

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

MÓDULOS COMPUTACIONAIS DE ANÁLISE GEOESTATÍSTICA E
RETIFICAÇÃO DE ZONAS DE MANEJO

NELSON MIGUEL BETZEK

CASCADEL - PARANÁ
FEVEREIRO – 2017

NELSON MIGUEL BETZEK

**MÓDULOS COMPUTACIONAIS DE ANÁLISE GEOESTATÍSTICA E
RETIFICAÇÃO DE ZONAS DE MANEJO**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Agrícola, em cumprimento parcial aos requisitos para obtenção do título de Doutor em Engenharia Agrícola, área de concentração: Engenharia de Sistemas Biológicos e Agroindustriais.

Orientador: Prof. Dr. Eduardo Godoy de Souza.
Coorientador: Prof. Dr. Claudio Leones Bazzi – UTFPR-MD.

**CASCADEL - PARANÁ
FEVEREIRO – 2017**

Dados Internacionais de Catalogação-na-Publicação (CIP)

B466m

Betzek, Nelson Miguel

Módulos computacionais de análise geoestatística e retificação de zonas de manejo./Nelson Miguel Betzek. Cascavel, 2017.
100 f.

Orientador: Prof. Dr. Eduardo Godoy de Souza

Orientador: Prof. Dr. Claudio Leones Bazzi

Tese (Doutorado) – Universidade Estadual do Oeste do Paraná, Campus de Cascavel, 2017

Programa de Pós-Graduação em Engenharia Agrícola

1. Agricultura de precisão. 2. Índice de suavização. 3. Krigagem ordinária. 4. Sistemas computacionais. I.Souza, Eduardo Godoy de. II. Bazzi, Claudio Leones. III. Universidade Estadual do Oeste do Paraná. IV. Título.

CDD 20.ed. 630

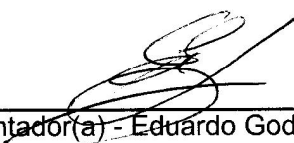
CIP-NBR 12899

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

NELSON MIGUEL BETZEK

Módulos Computacionais de Análise Geoestatística e Retificação de Zonas de Manejo

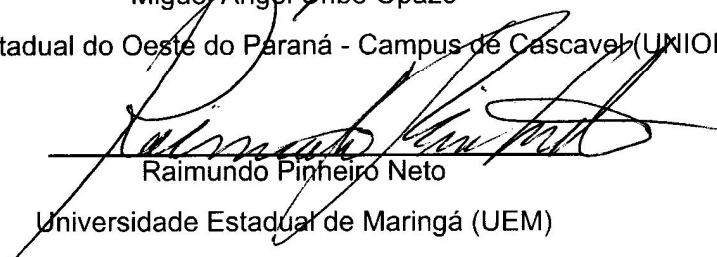
Tese apresentada ao Programa de Pós-Graduação em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de Doutor em Engenharia Agrícola, área de concentração Sistemas Biológicos e Agroindustriais, linha de pesquisa Geoprocessamento, Estatística Espacial e Agricultura de Precisão,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


Orientador(a) - Eduardo Godoy de Souza

Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)


Miguel Angel Uribe Opazo

Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)


Raimundo Pinheiro Neto

Universidade Estadual de Maringá (UEM)


Antônio Carlos Andrade Gonçalves

Universidade Estadual de Maringá (UEM)


Marcio Furlan Maggi

Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)

Cascavel, 17 de fevereiro de 2017

BIOGRAFIA

Nelson Miguel Betzek, graduado em Tecnologia em Processamento de Dados (1999) pelo Centro de Ensino Superior de Foz do Iguaçu – CESUFOZ, na cidade de Foz do Iguaçu, PR. Especialista em rede de computadores (2006) e em projeto e desenvolvimento de sistemas baseados em objetos (2005), pela Universidade Tecnológica Federal do Paraná - UTFPR, *Câmpus* Medianeira. Mestre (2012) e atualmente doutorando do Programa de Pós-Graduação Stricto Sensu em Engenharia Agrícola na área de concentração de Sistemas Agroindustriais na Universidade Estadual do Oeste do Paraná – UNIOESTE, *Câmpus* Cascavel. Docente do Núcleo de Ciência da Computação, na Universidade Tecnológica Federal do Paraná – UTFPR, *Campus* Medianeira.

DEDICATÓRIA

À minha esposa (Dulci) e filha (Bárbara),
pelo amor, carinho, apoio e compreensão
dispensados durante esta jornada.

Aos demais familiares que também
compreenderam o motivo de tal dedicação.

Aos amigos pelo incentivo.

AGRADECIMENTOS

Agradeço a Deus, por ter me concedido a vida e permitido realizar mais este sonho;

À Universidade Estadual do Oeste do Paraná (UNIOESTE), em especial ao PGEAGRI, pela oportunidade de realizar o Doutorado;

Ao professor Dr. Eduardo Godoy de Souza, pela orientação, sugestões, apoio, compreensão e confiança na realização deste trabalho;

Ao professor Dr. Claudio Leones Bazzi, pela coorientação, amizade, incentivo e apoio;

Ao professor Dr. Miguel Angel Uribe Opazo, pela colaboração;

À equipe de professores e colaboradores do PGEAGRI;

A todos os colegas do LAMAP pela amizade, companheirismo e colaboração no desenvolvimento deste trabalho;

Aos proprietários das áreas agrícolas utilizadas para coleta de amostras;

Aos meus familiares e amigos pela paciência, confiança, apoio e compreensão;

Aos colegas do departamento de computação da UTFPR e à direção do campus Medianeira pela compreensão e apoio;

A todos os demais professores, colegas e funcionários que, direta ou indiretamente, participaram da realização deste trabalho.

MÓDULOS COMPUTACIONAIS DE ANÁLISE GEOESTATÍSTICA E RETIFICAÇÃO DE ZONAS DE MANEJO

RESUMO

A utilização de sistemas computacionais para gerência de procedimentos agrícolas é necessidade premente. Porém, aplicativos utilizados com este objetivo geralmente possuem limitações quanto a funcionalidades, licenças de uso, sistema operacional, dentre outras. Por este motivo, foi desenvolvido o Software para Definição de Unidades de Manejo (SDUM). Esta tese teve como objetivo a implementação de rotinas computacionais que serão integradas ao SDUM, capazes de identificar automaticamente os melhores parâmetros para os métodos de interpolação Krigagem ordinária (KRI) e inverso da distância elevado a uma potência (IDW). Estas rotinas foram aplicadas aos dados amostrais de atributos selecionados para definir as ZMs em duas áreas agrícolas. Para cada conjunto de dados foram testados 300 diferentes ajustes para o semivariograma. Os melhores parâmetros foram utilizados para mensurar dados por KRI, e doze diferentes valores para expoente do IDW. A área A foi considerada homogênea por não apresentar média estatisticamente diferente entre as classes. E, para a área B, o melhor resultado obtido foi subdividi-la em duas classes distintas. As ZMs geralmente apresentam células isoladas ou manchas, dificultando a operacionalização da lavoura. Neste contexto, outro objetivo foi a implementação de rotinas capazes de retificar as ZMs a fim de suavizar e melhorar a continuidade. Três áreas amostrais foram subdivididas em duas, três, quatro e cinco classes, e então, retificadas por cinco vezes. Pixels isolados e praticamente todas as manchas foram eliminados, com destaque para os métodos com máscara 5×5 . Pode-se concluir que as rotinas computacionais implementadas foram eficientes e capazes de identificar o melhor ajuste para o semivariograma, bem como o melhor expoente para IDW, além de suavizar e melhorar a continuidade das ZMs.

PALAVRAS-CHAVE: agricultura de precisão; índice de suavização; krigagem ordinária; sistemas computacionais.

COMPUTATIONAL MODULES FOR GEOSTATISTIC ANALYSIS AND ADJUSTMENT OF MANAGEMENT ZONES

ABSTRACT

The use of computer systems for the management of agricultural procedures has been an emerging need. However, applications used for this purpose generally present restrictions on functionality, operating licenses, operating system and others. Thus, the Software for the Definition of Management Units (SDUM) was developed. This thesis aimed at implementing computational routines that will be integrated into SDUM, which are able to identify automatically the best parameters for the ordinary kriging (KRI) interpolation methods and inverse distance weighting (IDW). These routines were applied to the sample data of selected attributes to define MZs in two agricultural areas. For each dataset 300 different adjustments were tested for the semivariogram. The best parameters were used to measure data by KRI as well as twelve different values for IDW exponent. Area A was considered homogeneous because it did not present statistic different mean between classes. While for area B, the best result was subdivided into two distinct classes. MZs usually present isolated cells or patches, in such a way that crop management becomes difficult. In this context, another goal was to implement routines able of adjusting MZs, to smooth and to improve continuity. Three sampled areas were subdivided into two, three, four and five classes, and then adjusted for five times. Isolated pixels and almost all patches were eliminated, especially the methods with 5×5 mask. So, it can be concluded that the computational routines implemented were efficient and able to identify the best adjustment for the semivariogram as well as the best exponent for IDW, also smoothing and improving MZs continuity.

KEYWORDS: computer systems; precision agriculture; ordinary kriging; smoothing.

SUMÁRIO

LISTA DE FIGURAS	IX
LISTA DE TABELAS.....	XI
LISTA DE ABREVIATURAS	XII
1. INTRODUÇÃO GERAL	1
2. OBJETIVOS.....	3
2.1 Objetivo geral	3
2.2 Objetivos específicos	3
3. REVISÃO BIBLIOGRÁFICA.....	4
3.1 Culturas: soja e milho.....	4
3.2 Atributos do solo	4
3.3 Agricultura de precisão	5
3.4 Análise geoestatística.....	6
3.4.1 Semivariograma.....	6
3.4.2 Parâmetros do semivariograma	7
3.4.3 Modelagem do semivariograma experimental.....	8
3.4.4 Métodos de ajuste do semivariograma.....	10
3.5 Seleção de modelos.....	10
3.5.1 Validação cruzada.....	11
3.5.2 Índice de seleção de interpoladores (ISI)	11
3.6 Métodos de interpolação de dados.....	12
3.6.1 Inverso da distância elevado a uma potência (<i>inverse distance weighting</i> , IDW)	12
3.6.2 Krigagem ordinária.....	13
3.7 Zonas de manejo (ZMs)	14
3.8 Retificação de zonas de manejo	15
3.9 Softwares aplicados na agricultura de precisão.....	16
3.10 Desenvolvimento de software.....	17
4. REFERÊNCIAS BIBLIOGRÁFICAS	18
5. ARTIGO 1 - SELEÇÃO DE MÉTODOS DE INTERPOLAÇÃO DE DADOS.....	23
5.1 Introdução	24
5.2 Material e métodos.....	27
5.2.1 Localização da área, coleta de dados e seleção de atributos.....	27
5.2.2 Análise geoestatística	30
5.2.3 Análise do melhor expoente para IDW.....	31
5.2.4 Interpolação dos dados.....	31

5. 2. 5	Geração e avaliação das zonas de manejo (ZMs).....	32
5.3	Resultados e discussão	34
5. 3. 1	Análise estatística dos dados amostrais e seleção de atributos para gerar ZMs.....	35
5. 3. 2	Seleção do melhor ajuste do semivariograma (análise geoestatística).....	41
5. 3. 3	Seleção do melhor expoente para o método de interpolação inverso da distância elevado a uma potência (IDW)	43
5. 3. 4	Seleção do melhor método de interpolação pelo uso do índice de seleção de interpoladores (ISI).....	44
5. 3. 5	Definição, análise e avaliação das zonas de manejo (ZMs)	49
5.4	Conclusão.....	54
5.5	Agradecimentos	54
5.6	Referências bibliográficas.....	54
6.	ARTIGO 2 - MÉTODOS DE RETIFICAÇÃO PARA OTIMIZAÇÃO DE ZONAS DE MANEJO.....	58
6.1	Introdução	59
6.2	Materiais e métodos.....	60
6.3	Resultados e discussão	64
6.4	Conclusão.....	69
6.5	Agradecimentos	69
6.6	Referências bibliográficas.....	70
7.	CONSIDERAÇÕES FINAIS	73
	APÊNDICES.....	74

LISTA DE FIGURAS

Figura 1 Parâmetros do semivariograma.	7
---	---

ARTIGO 1

Figura 1 Parâmetros do semivariograma.	26
Figura 2 Localização dos pontos amostrais e contorno das áreas.	28
Figura 3 Fluxograma dos procedimentos executados neste trabalho.	35
Figura 4 Semivariogramas e mapas temáticos de atributos da área A gerados com quatro classes, mensurados pelos métodos de interpolação: KRI - krigagem ordinária; IDW ₁ – inverso da distância; IDW ₂ – inverso da distância elevado ao quadrado.	37
Figura 5 Semivariogramas e mapas temáticos de atributos da área B gerados com quatro classes, mensurados pelos métodos de interpolação: KRI - krigagem ordinária; IDW ₁ – inverso da distância; IDW ₂ – inverso da distância elevado ao quadrado.	39
Figura 6 Semivariogramas dos atributos elevação e RSP _{0,1-0,2_2015} (RSP – resistência do solo a penetração) da área A, elevação e RSP _{0,1-0,2_2013} da área B, ajustados por: todos os conjuntos de parâmetros e modelos testados (a); melhores parâmetros para cada um dos modelos testados (b); melhores parâmetros para o melhor modelo testado (c).	41
Figura 7 Índice de seleção de interpolador (ISI) para os 12 modelos de ajuste do semivariograma dos atributos elevação e resistência do solo à penetração (RSP) das áreas A (a) e B (b). Em destaque o melhor modelo de semivariograma correspondente ao menor ISI.	42
Figura 8 Gráfico de valores do índice de seleção de interpoladores (ISI) para escolha do melhor expoente para IDW (inverse distance weighting - inverso da distância elevado a uma potência) para os atributos selecionados para gerar zonas de manejo das áreas A (a) e B (b). Em destaque o melhor expoente correspondente ao menor ISI.	43
Figura 9 Mapas temáticos da área A dos atributos elevação e RSP _{0,1-0,2_2015} (RSP – resistência do solo a penetração) com quatro classes, mensurados pelos métodos de interpolação: KRI - krigagem ordinária (a); IDW ₁ – inverso da distância (b); IDW ₂ – inverso da distância elevado ao quadrado (c); IDW _{B(1,0, elevação; 0,5, RSP)} – inverso da distância elevado ao melhor expoente (d).	46

Figura 10 Mapas temáticos da área B dos atributos elevação e RSP_0,1-0,2_2013 (RSP– resistência do solo a penetração) com quatro classes, mensurados pelos métodos de interpolação: KRI - krigagem ordinária (a); IDW ₁ – inverso da distância (b); IDW ₂ – inverso da distância elevado ao quadrado (c); IDW _{B(5,5, elevação; 3,5, RSP)} – inverso da distância elevado ao melhor expoente (d).	48
Figura 11 Zonas de Manejo definidas para a área A com atributos elevação e RSP_0,1-0,2_2015 (RSP – resistência do solo a penetração), com duas, três e quatro classes utilizando Fuzzy C-Means, mensurados pelos métodos de interpolação: KRI - krigagem ordinária (a); IDW ₁ – inverso da distância (b); IDW ₂ – inverso da distância elevado ao quadrado (c); IDW _{B(1,0, elevação; 0,5, RSP)} – inverso da distância elevado ao melhor expoente (d).	49
Figura 12 Zonas de Manejo definidas para a área B com atributos elevação e RSP_0,1-0,2_2013 (RSP – resistência do solo a penetração), com duas, três e quatro classes utilizando Fuzzy C-Means, mensurados pelos métodos de interpolação: KRI - krigagem ordinária (a); IDW ₁ – inverso da distância (b); IDW ₂ – inverso da distância elevado ao quadrado (c); IDW _{B(5,5, elevação; 3,5, RSP)} – inverso da distância elevado ao melhor expoente (d).	50
Figura 13 Dispersão dos pontos interpolados por KRI - krigagem ordinária (a); IDW ₁ – inverso da distância (b); IDW ₂ – inverso da distância elevado ao quadrado (c); IDW _{B(5,5, elevação; 3,5, RSP)} – inverso da distância elevado ao melhor expoente (d), da área B subdividida em duas classes. Os centroides são destacados em preto.	53

ARTIGO 2

Figura 1 Localização dos pontos amostrais e contorno das áreas.	61
Figura 2 Definição da máscara de retificação: a) 8 pixels vizinhos (máscara 3 × 3); b) 24 pixels vizinhos (máscara 5 × 5).	63
Figura 3 Mapas temáticos das zonas de manejo originais, após a primeira e após a quinta retificação, considerando o método de retificação e o número de classes, para as Áreas A, B, e C.	65
Figura 4 Acréscimo do índice de suavidade após cada retificação, considerando cada método de retificação e o número de classes, para as Áreas A, B, e C.	67
Figura 5 Índice de suavidade (SI) após cada retificação, considerando cada método de retificação e o número de classes, para as Áreas A, B, e C.	68
Figura 6 Índice Kappa do mapa original e mapa retificado, considerando cada método de retificação e o número de classes, para as Áreas A, B, e C.	69

