diferentes amostras, o tratamento da redundância dos dados, a vizinhança

a ser considerada no procedimento inferencial e o erro associado ao valor estimado. Além disso, a krigagem também fornece os melhores estimadores exatos com propriedades de não tendenciosidade e eficiência.

Para que esta ferramenta seja usada, é necessário que exista a dependência espacial descrita pelo semivariograma. Assim, a krigagem é um processo no qual se estimam valores de propriedades espacialmente distribuídas, a partir de valores adjacentes (VIEIRA et al., 1983). O valor estimado da variável em estudo no ponto  $s_0$  é dado pela equação (2.31).

$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i Z(s_i) \quad (2.31)$$

em que  $N$  é o número de pontos adjacentes que foram medidos,  $Z(s_i)$ , e que foram utilizados nesta estimação para a propriedade em estudo e  $\lambda_i$  são os ponderadores aplicados a cada  $Z(s_i)$ . Observa-se que estes são selecionados de forma que a estimativa obtida seja não tendenciosa. Logo, duas condições são assumidas:

$$E[\hat{Z}(s_0) - Z(s_0)] = 0 \quad \text{e} \quad \sigma_k^2 = \text{mínimo}\{E[\hat{Z}(s_0) - Z(s_0)]^2\}. \quad (2.32)$$

A estimativa de valores para locais não amostrados, por meio da krigagem, permite estabelecer um mapa para a área de estudo, para as propriedades do solo e atributos das culturas em estudo, possibilitando a definição de isolinhas (ou linhas de isovalores), as quais podem ser de grande utilidade para o planejamento experimental (SOUZA, 1997).

Considerando-se o processo  $Z(\cdot)$  como intrinsecamente estacionário, tem-se que o erro médio é:

$$E[Z(s_0) - \hat{Z}(s_0)] = \mu - \mu \sum_{i=1}^n \lambda_i = \mu(1 - \sum_{i=1}^n \lambda_i) \quad (2.33)$$

Esse erro somente será nulo se:

$$\sum_{i=1}^n \lambda_i = 1 \quad (2.34)$$

Esta condição é necessária para que o estimador não seja tendencioso. Também a variância da estimativa é mínima, isto é:

$$\sigma_E^2 = \text{Var}[Z(s_0) - \hat{Z}(s_0)] = E[(Z(s_0) - \hat{Z}(s_0))^2] - \{E[\hat{Z}(s_0) - Z(s_0)]\}^2. \quad (2.35)$$

Essas duas condições garantem que o estimador de krigagem é "BLUE", Best Linear Unbiased Estimator, o melhor estimador linear não tendencioso de variância mínima.

Substituindo a equação (2.31) no segundo membro de (2.35), desenvolvendo a expressão resultante e minimizando-a sujeita à restrição expressa em (2.34), chega-se a que

os valores de  $\lambda_i$ ,  $i = 1, 2, 3, \dots, n$ , que garantem não tendenciosidade e variância mínima, são soluções do sistema de equações (2.36):

$$\begin{cases} \sum_{i=1}^n \lambda_j \gamma(h_{ij}) - \eta = \gamma(h_{i,0}), & i = 1, 2, \dots, n, & h_{ij} = \|s_i - s_j\|, & h_{i0} = \|s_0 - s_i\| \\ \sum_{i=1}^n \lambda_i = 1 \end{cases} \quad (2.36)$$

O sistema expresso em (2.36), constituído de  $n + 1$  equações e  $n + 1$  incógnitas é conhecido como sistema de krigagem ordinária. A incógnita  $\eta$  é um multiplicador de Lagrange, introduzido ao minimizar a variância do erro. Na forma matricial, o sistema pode ser escrito como na equação (2.37):

$$\gamma \lambda = A \quad (2.37)$$

em que:

$$\lambda = [\lambda_1 \quad \lambda_2 \quad \lambda_3 \quad \dots \quad \lambda_n \quad \eta]^T$$

$$\gamma = \begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \dots & \gamma(h_{1n}) & 1 \\ \gamma(h_{21}) & \gamma(h_{22}) & \dots & \gamma(h_{2n}) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(h_{n1}) & \gamma(h_{n2}) & \dots & \gamma(h_{nn}) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix}$$

$$A = [\gamma(h_{01}) \quad \gamma(h_{02}) \quad \dots \quad \gamma(h_{0n}) \quad 1]^T.$$

Os pesos  $\lambda_i$  podem, então, ser obtidos mediante o produto da matriz inversa de  $\gamma$  pela matriz  $A$ , ou seja,  $\lambda = \gamma^{-1}A$ .

Uma vez obtidos os pesos  $\lambda_i$  e o valor  $\eta$ , pode-se também calcular a variância do erro que é dada pela equação (2.38) (CRESSIE, 1993).

$$\sigma_E^2 = \sum_{i=1}^n \lambda_i \gamma(h_{0i}) + \eta. \quad (2.38)$$

## 2.6 Comparação de mapas temáticos

A grande quantidade de recursos existentes para a investigação de informações acaba por conduzir o pesquisador a construir diversos mapas temáticos. Logo, é necessário lançar

mão de técnicas que permitam a comparação desses mapas. Uma das maneiras de investigação da similaridade entre os mapas temáticos é a utilização de indicadores de acurácia, derivados da matriz de confusão que, por sua vez, é originada a partir dos elementos da matriz de contingência das classes dos mapas temáticos, conhecida como matriz dos erros. Pode-se calcular diversos índices a partir das informações contidas na matriz dos erros, como o índice Kappa,  $\kappa$  (COHEN, 1960), e o índice Tau,  $\tau$  (MA & REDMOND, 1995) e a Exatidão Global, *EG*.

Porém, é necessário que sejam definidos, inicialmente, os termos acurácia, exatidão e precisão, para que não sejam confusos.

Segundo Hush et al. (1982), precisão é o grau de concordância de uma série de observações ou medidas, enquanto a acurácia remete à proximidade de uma série de observações ou medida de seu valor real. Logo, precisão pode ser relacionada a erros aleatórios e sistemáticos conjuntamente (GEMAEL, 1994). Já a exatidão ocorre se não houve erro sistemático, pois a falta de exatidão das observações é, geralmente, oriunda dos erros sistemáticos e tendem a se acumular num mesmo sentido.

### 2.6.1 Matriz de erros

A matriz de erros é uma matriz  $m \times m$  e é representada pela Tabela 1, na qual  $i$  representa a linha,  $j$  representa a coluna,  $n_{.j}$  o total na coluna,  $n_i$  o total na linha:

Tabela 1 Matriz genérica dos erros

Mapa modelo	Mapa de Referência				Total
	1	2	...	$m$	
1	$n_{11}$	$n_{12}$	...	$n_{1m}$	$n_{1.}$
2	$n_{21}$	$n_{22}$	...	$n_{2m}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$m$	$n_{m1}$	$n_{m2}$	...	$n_{mm}$	$n_{m.}$
<b>Total</b>	$n_{.1}$	$n_{.2}$	...	$n_{.m}$	$n$

Nesta matriz, o total de pixels<sup>2</sup> do mapa de referência de cada classe é quantificado nas colunas, e o total de pixels do mapa modelo de cada classe é quantificado nas linhas. Cada célula apresenta o número total de pixels pertencentes à classe  $i$  do mapa modelo e a classe  $j$  do mapa de referência ( $n_{ij}$ ). A diagonal principal (quando  $i = j$ ) representa casos em que os pixels foram classificados por ambos os mapas ( $n_{ii}$ ), enquanto os elementos fora da diagonal principal representam as classificações diferenciadas dos mapas. Assim, se todos os mapas forem idênticos na classificação de seus pixels, os elementos fora da diagonal principal da matriz de erros serão nulos.

<sup>2</sup>Pixel é a contração das palavras picture element, e significa elemento de imagem, é o menor elemento num dispositivo de exibição, como um monitor, por exemplo, sendo o menor ponto que forma uma imagem digital. Um conjunto de pixels forma a imagem inteira

## 2.6.2 Índice Kappa

Um dos procedimentos que vem sendo utilizado para mensurar a exatidão das classificações temáticas é o índice Kappa. A estatística Kappa, introduzida por Cohen (1960), é uma medida largamente utilizada para a acurácia na classificação de mapas temáticos em estudos de pesquisadores como Lee, Grunes & Pottier (2001), Brites, Soares & Ribeiro (1996) e Lim et al. (2009). O índice Kappa é uma proporção de acerto depois da eliminação do acerto por acaso (ROSENFELD & FITZPATRICK-LINS, 1986). Em outras palavras, a análise do índice kappa é uma técnica multivariada discreta usada em estimativas de precisão para determinar estatisticamente se uma matriz de erros é significativamente diferente da outra.

Considere a situação de comparar dois mapas temáticos digitais referentes a uma mesma área de estudos, e que cada mapa é constituído por  $n$  pixels. O índice Kappa é definido por Cohen (1960) como um coeficiente de concordância para as escalas nominais que mede a proporção de concordância depois que a concordância atribuída à casualidade é retirada de consideração.

O índice de concordância Kappa é dado pela equação (2.39):

$$\hat{\kappa} = \frac{n \sum_{i=1}^m n_{ii} - \sum_{i=1}^m (n_i \cdot n_{\cdot i})}{n^2 - \sum_{i=1}^m (n_i \cdot n_{\cdot i})} \quad (2.39)$$

em que:

- $n_{ii}$  é o número de pixels na linha  $i$ , coluna  $i$ ;
- $n_i$  e  $n_{\cdot i}$  são os totais marginais da linha  $i$  e coluna  $i$ , respectivamente;
- $n$  é o número total de pixels.

A estatística Kappa, geralmente, varia entre 0 e 1, podendo ser transformado em percentual ao se multiplicar por 100, sendo os valores mais próximos de um (100%) representativos de maior concordância. É possível que existam valores negativos para Kappa, porém, não é comum. Na Tabela 2 observa-se a relação entre o índice Kappa e o desempenho de classificação obtido (KRIPPENDORFF, 1980):

Tabela 2 Relação entre o valor do índice Kappa e o desempenho de classificação obtido (KRIPPENDORFF, 1980)

Índice Kappa	Desempenho de classificação
maior que 0,80	concordância excelente
entre 0,67 e 0,80	concordância regular a boa
menor que 0,67	concordância ruim

Uma das vantagens do cálculo do índice Kappa é a possibilidade do uso desse valor como uma base para determinar a significância estatística de uma matriz de erros ou a diferença entre duas matrizes.

### 2.6.3 A estatística Tau

O índice Tau ( $\tau$ ), ou Kappa modificado, similar ao índice Kappa ( $\kappa$ ), fornece uma medida quantitativa relativamente precisa e intuitiva sobre a acurácia de classificação. A equação (2.40) apresenta o índice Tau, em que  $p_i$  indica a probabilidade a priori de cada classe:

$$\tau = \frac{\sum_{i=1}^m n_{ii} - p_i}{1 - p_i} \quad (2.40)$$

Quando as probabilidades a priori forem iguais para cada classe, tem-se  $p_i = \frac{1}{m}$ , em que  $m$  é o número de classes da matriz de erros, com  $\sum_{i=1}^m p_i = 1$

Segundo Ma & Redmond (1995), o coeficiente Tau é mais fácil de entender e interpretar que o Kappa. A concordância por chance, utilizada para calcular Tau, é uma medida direta da atribuição casual. Assim, o Tau prevê uma indicação clara de como a classificação comporta-se, quando comparada com o assinalamento casual dos pixels às classes. Com Kappa, entretanto, a concordância por chance pode resultar do julgamento do analista condicionado à casualidade, superestimando a concordância por chance.

### 2.6.4 Exatidão global, acurácia do produtor e do usuário

A acurácia global, também chamada de exatidão global,  $EG$ , é uma das medidas mais simples, sendo calculada pela soma do total de pixels corretamente classificados, dividida pelo número total de pixels da matriz de erros, conforme a equação (2.41):

$$EG = \sum_{i=1}^m \frac{n_{ii}}{n} \quad \text{com } 0 \leq EG \leq 1 \quad (2.41)$$

O índice de exatidão global (EG) acima de 0,85 indica uma precisão aceitável, segundo Anderson et al. (1976).

A acurácia do produtor, equação (2.42), indica a probabilidade de um pixel de referência ter sido corretamente classificado:

$$Ap_i = \frac{n_{ii}}{n_{.i}} \quad \text{com } 0 \leq Ap_i \leq 1. \quad (2.42)$$

A acurácia do usuário calcula a probabilidade de um pixel classificado dentro de uma determinada classe realmente pertencer àquela classe no campo (2.43) (STORY & CONGALTON, 1986):

$$Au_i = \frac{n_{ii}}{n_i}, \quad \text{com } 0 \leq Au_i \leq 1. \quad (2.43)$$

Ponzoni & Rezende (2002) afirmam que a exatidão global apresenta maiores valores em relação aos demais índices, por considerar somente a diagonal principal da matriz de contingência, ignorando a ocorrência de concordância casual entre as classes. Entretanto, o coeficiente Kappa, ao calcular a concordância casual, inclui nos cálculos os elementos da diagonal principal, fazendo com que esta seja superestimada, reduzindo o valor do índice.

### 2.6.5 Outras medidas de acurácia

Além da acurácia global, acurácia do produtor e acurácia do usuário, há outras estatísticas para globais e por categoria. Considerando uma amostragem, a acurácia seria expressa pelo desvio padrão da amostra do valor real da população. A precisão poderia ser relacionada ao desvio padrão da média da amostra. É possível que uma amostragem possua uma alta precisão, ou seja em pequeno desvio em relação a média, e ao mesmo tempo, devido a falta de exatidão, um grau acurácia inferior (GEMAEL, 1994).

A Tabela 3 apresenta outras medidas de acurácia para classes. Observa-se que cada medida de acurácia apresenta uma determinada faixa de valores possíveis de serem assumidos.

Tabela 3 Índices de acurácia por classe

Estatística	Equação	Intervalos	Referência
Kappa condicional (usuário)	$K_{cu_i} = \frac{Au_i - \frac{n_i}{n}}{1 - \frac{n_i}{n}}$	$] - \infty, 0]$	Rosenfield & Fitzpatrick-Lins (1986)
Kappa condicional (produtor)	$K_{cp_i} = \frac{Ap_i - \frac{n_i}{n}}{1 - \frac{n_i}{n}}$	$] - \infty, 0]$	Rosenfield & Fitzpatrick-Lins (1986)
Kappa condicional modificado (usuário)	$K_{cmu_i} = \frac{Au_i - \frac{1}{m}}{1 - \frac{1}{m}}$	$[-1, 1]$	Stehman (1997)
Kappa condicional modificado (produtor)	$K_{cmp_i} = \frac{Ap_i - \frac{1}{m}}{1 - \frac{1}{m}}$	$[-1, 1]$	Stehman (1997)

Também há outras estatísticas globais para acurácia, as quais estão expressas na Tabela 4.

### 2.6.6 Matriz de confusão

Grande parte dos métodos quantitativos para comparação entre mapas utiliza a matriz de confusão, obtida dos conjuntos de dados de classificação e referência, como ponto de partida para as medidas de acurácia (STORY & CONGALTON, 1986). A matriz de confusão apresenta uma visão tabulada da acurácia do mapa de classificação, além de permitir o cálculo de medidas específicas como a acurácia global, acurácia do usuário e acurácia do produtor (CONGALTON, 1991). A matriz de confusão é uma matriz quadrada, na qual seus elementos expressam o número de unidades de amostras de uma categoria particular relativa e inferida por classificador, ou regra de decisão, comparado com a categoria atual verificada no campo. Os elementos da diagonal principal indicam o nível de acerto, ou concordância, entre dois

Tabela 4 Índices de acurácia global

Estatística	Equação	Intervalos	Referência
Acurácia média pela perspectiva do usuário	$A_{mu} = \frac{1}{m} \sum_{i=1}^m Au_i$	[0, 1]	Fung &LeDrew (1998)
Acurácia média pela perspectiva do produtor	$A_{mp} = \frac{1}{m} \sum_{i=1}^m Ap_i$	[0, 1]	Fung &LeDrew (1998)
Média das acurácias médias pelas perspectivas do usuário e do produtor	$Maup = \frac{A_{mu} + A_{mp}}{2}$	[0, 1]	Fung &LeDrew (1998)
Índice de acurácia média de Hellden	$I_{amH} = \frac{1}{m} \sum_{i=1}^m \frac{2 \frac{n_{ii}}{n}}{\frac{n_{i.}}{n} + \frac{n_{.i}}{n}}$	[0, 1]	Hellden (1980), Rosenfield & Fitzpatrick-Lins (1986)
Índice de acurácia do mapeamento de Short	$I_{amS} = \frac{1}{m} \sum_{i=1}^m \frac{\frac{n_{ii}}{n}}{\frac{n_{i.}}{n} + \frac{n_{.i}}{n} - \frac{n_{ii}}{n}}$	[0, 1]	Short (1982), Rosenfield & Fitzpatrick-Lins (1986)
Acurácia combinada pelas perspectivas do produtor e usuário	$A_{cpu} = \frac{EG + I_{amH}}{2}$	[0, 1]	Fung &LeDrew (1998)

conjuntos de dados. A Tabela 5 apresenta de forma genérica a matriz de confusão da  $i$ -ésima classe:

Tabela 5 Matriz genérica de confusão da  $i$ -ésima classe (JENSEN, 1986)

		Dados de referência: reais ou controlados	
		verdadeiro	falso
Dados do modelo	verdadeiro	$a_i$	$b_i$
	falso	$c_i$	$d_i$

em que:

$a_i = n_{ii}$ : é o número de pixels pertencentes à classe  $i$  do mapa real que foram classificados corretamente como pertencentes à classe  $i$  do mapa modelo;

$b_i = n_{i.} - n_{ii}$ : é o número de pixels que não pertencem à classe  $i$  do mapa real, mas foram classificadas como pertencentes à classe  $i$  do mapa modelo;

$c_i = n_{.i} - n_{ii}$ : é o número de pixels pertencentes à classe  $i$  do mapa real e que pertencem a uma classe diferente de  $i$  no mapa modelo;

$d_i = n - (a_i + b_i + c_i)$ : é o número de pixels que não pertencem à classe  $i$  no mapa real e foram classificados como não pertencentes à classe  $i$  no mapa modelo.