LISTA DE TABELAS

Tabela 1 Classificação do índice de dependência espacial (IDE)	8
--	---

ARTIGO 1

Tabela 1 Atributos coletados em cada ponto amostral.....	28
Tabela 2 Análise estatística dos dados de atributos selecionados para gerar zonas de manejo para as áreas A e B.....	36
Tabela 3 Modelos e parâmetros de ajuste do semivariograma selecionados como melhores para interpolação por krigagem ordinária para os atributos elevação e resistência do solo a penetração (RSP) das áreas A e B	43
Tabela 4 Valores de erro médio (EM), desvio padrão do erro médio (DP_{EM}) e índice de seleção de interpolador (ISI) para cada método de interpolação utilizados para mensurar dados dos atributos selecionados para gerar zonas de manejo nas áreas A e B.....	44
Tabela 5 Índices de avaliação calculados considerando zonas de manejo definidas com valores mensurados pelos métodos de interpolação krigagem ordinária (KRI), inverso da distância (IDW_1), inverso da distância elevado ao quadrado (IDW_2), inverso da distância elevado ao melhor expoente (IDW_B), para as áreas A e B...	51
Tabela 6 Análise da variância (ANOVA) para produtividade e atributos do solo das áreas A e B, considerando duas classes definidas com valores mensurados pelos métodos de interpolação krigagem ordinária (KRI), inverso da distância (IDW_1), inverso da distância elevado ao quadrado (IDW_2) e inverso da distância elevado ao melhor expoente (IDW_B)	52

ARTIGO 2

Tabela 1 Atributos coletados em cada ponto amostral.....	62
--	----

LISTA DE ABREVIATURAS

AP	Agricultura de precisão
DP _{EM}	Desvio-padrão dos erros médios
EM	Erro médio
FPI	Índice de performance Fuzzy (Fuzziness performance index)
GPS	Sistema de posicionamento global (global position system)
ICVI	Índice de validação de cluster (Improved Cluster Validation Index)
IDE	Índice de dependência espacial
IDW	Inverso da distância elevado a uma potência (inverse distance weighting)
IDW ₁	Inverso da distância
IDW ₂	Inverso do quadrado da distância
IDW _B	Inverso da distância elevado ao melhor expoente
ISI	Índice de seleção de interpoladores
JVM	Máquina virtual java (java virtual machine)
KRI	Krigagem ordinária
ML	Máxima verossimilhança (maximum likelihood)
MPE	Entropia de partição modificada (modified partition entropy)
OLS	Quadrados mínimos ordinários (ordinary least squares)
RSP	Resistência do solo a penetração
SGBD	Sistema gerenciador de banco de dados
SI	Índice de suavidade (smooth index)
SIG	Sistemas de informações geográficas
SDUM	Software para Definição de Unidades de Manejo
VR	Redução da variância (variance reduction)
WLS	Quadrados mínimos ponderados pelo número de pares (weight least squares)
ZM	Zona de manejo

1. INTRODUÇÃO GERAL

A evolução que ocorre em todas as áreas do conhecimento também é evidente no âmbito da agricultura. Máquinas e implementos tecnologicamente equipados são utilizados no decorrer dos procedimentos agrícolas, e para que os benefícios propostos sejam usufruídos da melhor maneira, operadores necessitam conhecimento técnico e capacitação tanto na parte operacional como de manutenção. Estes equipamentos podem possuir diversos sensores, e principalmente receptores de GPS para a identificação exata da localização, para que o manejo seja realizado com precisão. Além dos equipamentos utilizados em campo, também é importante a necessidade de conhecimento em áreas diversas como a informática. Muitos aplicativos computacionais (softwares) são utilizados nos procedimentos relativos a uma safra agrícola, desde a análise do manejo do solo antes do plantio até a colheita. A preocupação com questões ambientais e financeiras exige a necessidade de utilização destes artefatos tecnológicos.

No período de uma safra agrícola, coleta-se uma elevada densidade de dados referente a diversos atributos do solo e da planta, então, os sistemas computacionais são utilizados para armazenar, manipular, avaliar, visualizar informações espaciais sob forma de mapas digitais. Estes sistemas, chamados de sistemas de informações geográficas (SIGs), são utilizados para proporcionar melhor interpretação dos dados armazenados a fim de auxiliar e agilizar a tomada de decisão sobre o uso e manejo do solo em áreas agrícolas. Porém, muitos destes softwares são proprietários ou fazem parte de um conjunto fechado de hardware e software, os quais limitam a própria utilização.

Segundo Gontijo et al. (2012), a limitação da produção nas lavouras está relacionada à baixa fertilidade natural do solo, causada por fatores ambientais ou antrópicos (resultado da atuação humana), que fazem com que haja variabilidade espacial dos atributos químicos do solo. Isto pode causar deficiência de nutrientes, logo são necessárias práticas adequadas de correção e manejo para minimizar as limitações (BURAK; PASSOS; ANDRADE, 2012). Geralmente, softwares são utilizados para auxiliar na determinação dessas práticas.

Devido à complexidade dos fatores que envolvem a composição do solo, é necessária a definição de áreas específicas de manejo, para a aplicação racional de adubos e corretivos, a fim de contemplar o contexto de manejo sustentável dos recursos agrícolas (CAMBARDELLA et al., 1994). É possível a redução da heterogeneidade e consequentemente da variabilidade espacial dos fatores relacionados à fertilidade do solo a partir da correção e do manejo localizado do solo com potencial aumento da produtividade (BURAK; PASSOS; ANDRADE, 2012). Para isso, são necessários planejamento e gerenciamento das áreas agrícolas que envolvem o conhecimento simultâneo de atributos do solo, do relevo e do rendimento dos talhões.

A amostragem georreferenciada dos atributos do solo e o monitoramento da produtividade são essenciais para obtenção de informações sobre a área. Softwares fazem uso dos dados amostrados para estimar e mapear a variabilidade espacial da produtividade e dos atributos do solo em locais não amostrados por interpolação de dados, a fim de que se realize o manejo agrícola diferenciado em áreas heterogêneas. Desta maneira, é possível definir zonas de manejo, ou seja, sub-regiões do campo que apresentam combinação de fatores limitantes de produtividade para a qual se pode aplicar dose uniforme de insumos.

Porém, esta área é carente de sistemas computacionais capazes de realizar todos os procedimentos relativos à definição e avaliação de zonas de manejo. Devido a esta necessidade, foi desenvolvido o Software para Definição de Unidades de Manejo (SDUM), o qual proporcionou o surgimento de outros projetos complementares, sendo um deles alvo deste estudo. O SDUM é uma aplicação computacional desenvolvida com ferramentas livres e em linguagem multiplataforma (Java), na qual é possível realizar procedimentos para definição e avaliação de zonas de manejo. Porém, possui algumas limitações como estar restrito a definir zonas de manejo baseadas em dados mensurados por interpoladores determinísticos.

Assim, foram desenvolvidas rotinas computacionais que irão contribuir para o aprimoramento do SDUM e que estão contextualizadas e organizadas nesta tese em forma de artigos científicos apresentados após uma revisão bibliográfica (Capítulos 3 e 4) sobre assuntos abordados no decorrer do estudo. O primeiro artigo (Capítulo 5) apresenta os resultados obtidos a partir de procedimentos computacionais implementados relativos à retificação de zonas de manejo por estatística moda e mediana, em três áreas agrícolas. O segundo artigo científico (Capítulo 6) apresenta os resultados obtidos a partir de procedimentos computacionais implementados que, além de automatizar a seleção do melhor expoente para o método de interpolação inverso da distância, também realiza a análise geoestatística e a interpolação por krigagem ordinária dos dados de duas áreas agrícolas. E para finalizar, as considerações finais do trabalho estão apresentadas no Capítulo 7.

2. OBJETIVOS

2.1 Objetivo geral

Desenvolver três módulos computacionais a serem incorporados ao software SDUM (Software para Definição de Unidades de Manejo): módulo de análise geoestatística e interpolação por krigagem ordinária; módulo de seleção do melhor expoente para o método de interpolação inverso da distância elevado a uma potência; e módulo de retificação de zonas de manejo.

2.2 Objetivos específicos

- Desenvolver rotinas computacionais implementadas em geoR e diretamente no banco de dados, capazes de realizar análise estatística e espacial dos dados, contemplando a análise geoestatística, seleção automática do melhor ajuste para o modelo do semivariograma e interpolação por krigagem ordinária;
- Desenvolver rotinas computacionais implementadas em geoR e diretamente no banco de dados, capazes de identificar por validação cruzada o melhor expoente para o método de interpolação inverso da distância elevado a uma potência;
- Implementar funções computacionais diretamente no banco de dados, capazes de realizar a retificação das zonas de manejo para adequação ao manejo mecanizado;
- Avaliar os módulos computacionais desenvolvidos pela análise dos dados de atributos de áreas agrícolas nos municípios de Serranópolis do Iguaçu e Céu Azul-PR.

3. REVISÃO BIBLIOGRÁFICA

3.1 Culturas: soja e milho

Soja (*Glycine max*) e milho (*Zea mays*) são duas das principais culturas agrícolas mundiais e juntas correspondem a quase 88% da produção de grãos no Brasil. O cultivo de soja é o que mais cresceu nos últimos 30 anos no Brasil, e corresponde a 49% da área plantada em grãos no País. Este acréscimo está relacionado a avanços nas pesquisas, nas tecnologias utilizadas, no manejo, na eficiência dos produtores, dentre outros fatores (MAPA, 2016).

Na safra 2015/2016, foram produzidas 312 milhões de toneladas de soja em 120 milhões de hectares no Planeta. O Brasil foi o segundo maior produtor com 95,4 milhões de toneladas em 33,2 milhões de hectares, média de 2,87 t ha⁻¹. O Estado do Paraná foi o segundo maior produtor brasileiro com 17,1 milhões de toneladas em 5,4 milhões de hectares, média de 3,14 t ha⁻¹ (CONAB, 2016).

A soja é um grão essencial para a fabricação de rações animais bem como de produtos para alimentação humana, porém, seu maior consumo como óleo refinado. Além dessas, a soja é utilizada como fonte alternativa de combustível, o biodiesel (EMBRAPA, 2016).

O cultivo do milho representa 38,8% das áreas plantadas com grãos no Brasil e é caracterizado, em algumas regiões do País, pelo plantio em duas épocas: primeira safra (ou safra de verão) e segunda safra (ou safrinha). Na safra 2015/2016, foram produzidas 346 milhões de toneladas de milho no Planeta. O Brasil foi o segundo maior produtor com 67,0 milhões de toneladas em 15,9 milhões de hectares, média de 4,21 t ha⁻¹. O Estado do Paraná foi o terceiro maior produtor brasileiro com 14,5 milhões de toneladas em 2,6 milhões de hectares, média de 6,52 ha⁻¹ (CONAB, 2016).

As indústrias de rações para animais são o principal destino da safra de milho com cerca de 70% da produção. Devido às suas qualidades nutricionais, o grão também é transformado em alimentação humana em forma de óleo, farinha, amido, margarina, xarope de glicose, flocos para cereais e outros (FIESP, 2016).

3.2 Atributos do solo

Os atributos químicos do solo sofrem variação ao longo do tempo ou até mesmo ao longo de uma safra, por isso, não são considerados estáveis, portanto, não são indicados para o estudo de ZMs. Por outro lado, a textura do solo tem pouca variabilidade, por conseguinte, é uma característica considerada estável (BUTTAUFUOCO et al., 2010; DOERGE, 2000). É importante citar que é possível determinar a proporção de areia, silte e argila, presentes no solo, a partir da análise granulométrica, fator que influencia na retenção e disponibilidade de

água às plantas (KLEIN; KLEIN, 2015). Por isso, realiza-se esta análise a fim de se identificar qual o nível de correlação espacial destes atributos com a produtividade.

A resistência do solo à penetração (RSP) é considerada um dos fatores limitante ao desenvolvimento da cultura, pois existe uma relação inversa entre o teor de água do solo e a RSP (RAWLS et al., 2003). Isso pode causar efeitos prejudiciais ao sistema radicular, e por isso tem sido utilizada como indicador de compactação do solo em diversos sistemas de produção.

Os aspectos topográficos do terreno podem ser bons indicadores da variação dos atributos do solo haja vista tal variabilidade ser causada por pequenas alterações do declive que influenciam o transporte e o armazenamento de água dentro do perfil do solo (BLANCO-CANQUI et al., 2015; CASTIONE et al., 2015). Por isso, os atributos físicos do solo juntamente com o levantamento topográfico das áreas são importantes variáveis para o estudo de ZMs.

3.3 Agricultura de precisão

O desenvolvimento natural das plantas ocorre por absorção de nutrientes e pode causar a variabilidade desses atributos no solo. Assim, é necessário realizar o manejo adequado para que a reposição dos nutrientes seja realizada em níveis corretos. A variabilidade espacial que ocorre na produtividade agrícola, e nos atributos físicos e químicos do solo, pode ser gerenciada pelo sistema de agricultura de precisão (AP). O AP é composto por tecnologias e procedimentos utilizados para a melhoria dos campos de cultivo e consiste no manejo localizado das variáveis de produção (COLAÇO et al., 2012; MONDO et al., 2012). E tem como propósito o desenvolvimento de atividades que permitem racionalizar o uso de insumos e, conseqüentemente, os custos de produção e o impacto ambiental, aliado ao incremento da produção, tornando-se assim um sistema de gestão da produção agrícola (MOLIN; RABELLO, 2011).

Segundo Doerge (2000), agronomicamente, faz sentido a aplicação de fertilizantes à taxa variada em campos heterogêneos e, para isso, é necessário estimar e mapear a variabilidade espacial dos atributos do solo e produtividades. Resultados de análises laboratoriais de atributos físicos e químicos do solo bem como dados de produtividade obtidos por amostras georreferenciadas são interpolados a fim de que se estimem valores para locais não amostrados e assim demonstrar o comportamento da variável a partir de um mapa. Logo, o método de interpolação é fator importante para a precisão com que os mapas são produzidos, pois podem influenciar os valores estimados e, conseqüentemente, a análise da variação espacial (COUTO; SCARAMUZZA; MARASCHINI, 2002; LUCAS et al., 2013; SAMIRA; AHMED; LHOUSSEINE, 2014; SCHENATTO et al., 2016).

3.4 Análise geoestatística

A geoestatística é baseada na teoria das variáveis regionalizadas, a qual foi formalizada por Matheron (1963) e faz uso de procedimentos que podem ser utilizados para analisar e modelar as relações entre determinadas características de fenômenos espaciais.

A análise geoestatística é utilizada para estudar a variabilidade espacial de atributos do solo e planta, considerando a localização geográfica dos pontos amostrais. Assim, não são considerados somente os valores obtidos para uma variável-resposta, mas também sua posição expressa por um sistema de coordenadas (CRESSIE, 1993).

Segundo Borssoi et al. (2011), a análise geoestatística é um método que considera a distribuição espacial das medidas. Portanto, permite determinar o raio de autocorrelação espacial entre as amostras e, conseqüentemente, a distância máxima até a qual as amostras são consideradas espacialmente dependentes. São calculadas semivariâncias pelo estimador clássico proposto por Matheron (1963) e plotadas em um gráfico chamado semivariograma que precisa ser ajustado por modelos matemáticos a fim de se representar da melhor maneira a distribuição das semivariâncias. Os parâmetros obtidos por este processo são utilizados como base para a interpolação por krigagem.

3.4.1 Semivariograma

A semivariância é calculada para cada par de observações cuja distância entre elas esteja compreendida dentro de um intervalo de distância (*lag*), e este processo é repetido para cada *lag*. Para pontos próximos (mais correlacionados), a diferença é pequena, portanto, a variância é pequena. Assim, os dados possuem comportamento mais semelhante entre si, ou seja, menor variabilidade (MELLO et al., 2005). Ao se aumentar a distância, os valores dos pontos tornam-se mais diferentes, conseqüentemente, a variância aumenta e tende a estabilizar-se em torno de uma variância máxima, a partir de certa distância (ISAACS; SRIVASTAVA, 1989). Assim, a ocorrência de correlação espacial entre as amostras remete a valores mais próximos em pontos vizinhos, diferenciando-se gradativamente à medida que os pontos são mais distantes, até tornarem-se independentes. Para grandes distâncias, há uma tendência pela redução do número de pares, tornando-se estatisticamente não significativos (CLARK, 1979).

Portanto, a estimação da dependência espacial considera os cálculos das semivariâncias amostrais e é representada pelo semivariograma. Logo, é possível se verificar e modelar o nível de dependência espacial entre amostras ao longo de um domínio, em forma de gráfico (WOODCOCK; STRAHLER; JUPP, 1988).

3. 4. 2 Parâmetros do semivariograma

O semivariograma (Figura 1) é composto de importantes parâmetros envolvidos no modelo, conhecidos como efeito pepita (C_0), contribuição (C_1), alcance (α) e patamar (C_0+C_1) (CAMPOS et al., 2013; ISAAKS; SRIVASTAVA, 1989). Segundo Druck et al. (2004), o semivariograma é um gráfico, no qual a função semivariância dos dados cresce em função da distância (h) até determinado ponto onde não se observa mais a continuidade espacial. Esse ponto no eixo x (h) é denominado de alcance (α) do semivariograma (*range*), ou seja, representa a distância observada até onde a variabilidade se estabiliza. No eixo y, este ponto é denominado patamar ($C_0 + C_1$) e, deste ponto em diante, considera-se que não existe mais dependência espacial entre as amostras, pois a variância da diferença entre pares de amostras torna-se aproximadamente constante (CLARK, 1979; WANDERLEY; AMORIM; CARVALHO, 2013).

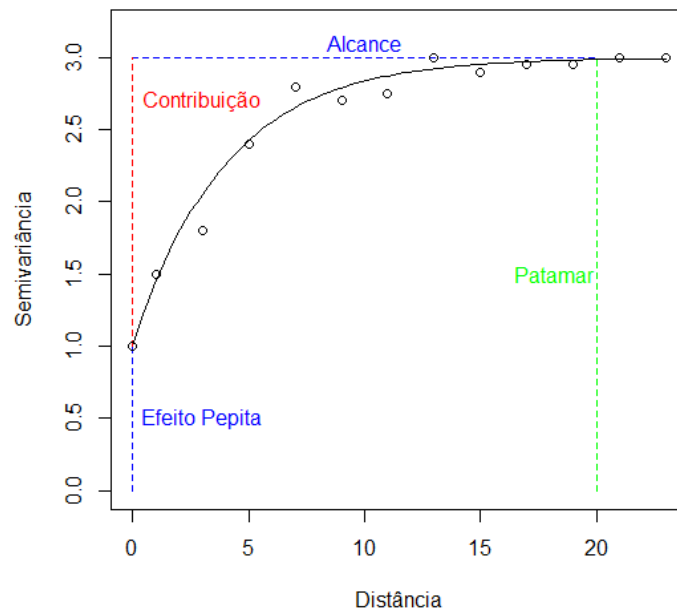


Figura 1 Parâmetros do semivariograma.

Alcance é a distância em que se atinge o patamar e a partir da qual as amostras passam a ser não correlacionadas. O efeito pepita (C_0) é o valor da semivariância para a distância zero e representa o componente da variação ao acaso, ou seja, a variabilidade para escalas menores que a distância entre os pontos amostrais. Segundo Druck et al. (2004), o parâmetro do semivariograma representa as variações locais de pequena escala, tais como erros de medição. O C_0 corresponde ao ponto onde o semivariograma toca o eixo das ordenadas. Esse ponto revela a descontinuidade do semivariograma para distâncias menores do que a menor distância entre os pontos amostrais.

A contribuição (C_1), também conhecida como variância da dispersão, representa as diferenças espaciais entre os valores de C_0 e o patamar, intervalo esse no qual o semivariograma cresce, e que existe dependência espacial (CRESSIE, 1993).

Depois de definidos, os parâmetros do semivariograma são avaliados pelo índice de dependência espacial (IDE), que corresponde à relação percentual entre o efeito pepita e o patamar (Equação 1), classificado por Cambardella et al. (1994) conforme a Tabela 1.

$$IDE = \frac{C_0}{C_0 + C_1} \times 100, \quad \text{Eq. (1)}$$

em que, IDE – índice de dependência espacial; C_0 – efeito pepita; C_1 – contribuição.

Tabela 1 Classificação do índice de dependência espacial (IDE)

IDE	Dependência espacial
até 25%	forte
de 25% a 75%	moderada
acima de 75%	fraca

Fonte: Cambardella et al. (1994).

3. 4. 3 Modelagem do semivariograma experimental

A qualidade da estimativa de valores para pontos não amostrados obtidos por krigagem depende do ajuste matemático do modelo experimental ao semivariograma teórico. Portanto, a modelagem da estrutura de dependência espacial pela abordagem da geoestatística é de fundamental importância para a definição de parâmetros que são utilizados na interpolação de valores em locais não amostrados, pela técnica de krigagem (BORSSOI; URIBE-OPAZO; GALEA, 2011; CRESSIE, 1993; GRZEGOZEWSKI et al., 2013; GONTIJO et al., 2012). Autores como Cressie (1993); Diggle e Ribeiro Jr. (2007); Isaaks e Srivastava (1989) destacam a importância e a dificuldade na escolha de um modelo teórico adequado e do ajuste ao semivariograma experimental, por isso, é um processo que necessita ser realizado muitas vezes, com diferentes ajustes.

O modelo de ajuste do semivariograma experimental que representar da melhor maneira a distribuição das semivariâncias deverá ser utilizado como base para a interpolação por krigagem. Embora existam vários modelos de semivariogramas teóricos, apenas alguns são considerados como os mais comuns na explicação da maioria dos fenômenos espaciais. São eles: esférico, exponencial, gaussiano e Família Matérn.

- a) Esférico (Equação 2): O patamar e o alcance são facilmente identificados e, geralmente, o efeito pepita é pequeno em relação ao patamar. Apresenta comportamento linear em sua origem e estrutura de correlação espacial que aumenta com a distância até certo ponto (alcance), a partir do qual, a semivariância torna-se constante, limitando a área de influência de cada amostra. O modelo esférico é linear até aproximadamente 1/3 do alcance, e o patamar é atingido no valor do alcance, o qual é igual a 3/2 da distância.

$$\begin{cases} Y(h) = C_0 + C_1 \left[1,5 \frac{h}{a} - 0,5 \left(\frac{h}{a} \right)^3 \right], & \text{para } h < a \\ Y(h) = C_0 + C_1, & \text{para } h \geq a. \end{cases} \quad \text{Eq. (2)}$$

- b) Modelo exponencial (Equação 3): é assintótico e tem comportamento linear em sua origem. Apresenta crescimento menos acentuado da origem para o patamar, e o modelo não atinge realmente o patamar. O modelo exponencial tende a representar melhor a variabilidade em pequena escala por apresentar maiores ordenadas do semivariograma para pequenas distâncias. O alcance prático é definido como a distância na qual o valor do modelo atinge aproximadamente 95% do patamar.

$$\begin{cases} Y(h) = 0, & \text{para } h = 0 \\ Y(h) = C_0 + C_1 \left[1 - \exp\left(-\frac{h}{a}\right) \right], & \text{para } h > 0. \end{cases} \quad \text{Eq. (3)}$$

- c) Modelo Gaussiano (Equação 4): é transitivo, muitas vezes usado para modelar fenômenos extremamente contínuos. Tende a representar melhor a variabilidade em grande escala e apresenta boa continuidade na variabilidade à medida que os pontos se afastam entre si. Também tende assintoticamente para o patamar, cuja amplitude é definida pela distância onde se atinge aproximadamente 95% do patamar, e o alcance prático é dado por $\sqrt{3}$ da distância. Segundo Cressie (1993), este modelo tem comportamento parabólico junto da origem e apresenta um ponto de inflexão para uma distância ligeiramente superior, e tem como expressão:

$$\begin{cases} Y(h) = 0, & \text{para } h = 0 \\ Y(h) = C_0 + C_1 \left[1 - \exp\left(-\left(\frac{h}{a}\right)^2\right) \right], & \text{para } h > 0. \end{cases} \quad \text{Eq. (4)}$$

- d) Família Matérn (Equação 5): a família dos modelos de Matérn é uma generalização de outros modelos teóricos e tem como característica fundamental a inclusão de um parâmetro (κ) que determina a suavização (MINASNY; MCBRATNEY, 2005). Para $\kappa = 0,5$, o modelo Matérn reduz-se ao modelo exponencial e, para κ tendendo a infinito, equivale ao modelo gaussiano (URIBE-OPAZO; BORSSOI; GALEA, 2012). Este parâmetro permite maior flexibilidade para caracterizar a variabilidade do semivariograma experimental, pois estabelece equilíbrio entre os dois extremos.

$$\begin{cases} Y(h) = 0, & \text{para } h = 0 \\ Y(h) = C_0 + C_1 \left[1 - \frac{2}{\Gamma(\kappa)} \left(\frac{h\sqrt{\kappa}}{a} \right)^\kappa B_\kappa \left(\frac{2h\sqrt{\kappa}}{a} \right) \right], & \text{para } h > 0, \end{cases} \quad \text{Eq. (5)}$$

em que B é a função Bessel, $\Gamma(\kappa)$ é a função Gama e κ é o parâmetro de suavização.

3. 4. 4 Métodos de ajuste do semivariograma

A sensibilidade dos semivariogramas para detectar a variabilidade espacial das amostras está diretamente relacionada ao melhor ajuste entre os dados amostrais e os modelos teóricos do semivariograma. Escolher o modelo adequado é obter os melhores estimadores dos parâmetros efeito pepita (C_0), patamar ($C_0 + C_1$) e alcance (a). Segundo Mello et al. (2005), os métodos utilizados para identificar os melhores ajustes dos modelos de semivariograma podem ser classificados em dois grupos. O primeiro grupo é o de ajuste dos modelos ao semivariograma experimental e é realizado pelo método dos quadrados mínimos ordinários (*ordinary least squares*, OLS) e do método dos quadrados mínimos ponderados (*weight least squares*, WLS). O segundo grupo é o método de ajuste de um modelo utilizando a distribuição da probabilidade dos dados, pelo método da máxima verossimilhança (*maximum likelihood*, ML).

O método OLS consiste em estimar o melhor conjunto de parâmetros por um modelo que minimiza a soma do quadrado da diferença entre os valores observados e os estimados pelo modelo teórico do semivariograma. WLS é semelhante ao OLS, porém utiliza um ponderador na estimativa dos parâmetros do semivariograma teórico. Os melhores valores para os parâmetros são os que minimizam a soma do quadrado das diferenças entre os valores experimentais e os preditos pelo modelo teórico do semivariograma, cuja diferença é ponderada diretamente pelo inverso da variância do semivariograma experimental (DIGGLE; RIBEIRO JR, 2007):

Estimar parâmetros pelo método ML pressupõe estacionaridade forte. Em condições regulares (amostra de dados suficientemente grande), este estimador é não viesado e eficiente, porém, pode requerer muito tempo computacional, pois a matriz de variância e covariância geradas podem possuir dimensões muito grandes (SILVA; RIBEIRO JR; ELMATZOGLOU, 2007). O estimador ML tende a ser um estimador não viciado de variância mínima (SHANKEN; ZHOU, 2007), no entanto, OLS e WLS são bastante confiáveis.

3.5 Seleção de modelos

A comparação do conjunto de parâmetros obtidos por modelos teóricos do semivariograma é utilizada para avaliar qual se ajusta melhor aos dados amostrais. Os erros

ou resíduos desta estimação são calculados e sua distribuição avaliada. Segundo Faraco et al. (2008), os principais critérios utilizados para avaliação do melhor modelo correspondem à validação cruzada, Jackknifing, Filliben e Akaike, os quais são métodos de verificação dos dados estimados. Os autores avaliaram os métodos e consideraram a validação cruzada, o mais adequado para a escolha do melhor ajuste do modelo de variabilidade espacial.

3. 5. 1 Validação cruzada

Usada para encontrar o melhor ajuste do modelo matemático para o semivariograma, a validação cruzada consiste em uma técnica de avaliação de erros de estimativa que permite comparar valores estimados e amostrados, usando somente a informação disponível na amostra de dados (ISAAKS; SRIVASTAVA, 1989).

O processo de validação cruzada pode ser utilizado para verificar qual dentre os modelos de semivariograma testados descreveu melhor o comportamento espacial das propriedades estudadas. Um valor de cada vez é retirado do conjunto de dados e faz-se a estimação para este ponto amostral a partir das observações remanescentes (FARACO et al., 2008). Repete-se o processo para todo o conjunto e, assim, para cada observação, é possível calcular o erro de estimação que poderá ser padronizado pelo desvio padrão da estimação. Obtêm-se como resultado o erro médio - EM (Equação 6), o desvio-padrão dos erros médios - DP_{EM} (Equação 7).

$$EM = \frac{1}{n} \sum_{i=1}^n Z(s_i) - \hat{Z}(s_i), \quad \text{Eq. (6)}$$

$$DP_{EM} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Z(s_i) - \hat{Z}(s_i))^2}, \quad \text{Eq. (7)}$$

em que, EM – erro médio; DP_{EM} – desvio padrão dos erros médios; n – número de dados; $Z(s_i)$ – valor observado no ponto s_i ; $\hat{Z}(s_i)$ – valor predito por krigagem no ponto s_i .

3. 5. 2 Índice de seleção de interpoladores (ISI)

A fim de aperfeiçoar a seleção do melhor modelo de ajuste do semivariograma, Bier (2015) propôs o índice de seleção de interpoladores – ISI (Equação 8). Tal índice considera informações do EM e DP_{EM} obtidas por validação cruzada em seu cálculo, em que é considerado o melhor modelo aquele que apresentar menor ISI.

$$ISI = \left\{ \frac{abs(EM)}{\max_{i=1}^J [abs(EM)]} + \frac{[DP_{EM} - \min_{i=1}^J (DP_{EM})]}{\max_{i=1}^J [abs(DP_{EM})]} \right\}, \quad \text{Eq. (8)}$$

em que, $abs(EM)$ – valor em módulo do erro médio da validação cruzada; $\min|_{i=1}^j$ – menor valor encontrado entre os j modelos comparados; $\max|_{i=1}^j$ – maior valor encontrado entre os j modelos comparados.

3.6 Métodos de interpolação de dados

A necessidade de se criar uma grade densa e regular para mapear a variabilidade espacial dos atributos do solo e produtividade remete à utilização de métodos de interpolação para que se estimem valores para locais não amostrados.

A interpolação espacial converte dados de observações pontuais em campos contínuos e gera mapas que devem representar a realidade da distribuição do atributo estudado da melhor forma possível. Os métodos de interpolação comumente utilizados em conjuntos de dados relacionados à AP são: vizinho mais próximo, inverso da distância elevado a uma potência (*inverse distance weighting*, IDW) e krigagem (KESTRING et al., 2015; SOUZA et al., 2016). A diferença entre os métodos de interpolação é a maneira como os pesos são atribuídos às diferentes amostras (DRUCK et al., 2004).

O método vizinho mais próximo estima o valor para determinado local como sendo o valor do ponto amostral mais próximo. Este método leva em consideração a escolha de apenas uma amostra vizinha para cada ponto da grade. É utilizado quando se deseja manter os valores das amostras na grade sem gerar valores intermediários (DRUCK et al., 2004).

3. 6. 1 Inverso da distância elevado a uma potência (*inverse distance weighting*, IDW)

Na interpolação usando IDW (Equação 9), os pesos são definidos como o inverso da distância elevado a uma potência, em que a distância separa o valor interpolado dos valores observados. Este método faz com que os pesos dos dados sejam avaliados durante o processo de interpolação e a influência de cada ponto amostrado é inversamente proporcional à distância do ponto a ser estimado.

$$\hat{Z}_i = \frac{\sum_{i=1}^n \left(\frac{1}{d_i^p} * Z_i \right)}{\sum_{i=1}^n \left(\frac{1}{d_i^p} \right)}, \quad \text{Eq. (9)}$$

em que, $\hat{Z}_i$ – valor interpolado; Z_i – valor do atributo amostrado; d_i^p – distância euclidiana entre o i -ésimo ponto de vizinhança e o ponto amostrado, elevado a uma potência; $p > 0$.

É um processo puramente matemático, no qual os dados são ponderados de tal forma que a influência entre eles diminui conforme a distância aumenta. O fator peso é pré-determinado pelo valor da potência escolhida, ou seja, quanto maior é esse valor, menor é a

influência dos pontos mais distantes. Este é um método rápido e que requer pouco custo computacional (MAZZINI; SCHETTINI, 2009).

O método IDW possui dois casos mais conhecidos: o inverso da distância (IDW_1) e o inverso do quadrado da distância (IDW_2). A diferença entre eles é o expoente ao qual o cálculo é submetido. Tanto IDW_1 como IDW_2 são métodos considerados simples e de fácil aplicação. No entanto, a acurácia obtida por esses métodos é menor do que pela krigagem, uma vez que não consideram o padrão da estrutura da dependência espacial dos dados amostrais para estimar valores para locais não amostrados. Entretanto, Couto et al. (2002) e Liu et al. (2016) relatam que tanto krigagem como IDW_2 e IDW_1 apresentam destaque de utilização na inferência de dados.

3. 6. 2 Krigagem ordinária

A qualidade do mapa depende da qualidade das inferências dos modelos ajustados para a interpolação por krigagem, pois os parâmetros encontrados são utilizados para definir os melhores pesos para as amostras e estimar valores para os locais não amostrados (MAZZINI; SCHETTINI, 2009).

A krigagem pondera os valores dos pontos vizinhos ao que será estimado, considera os critérios de não tendenciosidade e mínima variância do erro de estimativa, considera a estrutura de variabilidade espacial encontrada para o atributo, e isto proporciona precisão local (GRZEGOZEWSKI et al., 2013; ISAAKS; SRIVASTAVA, 1989). Porém, as estimativas são suavizadas e fazem com que os valores de máximo da distribuição estatística dos dados tenham a tendência de serem subestimados, enquanto os valores de mínimo podem ser superestimados (GRZEGOZEWSKI et al., 2013; KESTRING et al., 2015; ROCHA; LOURENÇO; LEITE, 2007; URIBE-OPAZO; BORSSOI; GALEA, 2012).

Krigagem baseia-se na ideia de que quanto maior a covariância entre uma amostra e o local que está sendo estimado, mais a amostra deve contribuir para a estimativa, e é definida pela Equação 10.