A Tabela 6 apresenta algumas medidas, obtidas da matriz de confusão, úteis na comparação entre os mapas modelos e o mapa real:

O índice de sensibilidade, ( $S$ ), é uma medida que indica a probabilidade de que um pixel no mapa modelo seja classificado como pertencente à classe  $i$  se ele de fato pertence

Tabela 6 Índices obtidos a partir da matriz de confusão por classe

Métrica	Expressão
Índice de sensibilidade (S)	$\frac{a_i}{a_i + c_i}$
Índice de especificidade (E)	$\frac{d_i}{b_i + d_i}$
Taxa de falso positivo (TFP) (Erro de comissão)	$\frac{b_i}{b_i + d_i}$
Taxa de falso negativo (TFN) (Erro de omissão)	$\frac{c_i}{a_i + c_i}$

à classe  $i$  no mapa real, ou seja, esta medida equivale a acurácia do produtor  $ap$  (FIELDING & BELL, 1997). Em outras palavras, a sensibilidade é a proporção de verdadeiros positivos, a capacidade do sistema de prever corretamente a condição para casos que realmente a têm. O índice de especificidade ( $E$ ) indica a probabilidade de um pixel não pertencente à classe  $i$  do mapa real ser classificado como não pertencente à classe  $i$  no mapa modelo (LURZ et al., 2001), ou seja, é a capacidade do sistema em prever corretamente a ausência da condição para casos que realmente não a têm. Os erros de comissão, também ditos erros de superestimativa, são os que indicam a proporção de pixels que não pertencem à classe  $i$  no mapa real, mas que são classificados como pertencentes à classe  $i$  no mapa modelo. Conforme Senna Jr et al. (2001), o erro de omissão indica a proporção de pixel que pertencem à classe  $i$  do mapa real e foram classificados em outras classes no mapa modelo.

É possível comparar as classes do mapa real com o mapa modelo a partir dos índices apresentados. Para uma comparação global dos mapas, Jenness & Wynne (2005) apresentam a matriz de confusão total, de acordo com a Tabela 7:

Tabela 7 Matriz genérica de confusão total			
Dados de referência: reais ou controlados			
		verdadeiro	falso
Dados do modelo	verdadeiro	$a = \sum_{i=1}^k n_{ii}$	$b = \sum_{i=1}^k \sum_{j \neq i}^k n_{ij}$
	falso	$c = \sum_{j=1}^k \sum_{i \neq j}^k n_{ij}$	$d = \sum_{i=1}^k \sum_{i \neq j}^k \sum_{j \neq i}^k n_{ij}$

Outra métrica obtida da matriz de confusão total, é conhecida como Coeficiente de correlação de Matthews - CCM (MATTHEWS, 1975), descrito pela expressão (2.44):

$$CCM = \frac{a \cdot d - b \cdot c}{\sqrt{(a + b)(a + c)(d + b)(d + c)}}. \quad (2.44)$$

O coeficiente de correlação de Matthews (CCM), segundo Gorodkin (2004), é uma versão discreta do coeficiente de correlação de Pearson, sendo seus valores pertencentes ao intervalo  $[-1, 1]$ . O valor 1 equivale a uma predição perfeita, zero a uma predição aleatória, e  $-1$  a uma predição inversa. Embora não haja uma métrica ideal para descrever a matriz de confusão como um único número, considera-se o CCM como uma das melhores métricas com este fim (SABATTINI, 1995).

### 3 MATERIAL E MÉTODOS

#### 3.1 Caracterização da área

Os dados da produtividade da soja ( $t\ ha^{-1}$ ) utilizados neste trabalho foram coletados no ano agrícola 2004/2005, numa área comercial de 57,16 ha, localizada no município de Cascavel, região Oeste do Paraná. A área possui coordenadas geográficas aproximadas de  $24,95^\circ$  Sul e  $53,57^\circ$  Oeste, datum SAD-69, com elevação média de 650 m em relação ao nível do mar. O solo da região é classificado como Latossolo Vermelho distroférico de textura argilosa e apresenta solos profundos, com boa capacidade de retenção de água, aeração e permeabilidade (IAPAR, 2007). O clima da região é classificado como mesotérmico superúmido, tipo climático Cfa (Köppen), temperatura anual média de  $21^\circ$ .

A Figura 6 apresenta um mapa da área, com coordenadas em UTM.



Figura 6 Mapa de localização da área experimental.

O levantamento topográfico e o posicionamento dos locais de amostragem foram feitos por meio de receptores GPS, pelo método estático, com correção diferencial pós-processada.

Foram utilizados pontos georreferenciados, com espaçamento médio de 75 m entre as parcelas. Um esboço da área em estudo pode ser visualizado na Figura 7. Os pontos correspondem às parcelas amostradas.

O monitor de colheita fornece a leitura da produtividade e da umidade dos grãos, durante a colheita, utilizando sensores instalados na colhedora. O sensor de produtividade realiza medições instantâneas de todo o fluxo de grãos de acordo com um sensor de impacto.

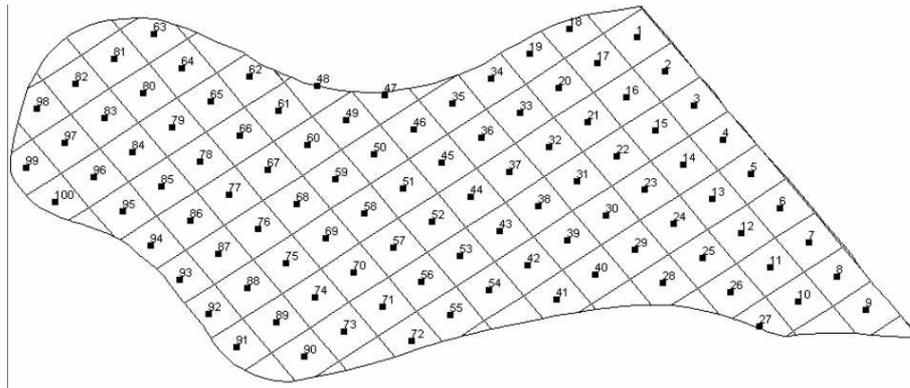


Figura 7 Área em estudo - esquema amostral. Fonte: Faraco (2006).

### 3.2 Análise exploratória de dados

Como análise inicial, os dados de produtividade da soja em  $t\ ha^{-1}$ , em todas as grades e monitor de colheita, foram submetidos à estatística descritiva, ou seja, foram calculadas as medidas de posição, em cada grade: média, mediana, máximo, mínimo, primeiro e terceiro quartis. A seguir, foram calculadas as medidas de dispersão: desvio-padrão e coeficiente de variação (CV, em porcentagem), também as medidas de forma: índices de assimetria e curtose. Os valores de assimetria e curtose encontrados para os dados, foram confrontados com os limites inferiores e superiores sugeridos por Jones (1969) a um nível de significância de 5%. A análise exploratória serviu para permitir a visualização geral do comportamento da variável produtividade da soja em  $t\ ha^{-1}$ , utilizando o programa estatístico R. Nessa análise, os dados foram considerados independentes entre si, ou seja, a sua localização espacial não é levada em conta.

### 3.3 Análise estatística espacial

Utilizou-se a variável produtividade da soja para ser submetida à análise geoestatística, para que fosse verificada a dependência espacial, e que fossem interpolados os dados e construídos os mapas temáticos de superfícies da variável.

A construção de envelopes simulados foi realizada, para identificar a continuidade espacial. O envelope simulado pode ser formulado partindo de simulações de Monte Carlo para cada grade amostral, conforme Guedes et al. (2008).

Para a realização da interpolação, utilizou-se o método da krigagem ordinária.

Embasando-se na Teoria das Variáveis Regionalizadas para dados não tendenciosos, foram construídos os semivariogramas experimentais para a análise de dependência da estrutura espacial e também função semivariância para as quatro grades amostrais, com distâncias de 25 m, 50 m, 75 m e 100 m entre os pontos. Na Figura 8 observa-se a localização e distribuição dos pontos nas quatro grades amostrais.

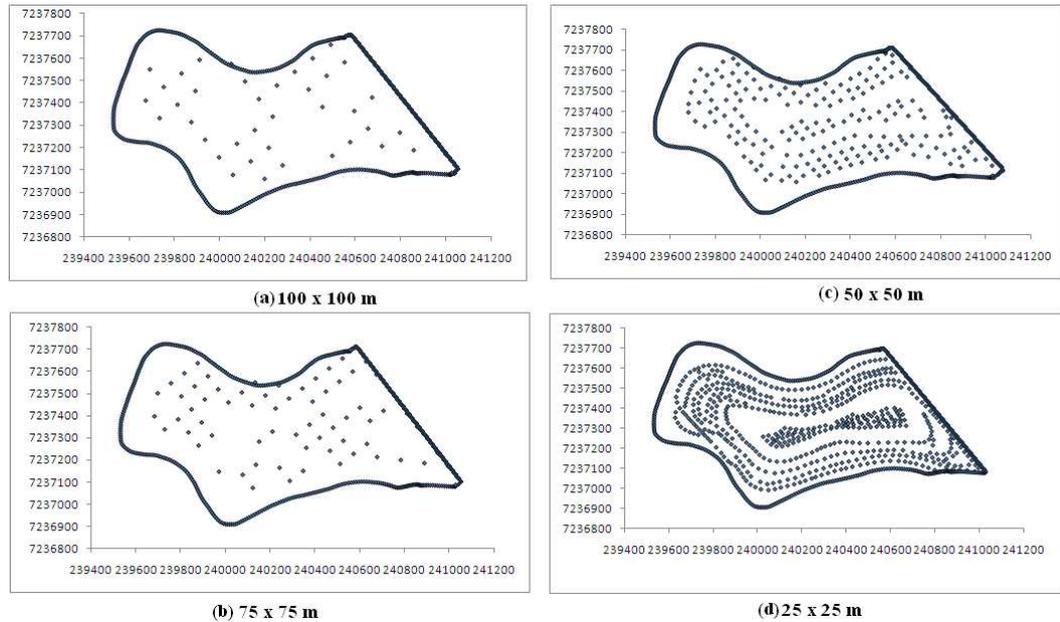


Figura 8 Localização dos pontos nas grades: (a) Grade 100x100 m, (b) Grade 75x75 m, (c) Grade 50x50 m, (d) Grade 25x25 m.

Os parâmetros dos modelos ajustados aos dados de produtividade da soja foram estimados pelos métodos de máxima verossimilhança (MV) e de máxima verossimilhança restrita (MVR). Os modelos utilizados para esses fins foram esférico, gaussiano, exponencial e Família Matérn.

### 3.4 Critérios de seleção de modelos

Após o ajustamento de modelos, procedeu-se a escolha do melhor modelo para cada grade. Os métodos de seleção de modelos de variabilidade espacial para a elaboração de mapas temáticos, estudados por Faraco et al. (2008), foram utilizados para esse fim, sendo eles: Validação Cruzada, Máximo do logaritmo da função verossimilhança e Critério de Akaike, considerados os métodos mais adequados para avaliação de ajuste. Com os modelos selecionados, ocorreu a confecção dos mapas temáticos, para as diferentes grades amostrais.

### 3.5 Comparação dos mapas temáticos

Os mapas construídos para as grades 100x100 m, 75x75 m e 50x50 m foram comparados, pixel a pixel, com o mapa construído para a grade 25x25 m, tido como referência. Não utilizou-se na íntegra os dados do monitor de colheita, uma vez que o esquema de amostragem para os dados da grade 25x25 m foram obtidos a partir dele.

Foi utilizado como instrumento computacional para a estimação, o ajuste de modelos e a krigagem o software R, (R Development Core Team, 2005), e o seu pacote geoR (RIBEIRO JR. & DIGGLE, 2001), que são softwares livres e possuem licença GPL (*General Public License*).

Para poder realizar a comparação entre os mapas temáticos, é necessário que os mesmos estejam construídos com as mesmas divisões para as classes. Como o mínimo krigado é  $2,59 \text{ t ha}^{-1}$ , isso na grade  $75 \times 75 \text{ m}$ , e o máximo valor krigado é  $3,91 \text{ t ha}^{-1}$ , na grade  $25 \times 25 \text{ m}$ , tem-se a amplitude de  $1,32 \text{ t ha}^{-1}$  que, dividida por 5 fornece uma amplitude de  $0,2164 \text{ t ha}^{-1}$ . Assim, os extremos dos intervalos de classe foram aproximados em duas casas decimais, definindo as seguintes classes:  $C1 = [2,59; 2,85]$ ,  $C2 = (2,85; 3,12]$ ,  $C3 = (3,12; 3,38]$ ,  $C4 = (3,38; 3,65]$  e  $C5 = (3,65; 3,91]$ . Observa-se que todos os mapas possuem 22867 pixels, ou seja, o mesmo número de pixels para as quatro grades.

Calcularam-se os índices globais de acurácia: índice Kappa  $K$ , índice Tau  $\tau$ , Exatidão global,  $EG$ , acurácia média pela perspectiva do usuário  $Amu$ , acurácia média pela perspectiva do produtor  $Amup$ , média das acurácias médias pelas perspectivas do usuário e do produtor  $Maup$ , índice de acurácia de Hellden  $IamH$ , índice de acurácia do mapeamento de Short  $IamS$  e acurácia combinada pelas perspectivas do produtor e do usuário  $Acpu$ .

Também foram calculados os índices de acurácia por classes no mapa: kappa condicional do usuário  $Kcu_i$ , kappa condicional do produtor  $Kcp_i$ , kappa condicional modificado do usuário  $Kcmu_i$  e kappa condicional modificado do produtor  $Kcmp_i$ .

Construíram-se as matrizes de confusão por classe e também as matrizes de confusão total. Com base nessa matriz, obtiveram-se o índice de sensibilidade  $S$ , o índice de especificidade  $E$ , a taxa de falsos positivos ou erro de comissão, TFP e a taxa de falsos negativos ou erro de omissão, TFN. Calculou-se, então, o coeficiente de correlação de Matthews.

## 4 RESULTADOS E DISCUSSÕES

### 4.1 Estatísticas descritivas

As estatísticas descritivas da produtividade da soja em  $t\ ha^{-1}$ , para as diferentes grades amostrais analisadas e dados do monitor de colheita, são apresentadas na Tabela 8.

Tabela 8 Resumo das estatísticas para as quatro grades analisadas e monitor de colheita

Estatística	25x25m	50x50m	75x75m	100x100m	Monitor de colheita
Nº de amostras	549	188	66	44	7582
Média	3,25	3,27	3,22	3,28	3,23
Mediana	3,26	3,28	3,19	3,32	3,27
Q1	3,00	3,00	2,95	3,11	2,98
Q3	3,54	3,55	3,48	3,49	3,55
Mínimo	1,56	1,63	2,09	2,26	0,68
Máximo	4,36	4,17	4,09	3,80	4,99
DP	0,40	0,45	0,38	0,32	0,47
CV(%)	12,36	12,27	11,71	9,72	14,50
Assimetria	-0,62	-0,54	-0,18	-0,89	4,62
Curtose	1,36	1,20	0,70	1,28	3,41

Q1: 1º quartil; Q3: 3º quartil; CV: Coeficiente de variação; DP: desvio padrão

É importante observar na Tabela 8, que algumas das estatísticas descritivas mostraram-se semelhantes para todas as grades amostrais e monitor de colheita. Os resultados nesta tabela, indicam que a variável regionalizada produtividade da soja, nas diferentes grades amostrais e obtidas pelo monitor de colheita, apresentaram baixa dispersão (DP), homogeneidade na distribuição dos dados nas grades 25x25, 50x50, 75x75, 100x100 m, e monitor de colheita, pelo  $CV \leq 20\%$  (GOMES, 2000). Observou-se também que os valores das médias são próximos entre si em todas as grades amostrais e nos dados do monitor de colheita, sendo que variou de  $3,22\ t\ ha^{-1}$  na grade 75x75 m até  $3,28\ t\ ha^{-1}$  na grade 100x100 m. Os valores para a mediana em todas as grades e monitor de colheita também apresentaram valores muito próximos, variando de  $3,19\ t\ ha^{-1}$  para a grade 75x75 m até  $3,32\ t\ ha^{-1}$  para a grade 100x100 m. O primeiro quartil variou de  $2,95\ t\ ha^{-1}$  na grade 75x75 m até  $3,11\ t\ ha^{-1}$  na grade 100x100 m. O terceiro quartil variou de  $3,48\ t\ ha^{-1}$  na grade 75x75 m até  $3,55\ t\ ha^{-1}$  na grade 25x25 m e monitor de colheita.

Os coeficientes de assimetria e curtose para os dados das grades 50x50 m e 75x75 m, que podem ser observados na Tabela 8, foram comparados com os da tabela de Jones

(1969), a qual apresenta intervalo de 95% de confiança para os coeficientes de curtose e assimetria, caracterizando a distribuição normal de probabilidades. Desta forma, não foram utilizados testes de normalidade, pois a variável produtividade da soja mostrou ter estrutura de dependência espacial.

É possível notar na Figura 9, a simetria na distribuição dos dados nos gráficos de box-plot, pela disposição de seus elementos, embora todas as grades amostrais e também o monitor de colheita tenham apresentado valores atípicos.

A Figura 9 apresenta os gráficos box-plot para a variável regionalizada produtividade da soja para as quatro grades amostrais e também para o monitor de colheita. Para a grade 100x100 m, apresenta um outlier, que é o ponto de mínimo, sendo o ponto de número 38, com coordenadas em UTM (239815 m, 7237392 m), cujo valor para produtividade é  $2,26 \text{ t ha}^{-1}$ . Para a grade 75x75 m, o ponto de mínimo, número 20, com coordenadas (240344 m, 7237524 m), é um outlier, cujo valor é uma produtividade da soja de  $2,09 \text{ t ha}^{-1}$ . Para a grade 50x50 m há dois outliers, sendo o valor menor o ponto número 33 apresentado a produtividade da soja de  $1,63 \text{ t ha}^{-1}$ , nas coordenadas (240895 m, 7237135 m); e o outro o ponto é o número 5, cuja produtividade da soja vale  $2,02 \text{ t ha}^{-1}$  e possui coordenadas geográficas (240737 m, 7237533 m). O gráfico box-plot para a grade 25x25 m apresentou 10 pontos classificados como outliers, sendo um ponto de máximo com produtividade de  $4,36 \text{ t ha}^{-1}$  e nove pontos menores que o limite inferior, sendo eles com medidas de produtividade de  $1,56; 1,63; 1,67; 1,74; 1,85; 1,99; 2,10; 2,12$  e  $2,17 \text{ t ha}^{-1}$ . Para o gráfico box-plot referente aos dados do monitor de colheita, obteve-se um número elevado de pontos classificados como outliers, principalmente valores indicando baixa produtividade.