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i * Z(x_i), \quad \text{Eq. (10)}$$

em que, $\hat{Z}(x_0)$ – valor estimado em determinado local; λ_i – peso atribuído aos valores amostrais; $Z(x_i)$ – valor do atributo amostrado; n – número de localidades vizinhas empregadas para a interpolação do ponto, em que o somatório dos pesos λ_i deve ser igual a um.

A eficiência da krigagem se deve ao fato de se calcular os erros associados aos resultados obtidos por um modelo contínuo de variação espacial. Este método de interpolação é considerado flexível e robusto e procura expressar as tendências direcionais que os dados

sugerem. Assim, evita-se o efeito de “olho-de-touro”, resultado de interpoladores como o IDW (CRESSIE, 1993).

3.7 Zonas de manejo (ZMs)

A efetiva utilização da AP necessita do desenvolvimento de equipamentos com sensores (medidores de clorofila, resistência à penetração, condutividade elétrica, índice de vegetação, além de monitores de produtividade) a fim de possibilitar a análise detalhada do solo e plantas, e assim, melhorar o processo de produção. Os resultados obtidos por estas análises possibilitam aplicar fertilizantes e defensivos a taxas variadas (BAZZI et al., 2013).

Porém, segundo Khosla et al. (2008), a questão econômica é um empecilho para a viabilidade da aplicação da AP. Assim, é necessário determinar a variabilidade espacial dos atributos de forma menos onerosa para a efetiva aplicação a taxa variada. Uma alternativa é dividir os talhões em unidades homogêneas menores (zonas de manejo) com o objetivo de aplicar insumos agrícolas com taxa constante, mas de maneira mais racional (DOERGE, 2000).

Zonas de manejo (ZMs) podem ser definidas como uma sub-região do campo que expressam combinação homogênea de fatores limitantes da produtividade. Desta maneira, a aplicação de insumos à taxa constante na ZM e variada entre elas tem como objetivo reduzir a heterogeneidade e conseqüentemente a variabilidade espacial dos fatores relacionados à fertilidade do solo, potencializando aumento da produtividade e redução do impacto ambiental (BOBRYK et al., 2016; DOERGE, 2000; MORAL; TERRÓN; SILVA, 2010; MOSHIA et al., 2014)

Neste contexto, a definição de ZMs é uma alternativa para se viabilizar economicamente a AP, pois facilita a aplicação à taxa variada e torna-se um indicador para amostragem do solo e da cultura (RODRIGUES JR et al., 2011; SUSZEK et al., 2011). É vantajoso definir as ZMs a partir de um conjunto de dados multivariados, que incluem propriedades que podem influenciar a produtividade e não variam significativamente ao longo do tempo (topografia, propriedades físicas do solo) (DOERGE, 2000).

Várias abordagens têm sido desenvolvidas para se definir uma classe, ou seja, um grupo de indivíduos que se assemelham mais fortemente (GUASTAFERRO et al., 2010). Xiang et al. (2007) sugerem a divisão das técnicas utilizadas para definição das zonas de manejo, considerando duas abordagens: a primeira é o método empírico, considerado mais simples, que usa a frequência de distribuição da produtividade (normalmente relativa) e conhecimento especializado, para dividir o talhão usualmente em três ou quatro ZMs (SUSZEK et al., 2011); a segunda corresponde ao método de análise de agrupamento (*cluster analysis*) como K-Means e Fuzzy C-Means (RODRIGUES JR et al., 2011; TAYLOR;

MCBRATNEY; WHELAN, 2007). Os métodos de agrupamento permitem maior grau de diferenciação entre classes, porém, são mais complexos.

Suszek et al. (2011) obtiveram resultados satisfatórios quanto à viabilização da AP utilizando técnicas empíricas (rendimento normalizado e padronizado) e dados de produtividade para definição de ZMs. Arno et al. (2011), Iliadis et al. (2010) e Tagarakis et al. (2013) recomendam a utilização de métodos de agrupamento para a definição de ZMs, os quais analisam, além da produtividade, outros atributos (condutividade elétrica, inclinação e textura do solo, nitrogênio). Portanto, o ideal é utilizar informação espacial de atributos estáveis temporalmente que sejam correlacionados com a produtividade.

3.8 Retificação de zonas de manejo

Frequentemente os mapas apresentam muitas células isoladas ou pequenas manchas dentro das classes, independente do método utilizado para interpolar os dados ou da técnica utilizada para definir as ZMs. Isso dificulta a mecanização para a aplicação de fertilizantes e defensivos à taxa variada. Muitas soluções foram propostas a fim de que se amenize este problema. Lowrance (2014) desenvolveu o software EZZone para suavizar ZMs pela fusão de pequenos polígonos com polígonos maiores de diferentes zonas. Outros pesquisadores, tais como Pramanik et al. (2013), propuseram a incorporação das manchas às zonas maiores pela atribuição de peso para pixels, isto é, a partir da análise da assimetria de pixels vizinhos. Também para a eliminação de células isoladas ou manchas, Xiang et al. (2007) utilizaram filtragem pela maior quantidade de valores iguais entre os vizinhos (moda). De acordo com Córdoba et al. (2016) e Dobermann et al. (2003), é recomendado o uso de filtros espaciais aplicados aos resultados para melhorar a continuidade do zoneamento. O filtro de mediana que substitui o valor do pixel pela mediana dos valores da vizinhança daquele pixel foi utilizado por Córdoba et al. (2016) na execução do protocolo por eles proposto.

Além dos diferentes filtros utilizados para retificar ZMs, a máscara de pixels também é fator relevante. A máscara é responsável por identificar a quantidade de pixels que serão considerados na filtragem. Por exemplo, o valor de três vizinhos na horizontal por três na vertical é considerado para a filtragem na máscara 3x3; todavia, para a máscara 5x5, são considerados cinco vizinhos na horizontal por cinco na vertical, e assim sucessivamente.

Para a aplicação da AP, quanto menor a variância dentro da zona de manejo, melhor. Por isso, é importante a retificação com o propósito de suavizar as ZMs, por meio da eliminação de pixels isolados e pequenas manchas, e assim delinearem-se as limitações das classes. A utilização de diferentes filtros e máscaras por parte de pesquisadores vem proporcionando bons resultados e moldando os mapas com maior qualidade.

3.9 Softwares aplicados na agricultura de precisão

A aplicação de técnicas de geoprocessamento e manipulação de imagens necessitam de aplicativos computacionais para serem processados. Antes e durante a coleta de dados, podem ser utilizados softwares que permitem comunicação com equipamentos receptores de sinal de sistemas de posicionamento global (GPS) e importação de dados georreferenciados, como o TracMaker® e o PathFinder® (KANNEMADUGU; PUNYASHESHUDU, 2014). Estes softwares permitem avaliar o posicionamento e magnitude dos dados e reposicioná-los, caso seja necessário.

As análises estatísticas dos dados amostrais podem ser realizadas por diversos softwares, como o Minitab®, o Statistica® e o R, além de planilhas eletrônicas. Cada software possui peculiaridades como funcionalidades diferenciadas, custo de aquisição, sistema operacional, facilidade de utilização, interface distinta, dentre outras. Para a análise espacial de dados e, portanto, que permitem a determinação dos parâmetros do semivariograma (efeito pepita, patamar e alcance), podem ser utilizados softwares como GS+®, Arcview® e R. Eles também são utilizados para mensurar dados para locais não amostrados e permitem elaborar mapas usando diferentes métodos de interpolação.

Outros softwares são desenvolvidos em universidades como o Campeiro da Universidade Federal de Santa Maria-RS que, em seu módulo de AP, realiza gestão de procedimentos que envolvem o uso de geotecnologias como GPS aplicadas no manejo de culturas agrícolas, por amostragem georreferenciada de solos; mapas de fertilidade e de produtividade; análises estatísticas e interpretação de mapas; geoestatística (CRCAMPEIRO, 2014).

O SDUM (Software para Definição de Unidades de Manejo; BAZZI et al., 2013) foi desenvolvido devido à necessidade de utilizar um software, gratuito, capaz de determinar e avaliar ZMs de forma amigável. Esta aplicação computacional foi projetada e implementada com ferramentas livres e em linguagem multiplataforma (Java), para ser executada sem a necessidade de sistema operacional específico. Para o armazenamento dos dados, é utilizado o sistema gerenciador de banco de dados (SGDB) PostgreSQL, que é um banco de dados gratuito e segue as normas internacionais dos bancos de dados comerciais. Além de contemplar as características de flexibilidade e suporte e permitir a utilização de dados espaciais. As funcionalidades disponibilizadas no SDUM são: análise estatística de dados amostrais, análise de correlação espacial dos atributos pela matriz de correlação espacial, interpolação de dados por inverso da distância elevado a uma potência, criação de zonas de manejo pelos métodos de normalização e padronização de dados (K-Means e Fuzzy C-Means), e avaliação destas zonas de manejo por cálculo da ANOVA e redução da variância.

3.10 Desenvolvimento de software

Softwares aplicativos são programas de computador desenvolvidos com o intuito de proporcionar ao usuário as funcionalidades e o desempenho por ele requeridos. Também devem ser confiáveis e fáceis de se manter e usar, pois a amigabilidade de um software com o usuário é importante bem como a redução do custo de desenvolvimento ou uso (SOMMERVILLE, 2011). Por isso, busca-se utilizar softwares de programação que criem aplicativos a partir de uma linguagem de programação e possibilitem desenvolver aplicativos com baixo ou sem custo, principalmente em relação às licenças de uso.

Após 1980, o desenvolvimento de software passou a utilizar nova abordagem, conhecida como programação orientada ao objeto. Neste contexto, pressupõe-se que o mundo é formado por objetos e que desenvolver um sistema é simular objetos e seus comportamentos (PRESSMAN, 2011). Uma das linguagens utilizadas para programação orientada a objetos é o Java. Um código desenvolvido em Java é portátil, ou seja, pode ser executado, sem alteração, em qualquer computador que possua uma máquina virtual java (JVM, *Java Virtual Machine*) instalada. A independência de plataforma é uma das vantagens mais significativas desta linguagem de programação (PRESSMAN, 2011).

Esta linguagem facilita o desenvolvimento para aplicativos portáteis, utilizando para isto uma linguagem de programação simplificada, segura e orientada a objetos (SILVEIRA, 2012). Java está presente em vários ramos da tecnologia, no mercado corporativo (Java EE) com soluções robustas para Web e aplicações distribuídas, no desenvolvimento para dispositivos móveis (Java ME), no Google Android, além do Desktop. Por sua portabilidade e segurança, a plataforma é utilizada por grandes bancos e empresas, pois não depende apenas de um fabricante, de único sistema operacional ou de apenas um banco de dados. A JVM abstrai a camada de *hardware* bem como a comunicação com o sistema operacional, assim, o código executável torna-se portátil (SILVEIRA, 2012). A tecnologia Java é de uso gratuito, de código aberto bem como de editores e ambientes de desenvolvimento (NetBeans, Eclipse, JCreator), além de servidores de aplicações (Tom Cat, Jetty e Jboss) (SILVEIRA, 2012).

O programa computacional R é gratuito e livremente distribuído, pois possui código aberto que pode ser reproduzido ou adaptado livremente. É utilizado como ambiente para análises estatísticas e possui recursos gráficos de qualidade. Os pacotes implementados no módulo geoestatístico geoR proporcionam a análise geoestatística e interpolação de dados por krigagem. O R tem fácil interação com outras linguagens de programação e banco de dados, inclusive com o PostgreSQL, utilizado pelo SDUM.

4. REFERÊNCIAS BIBLIOGRÁFICAS

- ARNO, J.; MARTINEZ-CASASNOVAS, J. A.; RIBES-DASI, M.; ROSELL, J. R. Clustering of grape yield maps to delineate site-specific management zones. **Spanish Journal of Agricultural Research**, v. 9, n. 3, p. 721-729, 2011.
- BAZZI, C. L.; SOUZA, E. G.; URIBE-OPAZO, M. A.; NÓBREGA, L. H. P.; ROCHA, D. M. Management zones definition using soil chemical and physical attributes in a soybean area. **Engenharia Agrícola**, v. 34, n. 5, p. 952-964, 2013.
- BIER V. A. **Construção e avaliação de mapas**. 2015. 82 f. Tese (Doutorado em Engenharia Agrícola) - Universidade Estadual do Oeste do Paraná, Cascavel, 2015.
- BLANCO-CANQUI, H.; SHAVER, T. M.; LINDQUIST, J. L.; SHAPIRO, C. A.; ELMORE, R. W.; FRANCIS, C. A.; HERGERT, G. W. Cover crops and ecosystem services: insights from studies in temperate soils. **Agronomy Journal**, v. 107, n. 1, p. 2449-2474, 2015.
- BOBRYK, C. W.; MYERS, D. B.; KITCHEN, N. R.; SHANAHAN, J. F.; SUDDUTH, K. A.; DRUMMOND, S. T.; GOMEZ RABOTEAUX, N. N. Validating a digital soil map with corn yield data for precision agriculture decision support. **Agronomy Journal**, v. 108, n. 3, p. 957-965, 2016.
- BORSSOI, J. A.; URIBE-OPAZO, M. A.; GALEA, M. Técnicas de diagnóstico de influência local na análise espacial da produtividade da soja. **Engenharia Agrícola**, v. 31, n. 2, p. 376-387, 2011.
- BURAK, D. L.; PASSOS, R. R.; ANDRADE, F. V. Variabilidade espacial de atributos químicos do solo sob cafeeiro Conilon : relação com textura, matéria orgânica e relevo. **Bragantia**, v. 71, p. 538-547, 2012.
- BUTTAUFUOCO, G.; CASTRIGNANO, A.; COLECCHIA, A. S.; RICCA, N. Delineation of management zones using soil properties and a multivariate geostatistical approach. **Italian Journal of Agronomy**, v. 5, n. 4, p. 323-332, 2010.
- 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 Sci. Soc. Am. J.**, v. 58, p. 1501-1511, 1994.
- CAMPOS, M. C. C.; SOARES, M. D. R.; SANTOS, L. A. C.; OLIVEIRA, I. A.; AQUINO, R. E., BERGAMIN, A. C. Variabilidade espacial dos atributos físicos em um Argissolo Vermelho sob floresta. **Comunicata Scientiae**, v. 4, n. 2, p. 168-178, 2013.
- CASTIONE, G. A. F.; SOUZA, Z. M.; SILVA, R. B.; CAMPOS, M. C. C.; CUNHA, J. M. M. Variabilidade espacial da textura do solo em área irrigada por pivô central em diferentes posições na paisagem. **Revista Agro@mbiente on-Line**, v. 9, n. 3, p. 219-226, 2015.
- CLARK, I. **Practical Geostatistics**. London: Applied Science, 1979. 120 p.
- COLAÇO, A. F.; POVH, F. P.; MOLIN, J. P.; ROMANELLI, T. L. Energy assessment for variable rate nitrogen application. **Agric. Eng.** v. 14, n. 3, p. 85-90, 2012.
- CONAB. - Companhia Nacional de Abastecimento. Acompanhamento da Safra Brasileira de Grãos. **Monitoramento agrícola** - Safra 2015/2016. v. 3, n. 9, p. 1-174, 2016.
- CORÁ, J. E.; ARAÚJO, A. V.; PEREIRA, G. T.; BERALDO, J. M. G. Variabilidade espacial de atributos do solo para adoção do sistema de agricultura de precisão na cultura de cana-de-

açúcar. **Revista Brasileira Ciência do Solo**, v. 1, n. 28, p. 1013-1021, 2004.

CÓRDOBA, M. A.; BRUNO, C. I.; COSTA, J. L.; PERALTA, N. R.; BALZARINI, M. G. Protocol for multivariate homogeneous zone delineation in precision agriculture. **Biosystems Engineering**, v. 143, p. 95-107, 2016.

COUTO, E. G.; SCARAMUZZA, J. F.; MARASCHINI, L. Influência dos métodos de interpolação dos dados nos mapas usados na agricultura de precisão. **Simpósio Internacional de Agricultura de Precisão**, 2002.

CRCAMPEIRO. **Projeto CR Campeiro - software para gerenciamento rural**. Disponível em: <<http://www.crcampeiro.net/>>.

CRESSIE, N. A. C. **Statistics for spatial data**. Revised ed. New York: Wiley-Interscience Publication, 1993. 900 p.

DIGGLE, P. J.; RIBEIRO JR, P. J. **Model-based geostatistics**. 1. ed. New York: Springer, 2007. 228 p.

DOBERMANN, A.; PING, J. L.; ADAMCHUK, V. I.; SIMBAHAN, G. C.; FERGUSON, R. B. Classification of crop yield variability in irrigated production fields. **Agronomy Journal**, v. 95, n. 5, p. 1105-1120, 2003.

DOERGE, T. A. Management Zone Concepts. **Site-Specific Management Guidelines**, p. 1-4, 2000.

DRUCK, S.; CARVALHO, M. S.; CÂMARA, G.; MONTEIRO, A. M. V. **Análise espacial de dados geográficos**. 1. ed. Brasília: Embrapa, 2004.

EMBRAPA. - Empresa Brasileira de Pesquisa Agropecuária. Soja. **Soja na alimentação**, 2016.

FARACO, M. A.; URIBE-OPAZO, M. A.; SILVA, E. A. A. DA, JOHANN, J. A.; BORSSOI, J. A. Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. **Revista Brasileira Ciência do Solo**, v. 32, n. 1, p. 463-476, 2008.