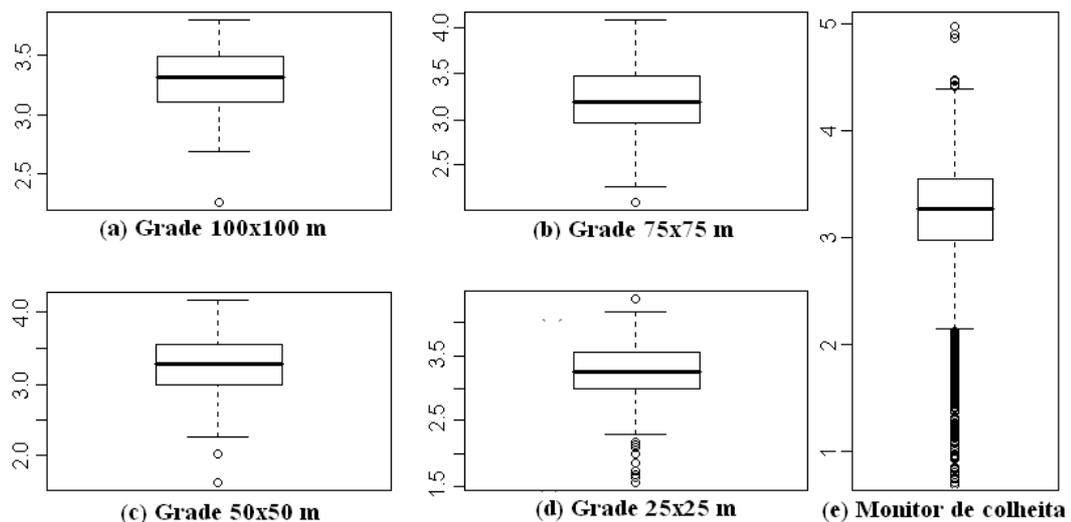


Figura 9 Gráficos box-plot: (a) 100x100 m; (b) 75x75 m; (c) 50x50 m; (d) 25x25 m; (e) monitor de colheita

## 4.2 Análise de tendência e estacionariedade

Na Figura 10 são apresentados os gráficos post-plot da área em estudo, classificada segundo os quartis, para a variável regionalizada produtividade da soja nas grades amostrais de 25x25 m, 50x50 m, 75x75 m, 100x100 m, e monitor de colheita. Pode-se verificar a não existência de tendência direcional em nenhum dos cinco gráficos post-plot, ou seja, os conjuntos de dados da produtividade da soja analisados atendem à hipótese intrínseca.

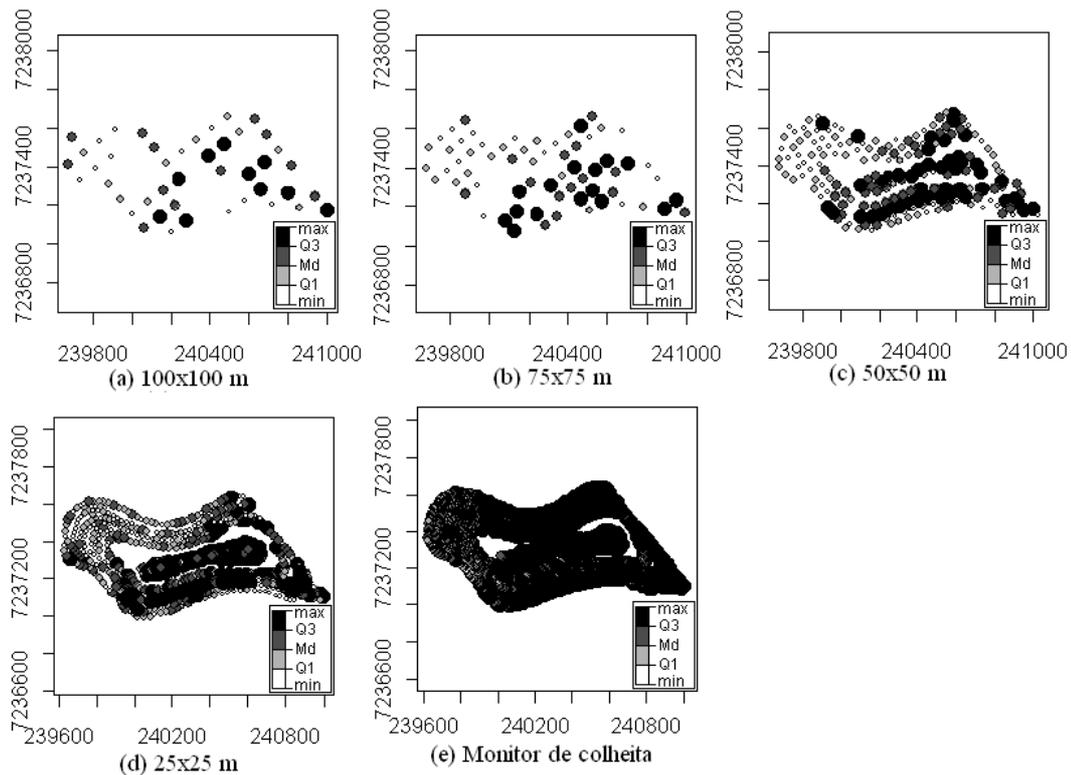


Figura 10 Gráficos post-plot: (a) 100x100 m; (b) 75x75 m; (c) 50x50 m; (d) 25x25 m; (e) Monitor de colheita para a variável produtividade da soja em  $t\ ha^{-1}$ . Md: mediana; Q1: 1º quartil, Q3: 3º quartil; min: valor mínimo; max: valor máximo

Embora tenha havido a existência de correlações diferentes de zero entre algumas variáveis e suas respectivas coordenadas, essas foram classificadas como relativamente fracas, o que não justificou nenhuma transformação dos dados, nem a eliminação de tendências.

## 4.3 Identificação da continuidade espacial

Para identificar a continuidade espacial em cada conjunto de dados, elaborou-se os gráficos de envelopes simulados, em que cada envelope é construído pelos valores mínimos a máximos obtidos em todos os semivariogramas dos conjuntos de dados modificados, definindo, assim, uma região de independência espacial na variável em estudo (DIGGLE & RIBEIRO JR, 2007).

Na Figura 11 são apresentados os semivariogramas experimentais de cada grade amostral e monitor de colheita, com seus respectivos envelopes simulados. Utilizou-se o software R para gerar os envelopes, sendo realizadas 99 simulações de Monte Carlo para cada grade amostral. Observa-se que, nos itens (b) grade 75x75 m, (c) 50x50 m e (d) 25x25 m, ao menos um ponto do semivariograma apresenta uma semivariância fora dos limites de variações do envelope, indicando que os dados possuem alguma continuidade espacial e que a dependência espacial realmente existe nas variáveis em estudo. O item (a) grade 100x100 m não apresentou pontos fora dos limites do envelope simulado, contudo, levando-se em conta a escala da semivariância, nota-se que há pontos muito próximos do limite inferior (3<sup>o</sup> e 9<sup>o</sup> pontos) e do limite superior (6<sup>o</sup> ponto), dando indícios de que possa haver continuidade espacial, mesmo que fraca.

Não foi possível gerar um gráfico com envelopes simulados para os dados do monitor de colheita devido à grande quantidade de pontos.

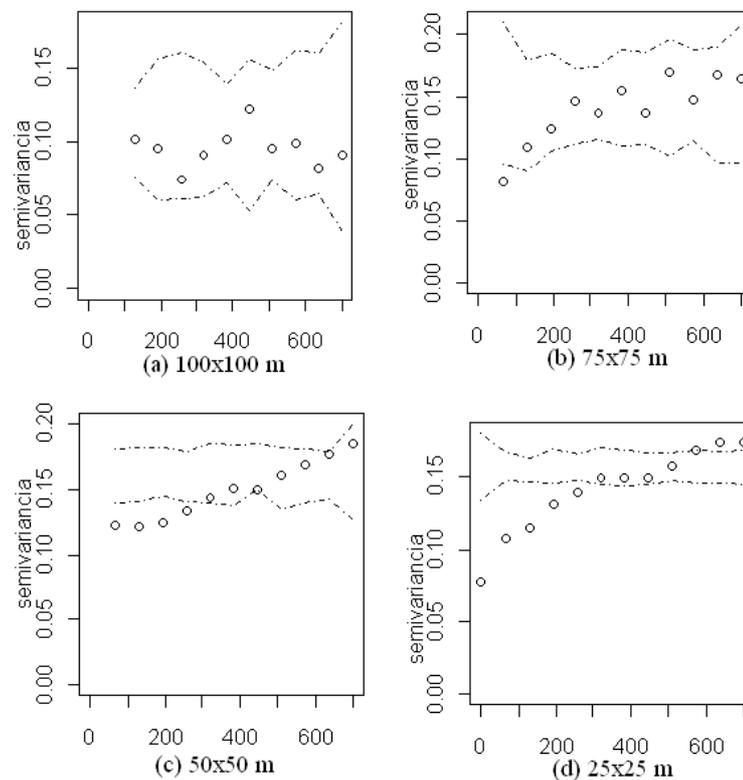


Figura 11 Semivariogramas experimentais para a variável regionalizada produtividade da soja nas grades (a) 100x100 m, (b) 75x75 m, (c) 50x50 m, (d) 25x25 m no ano agrícola 2004/2005 com envelopes simulados utilizando 99 simulações de Monte Carlo.

#### 4.4 Identificação da anisotropia

Investigou-se em cada caso, o estudo da anisotropia para a variável produtividade. Para tanto, foram construídos num mesmo sistema os semivariogramas direcionais, conforme Figura 12, utilizando o estimador de Matheron (Equação 2.6), nas direções 0<sup>o</sup>, 45<sup>o</sup>, 90<sup>o</sup> e 135<sup>o</sup>

(GUEDES et al., 2008), com cutoff igual a 50% da distância máxima (CLARK, 1979), que foi aproximadamente a mesma em todos os casos, ou seja, 700 m.

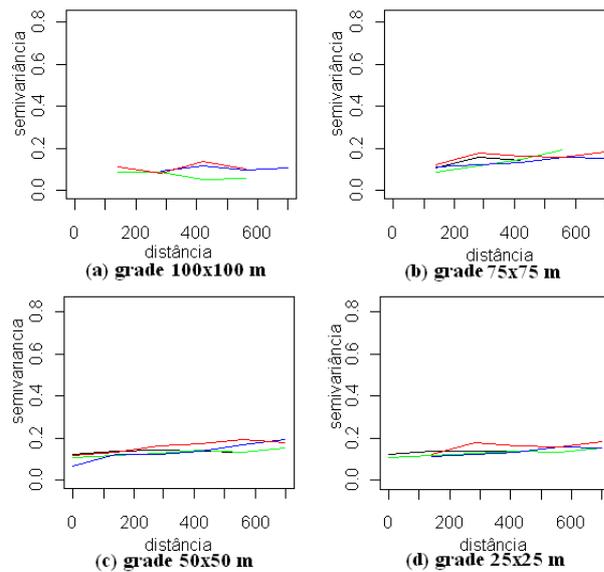


Figura 12 Semivariogramas direcionais para a variável produtividade da soja em  $t\ ha^{-1}$  para as grades: (a) 100x100 m, (b) 75x75 m, (c) 50x50 m, (d) 25x25 m, nas direções  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  e  $135^\circ$

Na Figura 12, verifica-se a similaridade entre os semivariogramas nas direções analisadas, em cada grade amostral. Logo, pode-se considerar que a distribuição espacial da variável regionalizada produtividade da soja é isotrópica, isto é, pode-se afirmar que não existe tendência direcional. Nesse caso, uma única direção é o suficiente para descrever a variabilidade espacial em cada grade amostral.

Partindo do pressuposto de que o fenômeno observado pode ser considerado isotrópico, construíram-se os semivariogramas omnidirecionais da variável regionalizada produtividade da soja para as quatro grades amostradas e também para os dados obtidos do monitor de colheita (Figura 13), apresentando a existência de dependência espacial, a qual será modelada.

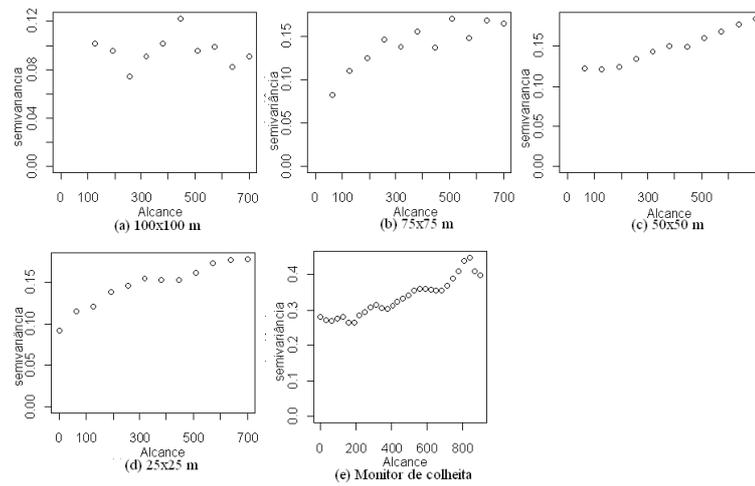


Figura 13 Semivariogramas omnidirecionais para as grades: (a) 100x100 m, (b) 75x75 m, (c) 50x50 m, (d) 25x25 m, (e) Monitor de colheita

#### 4.5 Modelagem espacial

A seguir são apresentados e discutidos os resultados da modelagem espacial para as grades amostrais 100x100 m, 75x75 m, 50x50 m e 25x25 m.

##### 4.5.1 Modelagem espacial para os dados da grade 100x100 m

A distância máxima entre os pontos na área em estudo é de 1362,08 m. O ponto de corte de no mínimo 50% como ponto de corte (CLARK, 1979), tem-se a distância máxima considerada como 700 m.

Os parâmetros do modelo espacial encontrados para a grade 100x100 m são apresentados na Tabela 9, para os diferentes modelos estudados e com os dois métodos de estimação de parâmetros utilizados: máxima verossimilhança, MV e máxima verossimilhança restrita, MVR.

Tabela 9 Determinação de parâmetros para os métodos de estimação e modelos - grade 100x100 m

Modelo	Métodos	$\varphi_1$	$\varphi_2$	$\varphi_3$	$a$	$EPR = \frac{C_0}{C_0 + C_1}$
Esférico	MV	0,0819	0,0173	312,0439	312,0439	0,8256
	MVR	0,0832	0,0197	389,2884	389,2884	0,8085
Gaussiano	MV	0,0861	0,0131	165,0576	285,8881	0,8679
	MVR	0,0871	0,0158	201,8991	349,4507	0,8465
Exponencial	MV	0,0678	0,0314	77,4357	232,3071	0,6835
	MVR	0,0799	0,0237	153,6634	460,3343	0,7712
Matérn $\kappa = 0,7$	MV	0,0738	0,0254	72,4749	249,8122	0,7439
	MVR	0,0819	0,0215	127,4088	439,1624	0,7921
Família Matérn $\kappa = 1,0$	MV	0,0782	0,0210	66,3746	265,4004	0,7883
	MVR	0,0918	0,0143	249,9914	999,5962	0,8652
Matérn $\kappa = 1,5$	MV	0,0814	0,0178	58,6757	278,3494	0,8206
	MVR	0,0850	0,0182	86,4213	409,9707	0,8236
Matérn $\kappa = 2,0$	MV	0,0829	0,0163	52,9814	284,4241	0,8357
	MVR	0,0857	0,0174	74,5869	400,4105	0,8009

$\varphi_1$ : efeito pepita;  $\varphi_2$ : contribuição;  $\varphi_3$ : função do alcance do modelo;  $a$ : alcance; EPR: efeito pepita relativo.

Pelas informações da Tabela 9, observa-se que o alcance para os ajuste encontrados variou de 232,3071 m para o modelo exponencial ajustado por máxima verossimilhança até 999,5962 m para o modelo Família Matérn, com  $\kappa = 1,0$ , ajustado por máxima verossimilhança restrita. Pela análise geoestatística, a variável produtividade da soja apresentou dependência espacial forte, considerando o efeito pepita relativo, EPR, exceto para os ajustes exponencial por máxima verossimilhança e Matérn,  $\kappa = 0,7$  por máxima verossimilhança, que demonstraram dependência espacial moderada (CAMBARDELLA et al., 1994).

Na Figura 14 são apresentados os modelos ajustados aos semivariogramas experimentais, segundo os parâmetros apresentados na Tabela 9.

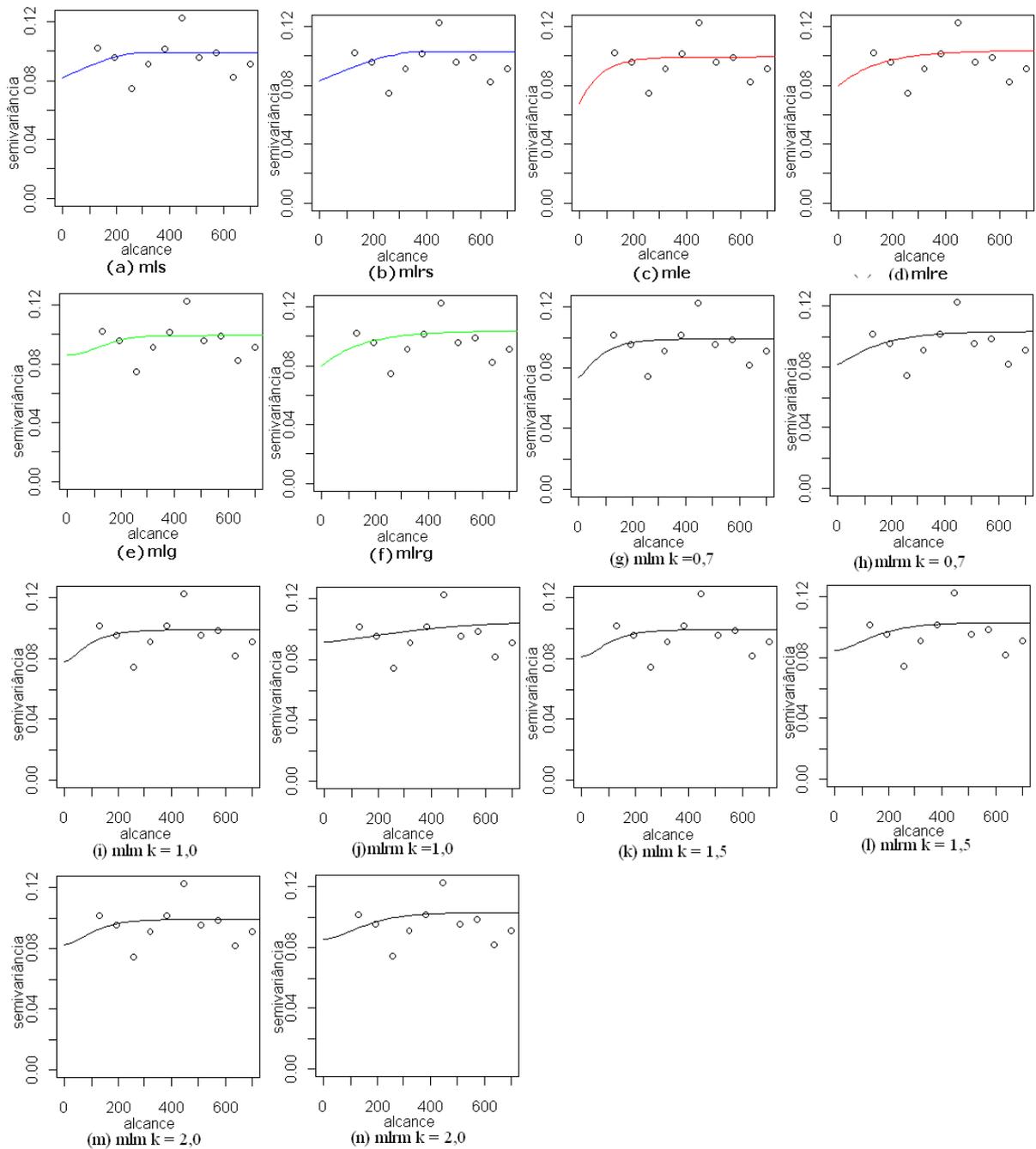


Figura 14 Semivariogramas ajustados à grade 100x100 m para: (a) máxima verossimilhança com modelo esférico; (b) máxima verossimilhança restrita com modelo esférico; (c) máxima verossimilhança com modelo exponencial; (d) máxima verossimilhança restrita com modelo exponencial; (e) máxima verossimilhança com modelo gaussiano; (f) máxima verossimilhança restrita com modelo gaussiano; (g) máxima verossimilhança com Família Matérn  $\kappa = 0,7$ ; (h) máxima verossimilhança restrita com Família Matérn  $\kappa = 0,7$ ; (i) máxima verossimilhança com Família Matérn  $\kappa = 1,0$ ; (j) máxima verossimilhança restrita com Família Matérn  $\kappa = 1,0$ ; (k) máxima verossimilhança com Família Matérn  $\kappa = 1,5$ ; (l) máxima verossimilhança restrita com Família Matérn  $\kappa = 1,5$ ; (m) máxima verossimilhança com Família Matérn  $\kappa = 2,0$ ; (n) máxima verossimilhança restrita com Família Matérn  $\kappa = 2,0$ .

Na Tabela 10 são apresentados os resultados da técnica de validação cruzada, critério de informação de Akaike,  $IAC$  e Máximo do valor do logaritmo da função verossimilhança ( $MVLV$ ) para a escolha do melhor modelo ajustado para a grade amostral 100x100 m.

Tabela 10 Seleção de melhor modelo para os dados de produtividade da soja na grade 100x100 m

Modelo	Método	EM	ER	$DP_{EM}$	$DP_{ER}$	EA	$IAC$	$MVLV$
Esférico	MV	0,00005	0,00008	0,31855	1,01776	<b>10,69</b>	23,94	-11,31
	MVR	-0,00004	-0,00006	0,31838	1,00965	10,71	28,57	-11,32
Gaussiano	MV	0,00004	0,00006	0,31871	1,01777	10,70	22,61	-11,32
	MVR	-0,00003	-0,00005	<b>0,31836</b>	<b>1,00960</b>	<b>10,69</b>	26,53	-11,32
Exponencial	MV	0,00007	0,00012	0,31949	1,01808	10,71	22,14	-11,36
	MVR	<b>-0,00001</b>	<b>-0,00002</b>	0,31918	1,00993	10,71	23,30	-11,35
Matérn $\kappa = 0,7$	MV	0,00007	0,00011	0,31943	1,01805	10,71	23,77	-11,36
	MVR	<b>-0,00001</b>	-0,00001	0,31907	1,00985	10,71	24,15	-11,35
Matérn $\kappa = 1,0$	MV	0,00007	0,00011	0,31935	1,01801	10,71	24,39	-11,36
	MVR	-0,00010	-0,00017	0,32038	1,01090	10,83	<b>17,73</b>	-11,38
Matérn $\kappa = 1,5$	MV	0,00006	0,00010	0,31925	1,01797	10,71	24,15	-11,35
	MVR	<b>-0,00001</b>	-0,00001	0,31885	1,00974	<b>10,69</b>	24,71	-11,34
Matérn $\kappa = 2,0$	MV	0,00006	0,00009	0,31917	1,01793	10,71	23,72	-11,35
	MVR	<b>-0,00001</b>	<b>-0,00002</b>	0,31877	1,00971	<b>10,69</b>	24,75	-11,34

EM: erro médio; ER: erro reduzido;  $DP_{EM}$ : desvio padrão do erro médio;  $DP_{ER}$ : desvio padrão do erro reduzido; EA: erro absoluto;  $IAC$ : informação de Akaike;  $MVLV$ : máximo do logaritmo da função verossimilhança.

O critério de validação cruzada, aplicado aos modelos em estudo, sugeriu que o menor valor da média dos erros de estimação indica o modelo com melhor ajuste. Para os dados de produtividade da soja na grade com distâncias de 100 m, em estudo, apresentado na Tabela 10, a escolha indicou o modelo Matérn, com  $\kappa = 2,0$ , ajustado pelo método da máxima verossimilhança restrita, visto que houve mesmo número de estatísticas escolhidas como melhor na validação cruzada que o modelo gaussiano ajustado por máxima verossimilhança restrita, porém, apresentou menor valor no Critério de Akaike,  $IAC = 24,75$  para o Matérn, contra  $IAC = 26,53$  para o gaussiano. A estatística  $MVLV$  mostrou-se similar para todos os ajustes realizados, não sendo um bom critério de seleção nesse caso.

Para o modelo escolhido, Matérn, com  $\kappa = 2,0$ , encontrou-se o coeficiente de efeito pepita relativo,  $EPR=0,8009$ , que indica que há uma forte dependência espacial. O modelo escolhido, conforme Tabela 9 e 10 apresentou como parâmetros:  $\varphi_1 = 0,0857$ ,  $\varphi_2 = 0,0174$ ,  $\varphi_3 = 74,5869$ , o que significa um alcance  $a = 400,4105$ . Colocando os parâmetros no modelo, tem-se:

$$\gamma(h) = 0,0857 + 0,0174 \left[ 1 - \left( 2\Gamma(2) \right)^{-1} \left( \frac{h}{74,5869} \right)^2 K_2 \left( \frac{h}{74,5869} \right) \right], \quad h > 0. \quad (4.1)$$

A interpolação dos dados de produtividade da soja  $t\ ha^{-1}$  na grade  $100 \times 100\ m$  foi realizada por krigagem ordinária. Fez-se a divisão da amplitude total em cinco classes iguais, considerando os valores krigados em todas as grades amostrais. Assim, a divisão em classes trouxe: C1 = [2,59; 2,85], C2 = (2,85; 3,12], C3 = (3,12; 3,38], C4 = (3,38; 3,65] e C5 = (3,65; 3,91]. O mapa temático da grade  $100 \times 100\ m$  é apresentado na Figura 15.

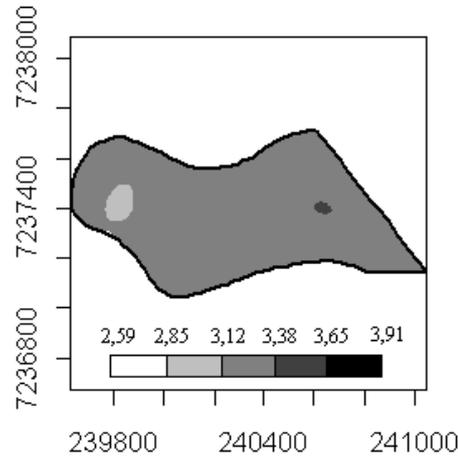


Figura 15 Mapa construído para a grade amostral  $100 \times 100\ m$

Nota-se visualmente, que o mapa da produtividade da soja para a grade  $100 \times 100\ m$  apresenta a maior parte de seus pixels na classe C3, ou seja, a maior parte dos valores krigados para a produtividade da soja variou de 3,12 a 3,38  $t\ ha^{-1}$ . Não percebe-se a presença de pixels pertencentes às classes C1 ou C5. A quantidade de pixels por classe está especificada na Tabela 11.

Tabela 11 Classificação dos pixels do mapa para a grade amostral  $100 \times 100\ m$

Classes	C1	C2	C3	C4	C5	Total
Nº de pixels	0	521	22247	99	0	22867
% do total de pixels	0	2,28	97,29	0,43	0	100
Área (ha)	0	1,30	55,61	0,25	0	57,16

Observa-se na Tabela 11, que para os valores krigados, considerando o modelo Matérn com  $\kappa = 2,0$ , não possui pixels com produtividade dentro da classe C1 = [2,59; 2,85], nem da classe C5 = (3,65; 3,91]. 55,61 dos 57,16 ha, ou seja, 97,29% dos pixels foram classificados com produtividade entre 3,12 a 3,38  $t\ ha^{-1}$ , evidenciando a homogeneidade dos valores preditos. Essa homogeneidade é comprovada ao observar a Tabela 12, na qual tem-se o coeficiente de variação,  $CV < 20\ %$  para os dados de produtividade da soja para a grade  $100 \times 100\ m$ .

Tabela 12 Estatísticas descritivas para os valores estimados pelo modelo Matérn, com  $\kappa = 2, 0$ , para a grade 100x100 m, em  $t\ ha^{-1}$

Grade	Mínimo	Q1	Mediana	Média	Q3	Máximo	DP	CV (%)
100x100 m	3,10	3,24	3,30	3,28	3,34	3,39	0,071	2,18

#### 4.5.2 Modelagem espacial para os dados da grade 75x75 m

Para os dados de produtividade da soja da grade 75x75 m, encontrou-se como distância máxima entre pontos a medida de 1337,60 m, e com o ponto de corte de pelo menos 50%, a medida de 668,80, a qual aproximou-se para 700 m. Procedeu-se com o ajuste de modelos, sendo que foram utilizados os modelos esférico, gaussiano, exponencial e a Família Matérn, para valores de  $\kappa$  em 0,7; 1,0; 1,5 e 2,0.

Os parâmetros para os modelos ajustados à grade 75x75 m são apresentados na Tabela 13.

Tabela 13 Determinação de parâmetros para os métodos de estimação e modelos - grade 75x75 m

Modelo	Métodos	$\varphi_1$	$\varphi_2$	$\varphi_3$	$a$	$EPR = \frac{C_0}{C_0 + C_1}$
Esférico	MV	0,0700	0,0669	275,8022	275,8022	0,5113
	MVR	0,0705	0,0712	291,6047	291,6047	0,4975
Gaussiano	MV	0,0846	0,0530	147,3874	255,1008	0,6148
	MVR	0,0860	0,0571	158,3992	274,1602	0,6010
Exponencial	MV	0,0356	0,1021	83,9565	251,5111	0,2585
	MVR	0,0474	0,0971	106,3093	318,4742	0,3280
Matérn $\kappa = 0,7$	MV	0,0513	0,0865	76,2370	262,7797	0,3723
	MVR	0,0597	0,0847	93,3779	321,8622	0,4134
Matérn $\kappa = 1,0$	MV	0,0627	0,0750	67,5403	270,0615	0,4553
	MVR	0,0687	0,0755	80,1473	320,4706	0,4764
Matérn $\kappa = 1,5$	MV	0,0711	0,0667	57,6047	273,2687	0,5160
	MVR	0,0753	0,0687	66,3390	314,7034	0,5229
Matérn $\kappa = 2,0$	MV	0,0751	0,0627	50,8724	273,1020	0,5450
	MVR	0,0784	0,0654	57,6130	309,2884	0,5452

$\varphi_1$ : efeito pepita;  $\varphi_2$ : contribuição;  $\varphi_3$ : função do alcance do modelo;  $a$ : alcance; EPR: efeito pepita relativo.

Na Tabela 13 observa-se que o alcance para os modelos ajustados na grade 75x75 m variou de 262,7797 m para o modelo Matérn com  $\kappa = 0,7$  ajustado por máxima verossimilhança até 321,8622 m, também para o modelo Matérn com  $\kappa = 0,7$  ajustado, porém, por máxima verossimilhança restrita. Os valores para o alcance ficaram próximos, se comparados aos valores de alcance para os ajustes da grade 100x100 m (Tabela 9). EPR, classifica-se como moderada dependência espacial (CAMBARDELLA et al., 1994).

Pode-se observar na Figura 16 os semivariogramas ajustados para a grade 75x75 m.

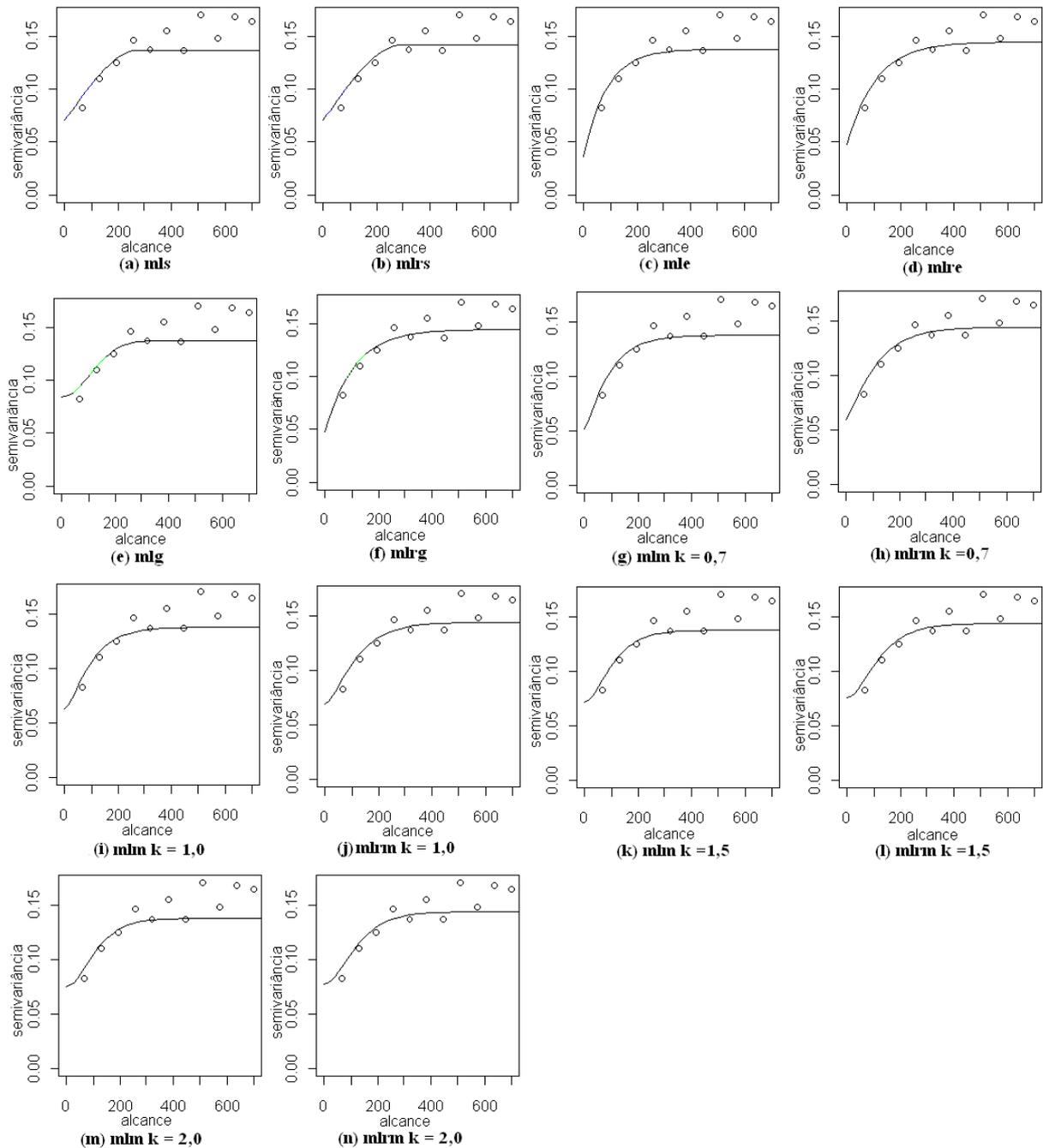


Figura 16 Semivariogramas ajustados à grade 75x75 m para (a)máxima verossimilhança com modelo esférico; (b) máxima verossimilhança restrita com modelo esférico; (c) máxima verossimilhança com modelo exponencial; (d) máxima verossimilhança restrita com modelo exponencial; (e) máxima verossimilhança com modelo gaussiano; (f) máxima verossimilhança restrita com modelo gaussiano; (g) máxima verossimilhança com modelo Família Matérn  $k=0,7$ ; (h) máxima verossimilhança restrita com modelo Família Matérn  $k=0,7$ ; (i) máxima verossimilhança com Família Matérn  $k=1,0$ ; (j) máxima verossimilhança restrita com Família Matérn  $k=1,0$ ; (k) máxima verossimilhança com Família Matérn  $k=1,5$ ; (l) máxima verossimilhança restrita com Família Matérn  $k=1,5$ ; (m) máxima verossimilhança com Família Matérn  $k=2,0$ ; (n) máxima verossimilhança restrita com Família Matérn  $k=2,0$ .