FIESP - Sindicato da Indústria do Milho, Soja e seus Derivados no Estado de São Paulo. **Milho e suas riquezas – História**. 2016.

GONTIJO, I.; NICOLE, L. R.; PARTELLI, F. L.; BONOMO, R.; SANTOS, E. O. J. Variabilidade e correlação espacial de micronutrientes e matéria orgânica do solo com a produtividade da pimenta-do-reino. **Revista Brasileira Ciência do Solo**, v. 36, n. 3, p. 1093-1102, 2012.

GRZEGOZEWSKI, D. M.; URIBE-OPAZO, M. A.; BASTIANI, F. DE; GALEA, M. Local influence when fitting Gaussian spatial linear models: an agriculture application. **Ciencia e Investigacion Agraria**, v. 40, n. 3, p. 523-535, 2013.

GUASTAFERRO, F.; CASTRIGNANÒ, A.; BENEDETTO, D.; SOLLITTO, D.; TROCCOLI, A.; CAFARELLI, B. A comparison of different algorithms for the delineation of management zones. **Precision Agriculture**, v. 11, n. 6, p. 600-620, 2010.

ILIADIS, L. S.; VANGELOUDH, M.; SPARTALIS, S. An intelligent system employing an enhanced fuzzy c-means clustering model: Application in the case of forest fires. **Computers and Electronics in Agriculture**, v. 70, n. 2, p. 276-284, 2010.

ISAAKS, E. H.; SRIVASTAVA, R. M. **Applied Geostatistics**. 1. ed. New York: Oxford, 1989. 561 p.

KANNEMADUGU, J.; PUNYASHESHUDU, D. GPS Based surveying for mapping of canal alignment in GIS environment. **Oriental Journal of Computer Science & Tecnology**, v. 7, n. 1, p. 53-59, 2014.

KESTRING, F. B. F.; GUEDES, L. P. C.; BASTIANI, F. DE; URIBE-OPAZO, M. A. Comparação de mapas temáticos de diferentes grades amostrais para a produtividade da soja. **Engenharia Agrícola**, v. 35, n. 4, p. 733-743, 2015.

KHOSLA, R.; INMAN, D.; WESTFALL, D. G.; REICH, R. M.; FRASIER, M.; MZUKU, M.; HORNUNG, A. A Syntesis of multi-disciplinary research in precision agriculture: site-specific management zones in the semi-arid western great plains of the USA. **Precision Agriculture**, v. 9, p. 85-100, 2008.

KLEIN, C.; KLEIN, V. A. Estratégias para potencializar a retenção e disponibilidade de água no solo. **Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental**, v. 19, n. 1, p. 21-29, 2015.

LIU, W.; DU, P.; ZHAO, Z.; ZHANG, L. An Adaptive Weighting Algorithm for Interpolating the Soil Potassium Content. **Scientific Reports**, v. 6, p. 1-11, 2016

LOWRANCE, C. **Open source hardware and software in agriculture: an autonomous sap flow measurement wireless network & a user friendly management zone delineation tool**. 2014. 121 p. Thesis (Master's degree in crop and soil sciences) - University of Georgia, 2014.

LUCAS, T. P. B.; PLEC, D.; ABREU, M. L.; PARIZZI, M. G. Identificação de interpoladores adequados a dados de chuva a partir de parâmetros estatísticos. **Revista Brasileira de Climatologia**, v. 13, n. 9, p. 7-21, 2013.

MAPA - Ministério da Agricultura, Pecuária e Abastecimento. **Soja**. Brasília: MAPA, 2016. Disponível em: <<http://www.agricultura.gov.br/vegetal/culturas/soja>>

MATERN, B. **Spatial Variation**: Stochastic models and their application to some problems in forest surveys and other sampling investigations. Stockholm: States Skogs for Sknings Institut, 1960. 144 p.

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

MAZZINI, P. L. F.; SCHETTINI, C. A. F. Avaliação de metodologias de interpolação espacial aplicadas a dados hidrográficos costeiros quase-sinóticos. **Braz. J. Aquat. Sci. Technol.**, v. 13, n. 1, p. 53-64, 2009.

MELLO, J. M.; BATISTA, J. L. F.; RIBEIRO JR, P. J.; OLIVEIRA, M. S. Ajuste e seleção de modelos espaciais de semivariograma visando à estimativa volumétrica de Eucalyptus grandis. **Scientia Forestalis**, v. 1, n. 69, p. 25-37, 2005.

MINASNY, B.; MCBRATNEY, A. B. The Matérn function as a general model for soil variograms. **Geoderma**, v. 128, n. 1, p. 192-207, 2005.

MOLIN, J. P.; RABELLO, L. M. Estudos sobre a mensuração da condutividade elétrica do solo. **Engenharia Agrícola**, v. 31, n. 1, p. 90-101, 2011.

MONDO, H. V. V.; GOMES JUNIOR, F. G.; PINTO, T. L. F.; MARCHI, J. L., MOTOMIYA, A. V. A.; MOLIN, J. P.; CÍCERO, S. M. Spatial variability of soil fertility and its relationship with seed physiological potential in a soybean production area. **Revista Brasileira de Sementes**, v. 34, n. 2, p. 193-201, 2012.

MORAL, F. J.; TERRÓN, J. M.; SILVA, J. R. M. DA. Delineation of management zones using

mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. **Soil and Tillage Research**, v. 106, n. 2, p. 335-343, 2010.

MOSHIA, M. E.; KHOSLA, R.; LONGCHAMPS, L.; REICH, R.; DAVIS, J. G.; WESTFALL, D. G. Precision manure management across site-specific management zones: grain yield and economic analysis. **Agronomy Journal**, v. 106, n. 6, p. 2146-2156, 2014.

PRAMANIK, S.; PRUSTY, S.; BHATTACHARJEE, D.; BHUNRE, P. K. A Region-to-pixel based multi-sensor image fusion. **Procedia Technology**, v. 10, p. 654-662, 2013.

PRESSMAN, R. S. **Engenharia de software**. 7. ed. São Paulo: Pearson Education do Brasil, 2011. 780 p.

RAWLS, W. J.; PACHEPSKY, Y. A.; RITCHIE, J. C.; SOBECKI, T. M.; BLOODWORTH, H. Effect of soil organic carbon on soil water retention. **Geoderma**, v. 116, n. 1, p. 61-76, 2003.

ROCHA, M. M.; LOURENÇO, D. A.; LEITE, C. B. B. Aplicação de krigagem com correção do efeito de suavização em dados de potenciometria da cidade de Pereira Barreto - SP. **Geologia USP Série Científica, São Paulo**, v. 7, n. 2, p. 37-48, 2007.

RODRIGUES JR, F.; VIEIRA, L. B.; QUEIROZ, D. M.; SANTOS, N. T. Geração de zonas de manejo para cafeicultura empregando-se sensor SPAD e análise foliar. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 15, n. 8, p. 778-787, 2011.

SAMIRA, I.; AHMED, D.; LHOUSSAINE, M. Soil fertility mapping : comparison of three spatial interpolation techniques. **International Journal of Engineering Research & Technology**, v. 3, n. 11, p. 1635–1643, 2014.

SCHENATTO, K.; SOUZA, E. G.; BAZZI, C. L.; BIER, V. A.; BETZEK, N.; GAVIOLI, A. Data Interpolation in the definition of management zones. **Acta Scientiarum**, v. 38, n. 1, p. 31-40, 2016.

SHANKEN, J.; ZHOU, G. Estimating and testing beta pricing models: alternative methods and their performance in simulations. **Journal of Financial Economics**, v. 84, n. 1, p. 40-86, 2007.

SILVA, A. S.; RIBEIRO JR, P. J.; ELMATZOGLOU, I. Modelagem geoestatística utilizando a família de gneiting de funções de covariância espaço-temporais. **Rev. Mat. Estat.**, v. 25, n. 1, p. 65-83, 2007.

SILVEIRA, P. **Introdução à arquitetura e design de software**: uma visão sobre a plataforma Java. 1. ed. Rio de Janeiro: Elsevier, 2012. 257 p.

SOMMERVILLE, I. **Engenharia de Software**. 9. ed. São Paulo: Pearson Education do Brasil, 2011. 529 p.

SOUZA, E. G.; BAZZI, C. L.; KHOSLA, R.; URIBE-OPAZO, M. A.; REICH, R. M. Interpolation type and data computation of crop yield maps is important for precision crop production. **Journal of Plant Nutrition**, v. 39, n. 4, p. 531-538, 2016.

SUSZEK, G.; SOUZA, E. G.; URIBE-OPAZO, M. A.; NOBREGA, L. H. P. Determination of management zones from normalized and standardized equivalent productivity maps in the soybean culture. **Eng. Agríc.**, v. 31, n. 5, p. 895-905, 2011.

TAGARAKIS, A.; LIAKOS, V.; FOUNTAS, S.; KOUNDOURAS, S.; GEMTOS, T. A. Management zones delineation using fuzzy clustering techniques in grapevines. **Precision Agriculture**, v. 14, n. 1, p. 18-39, 2013.

TAYLOR, J. A.; MCBRATNEY, A. B.; WHELAN, B. M. Establishing Management Classes for Broadacre Agricultural Production. **Agronomy Journal**, v. 99, n. 5, p. 1366-1376, 2007.

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

WANDERLEY, H. S.; AMORIM, R. F. C.; CARVALHO, F. O. Interpolação espacial da precipitação no estado de Alagoas utilizando técnicas geoestatísticas. **Revista Campo Digital**, v. 8, n. 1, p. 34-42, 2013.

WOODCOCK, C. E.; STRAHLER, A. H.; JUPP, D. L. The use of variograms in remote sensing: I. Scene models and simulated images. **Remote Sensing of Environment**, v. 25, p. 323-348, 1988.

XIANG, L.; YU-CHUN, P.; ZHONG-QIANG, G.; CHIN-JIANG, Z. Delineation and Scale Effect of Precision Agriculture Management Zones Using Yield Monitor Data Over Four Years. **Agricultural Sciences in China**, v. 6, n. 2, p. 180-188, 2007.

5. ARTIGO 1 - SELEÇÃO DE MÉTODOS DE INTERPOLAÇÃO DE DADOS²

RESUMO: A krigagem ordinária (KRI) e o inverso da distância elevado a uma potência (IDW) são métodos de interpolação comumente utilizados para mensurar dados para a geração de mapas temáticos e zonas de manejo (ZMs). Porém, a utilização de KRI é complexa e a definição de qual valor utilizar como expoente para IDW é questionável. Desta maneira, o objetivo deste trabalho foi desenvolver e aplicar rotinas computacionais capazes de identificar automaticamente os melhores parâmetros para cada método de interpolação, por índice de seleção de interpoladores (ISI). As rotinas foram aplicadas aos dados amostrais de produtividade de soja e milho, bem como de atributos químicos e físicos do solo coletados em duas áreas agrícolas localizadas no município de Serranópolis do Iguaçu-PR. Os atributos que apresentaram correlação espacial com a produtividade média normalizada foram elevação e resistência do solo a penetração (RSP), os quais foram selecionados para definir as ZMs. Para cada atributo selecionado, foram testados 300 diferentes ajustes para o semivariograma a fim de se identificarem os melhores parâmetros utilizados para mensurar dados por KRI e doze diferentes valores para expoente do IDW, cujo melhor foi identificado como IDW_B. ZMs com duas, três e quatro classes foram definidas com base nos dados mensurados pelos diferentes métodos de interpolação (KRI, IDW₁, IDW₂, e IDW_B). A área A foi considerada homogênea por não apresentar média estatisticamente diferente entre as classes, e para a área B, a melhor maneira de agrupamento foi subdividir a área em duas classes distintas. Os melhores resultados foram obtidos pelo método de interpolação KRI. Pode-se concluir que as rotinas computacionais implementadas foram eficientes e capazes de identificar o melhor ajuste para o semivariograma a ser utilizado na KRI bem como o melhor expoente a ser utilizado pelo método de interpolação IDW. A automatização dos processos implementados facilitou e agilizou a definição dos parâmetros utilizados pelos métodos de interpolação.

PALAVRAS-CHAVE: inverso da distância; krigagem; semivariograma; zonas de manejo.

SELECTION OF DATA INTERPOLATION METHODS

ABSTRACT: The ordinary kriging (KRI) and the inverse distance weighting (IDW) are interpolation methods commonly used before generating thematic maps and management zones (MZs). However, KRI application is complex and defining which value is going to use as an exponent for IDW is debatable. Thus, this study aimed at developing and applying computational algorithms able to identify automatically the best parameters for each interpolation method, using the interpolation selection index (ISI). The routines were applied to soybean and maize yield data, as well as soil chemical and physical attributes, collected in two agricultural areas in Serranópolis do Iguaçu county, Paraná. The attributes that presented spatial correlation whose average yield was normalized were elevation and soil penetration resistance (SPR), selected to define the MZs. For each selected attribute, 300 different adjustments were tested for the semivariogram in order to identify the best parameters that were used to measure data by KRI and twelve different values for IDW exponent, which the best one was identified as IDW_B. MZs with two, three and four classes were defined based on recorded data by four different interpolation methods (KRI, IDW₁, IDW₂, and IDW_B). Area A was considered homogeneous because it did not present statistic difference in mean between classes. While for area B, the best way of clustering was to subdivide the area into two distinct

² Artigo a ser enviado para o periódico internacional Computers and Electronics in Agriculture, com classificação A2 no QUALIS/CAPES na área de Ciências Agrárias.

classes. The best results were obtained by the KRI interpolation method. Thus, it can be concluded that the implemented computational functions were efficient and able to identify the best adjustment for the semivariogram to be used in KRI, as well as the best exponent to be used by IDW. The automation of the implemented processes facilitated and speeded up the parameters definition applied by interpolation methods.

KEYWORDS: inverse of the distance; kriging; management zones; semivariogram.

5.1 Introdução

A agricultura de precisão (AP) é um sistema de gerenciamento da variabilidade espacial e temporal da produção agrícola e dos fatores nela envolvidos. Abrange o uso de tecnologias e procedimentos com o objetivo de realizar manejo localizado na lavoura, a fim de se racionalizar o uso de insumos e, conseqüentemente, reduzirem-se custos de produção e o impacto ambiental (COLAÇO et al., 2012; MOLIN; RABELLO, 2011; MONDO et al., 2012; REZA et al., 2010).

Os mapas temáticos são usados para ilustrar temas, além de representarem o terreno. Eles também identificam diferentes representações cartográficas tanto da área quanto de características a ela associadas. O desenvolvimento de mapas temáticos está ligado à coleta, análise e interpretação de dados, além da representação das informações em um mapa, o que facilita a identificação de semelhanças. Um caso específico de mapas temáticos são mapas de contorno, os quais são construídos com a união de pontos de mesmo valor, e aplicáveis a fenômenos geográficos que apresentam continuidade no espaço geográfico.

Os mapas de contorno podem ser construídos a partir de dados absolutos (elevação, temperatura, precipitação, umidade, pressão atmosférica) ou dados relativos (densidade, porcentagens, índices). Os mapas temáticos são gerados para identificar a variabilidade das propriedades do solo, das plantas e da produtividade com base em amostras coletadas ao longo de uma safra. É importante interpolar os dados em uma grade densa e regular para o fornecimento de valores para locais que não foram amostrados.

A interpolação é um procedimento que se baseia em dados de atributos obtidos em pontos georreferenciados, e que permite estimar valores para locais não amostrados. O inverso da distância elevado a uma potência (*inverse distance weighting*, IDW) e a krigagem são os métodos de interpolação comumente utilizados em AP e diferenciam-se pela maneira como os pesos são atribuídos às diferentes amostras, o que pode influenciar os valores estimados (COUTO; SCARAMUZZA; MARASCHINI, 2002; DRUCK et al., 2004; REZA et al., 2010; SILVA et al., 2008; SOUZA et al., 2016).

Na interpolação usando KRI e IDW, os pesos são definidos com base na distância que separa o valor interpolado dos valores observados. IDW é um processo puramente matemático, no qual os dados são ponderados de tal maneira que a influência entre eles

diminui conforme aumenta a distância, ou seja, a influência de cada ponto amostrado é inversamente proporcional à distância do ponto a ser estimado (MAZZINI; SCHETTINI, 2009). O fator peso é predeterminado pelo valor da potência escolhida, em que quanto maior o expoente, menor é a influência dos pontos mais distantes. Os modelos mais utilizados são o inverso da distância e o inverso da distância elevado ao quadrado. IDW corresponde a um método simples e de fácil aplicação, que requer pouco custo computacional.

A krigagem (KRI) é considerada um método de interpolação mais robusto, pois pondera os valores dos pontos vizinhos ao que será estimado. Ela também considera os critérios de não tendenciosidade e mínima variância do erro de estimativa, além da estrutura de variabilidade espacial encontrada para o atributo (GRZEGOZEWSKI et al., 2013; ISAAKS; SRIVASTAVA, 1989), logo proporciona precisão local. A utilização da KRI resulta em estimativas suavizadas que fazem com que os valores de máximo da distribuição estatística dos dados tenham a tendência de serem subestimados, enquanto os valores de mínimo tendem a ser superestimados (GRZEGOZEWSKI et al., 2013; KESTRING et al., 2015; ROCHA; LOURENÇO; LEITE, 2007; URIBE-OPAZO; BORSSOI; GALEA, 2012).

Teoricamente, a krigagem é considerada o interpolador ideal, porém, exige conhecimento em geoestatística, que é uma etapa de análise necessária antes do processo de interpolação. Como a geoestatística não considera somente o valor obtido para uma variável resposta, mas também sua posição expressa por um sistema de coordenadas, a função semivariância é utilizada como ferramenta para apresentar a medida do grau de dependência espacial entre amostras (MELLO et al., 2005) apresentada na forma de semivariograma. A semivariância é definida como a metade da variância de diferenças entre observações de uma variável aleatória, separadas por determinada distância. Assim, valores baixos indicam menor variabilidade, ou seja, maior similaridade.

O desempenho de Krigagem pode ser influenciado pela variabilidade e estrutura espacial dos dados e pela escolha do modelo de semivariograma, o raio de busca e o número dos pontos vizinhos mais próximos utilizados nas estimativas (GUASTAFERRO et al., 2010; ISAAKS; SRIVASTAVA, 1989; REZA et al., 2010; TOMCZAK, 1998).

Para pontos próximos (mais correlacionados), a variância é pequena, ao se aumentar a distância, os valores dos pontos tendem a se tornarem mais diferentes, consequentemente, a variância aumenta (Figura 1). Normalmente, há uma tendência a estabilizar-se em torno de uma variância máxima a partir de certa distância (patamar). Deste ponto em diante, considera-se que não existe mais dependência espacial entre as amostras (independência entre amostras), pois a variância da diferença entre pares de amostras torna-se aproximadamente constante. Esse ponto no eixo X (distância) é denominado de alcance (α) do semivariograma (CAMPOS et al., 2013; WANDERLEY; AMORIM; CARVALHO, 2013). Efeito pepita (C_0) é o parâmetro do semivariograma que representa as variações locais de pequena escala, tais como erros de medição. A contribuição (C_1) também conhecida como a variância da

dispersão, representa a diferença entre os valores de C_0 e o patamar (C), intervalo esse no qual o semivariograma cresce quando há dependência espacial (CRESSIE, 1993).

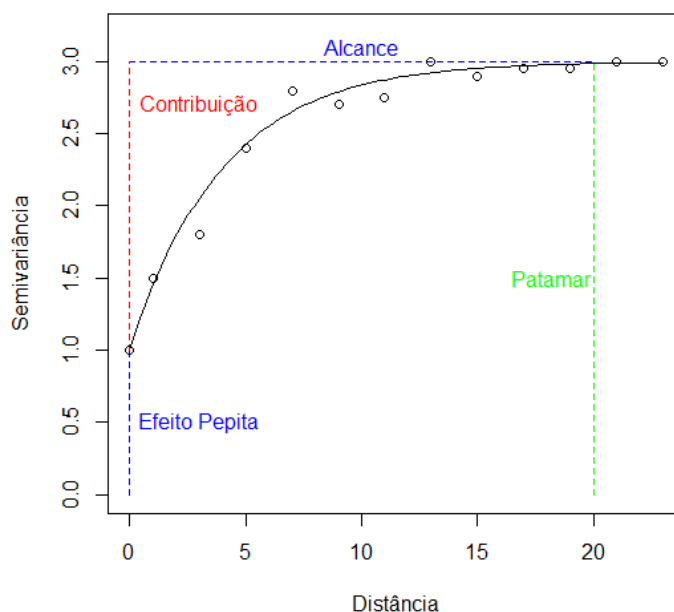


Figura 1 Parâmetros do semivariograma.

A escolha do modelo representa a interpretação da estrutura de correlação espacial a ser utilizada nos procedimentos inferenciais da krigagem. Todavia, a sensibilidade para detectar a variabilidade espacial das amostras está diretamente ligada ao melhor ajuste entre os dados experimentais e os modelos teóricos do semivariograma (CRESSIE, 1993; ODA-SOUZA et al., 2008; YAMAMOTO; LANDIM, 2013). Os modelos geralmente avaliados são: esférico, exponencial, gaussiano, e Família Matérn com diferentes valores no parâmetro de forma κ .

Segundo Mello et al. (2005), os métodos de ajuste dos modelos de semivariograma podem ser classificados em dois grupos. O primeiro grupo é o de ajuste dos modelos ao semivariograma experimental e é realizado pelo método dos quadrados mínimos ordinários (*ordinary least squares*, OLS) e do método dos quadrados mínimos ponderados (*weight least squares*, WLS). O segundo grupo é o método de ajuste de um modelo considerando a distribuição de probabilidade dos dados, pelo método da máxima verossimilhança (*maximum likelihood*, ML) (BORSSOI; URIBE-OPAZO; GALEA, 2011; GRZEGOZEWSKI et al., 2013; KESTRING et al., 2015; URIBE-OPAZO; BORSSOI; GALEA, 2012).

OLS consiste em estimar os parâmetros do semivariograma por um modelo que minimiza a soma do quadrado da diferença entre os valores observados e os estimados. WLS é semelhante ao OLS, porém utiliza um ponderador na estimativa dos parâmetros do semivariograma teórico. Os melhores valores para os parâmetros são os que minimizam a soma do quadrado das diferenças entre os valores experimentais e os preditos pelo modelo teórico do semivariograma. Assim, a diferença é ponderada diretamente pelo inverso da variância do semivariograma experimental (DIGGLE; RIBEIRO JR, 2007).

A estimação de parâmetros pelo método da máxima verossimilhança pressupõe estacionaridade forte. Em condições regulares (amostra de dados suficientemente grande), este estimador é assintoticamente não viesado e eficiente. Embora as estimativas geradas por OLS e WLS sejam menos precisas, são bastante confiáveis e tendem a ser menos tendenciosas, além de apresentarem maior desempenho (DIGGLE; RIBEIRO JR, 2007; MELLO et al., 2005; SHANKEN; ZHOU, 2007).

A distribuição dos erros ou resíduos pode ser utilizada para avaliar qual o melhor modelo de ajuste para o semivariograma e de estimação dos parâmetros. Segundo Faraco et al. (2008), validação cruzada, Jackknifing, Filliben e Akaike são os principais critérios utilizados para avaliação dos modelos de variabilidade espacial e consideraram em seu estudo que a validação cruzada é mais adequada. A validação cruzada consiste de uma técnica de avaliação de erros de estimativa que permite comparar valores estimados e amostrados, usando-se somente a informação disponível na amostra de dados (BORSSOI; URIBE-OPAZO; GALEA, 2011; ISAAKS; SRIVASTAVA, 1989). Os valores calculados do erro médio e do desvio-padrão do erro médio são utilizados pelo critério da validação cruzada para identificar o melhor ajuste para o semivariograma (FARACO et al., 2008).

A fim de facilitar a escolha e utilização de métodos de interpolação, o objetivo deste estudo foi desenvolver e aplicar rotinas computacionais capazes de identificar automaticamente os melhores parâmetros para os métodos de interpolação krigagem e inverso da distância elevado a uma potência, pelo índice de seleção de interpoladores. Os modelos selecionados foram testados na definição de zonas de manejo.

5.2 Material e métodos

5.2.1 Localização da área, coleta de dados e seleção de atributos

Dados amostrais de produtividade de soja e milho bem como de atributos químicos e físicos do solo foram coletados com base em grade amostral irregular em duas áreas agrícolas (Figura 2), localizadas na área rural do município de Serranópolis do Iguçu-PR, na região Oeste do estado do Paraná – Brasil. As áreas são cultivadas sob sistema de plantio direto há mais de 10 anos com rotação de cultura soja/milho, para fins comerciais. Área A, com centro geográfico sob coordenadas 25°24'30"S e 54°00'10"O e elevação média de 363 m, possui aproximadamente 9,9 ha onde foram definidos 42 pontos amostrais. Área B, com centro geográfico sob coordenadas 25°24'17"S e 54°00'18"O e elevação média de 363 m, possui aproximadamente 20,9 ha onde foram definidos 73 pontos amostrais. A densidade amostral de 4,2 pontos ha⁻¹ na área A e de 3,7 pontos ha⁻¹ na área B foi considerada suficiente para mensurar a variabilidade espacial dos atributos destas áreas (DEMATTÊ et al., 2014;

FERGUSON; HERGERT, 2009; FRANZEN et al., 2002; WOLLENHAUPT; WOLKOWSKI; CLAYTON, 1994).

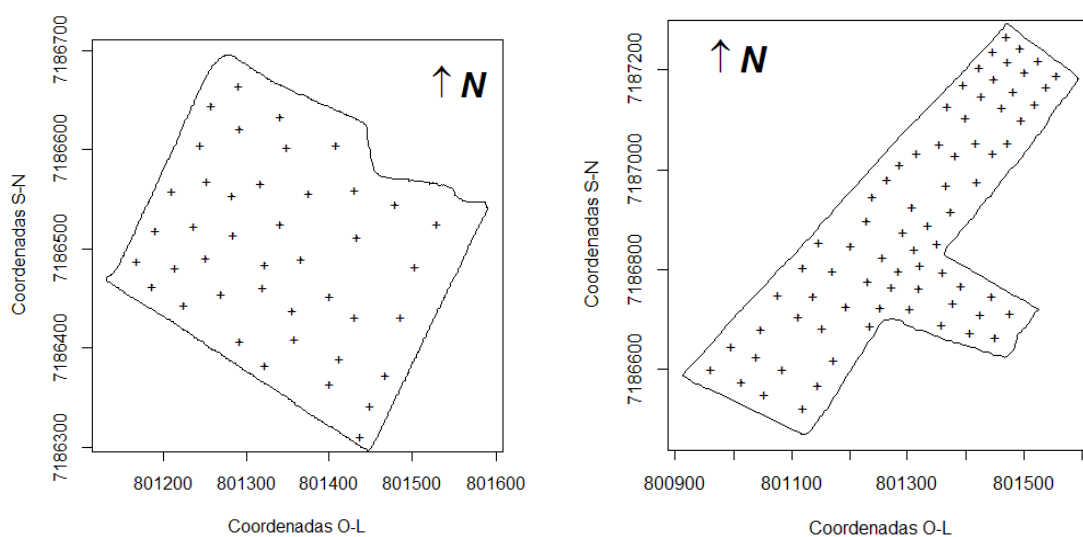


Figura 2 Localização dos pontos amostrais e contorno das áreas.

A qualidade dos dados interpolados e das ZMs definidas a partir destes dados foram analisadas para a validação das técnicas de interpolação. Por este motivo, neste trabalho, foram utilizados dados de produtividade e de atributos considerados temporalmente estáveis (Tabela 1), seguindo a recomendação de que ZMs sejam criadas a partir de variáveis estáveis (DOERGE, 2000).

Tabela 1 Atributos coletados em cada ponto amostral

Atributos	Área A			Área B		
	2013	2014	2015	2013	2014	2015
Produtividade de soja	X	X	X	X	X	X
Produtividade de milho	X	X		X	X	
RSP_0,0-0,1 m	X	X	X	X	X	X
RSP_0,1-0,2 m	X	X	X	X	X	X
RSP_0,2-0,3 m	X	X	X	X	X	X
Argila	X			X		
Silte	X			X		
Areia	X			X		
Elevação	X					X

RSP: resistência do solo à penetração

A localização dos pontos amostrais foi obtida a partir de um aparelho receptor GPS da marca Trimble, modelo Juno SB e a elevação obtida pela utilização de estação total eletrônica Topcon GPT-7505. No entorno de cada ponto definido na grade amostral (a distância máxima de 3 m), foram realizadas quatro medições da resistência do solo à penetração (RSP) com a utilização de um penetrômetro eletrônico da marca Falker PenetroLOG. Calcularam-se posteriormente as médias das medições para representação do ponto amostral nas profundidades de 0-0,1; 0,1-0,2; e 0,2-0,3 m. Nos mesmos locais, foram coletadas oito

subamostras de solo na profundidade de 0-0,2 m em um raio de 3 metros do ponto determinado na grade (adaptado de WOLLENHAUPT et. al., 1994). Posteriormente, as amostras foram encaminhadas ao laboratório para análise e obtenção de dados de textura (argila, silte e areia), e características químicas do solo (C, pH, Ca, Mg, K, Cu, Zn, Fe, Mn).

As produtividades da soja e do milho foram determinadas nos mesmos pontos em que foram realizadas as amostragens de solo, cujas colheita e a debulha ocorreram de forma manual. Cada ponto amostral representa uma área de aproximadamente 0,9 m². A partir do ponto amostral, foram colhidas duas linhas em um percurso de um metro, cujo espaçamento entre linhas é de 0,45 m. Posteriormente, foram calculados os valores da produtividade, corrigidos para o teor de água de 13%.

Os dados de produtividade foram normalizados pela amplitude (Equação 1) com o objetivo de remover a variabilidade sazonal e da cultura utilizada, a fim de transformar o valor obtido em cada ponto amostral em cada uma das cinco safras (P_{ij}) utilizadas neste estudo, em apenas um valor normalizado ($P_{ij_Amplitude}$).

$$P_{ij_Amplitude} = \frac{P_{ij} - \tilde{x}_j}{A_j}, \quad \text{Eq. (1)}$$

em que, $P_{ij_Amplitude}$ – produtividade normalizada pela amplitude no ponto i no ano j ; P_{ij} – produtividade no ponto i no ano j ; $\tilde{x}_j$ – mediana da produtividade no ano j ; A_j – amplitude das amostras no ano j .

Os atributos utilizados para gerar as ZMs foram selecionados pelo método de seleção de variáveis, proposto e implementado por Bazzi et al. (2013) no *software* SDUM (Software para definição de unidades de manejo). A análise de correlação espacial entre os atributos faz uso da correlação cruzada entre variáveis e gera a matriz de correlação espacial, com base no cálculo da estatística de autocorrelação bivariada de Moran (Equação 2; CZAPLEWSKI; REICH, 1993). Nesta matriz, é possível a identificação dos atributos que influenciaram de forma positiva ou negativa a produtividade, e se uma amostra está correlacionada espacialmente (autocorrelação espacial).

$$I_{YZ} = \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij} \times Y_i \times Z_j}{W \sqrt{m_Y^2 \times m_Z^2}}, \quad \text{Eq. (2)}$$

em que, I_{YZ} – nível de associação entre a variável Y e Z , variando de -1 a 1, sendo: correlação positiva $I_{YZ} > 0$, e correlação negativa $I_{YZ} < 0$; W_{ij} – matriz de associação espacial, calculado por $W_{ij} = (1/(1 + D_{ij}))$, D_{ij} é a distância entre os pontos i e j ; Y_i é valor da variável Y transformada no ponto i . A transformação ocorre, para se obter média zero, pela fórmula: $Y_i = (Y_i - \bar{Y})$, em que $\bar{Y}$ é a média amostral da variável Y ; e Z_j é o valor da variável Z

transformada no ponto j . A transformação ocorre, para se obter média zero, pela fórmula: $Z_i = (Z_i - \bar{Z})$, em que $\bar{Z}$ é a média amostral da variável Z ; W é a soma dos graus de associação espacial, obtidos através da matriz W_{ij} , para $i \neq j$; e m_Y^2 é a variância amostral da variável Y ; m_Z^2 – variância amostral da variável Z .

5. 2. 2 Análise geoestatística

Foram implementados procedimentos no software geoR, capazes de realizar análise geoestatística dos dados e identificar automaticamente o melhor ajuste para o semivariograma experimental (Apêndice A). Esta funcionalidade será integrada ao software SDUM.

Embora existam vários modelos de semivariogramas teóricos, apenas alguns são considerados como os mais comuns na explicação da maioria dos fenômenos espaciais. São eles: esférico, exponencial, e gaussiano e Família Matérn. A família dos modelos de Matérn é uma generalização de outros modelos teóricos e tem como característica fundamental a inclusão de um parâmetro (κ) que determina a suavização (MINASNY; MCBRATNEY, 2005). Para $\kappa = 0.5$, o modelo Matérn reduz-se ao modelo exponencial e para κ tendendo a infinito, equivale ao modelo gaussiano (URIBE-OPAZO; BORSSOI; GALEA, 2012). Tal parâmetro permite maior flexibilidade para caracterizar a variabilidade do semivariograma experimental e estabelece um equilíbrio entre os dois extremos.

Desta maneira, foram analisados seis diferentes modelos (esférico, gaussiano, exponencial, Matérn 1.0, Matérn 1.5 e Matérn 2.0) e dois métodos estatísticos de otimização de ajuste do semivariograma (OLS e WLS), totalizando então, 12 modelos diferentes. Para cada um dos modelos, 25 conjuntos diferentes de parâmetros (cinco valores iniciais para o parâmetro contribuição e cinco para alcance) foram utilizados, totalizando 300 diferentes ajustes analisados para identificar-se o melhor. Neste trabalho, WLS representa quadrados mínimos ponderados pelo número de pares.