Na Figura 16, observa-se como se comporta cada semivariograma ajustado ao semivariograma experimental.

Na Tabela 14, tem-se os valores calculados para o erro médio,  $EM$ ; erro reduzido,  $ER$ ; desvio padrão do erro médio,  $DP_{EM}$ ; desvio padrão dos erros reduzidos,  $DP_{ER}$ , erro absoluto,  $EA$ , critério de informação de Akaike,  $IAC$  e máximo do logaritmo da função verossimilhança,  $MVLV$ .

Tabela 14 Validação cruzada, critério de informação de Akaike,  $IAC$ , e máximo do logaritmo da função verossimilhança para os dados de produtividade da soja para a grade 75x75 m

Modelo	Método	EM	ER	$DP_{EM}$	$DP_{ER}$	EA	$IAC$	$MVLV$
Esférico	MV	-0,00248	-0,00351	<b>0,34332</b>	1,01696	16,082	-4423,755	-23,65
	MVR	-0,00272	-0,00382	0,34412	<b>1,01465</b>	16,148	-4385,781	-23,23
Gaussiano	MV	-0,00257	-0,00364	0,34656	1,01725	<b>16,079</b>	-4447,821	-23,69
	MVR	-0,00284	-0,00401	0,34426	1,01503	16,139	-4403,407	-23,22
Exponencial	MV	<b>-0,00236</b>	<b>-0,00329</b>	0,34479	1,01736	16,169	-4515,250	-23,84
	MVR	-0,00271	-0,00378	0,34547	1,01538	16,157	-4447,821	-23,31
Matérn $\kappa = 0,7$	MV	-0,00239	-0,00334	0,34469	1,01736	16,149	<b>-4660,267</b>	-23,82
	MVR	-0,00273	-0,00382	0,34535	1,01537	16,153	-4447,821	-23,30
Matérn $\kappa = 1,0$	MV	-0,00242	-0,00339	0,34456	1,01738	16,133	-4423,755	-23,80
	MVR	-0,00276	-0,00386	0,34521	1,01536	16,147	-4515,250	-23,29
Matérn $\kappa = 1,5$	MV	-0,00246	-0,00346	0,34440	1,01739	16,123	-4515,250	-23,78
	MVR	-0,00279	-0,00391	0,34503	1,01534	16,139	-4423,755	-23,27
Matérn $\kappa = 2,0$	MV	-0,00250	-0,00350	0,34428	1,01739	16,117	-4423,755	-23,77
	MVR	-0,00281	-0,00395	0,34491	1,01532	16,132	-4515,250	-23,26

EM: erro médio; ER: erro reduzido;  $DP_{EM}$ : desvio padrão do erro médio;  $DP_{ER}$  desvio padrão do erro reduzido; EA: erro absoluto; IAC: informação de Akaike;  $MVLV$ : máximo do logaritmo da função verossimilhança.

De acordo com a Tabela 14, o modelo Matérn, com  $\kappa = 0,7$  ajustado por máxima verossimilhança foi o modelo que teve, em geral, os melhores valores para os dados de validação cruzada, e também o menor valor para o critério de informação de Akaike,  $IAC = -4660,267$ . O valor do  $MVLV$  demonstrou pequena variação em todos os ajustes.

A partir do modelo Matérn, com  $\kappa = 0,7$ , ajustado por máxima verossimilhança, aceito como o melhor modelo, cujos parâmetros são  $\varphi_1 = 0,0513$ ,  $\varphi_2 = 0,0865$ ,  $\varphi_3 = 76,2370$ ,  $a = 262,7797$ , obtém-se a equação 4.2.

$$\gamma(h) = 0,0513 + 0,0865 \left[ 1 - \left( 0,8123\Gamma(0,7) \right)^{-1} \left( \frac{h}{76,2370} \right)^{0,7} K_{0,7} \left( \frac{h}{76,2370} \right) \right], \quad h > 0. \quad (4.2)$$

Observa-se na Figura 17, o mapa gerado para a grade 75x75 m, bem como a quantidade de pixels em cada uma das classes, sendo elas: C1 = [2,59; 2,85], C2 = (2,85; 3,12], C3 = (3,12; 3,38], C4 = (3,38; 3,65] e C5 = (3,65; 3,91]

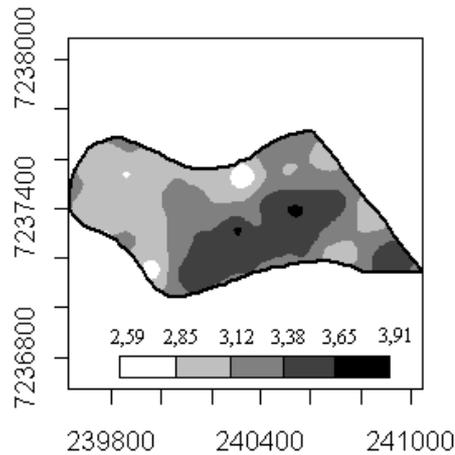


Figura 17 Mapa construído para a grade 75x75 m

O mapa temático para a produtividade da soja, referente à grade 75x75 m, apresentou regiões pertencentes a todas as classes, diferentemente do mapa para a grade 100x100 m. Observa-se que a região mais à esquerda apresentou cores mais claras, representando produtividades mais baixas. A Tabela 15 apresenta a quantidade de pixels pertencentes a cada classe no mapa da produtividade da soja para a grade 75x75 m.

Tabela 15 Classificação dos pixels do mapa para a grade amostral 75x75 m

Classes	C1	C2	C3	C4	C5	Total
Nº de pixels	467	8348	7692	6224	136	22867
% do total de pixels	2,04	36,51	33,64	27,22	0,59	100
Área (ha)	1,17	20,87	19,22	15,56	0,34	57,16

O mapa temático da produtividade da soja em  $t\ ha^{-1}$  para a grade amostral 75x75 m apresentou uma pequena quantidade de pixels classificados nas classes extremas, sendo apenas 2,04% dos pixels pertencentes à classe C1 e 0,59% pertencendo à classe C5. Da mesma maneira que no mapa para a grade 100x100 m, as classes C2 e C3 foram aquelas às quais mais pixels foram classificados como pertencentes.

Tabela 16 Estatísticas descritivas para os valores preditos pelo modelo Matérn, com  $\kappa = 0,7$ , para a grade 75x75 m, em  $t\ ha^{-1}$

Grade	Mínimo	Q1	Mediana	Média	Q3	Máximo	DP	CV (%)
75x75 m	2,59	3,06	3,18	3,22	3,41	3,73	0,202	6,28

As estatísticas descritivas para os valores preditos pelo modelo Matérn,  $\kappa = 0,7$ , por MV, que foi ajustado aos dados da grade 75x75 m, mostram que há homogeneidade nos dados (CV < 20%).

### 4.5.3 Modelagem espacial para os dados da grade 50x50 m

A partir dos dados de produtividade da soja em  $t\ ha^{-1}$ , calculou-se a distância máxima entre os pontos como 1383,10 m e aproximou-se para 700 m o ponto de corte de 50%. Realizou-se o ajuste desses dados por meio da máxima verossimilhança e máxima verossimilhança restrita aos modelos exponencial, esférico e Matérn, este último com  $\kappa$  variando em 0,7, 1,0, 1,5 e 2,0. Os coeficientes calculados para tais ajustes estão dispostos na Tabela 17.

Tabela 17 Determinação de parâmetros para os métodos de estimação e modelos - grade 50x50 m

Modelo	Métodos	$\varphi_1$	$\varphi_2$	$\varphi_3$	$a$	$EPR = \frac{C_0}{C_0 + C_1}$
Esférico	MV	0,1012	0,0631	392,0017	392,0017	0,6159
	MVR	0,0996	0,0706	397,6015	397,6015	0,5852
Exponencial	MV	0,0894	0,0676	117,0346	350,6043	0,5694
	MVR	0,0925	0,0720	147,6119	442,2057	0,5623
Família Matérn $\kappa = 0,7$	MV	0,0971	0,0602	103,5097	356,7852	0,6173
	MVR	0,0994	0,0653	126,4875	435,9870	0,6035
Matérn $\kappa = 1,0$	MV	0,1027	0,0549	89,7617	358,9141	0,6516
	MVR	0,1044	0,0606	106,6412	426,4073	0,6327
Matérn $\kappa = 1,5$	MV	0,1070	0,0512	75,6545	358,8945	0,6764
	MVR	0,1082	0,0576	87,6317	415,7128	0,6526
Matérn $\kappa = 2,0$	MV	0,1090	0,0496	66,7865	358,5349	0,6873
	MVR	0,1100	0,0564	76,2462	409,3180	0,6571

$\varphi_1$ : efeito pepita;  $\varphi_2$ : contribuição;  $\varphi_3$ : função do alcance do modelo;  $a$ : alcance; EPR: efeito pepita relativo.

Pelos resultado da Tabela 17, observa-se que o alcance dos modelos ajustados está variando de 350,6043 para o modelo exponencial ajustado por máxima verossimilhança, até 442,2057 m, para o modelo exponencial ajustado por máxima verossimilhança restrita, caracterizando uma variação de 91,6 m nos valores estimados como o alcance pelos modelos utilizados.

Os valores para o efeito pepita relativo variaram de 0,5623 para o modelo exponencial ajustado por máxima verossimilhança restrita até 0,6873, para o modelo Matérn, com  $\kappa = 2,0$ , ajustado por máxima verossimilhança. Esses valores para o efeito pepita relativo, EPR, indicam a existência de moderada dependência espacial em todos os casos.

Na Figura 18, observa-se os modelos ajustados ao semivariograma experimental para os dados de produtividade da soja na grade 50x50 m:

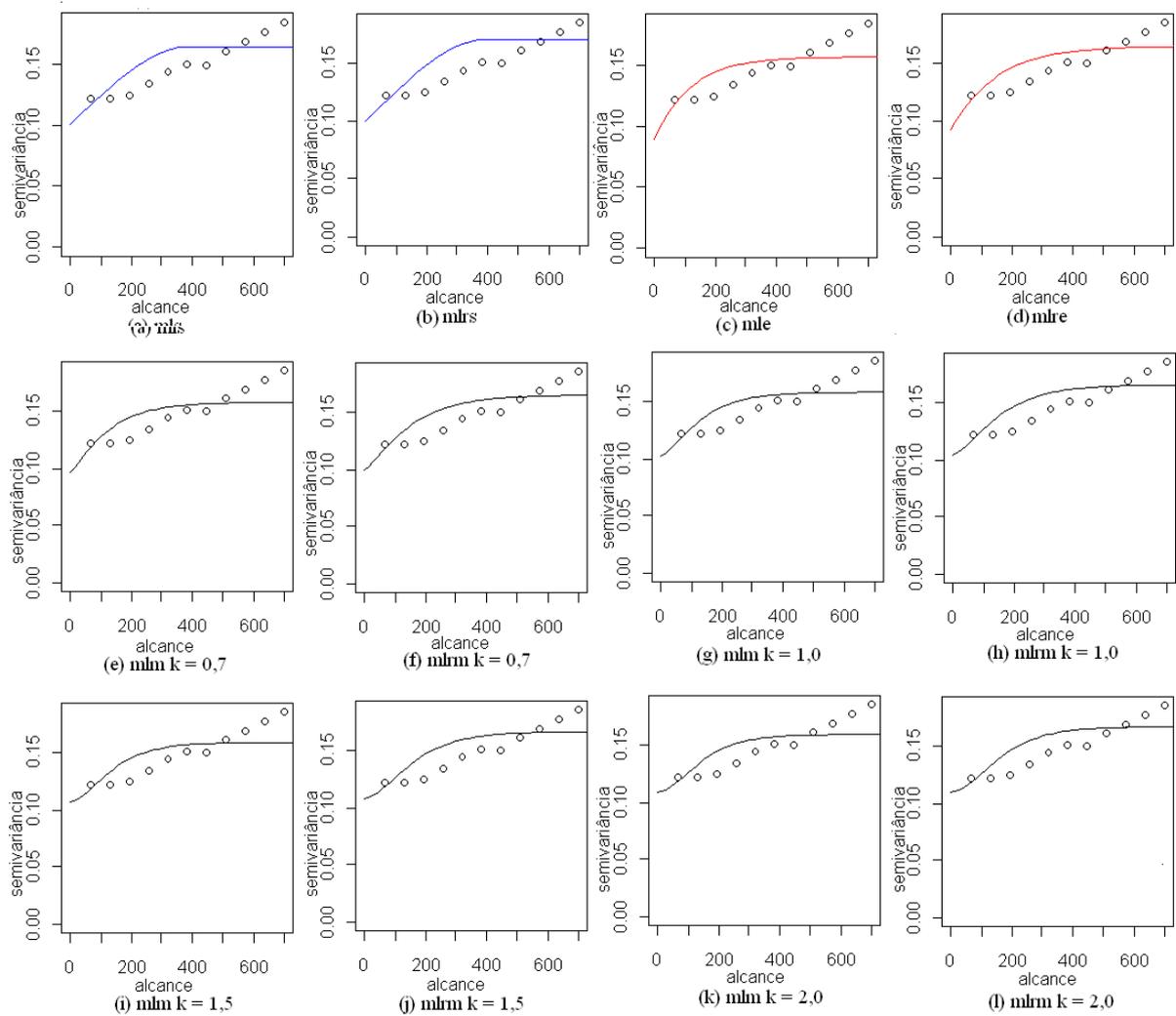


Figura 18 Semivariogramas ajustados à grade 50x50 m para (a)máxima verossimilhança com modelo esférico; (b) máxima verossimilhança restrita com modelo esférico; (c) máxima verossimilhança com modelo exponencial; (d) máxima verossimilhança restrita com modelo exponencial; (e)máxima verossimilhança com modelo Família Matérn  $\kappa=0,7$ ; (f)máxima verossimilhança restrita com modelo Família Matérn  $\kappa=0,7$ ; (g) máxima verossimilhança com Família Matérn  $\kappa=1,0$ ; (h) máxima verossimilhança restrita com Família Matérn  $\kappa=1,0$ ; (i)máxima verossimilhança com Família Matérn  $\kappa=1,5$ ; (j)máxima verossimilhança restrita com Família Matérn  $\kappa=1,5$ ; (k) máxima verossimilhança com Família Matérn  $\kappa=2,0$ ; (l) máxima verossimilhança restrita com Família Matérn  $\kappa=2,0$ .

Na Tabela 18, observa-se a validação cruzada para a grade 50x50 m. Esse método de seleção escolheu o modelo Família Matérn, com  $\kappa = 2,0$ , ajustado por máxima verossimilhança, como sendo o melhor modelo.

Tabela 18 Seleção do melhor modelo para os dados de produtividade da soja para a grade 50x50 m

Modelo	Método	EM	ER	$DP_{EM}$	$DP_{ER}$	EA	IAC	MVLV
Esférico	MV	-0,00100	-0,00142	<b>0,35424</b>	1,00638	47,8263	-12231,73	-76,82
	MVR	-0,00110	-0,00149	0,35402	1,00617	<b>47,8602</b>	-12106,84	-75,76
Exponencial	MV	-0,00097	-0,0014	0,35584	1,00546	48,2705	12307,18	-77,58
	MVR	-0,00100	-0,00139	0,35610	<b>1,00541</b>	48,2378	12280,31	-76,54
Matérn $\kappa = 0,7$	MV	-0,00096	-0,0016	0,35609	1,00562	48,2563	12459,63	77,61
	MVR	-0,00098	-0,00138	0,35634	1,00556	48,2212	12150,11	76,57
Matérn $\kappa = 1,0$	MV	-0,00096	-0,00134	0,35641	1,00580	48,2331	12420,02	77,64
	MVR	-0,00098	-0,00137	0,35663	1,00574	48,1903	12351,47	76,60
Matérn $\kappa = 1,5$	MV	-0,00096	-0,00134	0,35678	1,00600	48,1936	12336,12	77,68
	MVR	-0,00099	-0,00137	0,35695	1,00594	48,1463	12209,60	76,64
Matérn $\kappa = 2,0$	MV	<b>-0,00095</b>	<b>-0,00133</b>	0,35700	1,00613	48,1645	12293,50	77,07
	MVR	-0,00098	-0,00136	0,35713	1,00606	48,1368	12,612,09	76,65

*EM*: erro médio; *ER*: erro reduzido; *DP<sub>EM</sub>*: desvio padrão do erro médio; *DP<sub>ER</sub>*: desvio padrão do erro reduzido; *EA*: erro absoluto; *IAC*: informação de Akaike; *MVLV*: máximo do logaritmo da função verossimilhança.

O modelo Matérn, com  $\kappa = 2,0$ , ajustado por máxima verossimilhança foi considerado como o melhor modelo, de acordo com informações extraídas da Tabela 18. Observa-se que este modelo apresentou o valor mais próximo de zero para o erro médio, EM, e para o erro reduzido, ER, sendo -0,00095 e -0,00133 respectivamente. Utilizando o modelo Família Matérn com  $\kappa = 2,0$ , ajustado pelo método de máxima verossimilhança, tem-se o mapa da Figura 19.

A partir do modelo Matérn, com  $\kappa = 2,0$ , ajustado por máxima verossimilhança, aceito como o melhor modelo, cujos parâmetros são  $\varphi_1 = 0,1090$ ,  $\varphi_2 = 0,0496$ ,  $\varphi_3 = 66,7865$ ,  $a = 358,5349$ , obtém-se a equação 4.3.

$$\gamma(h) = 0,1090 + 0,0496 \left[ 1 - \left( 0,5\Gamma(2) \right)^{-1} \left( \frac{h}{66,7865} \right)^2 K_2 \left( \frac{h}{66,7865} \right) \right], \quad h > 0. \quad (4.3)$$

Na Figura 19, observa-se o mapa temático para a variável regionalizada produtividade da soja para a grade amostral com distância de 50 m.

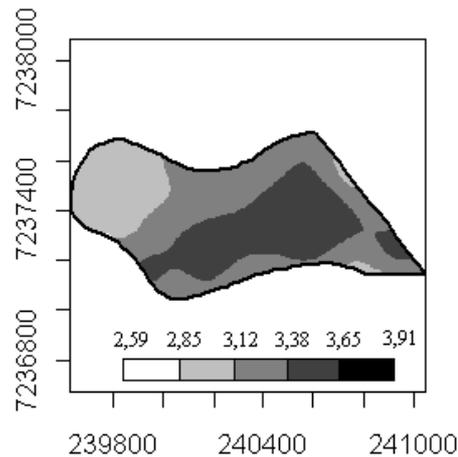


Figura 19 Mapa temático para a grade 50x50 m.

Visualmente, o mapa temático para os dados de produtividade da soja para a grade 50x50 m, Figura 19, demonstra que não houve pixels classificados nas classes C1 e C5, por não haver regiões brancas (menor produtividade), nem pretas (maior produtividade), da mesma maneira que os dados da grade 100x100 m. Para quantificar que região apresenta cada classe, construiu-se a Tabela 19.

Tabela 19 Classificação dos pixels do mapa para a grade amostral 50x50 m

Classes	C1	C2	C3	C4	C5	Total
Nº de pixels	0	4899	9772	8196	0	22867
% do total de pixels	0	21,42	42,73	35,85	0	100
Área (ha)	0	12,25	24,43	20,48	0	57,16

A classe à qual houve mais pixels pertencendo foi a classe C3, com 42,73% dos pixels, seguida pela classe C4, com 35,85% do total de pixels do mapa.

Calcularam-se as estatísticas descritivas para os valores preditos pelo modelo Matérn, com  $\kappa = 2,0$ , as quais podem ser observadas na Tabela 20.

Tabela 20 Estatísticas descritivas para os valores preditos pelo modelo Matérn, com  $\kappa = 0,7$ , ajustado por máxima verossimilhança, para a grade 50x50 m, em  $t\ ha^{-1}$

Grade	Mínimo	Q1	Mediana	Média	Q3	Máximo	DP	CV (%)
50x50 m	2,86	3,31	3,31	3,29	3,44	3,63	0,190	5,78

Pelo coeficiente de variação, CV, em 5,78%, pode-se dizer que esses valores são homogêneos. Isso justifica que a krigagem não tenha classificado para o mapa temático pixels nas classes C1 e C5, mantendo o mapa apenas com três classes, que variam de 2,85 a 3,65  $t\ ha^{-1}$ .

#### 4.5.4 Modelagem espacial para os dados da grade 25x25 m

A Figura 20 e a Tabela 21 referem-se aos ajustes realizados para a variável regionalizada produtividade da soja para a grade de 25x25 m.

Tabela 21 Determinação de parâmetros para os métodos de estimação e modelos - grade 25x25 m

Modelo	Métodos	$\varphi_1$	$\varphi_2$	$\varphi_3$	$a$	$EPR = \frac{C_0}{C_0 + C_1}$
Esférico	MV	0,0712	0,0985	202,9154	202,9154	0,4196
	MVR	0,0708	0,1020	205,7238	205,7238	0,4097
Exponencial	MV	0,0657	0,0998	97,8487	293,1286	0,3670
	MVR	0,0668	0,1077	112,2458	336,2583	0,3828
Matérn $\kappa = 0,7$	MV	0,0721	0,0929	77,0467	265,5704	0,4370
	MVR	0,0729	0,0992	84,7438	292,1016	0,4236
Matérn $\kappa = 1,0$	MV	0,0764	0,0877	59,2579	236,9440	0,4656
	MVR	0,0770	0,0925	63,2098	252,7458	0,4543
Matérn $\kappa = 1,5$	MV	0,0793	0,0837	44,2655	209,9895	0,4865
	MVR	0,0797	0,0875	46,2707	219,5017	0,4938
Matérn $\kappa = 2,0$	MV	0,0806	0,0817	36,3558	195,1718	0,4865
	MVR	0,0809	0,0851	37,6852	202,3080	0,4873

$\varphi_1$ : efeito pepita;  $\varphi_2$ : contribuição;  $\varphi_3$ : função do alcance do modelo;  $a$ : alcance; EPR: efeito pepita relativo.

Observa-se na Tabela 21 que o alcance dos modelos ajustados variou de 195,1718 m no modelo Matérn com  $\kappa = 1,0$ , ajustado por máxima verossimilhança, até 336,2583 m no modelo exponencial ajustado por máxima verossimilhança restrita. O efeito pepita relativo, EPR, em todos os casos, esteve maior que 0,20 e menor que 0,75, indicando a existência de moderada dependência espacial.

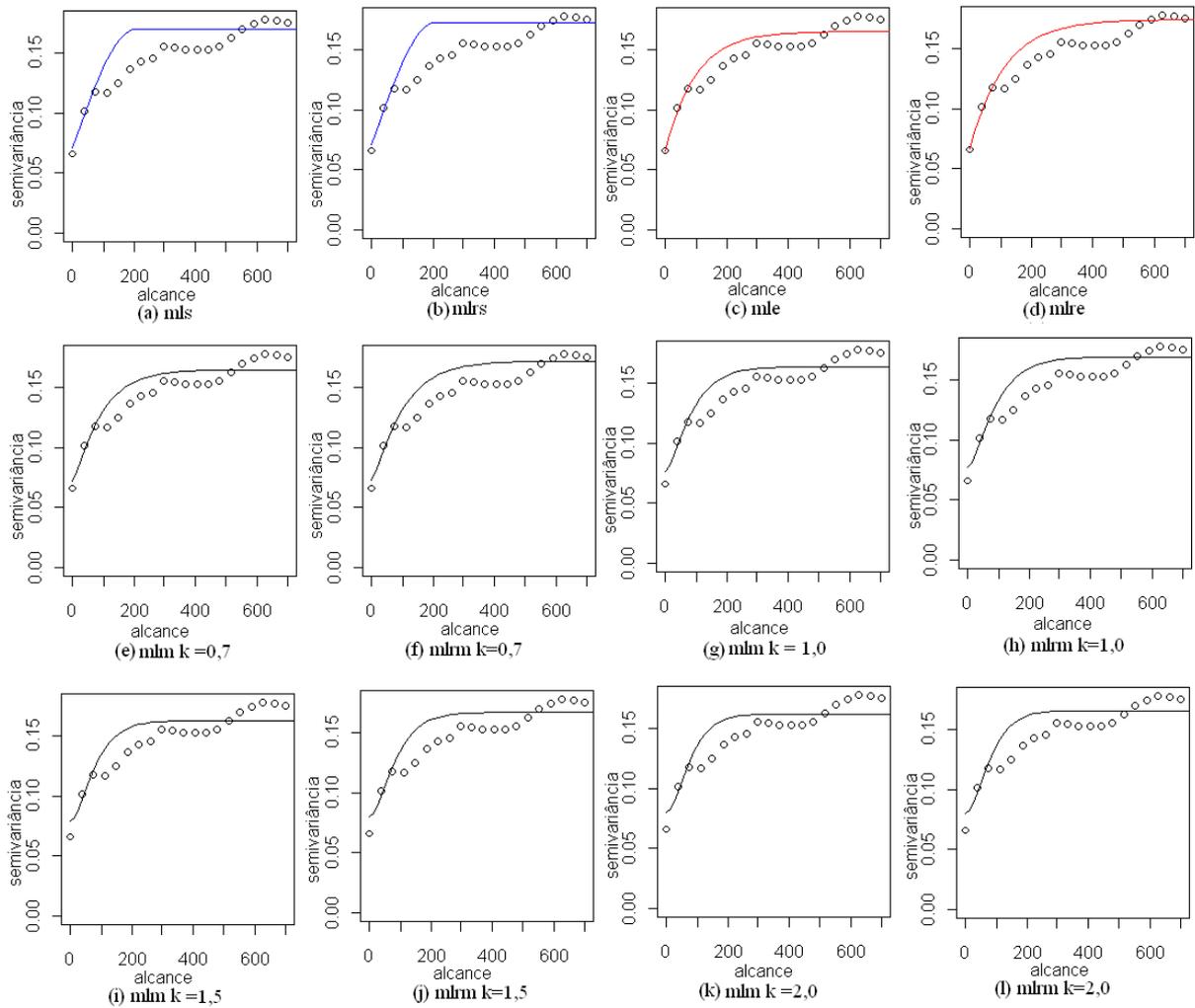


Figura 20 Semivariogramas ajustados à grade 25x25 m para (a) máxima verossimilhança com modelo esférico; (b) máxima verossimilhança restrita com modelo esférico; (c) máxima verossimilhança com modelo exponencial; (d) máxima verossimilhança restrita com modelo exponencial; (e) máxima verossimilhança com modelo Família Matérn  $\kappa=0,7$ ; (f) máxima verossimilhança restrita com modelo Família Matérn  $\kappa=0,7$ ; (g) máxima verossimilhança com Família Matérn  $\kappa=1,0$ ; (h) máxima verossimilhança restrita com Família Matérn  $\kappa=1,0$ ; (i) máxima verossimilhança com Família Matérn  $\kappa=1,5$ ; (j) máxima verossimilhança restrita com Família Matérn  $\kappa=1,5$ ; (k) máxima verossimilhança com Família Matérn  $\kappa=2,0$ ; (l) máxima verossimilhança restrita com Família Matérn  $\kappa=2,0$ .

Tabela 22 Validação cruzada para a grade 25x25 m

Modelo	Método	EM	ER	$DP_{EM}$	$DP_{ER}$	EA	IAC	MVLV
Esférico	MV	-0,00027	-0,00041	0,31517	1,00144	126,28	-34066,26	-178,5
	MVR	-0,00025	-0,00040	0,31503	1,00131	126,27	-33788,25	-177,3
Exponencial	MV	<b>-0,00010</b>	<b>-0,00016</b>	0,31527	1,00124	126,33	-33862,58	-177,0
	MVR	-0,00011	-0,00018	0,31496	1,00038	126,05	-34250,75	-175,5
Matérn $\kappa = 0,7$	MV	-0,00014	-0,00023	0,31486	<b>1,00061</b>	125,95	-34154,63	-176,9
	MVR	-0,00014	-0,00023	0,31486	1,00070	125,96	-34397,10	-175,4
Matérn $\kappa = 1,0$	MV	-0,00017	-0,00028	0,31476	1,00094	125,87	<b>-34905,98</b>	-176,8
	MVR	-0,00018	-0,00028	0,31478	1,00101	125,88	-34518,39	-175,4
Matérn $\kappa = 1,5$	MV	-0,00021	-0,00034	<b>0,31472</b>	1,00128	125,79	-34214,82	-176,9
	MVR	-0,00022	-0,00035	0,31475	1,00133	125,80	-34297,40	-175,6
Matérn $\kappa = 2,0$	MV	-0,00024	-0,00038	0,31474	1,00150	<b>125,74</b>	-34397,10	-177,0
	MVR	-0,00024	-0,00039	0,31478	1,00153	<b>125,74</b>	-34518,39	-175,8

EM: erro médio; ER: erro reduzido;  $DP_{EM}$ : desvio padrão do erro médio;  $DP_{ER}$ : desvio padrão do erro reduzido; EA: erro absoluto; IAC: informação de Akaike; MVLV: máximo do logaritmo da função verossimilhança.

Observa-se que na Tabela 22, o valor de MVLV manteve-se estabilizado entre -178,5 a -175,4, não sendo a melhor forma de decidir pelo melhor modelo. Observando-se o modelo Matérn com  $\kappa = 1,0$ , apresentou valores baixos para o desvio-padrão dos erros reduzidos,  $DP_{ER}$ , erro absoluto, EA. O critério de Akaike selecionou o modelo Família Matérn, com  $\kappa = 1,0$ , ajustado por máxima verossimilhança. Como as demais medidas diferenciavam-se apenas na quarta ou quinta casa decimal, optou-se pelo modelo Matérn como o mais adequado.

A partir do modelo Matérn, com  $\kappa = 1,0$ , ajustado por máxima verossimilhança, aceito como o melhor modelo, cujos parâmetros são  $\varphi_1 = 0,0764$ ,  $\varphi_2 = 0,0877$ ,  $\varphi_3 = 59,2579$ ,  $a = 236,9440$ , obtém-se a equação 4.4.

$$\gamma(h) = 0,0764 + 0,0877 \left[ 1 - \left( \Gamma(1) \right)^{-1} \left( \frac{h}{59,2579} \right) K_1 \left( \frac{h}{59,2579} \right) \right], \quad h > 0. \quad (4.4)$$

Utilizando o modelo Família Matérn, com  $\kappa = 1,0$ , ajustado pelo método de máxima verossimilhança, tem-se o mapa da Figura 21:

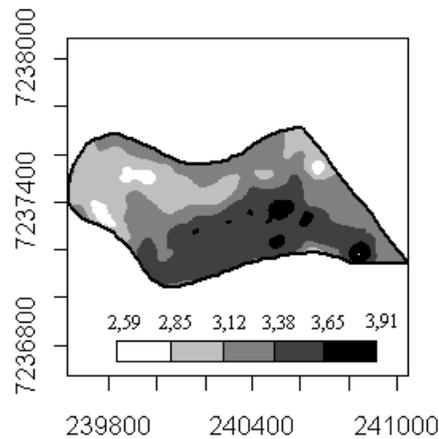


Figura 21 Mapa temático da produtividade da soja para a grade 25x25 m

Observa-se pelo mapa temático da Figura 21 que há regiões classificadas nas cinco classes, sendo menores as áreas classificadas nas classes C1 e C5. Para quantificar estes valores, tem-se a Tabela 23, que apresenta a quantidade de pixels em cada classe e que área isso representa.

Tabela 23 Classificação dos pixels do mapa para a grade amostral 25x25 m

Classes	C1	C2	C3	C4	C5	Total
Nº de pixels	647	5662	8075	7627	856	22867
% do total de pixels	2,83	24,76	35,31	33,35	3,75	100
Área (ha)	1,62	14,15	20,18	19,07	2,14	57,16

Nota-se, que as classes C3 e C4 foram predominantes no mapa temático da produtividade da soja para a grade 25x25 m, sendo que a soma de ambas representa quase 70% do total da área, ou seja, 39,25 dos 57,16 ha da área em estudo. Poucos valores ( 3,76% do total) foram classificados nas classes C1 e C5.

Calcularam-se as estatísticas descritivas para os valores preditos pelo modelo Matérn, com  $\kappa = 2,0$ , as quais poder ser observadas na Tabela 24.

Tabela 24 Estatísticas descritivas para os valores preditos pelo modelo Matérn, com  $\kappa = 0,7$ , ajustado por máxima verossimilhança, para a grade 25x25 m, em  $t\ ha^{-1}$

Grade	Mínimo	Q1	Mediana	Média	Q3	Máximo	DP	CV (%)
50x50 m	2,69	3,09	3,27	3,28	3,47	3,91	0,042	6,78

#### 4.6 Comparação dos mapas gerados

Como pode-se observar na Figura 22, foram construídos novos mapas, para as grades 100x100 m, 75x75 m, 50x50 m e 25x25 m, todos com a mesma legenda e cinco classes de mesma amplitude.

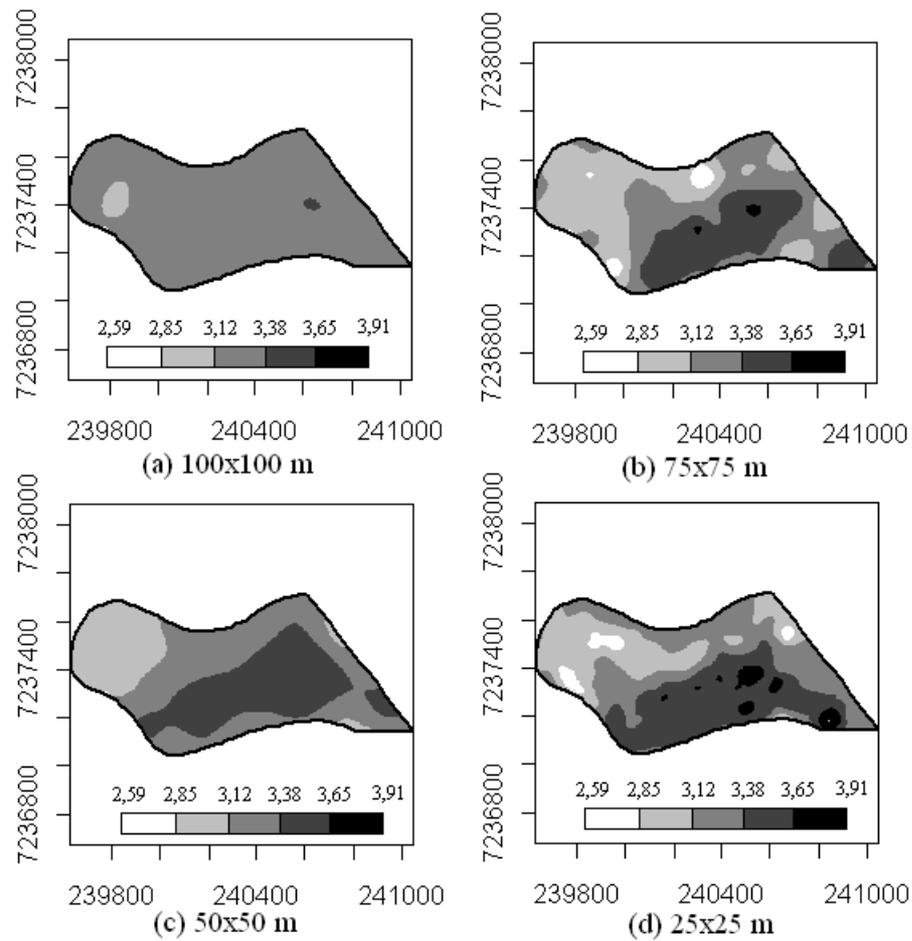


Figura 22 Mapas temáticos da produtividade da soja para as grades (a) 100x100 m; (b) 75x75 m; (c) 50x50 m e (d) 25x25 m

Na Tabela 25 observa-se a quantidade de pixels por classe e suas respectivas porcentagens, para os mapas das grades 100x100 m, 75x75 m, 50x50 m, 25x25 m.