Esta análise foi realizada por meio da implementação de procedimentos que realizam para cada modelo de ajuste, variação de valores dos parâmetros do semivariograma, e por validação cruzada (ISAAKS; SRIVASTAVA, 1989) calcula-se o erro médio (EM – Equação 4) e o desvio padrão do erro médio (DP_{EM} – Equação 5). Os valores de EM e DP_{EM} calculados para cada conjunto de parâmetros foram armazenados e utilizados para determinar o índice de seleção de interpoladores (ISI; Bier, 2015), possibilitando assim a identificação do melhor ajuste para cada um dos modelos analisados. Foi considerado o melhor modelo de ajuste para o semivariograma aquele que apresentou menor valor para o ISI calculado por meio da Equação 3.

$$ISI = \left\{ \frac{abs(EM)}{máx|_{i=1}^j[abs(EM)]} + \frac{[DP_{EM} - mín|_{i=1}^j(DP_{EM})]}{máx|_{i=1}^j[abs(DP_{EM})]} \right\}, \quad \text{Eq. (3)}$$

em que:

$$EM = \frac{1}{n} \sum_{i=1}^n Z(s_i) - \hat{Z}(s_i), \quad \text{Eq. (4)}$$

$$DP_{EM} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Z(s_i) - \hat{Z}(s_i))^2}, \quad \text{Eq. (5)}$$

em que, EM – erro médio; DP_{EM} – desvio-padrão dos erros médios; n – número de dados; $Z(s_i)$ – valor observado no ponto S_i ; $\hat{Z}(s_i)$ – valor predito por krigagem no ponto S_i ; $abs(EM)$ – valor em módulo do erro médio da validação cruzada; $mín|_{i=1}^j$ – menor valor encontrado entre os j modelos comparados; $máx|_{i=1}^j$ – maior valor encontrado entre os j modelos comparados.

5. 2. 3 Análise do melhor expoente para IDW

Foram desenvolvidas funções computacionais implementadas diretamente no banco de dados PostgreSQL por meio da linguagem procedural PL/pgSQL, com o objetivo de analisar e identificar o melhor expoente para o interpolador inverso da distância elevado a um expoente (IDW). IDW é um estimador determinístico que considera que os pontos mais próximos dos locais a serem mensurados são mais representativos do que aqueles mais afastados. Assim, a ponderação muda de acordo com a distância linear das amostras aos pontos não amostrados. A escolha do parâmetro do expoente de ponderação da distância do IDW pode afetar significativamente o resultado da interpolação.

Neste estudo, foram analisados doze diferentes valores utilizados como expoente para o IDW (0,5; 1,0; 1,5; 2,0; 2,5; 3,0; 3,5; 4,0; 4,5; 5,0; 5,5; e 6,0), e tal como na seleção do melhor modelo de semivariograma, o índice ISI (Equação 3) foi utilizado para identificar o melhor valor para o expoente do método de interpolação IDW.

5. 2. 4 Interpolação dos dados

É possível criar uma grade mais densa do que a amostral a partir da interpolação dos dados, o que possibilita a construção de mapas temáticos e ZMs com contornos mais suaves e maior redução da variância dos dados. Por isto, os dados dos atributos selecionados foram interpolados para uma grade de pixels de tamanho 5x5 m, por quatro diferentes métodos: krigagem ordinária (KRI), inverso da distância (IDW_1), inverso da distância elevado ao quadrado (IDW_2), e inverso da distância elevado ao melhor expoente (IDW_B), método proposto neste trabalho.

Utilizou-se o estimador clássico de Matheron (1963) para o cálculo das semivariâncias, e o alcance foi limitado à metade da distância máxima entre dois pontos (cutoff = 50%). O ajuste do semivariograma experimental foi realizado pelo melhor modelo identificado na análise geoestatística e os parâmetros obtidos foram utilizados no processo de interpolação por KRI (Apêndice B). E foram considerados os dados dos 10 vizinhos mais próximos ao pixel a ser estimado para a mensuração de valores para locais não amostrados realizada pelo método IDW.

A análise geoestatística e a escolha do melhor expoente para o método de interpolação inverso da distância foram realizadas automaticamente com a utilização de procedimentos implementados no software geoR, utilizando informações armazenadas no banco de dados PostgreSQL.

5. 2. 5 Geração e avaliação das zonas de manejo (ZMs)

É importante diferenciar os termos classe e zona de manejo. De acordo com Pedroso et al. (2010), ZM é uma área espacialmente contínua à qual pode ser aplicado um tratamento único, ao passo que uma classe de manejo é a área sobre a qual determinado tratamento pode ser aplicado, e esse pode ser composto por mais do que uma ZM.

As ZMs foram geradas no software SDUM com base nos valores interpolados por KRI, IDW₁, IDW₂, e IDW_B, para os atributos selecionados pela análise da matriz de correlação espacial. Foram definidos mapas com duas, três, e quatro classes pelo método de agrupamento Fuzzy C-Means (RUSPINI, 1969), usando-se o valor de ponderação de 1,3 conforme sugerido por Fridgen et al. (2004) e Odeh et al. (1992); e um erro que deve ser o mínimo possível, aqui definido em 0,0001, que corresponde ao critério de parada do algoritmo. Geralmente, são utilizados índices como FPI (Equação 7) e MPE (Equação 8) para identificar o número ideal de classes para subdividir uma área cujo agrupamento foi realizado por Fuzzy C-Means. Porém, os resultados dos índices podem ser contrastantes e necessitam de averiguações adicionais (FRIDGEN et al., 2004; VALENTE et al., 2012; XIANG et al., 2007). Neste trabalho, a identificação do número de classes estatisticamente diferentes para as áreas analisadas foi realizada pela análise de variância (ANOVA). Desta maneira, foi possível identificar se existe diferença média significativa entre as classes.

As ZMs foram avaliadas pela ANOVA e também pelos seguintes índices:

1) *Variance Reduction* (VR), Equação 6: este índice foi utilizado para verificar se, para a variável produtividade média normalizada, a soma das variâncias de cada classe é menor que a variância total (DOBERMANN et al., 2003; XIANG et al., 2007).

$$VR = \left(1 - \frac{\sum_{i=1}^c W_i \times V_{ZM_i}}{V_{\text{área}}} \right) \times 100, \quad \text{Eq. (6)}$$

em que, c – número de ZMs; W_i – proporção da área em cada ZM; V_{ZM_i} – variância dos dados de cada ZM; $V_{\text{área}}$ – variância da amostra dos dados para toda a área.

2) *Fuzziness Performance Index* (FPI), Equação 7: este índice permite determinar o grau de separação de um conjunto de dados entre os clusters difusos. Quando o valor de FPI é próximo a 0 (zero), classes distintas possuem pequeno grau de partilha dos dados, enquanto valores próximos a 1 indicam classes distintas com alto grau de partilha dos dados (FRIDGEN et al., 2004).

$$FPI = 1 - \frac{c}{(c-1)} \left[1 - \sum_{j=1}^n \sum_{i=1}^c (u_{ij})^2 / n \right], \quad \text{Eq. (7)}$$

em que, c – número de clusters; n – número de observações; u_{ij} – elemento ij da matriz de pertinência fuzzy u .

3) *Modified Partition Entropy* (MPE), Equação 8: este índice estima a quantidade de desorganização criada por um número específico de clusters. Valores de MPE próximos de 0 (zero) indicam melhor organização, enquanto valores próximos de 1 indicam maior desorganização (FRIDGEN et al., 2004).

$$MPE = \frac{- \sum_{j=1}^n \sum_{i=1}^c u_{ij} \log(u_{ij}) / n}{\log c}, \quad \text{Eq. (8)}$$

em que, c – número de clusters; n – número de observações; u_{ij} – elemento ij da matriz de pertinência fuzzy u .

4) *Smooth Index* (SI), Equação 9: este índice calcula a frequência de mudanças das classes do mapa temático nas direções horizontal, vertical e das diagonais, pixel a pixel. Para mapas com classes mais uniformes, o índice tende a valores próximos a 100%, e mapas com muitas alterações entre classes tendem a valores menores.

$$SI = 100 - \left(\frac{\sum_{i=1}^k C_{H_i}}{4P_H} + \frac{\sum_{j=1}^k C_{V_j}}{4P_V} + \frac{\sum_{l=1}^k C_{DR_l}}{4P_{DR}} + \frac{\sum_{m=1}^k C_{DL_m}}{4P_{DL}} \right) \times 100, \quad \text{Eq. (9)}$$

em que, C_{H_i} – número de alterações na linha i (horizontal); C_{V_j} – número de alterações na coluna j (vertical); C_{DR_l} – número de alterações na diagonal l (diagonal direita - DR);

C_{DL_m} – número de alterações na diagonal m (diagonal esquerda - DL); k – número máximo de pixels numa linha, coluna ou diagonal; P_H – possibilidade de alteração de pixels na horizontal; P_V – possibilidade de alteração de pixels na vertical; P_{DR} – possibilidade de alteração de pixels na diagonal direita (DR); P_{DL} – possibilidade de alteração de pixels na diagonal esquerda (DL).

5) *Improved Cluster Validation Index* (ICVI), Equação 10: Gavioli et al. (2016) propuseram o índice ICVI para resolver um possível problema quanto às estimativas de FPI, MPE e VR não indicarem resultado semelhante à definição de ZMs. Quanto maior o valor de VR e menores os valores de FPI e MPE, mais próximo de 0 (zero) será o ICVI. Quanto menor o ICVI melhor o método de agrupamento.

$$ICVI_i = \frac{1}{3} * \left(\frac{FPI_i}{Max\{FPI\}} + \frac{MPE_i}{Max\{MPE\}} + \left(1 - \frac{VR_i}{Max\{VR\}} \right) \right), \quad \text{Eq. (10)}$$

em que, FPI_i – valor de FPI do i -ésimo método de seleção de variáveis; MPE_i – é o valor do MPE do i -ésimo método de seleção de variáveis; VR_i – é o valor do VR do i -ésimo método de seleção de variáveis; $Max\{Índice X\}$ – representa o valor máximo do *Índice X* entre os n métodos de seleção de variáveis.

6) Análise de variância (ANOVA): os valores da produtividade foram comparados entre ZMs pelo teste de Tukey a fim de identificar se as sub-regiões geradas apresentaram diferenças significativas na produtividade média normalizada (primeiro confirmamos que não havia dependência espacial dentro de cada classe, por semivariogramas).

5.3 Resultados e discussão

O fluxograma ilustrado na Figura 3 apresenta a sequência de procedimentos executados neste trabalho: 3.1 Análise estatística dos dados amostrais e seleção de atributos para gerar ZMs (matriz de correlação espacial); 3.2 Seleção do melhor ajuste do semivariograma (análise geoestatística); 3.3 Seleção do melhor expoente para IDW; 3.4 Interpolação dos dados; 3.5 Definição, análise e avaliação de ZMs.

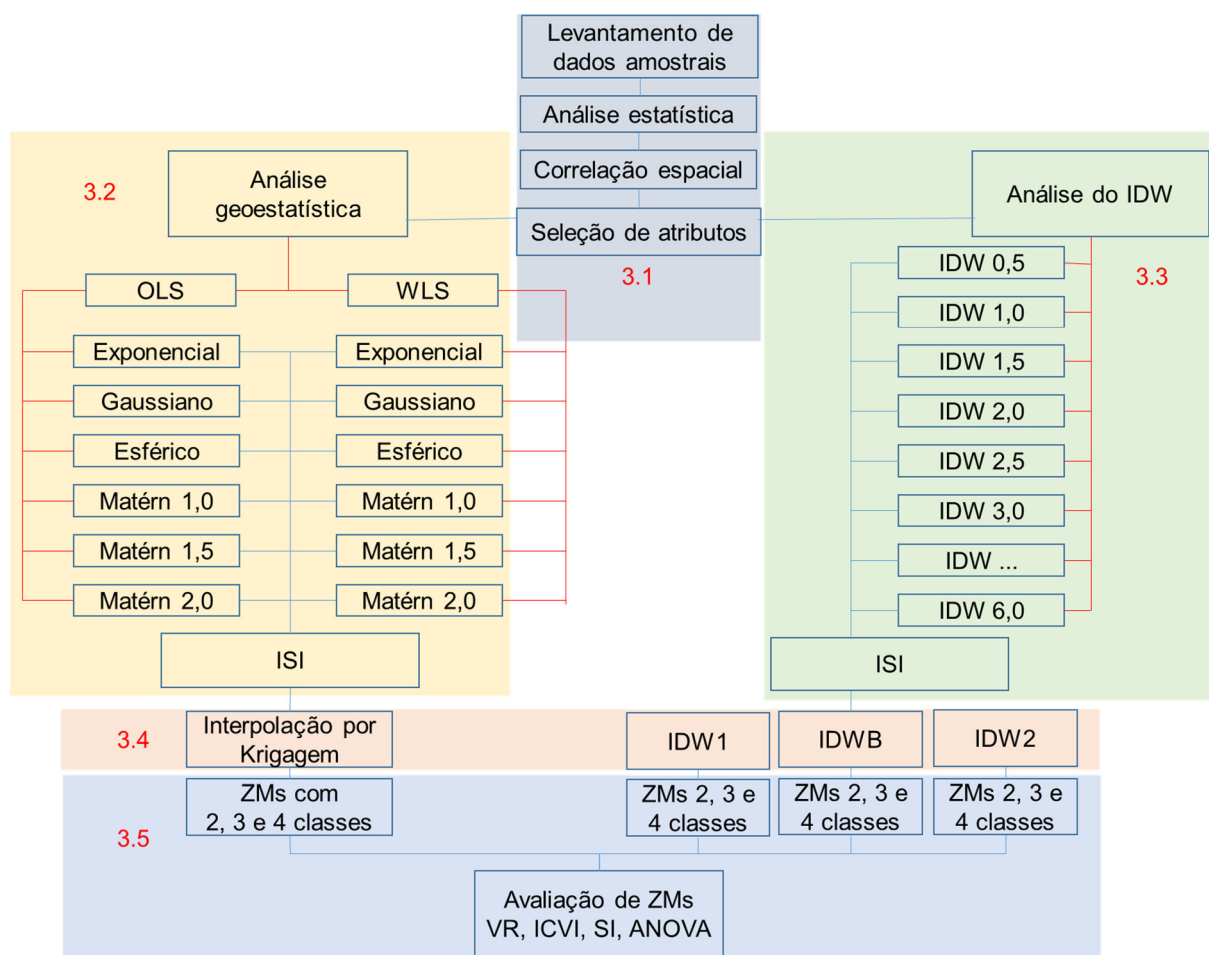


Figura 3 Fluxograma dos procedimentos executados neste trabalho.

5. 3. 1 Análise estatística dos dados amostrais e seleção de atributos para gerar ZMs

Os dados considerados *outliers* foram eliminados após análise estatística dos dados amostrais dos atributos estudados, isto é, com valores fora da faixa de média ± 3 desvios-padrão (CÓRDOBA et al., 2016). Foi então calculada a estatística de autocorrelação bivariada de Moran e construída a matriz de correlação espacial de Moran, visando identificar a autocorrelação espacial dos atributos e a correlação cruzada entre eles. Os atributos que apresentaram autocorrelação na área A foram: elevação, argila, silte, resistência do solo a penetração (RSP_0,0-0,1_2014), RSP_0,1-0,2_2015; na área B: elevação, argila, silte, RSP_0,0-0,1_2014 e RSP_0,0-0,1_2015, RSP_0,1-0,2_2013. Seguindo a metodologia apresentada por Bazzi et al. (2013), foram eliminados argila, silte e RSP_0,0-0,1_2014 na área A, e argila, silte, RSP_0,0-0,1_2014 e RSP_0,0-0,1_2015 na área B, por não apresentarem correlação com a produtividade média normalizada. Além destes atributos, na sequência, foi eliminada a areia na área B por apresentar correlação com a elevação posto que o valor da correlação da elevação com a produtividade era maior. Desta maneira, para gerar as ZMs, foram utilizados além dos dados de elevação também os dados dos atributos RSP_0,1-0,2_2015 (área A), RSP_0,1-0,2_2013 (área B).

De acordo com a análise estatística dos dados amostrais dos atributos selecionados para gerar as ZMs das duas áreas (Tabela 2), foi possível observar que o coeficiente de variação (CV) foi considerado baixo (PIMENTEL-GOMES, 2009). O atributo elevação foi o mais homogêneo com CV de variação abaixo de 2%, enquanto a RSP da área A foi considerada com CV médio. Também é possível observar que o solo da área A é mais compactado, porém, há de se considerar que apesar da mesma profundidade as medidas são de anos diferentes.

Tabela 2 Análise estatística dos dados amostrais de atributos selecionados para gerar zonas de manejo para as áreas A e B

Área	Atributo/unidade	n	Mínimo	Média	Mediana	Máximo	DP	CV(%)
A	Elevação (m)	42	351,7	362,8	363,2	372,7	5,46	1,5
A	RSP_0,1-0,2_2015 (MPa)	42	2,03	3,43	3,36	4,72	0,69	20,0
B	Elevação (m)	73	344,4	363,6	365,8	372,9	7,38	2,0
B	RSP_0,1-0,2_2013 (MPa)	73	1,95	2,56	2,58	3,07	0,26	10,1

n – número de amostras; DP – Desvio Padrão; CV (%) – Coeficiente de Variação; RSP – resistência do solo a penetração; m – metros; MPa – megapascal.