Tabela 25 Quantidade de pixels por classe para os mapas temáticos da produtividade da soja e porcentagem do total de pixels

Mapa	Intervalos de classes				
	C1	C2	C3	C4	C5
100x100 m	0	521	22247	99	0
%	0	2,29	97,29	0,43	0
75x75 m	467	8348	7692	6224	136
%	2,04	36,51	33,64	27,22	0,59
50x50 m	0	4899	9772	8196	0
%	0	21,42	42,76	35,84	0
25x25 m	647	5662	8075	7627	856
%	2,82	24,76	35,31	33,35	3,74

C1 = [2,59; 2,85], C2 = (2,85; 3,12], C3 = (3,12; 3,38], C4 = (3,38; 3,65] e C5 = (3,65; 3,91].

A matriz de erros pode quantificar o grau de semelhança entre dois mapas que estão sendo comparados e, a partir desta matriz, encontram-se diversos outros índices, úteis na análise e comparação de mapas temáticos.

A Tabela 26 representa a matriz de erros considerando o mapa gerado a partir da grade 25x25 m como o mapa padrão e o mapa da grade 50x50 m como o mapa modelo.

Tabela 26 Matriz de erros para comparação dos mapas da grade 25x25 m (referência) e 50x50 m (modelo)

		Mapa padrão - 25x25 m					Total
		Classes	C1	C2	C3	C4	
Mapa modelo 50x50	C1	0	0	0	0	0	0
	C2	559	3099	1111	71	59	4899
	C3	88	2499	4782	2205	198	9772
	C4	0	64	2182	5351	599	8196
	C5	0	0	0	0	0	0
Total		647	5662	8075	7627	856	22867

C1 = [2,59; 2,85], C2 = (2,85; 3,12], C3 = (3,12; 3,38], C4 = (3,38; 3,65] e C5 = (3,65; 3,91]

A partir da matriz de erros, para as grades 25x25 m como referência e 50x50 m como modelo (Tabela 26), obteve-se as estatísticas de Exatidão Global, EG = 56,65%, índice kappa,  $\kappa = 39,10\%$ . Pela classificação de Landis & Koch (1977), considera-se que o desempenho de classificação foi razoável. Pela classificação de Foody (2002), considera-se um desempenho de classificação ruim.

Na Tabela 27, são apresentados os índices de acurácia por classe.  $Au_i$ : acurácia do usuário;  $Ap_i$ : acurácia do produtor;  $Kcu_i$ : kappa condicional do usuário;  $Kcp_i$ : kappa condicional do produtor;  $Kcmu_i$ : kappa condicional modificado do usuário e  $Kcmp_i$ : kappa condicional modificado do produtor.

Tabela 27 Índice de Acurácia para as classes na produtividade da soja na comparação entre os mapas temáticos nas grades 25x25 m (referência) e 50x50 m (modelo), em percentual

Classes	$Au_i$	$Ap_i$	$Kcu_i$	$Kcp_i$	$Kcmu_i$	$Kcmp_i$
C1	0	0	-2,91	ind	ind	ind
C2	63,26	54,73	51,17	42,39	54,08	43,41
C3	48,94	59,22	21,07	28,79	36,18	49,03
C4	65,29	70,16	47,92	53,49	56,61	62,70
C5	0	0	-3,89	ind	ind	ind

C1 = [2,59; 2,85], C2 = (2,85; 3,12], C3 = (3,12; 3,38], C4 = (3,38; 3,65] e C5 = (3,65; 3,91]

Observa-se pelos dados da Tabela 27 que os resultados são todos positivos ou nulos, exceto os valores de  $Kcu_i$  para as classes C1 e C5. A classe que apresentou maior acurácia do usuário, AU, foi a classe C4, com 65,29%.

A Tabela 28 representa a matriz de erros, considerando o mapa gerado a partir da grade 25x25 m como o mapa padrão e o mapa da grade 75x75 m como o mapa modelo.

Tabela 28 Matriz de erros para comparação dos mapas da grade 25x25 m (referência) e 75x75 m (modelo)

		Mapa padrão - 25x25 m						
		Classes	C1	C2	C3	C4	C5	Total
Mapa modelo 75x75 m	C1		10	85	212	160	0	467
	C2		580	3559	3657	552	0	8348
	C3		57	2018	3311	2137	169	7692
	C4		0	0	895	4743	586	6224
	C5		0	0	0	35	101	136
Total			647	5662	8075	7627	856	22867

C1 = [2,59; 2,85], C2 = (2,85; 3,12], C3 = (3,12; 3,38], C4 = (3,38; 3,65] e C5 = (3,65; 3,91]

A partir da matriz de erros, para as grades 25x25 m como referência e 75x75 m como modelo (Tabela 28), obteve-se as estatísticas de exatidão global, EG = 47,57% e índice kappa,  $\kappa = 25,91\%$ . Pela classificação de Landis & Koch (1977), considera-se que o desempenho de classificação foi razoável. Pela classificação de Foody (2002), considera-se um desempenho de classificação ruim.

Calculou-se, utilizando informações da matriz de erros, Tabela 28, as estatísticas de Acurácia do Usuário ( $Au_i$ ), e Acurácia do produtor ( $Ap_i$ ).

Tabela 29 Índice de Acurácia para as classes na produtividade da soja na comparação entre os mapas temáticos nas grades 25x25 m (referência) e 75x75 m (modelo)

Classes	$Au_i$	$Ap_i$	$Kcu_i$	$Kcp_i$	$Kcmu_i$	$Kcmp_i$
C1	2,14	1,55	-0,71	-0,50	-22,33	-23,06
C2	42,63	62,86	23,75	41,51	28,29	53,58
C3	43,04	41,00	11,96	11,09	28,80	26,25
C4	76,21	62,19	64,30	48,05	70,26	57,74
C5	74,26	11,80	73,26	11,27	67,83	-10,25

C1 = [2,59; 2,85], C2 = (2,85; 3,12], C3 = (3,12; 3,38], C4 = (3,38; 3,65] e C5 = (3,65; 3,91]

Na Tabela 29 é possível notar que apenas a classe C1 apresentou índices negativos, sendo eles para  $Kcu_1$ ,  $Kcp_1$ ,  $Kcmu_1$  e  $Kcmp_1$ , e também a classe C5, para o índice  $Kcmp_5$ .

A Tabela 30 representa a matriz de erros considerando o mapa gerado a partir da grade 25x25 m como o mapa padrão e o mapa da grade 100x100 m como o mapa modelo.

Tabela 30 Matriz de erros para comparação dos mapas da grade 25x25 m (referência) e 100x100 m (modelo)

		Mapa padrão - 25x25 m					
Classes		C1	C2	C3	C4	C5	Total
Mapa modelo 100x100	C1	0	0	0	0	0	0
	C2	46	463	12	0	0	521
	C3	601	5199	7996	7595	856	22247
	C4	0	0	67	32	0	99
	C5	0	0	0	0	0	0
Total		647	5662	8075	7627	856	22867

C1 = [2,59; 2,85], C2 = (2,85; 3,12], C3 = (3,12; 3,38], C4 = (3,38; 3,65] e C5 = (3,65; 3,91]

Os resultados apresentam a dificuldade do método preditivo identificar as classes C1 e C5. Isso pode ser explicado pelo fato de as áreas dessas classes serem pequenas em relação ao total da área. O total de pixels da classe C1, 647, representa 2,83 % do total de pixels do mapa, e da classe C5, que são 856 pixels, representam 3,74 % do total do mapa.

Tabela 31 Índices de Acurácia para as classes na produtividade da soja na comparação entre os mapas temáticos nas grades 25x25 m (referência) e 100x100 m (modelo)

Classes	$Au_i$	$Ap_i$	$Kcu_i$	$Kcp_i$	$Kcmu_i$	$Kcmp_i$
C1	ind	0	ind	0	ind	-25,00
C2	88,87	8,18	85,21	6,04	86,09	-14,78
C3	35,94	99,02	0,97	63,86	19,93	98,78
C4	32,32	0,42	-1,55	-0,01	15,40	-24,48
C5	ind	0	ind	0	ind	-25,00

C1 = [2,59; 2,85], C2 = (2,85; 3,12], C3 = (3,12; 3,38], C4 = (3,38; 3,65] e C5 = (3,65; 3,91], ind = indeterminado.

A classe C2, de acordo com os dados da Tabela 32, apresentou acurácia do usuário em 88,87%, o que indica que apenas 11,13% dos pixels classificados nessa classe no mapa modelo pertencem a outras classes no mapa referência. Esses resultados podem ser explicados pelo fato de que os dados apresentaram homogeneidade e a predição no mapa classificou a maior parte dos pontos (97,29%, conforme Tabela 25) numa mesma classe, que seria a C3.

Tabela 32 Índices de acurácia na produtividade da soja na comparação entre os mapas temáticos nas grades 25x25 m (referência) e os mapas 100x100, 75x75 e 50x50 m (modelo), valores em percentual

Mapa modelo	$\hat{\kappa}$	$\tau$	$EG$	$Amu$	$Amp$	$Maup$	$IamH$	$IamS$	$Acpu$
100x100 m	3,18	13,71	37,13	31,43	21,52	26,48	13,71	8,86	25,42
75x75 m	30,31	39,09	51,27	47,66	35,88	41,77	36,69	23,65	43,98
50x50 m	37,72	47,33	57,87	35,50	36,82	36,16	35,98	11,86	46,93

Observa-se que o mapa construído a partir da grade 100x100 m não é representativo da área experimental, pois concordou apenas 3,18 % com o mapa da grade 25x25 m, de acordo com o índice Kappa. A partir da matriz de erros para as grades 25x25 m (referência) e 100x100 m (modelo) (Tabela 30), obteve-se as estatísticas de exatidão global,  $EG = 41,43\%$  índice kappa,  $\kappa = 13,71\%$  - Tabela 32. Pela classificação de Landis & Koch (1977), considera-se que o desempenho de classificação foi mau e pelas classificações de Foody (2002) e Krippendorff (1980), considera-se um desempenho de classificação ruim.

Embora o índice de Exatidão Global, EG, apresente valores mais elevados para as comparações entre os mapas analisados, os coeficientes de concordância Kappa e Tau são mais consistentes, pois utilizam-se de todas as células da matriz de erros.

Na Tabela 33 são apresentados os índices obtidos a partir da matriz de confusão para a grade 100x100 m.

Tabela 33 Índices obtidos a partir da matriz de confusão para a grade 100x100 m (%)

Métricas	C1	C2	C3	C4	C5
Índice de sensibilidade (S)	0	8,18	99,02	0,42	0
Índice de especificidade (E)	100	99,66	3,66	99,56	100
Taxa de falso positivo (TFP) (Erro de comissão)	0	0,34	96,34	0,44	0
Taxa de falso negativo (TFN) (Erro de omissão)	100	91,82	0,98	99,58	100

Para a grade 100x100 m, obteve-se como maiores índices de especificidade as classes C1 e C5. Isso significa que é certo que um pixel que não pertença à classe C1 ou à classe C5 seja classificado como não pertencente a qualquer dessas classes, uma vez que, de acordo com o observado na Tabela 30, o mapa temático para a variável regionalizada produtividade da soja para a grade 100x100 m não classificou pixels nessas classes.

Na Tabela 34 são apresentados os índices obtidos a partir da matriz de confusão para a grade 75x75 m.

Tabela 34 Índices obtidos a partir da matriz de confusão para a grade 75X75 m (%)

Métricas	C1	C2	C3	C4	C5
Índice de sensibilidade (S)	0	15,12	41,00	62,19	11,80
Índice de especificidade (E)	97,94	72,17	70,38	90,28	99,84
Taxa de falso positivo (TFP) (Erro de comissão)	2,06	27,83	29,62	9,72	0,16
Taxa de falso negativo (TFN) (Erro de omissão)	100	84,88	59,00	37,81	88,20

A classe C1 apresentou índice de sensibilidade igual a zero e TFN = 100. A classe C4 apresentou maior índice de sensibilidade, sendo ele de 62,19 %. Já a classe C5 retornou o maior índice de especificidade,  $E = 99,84$ , ou seja, há uma chance de 99,84% de um pixel não pertencente à classe C5 ser de fato classificado como não pertencente à classe C5.

Na Tabela 35 são apresentados os índices obtidos a partir da matriz de confusão para a grade 50x50 m.

Tabela 35 Índices obtidos a partir da matriz de confusão para a grade 50X50 m (%)

Métricas	C1	C2	C3	C4	C5
Índice de sensibilidade (S)	0	54,73	59,22	70,16	0
Índice de especificidade (E)	100	89,54	66,27	81,33	100
Taxa de falso positivo (TFP) (Erro de comissão)	0	10,46	33,73	18,67	0
Taxa de falso negativo (TFN) (Erro de omissão)	100	45,27	40,78	29,84	100

Destaca-se que não houve pixels classificados como pertencentes às classes C1 e C5, para o mapa da grade 50x50 m, dessa forma, o índice de sensibilidade S e a taxa de falsos positivos, TFP, apresentaram valor zero para essas classes. Repara-se que o maior erro de

comissão para os dados da grade 50x50 m ocorre na classe C3, e o maior erro de omissão nas classes C1 e C5.

Tabela 36 Matriz de confusão total para a grade 100x100 m

		Mapa 25x25 m	
		verdadeiro	falso
Mapa 100x100 m	verdadeiro	8491	8463
	falso	5913	14376

O coeficiente de correlação de Matthews calculado para os dados da grade 100x100 m foi de 21,41 %, o que indica uma fraca correlação positiva.

Tabela 37 Matriz de confusão total para a grade 75x75 m

		Mapa 25x25 m	
		verdadeiro	falso
Mapa 75x75 m	verdadeiro	11724	7558
	falso	3585	11143

O coeficiente de correlação de Matthews, *CCM*, calculado para os dados da grade 75x75 m foi de 36,31 %, o que indica uma correlação moderada positiva.

Tabela 38 Matriz de confusão total para a grade 50x50 m

		Mapa 25x25 m	
		verdadeiro	falso
Mapa 50x50 m	verdadeiro	13232	4243
	falso	5392	9635

Para os dados da grade 50x50 m, calculou-se um *CCM* de 45,42 %, que indica uma moderada correlação positiva.

Na Tabela 39, observam-se os valores estimados por cada modelo para cada classe de produção média, em toneladas.

Tabela 39 Produção (t) estimada para cada modelo, por classes

Classes	C1	C2	C3	C4	C5	Total
Referência	4,41	42,31	65,59	67,13	8,09	187,53
100x100 m	0	3,89	180,73	0,88	0	185,50
75x75 m	3,18	62,40	62,47	54,77	1,28	184,10
50x50 m	0	36,63	79,40	72,09	0	188,12

Observa-se na Tabela 39, que a produção total estimada pelo mapa de referência foi de 187,53 t, ou seja, 2,03 t a mais que a estimativa do mapa da grade 100x100 m, 3,43 t a mais

que o mapa da grade 75x75 m e -0,59 t a menos que a estimativa do mapa para a grade 50x50 m.

## 5 CONCLUSÕES

Além de uma melhor precisão dos mapas de variabilidade espacial gerados, a análise dos índices de acurácia possibilita um melhor planejamento das malhas amostrais, para futuros estudos na mesma área experimental.

A 100x100 dos dados de produtividade da soja apresentam baixa variabilidade. Isso demonstra um equívoco, pois ao analisar por uma grade mais densa, como a 75x75 m, identifica-se essa variabilidade espacial.

Os mapas construídos para cada grade amostral permitiram visualizar a situação da área em estudo, possibilitando ao produtor ou pesquisador identificar quais as áreas com maior e menor produtividade. Apesar de, visualmente, o mapa da grade 75x75 m apresentar maior proximidade com os dados do mapa modelo, os índices de acurácia calculados, mostraram que, na verdade, o mapa da grade 50x50 m é que melhor representou a área experimental, seguido do mapa gerado pela grade 75x75 m.

## 6 CONSIDERAÇÕES FINAIS

Para os dados do monitor de colheita, não conseguiu-se realizar a krigagem universal, impossibilitando realizar a comparação por meio da matriz de erros, matriz de confusão e índices de acurácia, que exigem que haja o mesmo número de pixels para os mapas comparados. Assim, para a comparação que, inicialmente, pretendia-se realizar com o monitor de colheita (real), foi tomado como mapa de referência o krigado, para os dados da grade 25x25 m, sendo que esses pontos foram extraídos do monitor de colheita.

Os resultados mostraram que as medidas de acurácia, obtidas por meio da matriz de erros, são opções significativas para realizar a comparação entre mapas temáticos, uma vez que fornecem índices globais e também por classes.

## 7 REFERÊNCIAS

- AKAIKE, H. Information theory as an extension of the maximum likelihood principle. In: KOTZ, S.; JOHNSON, N. L. **Breaktroughs in Statistics**, v.1: Foundations and Basic Theory. New York: Springer-Verlag, p. 610-624, 1992.
- ANDERSON, J. F.; HARDY, E. E.; ROACH, J. T.; WITMER, R. E. A land use and land cover classification system for use with remote sensor data. U.S. **Geological Survey Professional Paper**, n. 964. USGS, Washington, D.C, 1976. 28p.
- BAZZI, C. L.; SOUZA, E. G.; URIBE-OPAZO, M. A.; NÓBREGA, L. H. P.; NETO, R. P. Influência da distância entre passadas de colhedora equipada com monitor de colheita na precisão dos mapas de produtividade na cultura do milho. **Engenharia Agrícola**, Jaboticabal, v. 28, n. 2, p. 355-363, 2008.
- BORSSOI, J. A.; URIBE-OPAZO, M. A.; GALEA, M. Diagnostic techniques applied in geostatistics for agricultural data analysis. In: **Revista Brasileira de Ciência do Solo**, Viçosa, v. 33, n. 6, p. 1561-1570, nov/dez 2009.
- BORSSOI, J. A.; URIBE-OPAZO, M. A.; GALEA, M. Técnicas de diagnósticos de influência local na análise espacial da produtividade da soja. **Engenharia Agrícola** (impresso), v.31, p.376-387, 2011.
- BOURGAULT, G.; JOURNAL, A. G.; RHOADES, J. D.; CORWIN, J. D.; LESCH, S. M. Geostatistical Analysis of a soil salinity data set. **Advances in Agronomy**, New York, v. 58, p. 241 - 292, 1996.
- BRITES, R. S.; SOARES, V. P.; RIBEIRO, C. A. A. S. Verificação da exatidão em classificações de uma imagem mediante a utilização de três índices. **Revista Árvore**, Viçosa -MG, v. 20, n. 3, p. 415-424, 1996.
- 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**, Madison, v. 58, n. 5, p. 1501-1511, 1994.
- CARVALHO, J. R. P.; ASSAD, E. D. Análise espacial da precipitação pluviométrica no estado de São Paulo: comparação de métodos de interpolação. **Engenharia Agrícola**, v. 25, n. 2, p. 377-384, 2005.
- CHRISTENSEN, R.; JOHNSON, W.; PEARSON, L. Covariance function diagnostics for spatial linear models. **International Association of Mathematical geology**, New York, v. 25, n. 2, p. 145-160, 1993.
- CLARCK, I. **Practical geostatistics**. London: Applied science publishers Ltda, 1979. 123p.

- COHEN, J. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, v. 20, n. 1, p. 37-46, 1960.
- CONAB - Companhia Nacional de Abastecimento. **Indicadores da agropecuária**, Brasília, ano XX, n.08, 2011.
- CONGALTON, R. G. A review of assessing the accuracy of classifications of remotely sensed data. **Remote sensing of environment**. v. 37, n. 1, p. 35-46, 1991.
- COOK, R. D. Assessment of local influence (with discussion). **Journal of the Royal Statistical Society**, Series B, London, v. 48, n. 2, p. 133-169, 1986.
- CRESSIE, N. **Statistics for spatial data**. New York: John Wiley, 1993. 900 p.
- DAVIS, B. M. Uses and abuses of Cross-Validation in geostatistics. **Mathematical Geology**, New York, v. 19, n. 3, p. 241-248, 1987.
- DELFINER, P.; DELHOME, J.P. Optimum interpolation by kriging. In: DAVIS, J.C.; McCULLAGH, M.J.(eds) **Display and analysis of spatial data**. New York, p. 96-114, 1975.
- FARACO, M. A.; URIBE-OPAZO, M. A.; SILVA, E. A. A.; JOHANN, J. A.; BORSSOI, J. A. Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. In: **Revista Brasileira de Ciência do Solo**. [online]. v. 32, n.2, p. 463-476, 2008.
- FIELDING, A. H.; BELL, J. F. A review of methods for the assessment of prediction errors in conservation presence/absence models. In: **Foundation for Environmental Conservation**. [online] v. 24, n. 1, p. 38-49 mar/1997.
- FIGUEIREDO, S. M. M.; CARVALHO, L. M. T. Avaliação da exatidão do mapeamento da cobertura da terra em Capixaba, Acre, utilizando classificação por árvore de decisão. **Cerne**, Lavras, v. 12, n. 1, p. 38-47, jan./mar. 2006.
- FOODY, G. M. Status of land cover classification accuracy assessment. **Remote Sensing of Environment**, v. 80, n. 1, p. 185-201, 2002.
- GEMAEL, C. **Introdução ao ajustamento de observações: Aplicações geodésicas**. Curitiba: Editora da UFPR. 1994. 320p.
- GOMES, F. P. **Curso de estatística experimental**. 14.ed. Piracicaba: Nobel, 2000. 477p.
- GOMES, J. B. V.; BOLFE, E. L.; CURI, N.; FONTES, H. R.; BARRETO, A. C.; VIANA, R. D. Variabilidade espacial de atributos de solos em unidades de manejo em área piloto de produção integrada de coco. **Revista Brasileira de Ciência do Solo**, Viçosa, v.32, n.6, p. 2471-2482, nov/dez 2008.
- GORODKIN, J. Comparing two K-category assignments by a K-category correlation coefficient. **Computational Biology and Chemistry**, v. 28, n.5-6, p. 367-374, 2004.

GUEDES, L. P. C. **Otimização de amostragem espacial**. Tese (Doutorado em Agronomia) Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2008a, 143p.

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, p. 2217-2226, 2008b.

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

HELLDEN, U. **A test of Landsat-2 imagery and digital data for thematic mapping**. Sweden, Lund University Natural Geography Institute Report, 1980. 63p.

HUSCH, B.; MILLER, C. I.; BEERS, T. W. **Forest mensuration**. New York: John Wiley & Sons, 1982. 402p.

IAPAR e EMBRAPA. CD-ROM: **Mapas de Solos do Paraná**, 2007.

JENSEN, J. **Introductory digital image processing**. New Jersey: Prentice-Hall, 1986.

JONES, T. A. Skewness and kurtosis as criteria of normality i observed frequency distributions. **Journal Sedimentar Petrology**, Colorado, v. 39, n. 4, p.1622-1627, 1969. 379p.

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

KIRKPATRICK, S. C.; GELATT, C. D.; VECCHI, J. M. P. Optimization by simulated annealing. **Science**, Oxford, v. 220, n. 4598, p. 671-680, may. 1983.

KRIPPENDORFF, K. **Content Analysis: An Introduction to Its Methodology**. Newbury Park, CA: Sage, 1980. 189p.

LANDIM, P. M. B. Sobre Geoestatística e Mapas. **Terrae Didatica**, Campinas, v. 2, n. 1, p. 19-33, 2006.

LEE, J. S.; GRUNES, M. R.; POTTIER, E. Quantitative comparison of classification capability: fully polametric versus dual and single-polarization SAR. **IEEE Transactions on Geoscience and Remote Sensing**, Michigan, v. 39, n. 11, p. 2343-2351, 2001.

LILLESAND, T. M.; KIEFER, R. W.; CHIPMAN, J. W. **Remote sensing and image interpretation**. 5. ed. New York: Wiley & Sons, 2004. 763p.

LIM, H. S.; MATJAFRI, M. Z.; ABDULLAH, Land cover classification over penang Island, Malaysia using SPOT data. In: **IJCIM**, v. 17, n. SP1, p. 15.1-15.5, Bangkok Metro, Thailand, mar/2009.

LOPHAVEN, S. **Design and analysis of environmental monitoring programs**. 2004. 190f. Thesis (Ph.D. in Engineerin) - Technical University of Denmark, Kongens Lyngby, 2004.

LURZ, P. W. W.; RUSHTON, S. P.; WAUTERS, L. A.; BERTOLINO, S.; CURRADO, I.; MAZZOGLIO, P.; SHIRLEY, M. D. F. Predicting grey squirrel expansion in North Italy: a spatially explicit modelling approach. **Landscape Ecology**, Tempe, Arizona, v.16, n.5, p. 407-420. jul/2001.

MA, Z.; REDMOND, R. L. Tau coefficients of accuracy assessment of classification of remote sensing data. **Photogrammetry Engineering and Remote Sensing**. Bethesda, v. 61, n. 4, p. 453-439, 1995.

MATÉRN, B. **Spatial variation**. Lecture Notes in Statistics, n. 36, Berlin: Springer-Verlag, 1986.

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

MATHERON, G. **Principles of Geostatistics**. Economic Geology, Lancaster, v. 58, n. 1, p. 1246-1266, 1963.

MATTHEWS, B. W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. **Biochim. Biophys. Acta**, 405, p. 442-451, 1975.

MELLO, C. R.; LIMA, J.M.; SILVA, A. M.; MELLO, J. M.; OLIVEIRA, M. S. Krigagem e inverso do quadrado da distância para interpolação dos parâmetros da Equação das chuvas intensas. **Revista Brasileira de Ciência do Solo**, Viçosa, v.27, n. 5, p. 925-933, 2003.

MERCANTE, E.; URIBE-OPAZO, M. A.; SOUZA, E. G. Variabilidade espacial e temporal da resistência mecânica do solo à penetração em áreas com e sem manejo químico localizado. **Revista Brasileira de Ciência do Solo**, Viçosa, MG, v. 27, n. 06, p. 1149-1160, 2003.

METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M.; TELLER, A. H.; and TELLER, E. Equation of State Calculations by Fast Computing Machines. **J. Chem. Phys.** 21, 1087-1092, 1953.

MINASNY, B; MCBRATNEY, A. The Matérn function as a general model for soil variograms. **Geoderma**. v. 128, n. 3-4, p. 192-207, 2005.

MOLIN, J. P. **Agricultura de Precisão: o gerenciamento da variabilidade**. Piracicaba: o autor, 2008. 83p.

PONZONI, F. J.; REZENDE, A. C. P. Influência da resolução espacial de imagens orbitais na identificação de elementos da paisagem em Altamira-PA. **Revista Árvore**, Viçosa, v. 26, n. 4, p.403-410, 2002.

RIBEIRO JR, P. J.; DIGGLE, P. J. **geoR: A package for geostatistical analysis**. R-NEWS, 01, 2001. Disponível em: <http://cran.r-project.org/doc/Rnews>.

ROSENFELD, G. H.; FITZPATRICK-LINS, K. A coefficient of agreement as measure of thematic classification accuracy. **Photogrammetric Engineering and Remote Sensing**, Bethesda, Maryland, v.52, n. 2, p. 223-227, 1986.

SABATTINI, R. M. E. Um programa para o cálculo da acurácia, especificidade e sensibilidade de testes médicos. In: **Revista Informédica** [online]. v. 2, n. 12, p. 19-21, 1995.

SCHABENBERGER, O.; GOTWAY, C. A. **Statistical Methods for Spatial Data Analysis**. Boca Raton, FL: Chapman & Hall/CRC, 2005. 488p.

SENNA JR, D. G.; PINTO, F. A. C.; QUEIROZ, D. M.; MANTOVANI, E. C. Algoritmo para classificação de plantas de milho atacadas pela lagarta do cartucho (*Spodoptera frugiperda*, Smith) em imagens digitais. **Revista Brasileira de Engenharia Agrícola e Ambiental**. v. 5, n.3. Campina Grande, p. 502-509, set.dez/2001.

SILVA, S. A.; LIMA, J. S. S.; SOUZA, G. S.; OLIVEIRA, R. B. Avaliação de interpoladores estatísticos e determinísticos na estimativa de atributos do solo em agricultura de precisão. **Idesia**, v. 26, n. 2, p. 75-81, 2008.

SHORT, N. Principles of computer processing of Landsat data. **NASA Technical Reference Publication 1078**. Washington, NASA. p. 421-453,1982.

SOUZA, E. G.; JOHANN, J. A.; ROCHA, J. U.; RIBEIRO, S. R. A.; SILVA, M. S.; URIBE-OPAZO, M. A.; MOLIN, J. P.; OLIVEIRA, E. F.; NÓBREGA, L. H. P. Análise da variabilidade espacial dos atributos de um solo em uma área experimental. Parte I: Análise do P, K, MO e pH. **Revista Brasileira de Engenharia Agrícola**, p. 80-92, 1999.

SOUZA, L. S.; COGO, N. P.; VIEIRA, S. R. Variabilidade de propriedades físicas e químicas do solo em um pomar cítrico. **Revista Brasileira de Ciência do Solo**, Campinas, v. 21, n. 3, p. 1-10, 1997.

SOUZA, L. S. Variabilidade de fósforo, potássio e matéria orgânica no solo, em relação a sistemas de manejo. **Revista Brasileira de Ciência do Solo**. Viçosa, v. 21, n. 3, p. 367-372, 1997.

STEHMAN, S. V. Selecting and interpreting measures of thematic classification accuracy. **Remote Sensing of Environment**, v. 62,n. 1, p. 77-89. 1997.

STORY, M.; CONGALTON, R. G. Accuracy assessment: A user's perspective. **Photogrammetric Engineering and Remote Sensing**, Bethesda, Maryland, v. 52, n. 3, p.397-399. 1986.

THOMPSON, S. K. **Sampling**. New York, Wiley - Interscience, 1992. 343p.

URIBE-OPAZO, M. A.; KAVANAG, C. R. Estimção de parâmetros no ajuste de modelos teóricos a semivariogramas experimentais. In: **Revista Brasileira de Estatística**. Rio de Janeiro, v. 67, n.227, p. 97-122, jul/dez 2006.

URIBE-OPAZO, M. A.; JOHANN, J. A.; BOAS, M. A. V.; LUNKES, C.; BORSSOI, J. A. Métodos de ajuste à semivariogramas experimentais utilizando diferentes grades amostrais na produtividade da soja. **Engenharia na Agricultura**, Viçosa, v. 15, n. 4, p. 319-330, 2007.

VIEIRA, S. R.; NIELSEN, D. R.; BIGGAR, J. W. Spatial variability of field-measured infiltration rate, In: **Soil Science Society of America Journal**, Madison, v. 45, n. 4; 1040-1048, 1981.

VIEIRA, S. R.; HATFIELD, T. L.; NIELSEN, D. R.; BIGGAR, J. W. Geostatistical theory and application to variability of some agronomical properties. **Hilgardia**, Berkeley, v. 52, n. 3, p. 1-75, 1983.

VIEIRA, S. R.; TILLOTSON, P. M.; BIGGAR, J. W.; NIELSEN, D. R. Scaling of semivariograms and the kriging estimation of field-measured properties. **Revista Brasileira de Ciência do Solo**, Viçosa, v.21, n.3, p.525-33, 1997.

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

VOLTZ, M.; WEBSTER, R. A comparasion of kriging, cubc spiles and classification for predicting soil properties from sample information. **Journal of Soil Science**, Oxford, v. 41, n. 3, p. 473-490, 1990.

WEBSTER, R; McBRATNEY, A. B. On the Akaike Information Criterion for choosing models for variograms of soil properties. **European Journal of Soil Science**, Oxford, v. 40, n. 3, p. 493-496, 1989.

WOLLENHAUPT, N. C.; MULLA, D. J.; CRAWFORD, G. Soil sampling and interpolation techniques for mapping spatial variability of soil properties. In: PIERCE, F. J.; SADLER, E. J. (ed). **The state of site-specific management for agriculture**. Madison: ASA, CSSA, SSSA, 1997. p.19-53.

ZIMMERMAN, D. Computationally efficient restricted maximum likelihood estimation of generalized covariance functions. **International Association for Mathematical Geology**, New York, v. 21, n.7, p. 655-672, 1989.

## 8 ANEXOS

Matrizes de confusão por classes.

Tabela 40 Matriz de confusão da primeira classe 100x100 m

		Mapa 25x25 m	
		verdadeiro	falso
Mapa 50x50 m	verdadeiro	0	0
	falso	647	22220

Tabela 41 Matriz de confusão da segunda classe 100x100 m

		Mapa 25x25 m	
		verdadeiro	falso
Mapa 50x50 m	verdadeiro	463	58
	falso	5199	17147

Tabela 42 Matriz de confusão da terceira classe 100x100 m

		Mapa 25x25 m	
		verdadeiro	falso
Mapa 50x50 m	verdadeiro	7996	14251
	falso	79	541

Tabela 43 Matriz de confusão da quarta classe 100x100 m

		Mapa 25x25 m	
		verdadeiro	falso
Mapa 50x50 m	verdadeiro	32	67
	falso	7595	15173

Tabela 44 Matriz de confusão da quinta classe 100x100 m

		Mapa 25x25 m	
		verdadeiro	falso
Mapa 50x50 m	verdadeiro	0	0
	falso	856	22011

Tabela 45 Matriz de confusão da primeira classe 75x75 m

		Mapa 25x25 m	
		verdadeiro	falso
Mapa 50x50 m	verdadeiro	10	457
	falso	637	21763

Tabela 46 Matriz de confusão da segunda classe 75x75 m

		Mapa 25x25 m	
		verdadeiro	falso
Mapa 50x50 m	verdadeiro	3559	4789
	falso	2103	12416

Tabela 47 Matriz de confusão da terceira classe 75x75 m

		Mapa 25x25 m	
		verdadeiro	falso
Mapa 50x50 m	verdadeiro	3311	4381
	falso	4764	10411

Tabela 48 Matriz de confusão da quarta classe 75x75 m

		Mapa 25x25 m	
		verdadeiro	falso
Mapa 50x50 m	verdadeiro	4743	1481
	falso	2884	13759

Tabela 49 Matriz de confusão da quinta classe 75x75 m

		Mapa 25x25 m	
		verdadeiro	falso
Mapa 50x50 m	verdadeiro	101	35
	falso	755	21976

Tabela 50 Matriz de confusão da primeira classe 50x50 m

		Mapa 25x25 m	
		verdadeiro	falso
Mapa 50x50 m	verdadeiro	0	0
	falso	647	22220

Tabela 51 Matriz de confusão da segunda classe 50x50 m

		Mapa 25x25 m	
		verdadeiro	falso
Mapa 50x50 m	verdadeiro	3099	1800
	falso	2563	15405

Tabela 52 Matriz de confusão da terceira classe 50x50 m

		Mapa 25x25 m	
		verdadeiro	falso
Mapa 50x50 m	verdadeiro	4782	4990
	falso	3293	9802

Tabela 53 Matriz de confusão da quarta classe 50x50 m

		Mapa 25x25 m	
		verdadeiro	falso
Mapa 50x50 m	verdadeiro	5351	2845
	falso	2276	12395

Tabela 54 Matriz de confusão da quinta classe 50x50 m

		Mapa 25x25 m	
		verdadeiro	falso
Mapa 50x50 m	verdadeiro	0	0
	falso	856	22011