Mapas temáticos foram gerados para os principais atributos estudados, mas que não foram selecionados para definir as ZMs: produtividade média de soja, produtividade média do milho, RSP_média_0,0-0,1 m, RSP_média_0,1-0,2 m, RSP_média_0,2-0,3 m, areia, argila e silte. Os semivariogramas gerados para os oito atributos da área A (Figura 4) demonstram fraca ou nenhuma dependência espacial entre as amostras, ou seja, as amostras são independentes entre si. Ocorreu efeito pepita puro para os dados dos atributos produtividade média de soja, areia, argila e silte, portanto, os dados não são correlacionados, não existe dependência espacial. Assim, para esses atributos, os mapas temáticos não foram gerados por krigagem e sim, por média móvel, cujos mapas são compostos por apenas uma classe. Os mapas gerados para a produtividade média da soja na área A pelos métodos IDW₁ e IDW₂ demonstraram semelhança entre si e apresentaram maior produtividade na faixa central da área. Já a maior produtividade média do milho foi observada na região mais elevada (Leste), a qual reduziu conforme a elevação diminuiu. Os mapas gerados pelos métodos IDW apresentaram semelhança entre si também para os demais atributos. Os mapas gerados com dados interpolados por KRI foram mais suavizados e os limites entre classes melhor definidos.

Na área B, com exceção do atributo RSP_média_0,2-0,3 metros, em que ocorreu efeito pepita puro, a interpolação foi realizada por média móvel. Os semivariogramas gerados para todos os outros atributos apresentaram moderada ou forte dependência espacial sendo então previamente indicada a utilização de KRI como método de interpolação (Figura 5). As produtividades médias de milho e soja apresentam mapas com alguma semelhança. Logo, isso indica que, independente da cultura, a produtividade na mesma região da área foi equivalente. Também foi possível identificar mapas opostos de argila e areia.

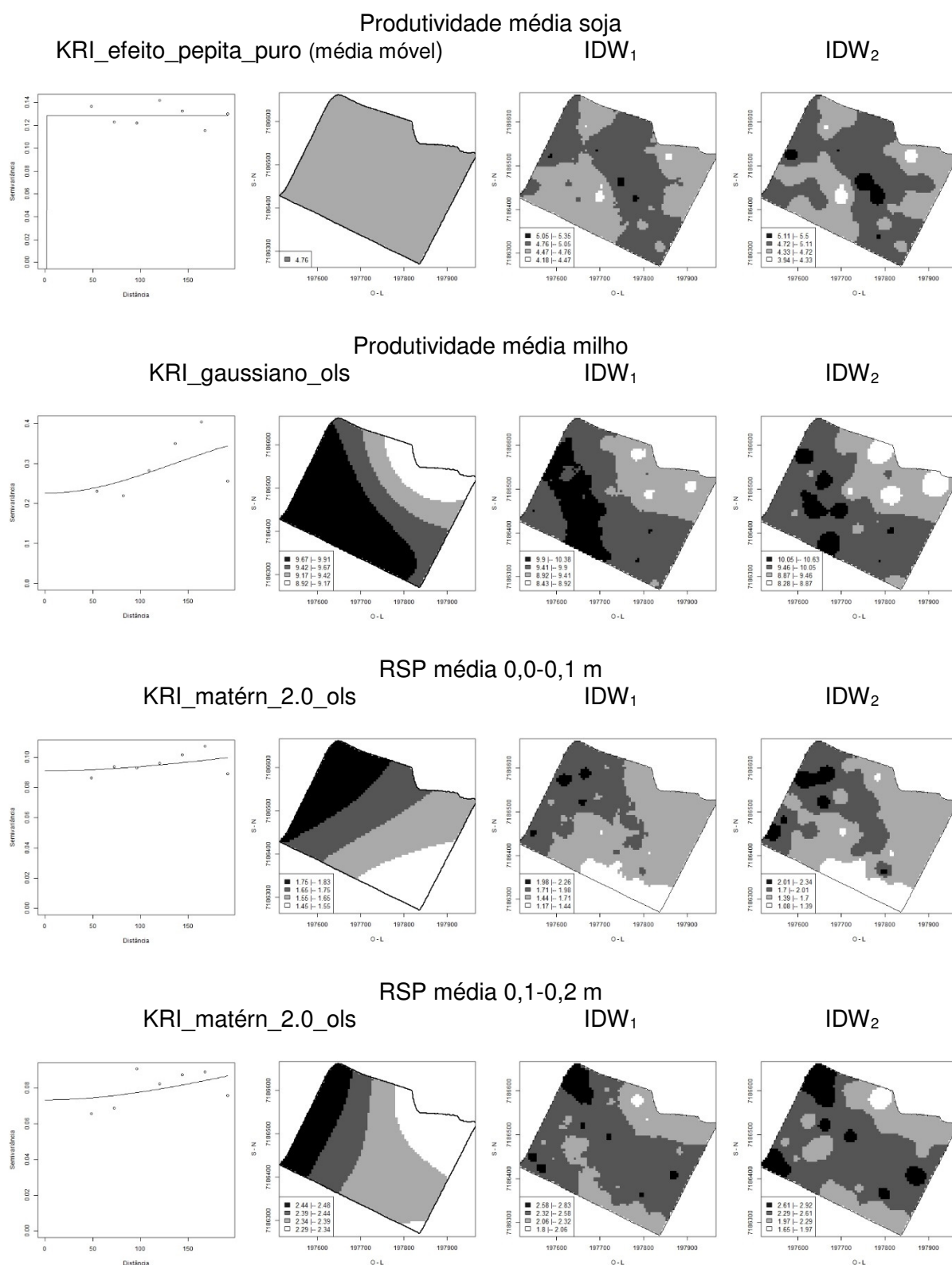


Figura 4 Semivariogramas e mapas temáticos de atributos da área A gerados com quatro classes, mensurados pelos métodos de interpolação: KRI - krigagem ordinária; IDW₁ – inverso da distância; IDW₂ – inverso da distância elevado ao quadrado (continua...).

(continuação)

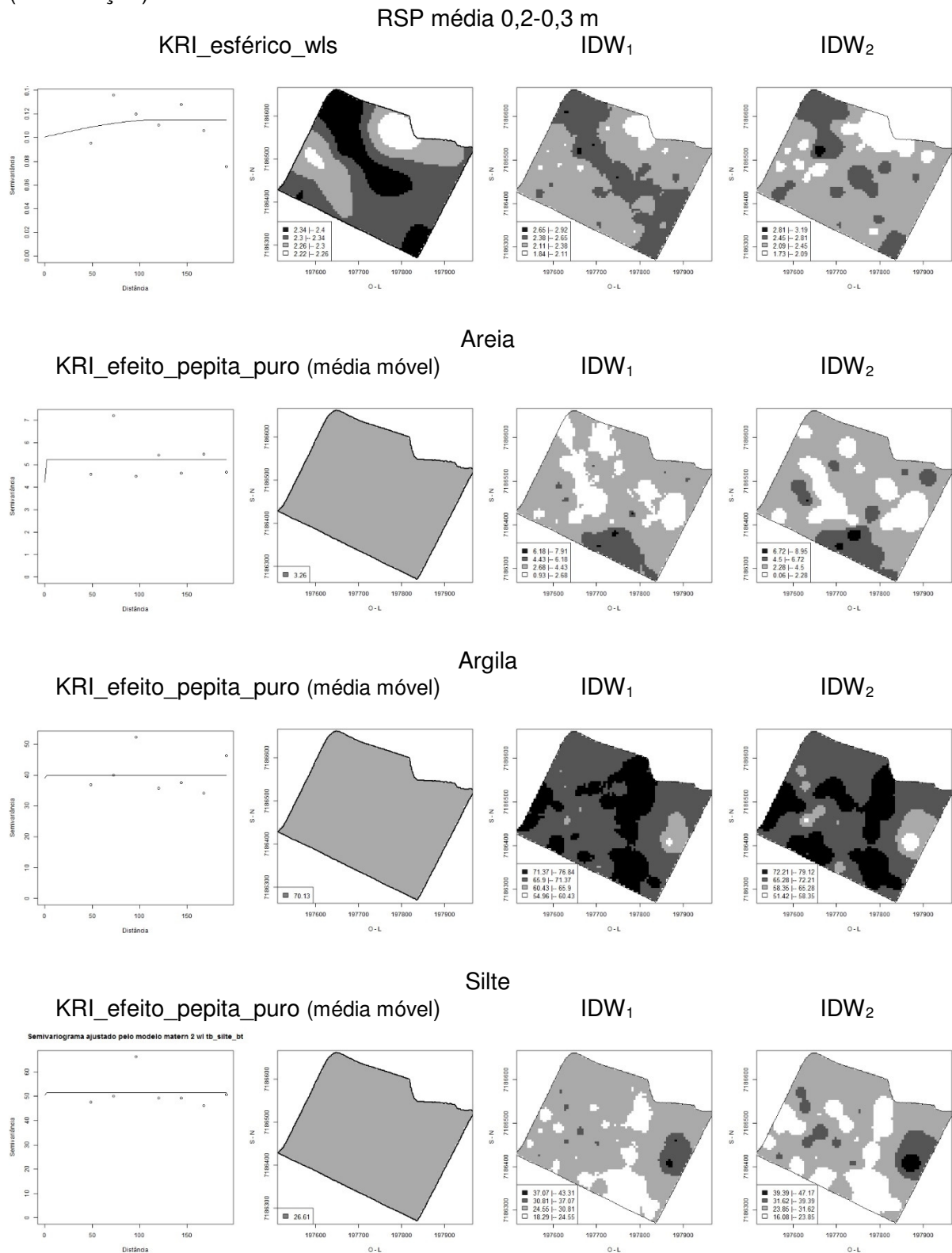


Figura 4 Semivariogramas e mapas temáticos de atributos da área A gerados com quatro classes, mensurados pelos métodos de interpolação: KRI - krigagem ordinária; IDW₁ – inverso da distância; IDW₂ – inverso da distância elevado ao quadrado.

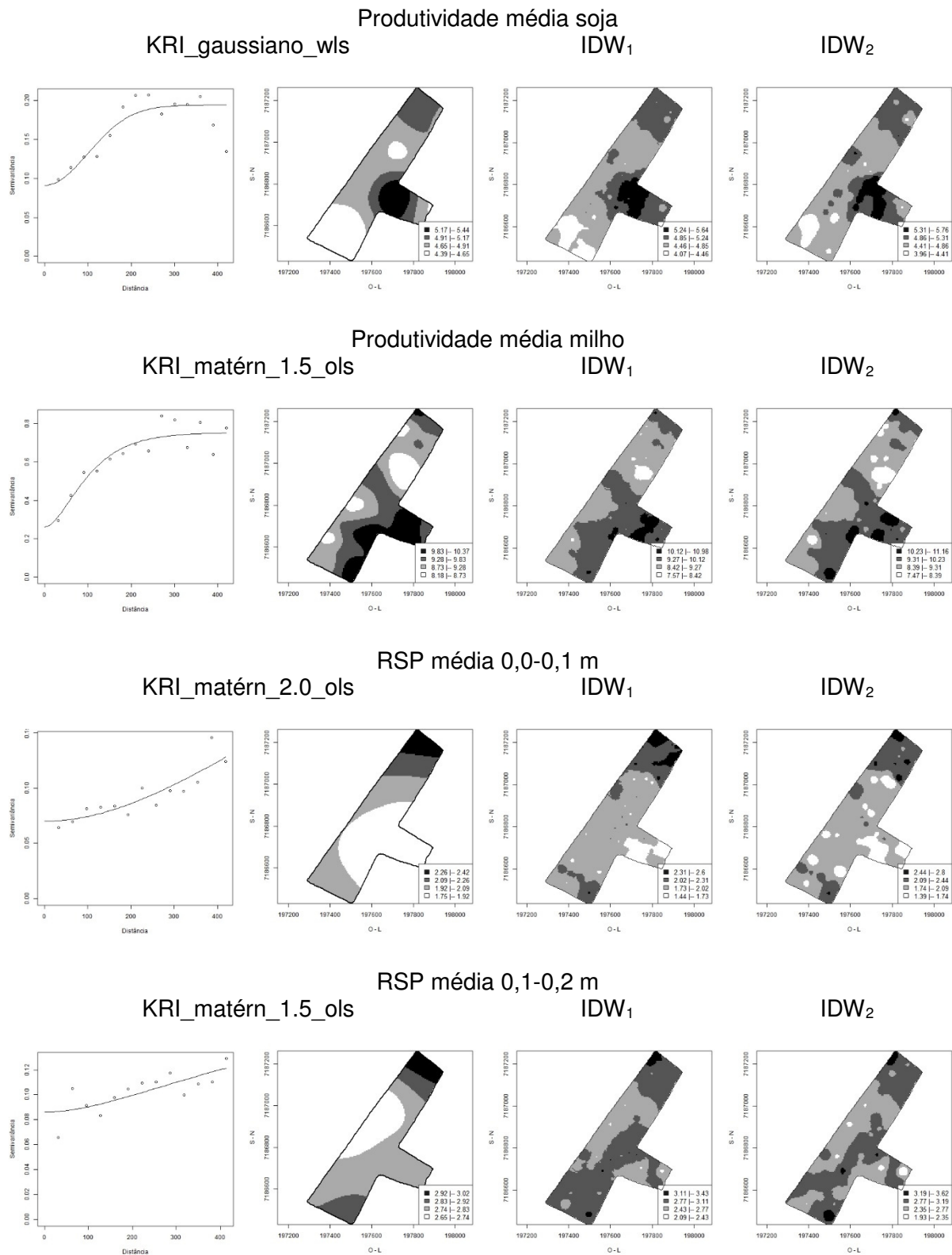


Figura 5 Semivariogramas e mapas temáticos de atributos da área B gerados com quatro classes, mensurados pelos métodos de interpolação: KRI - krigagem ordinária; IDW₁ – inverso da distância; IDW₂ – inverso da distância elevado ao quadrado (continua...).

(continuação)

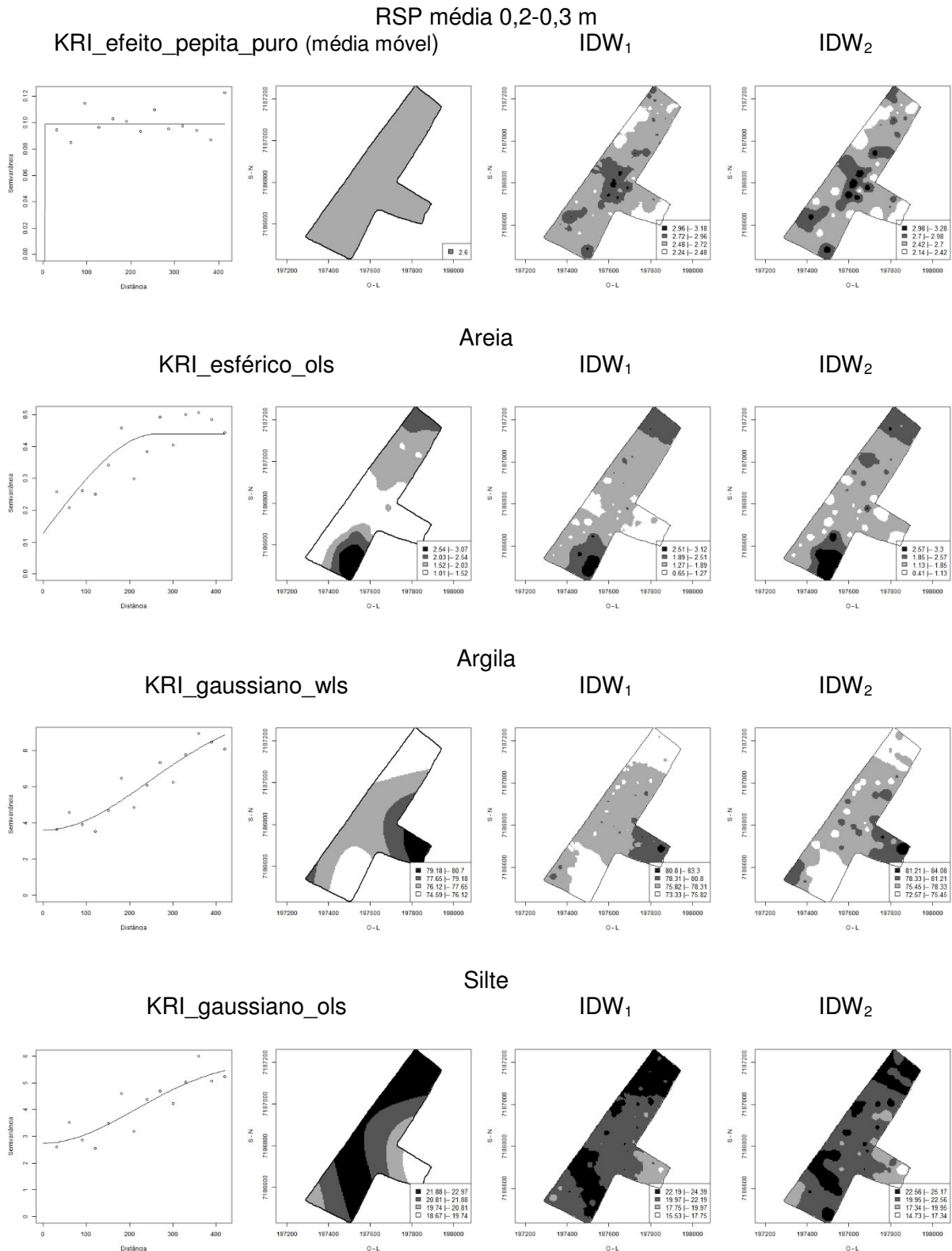
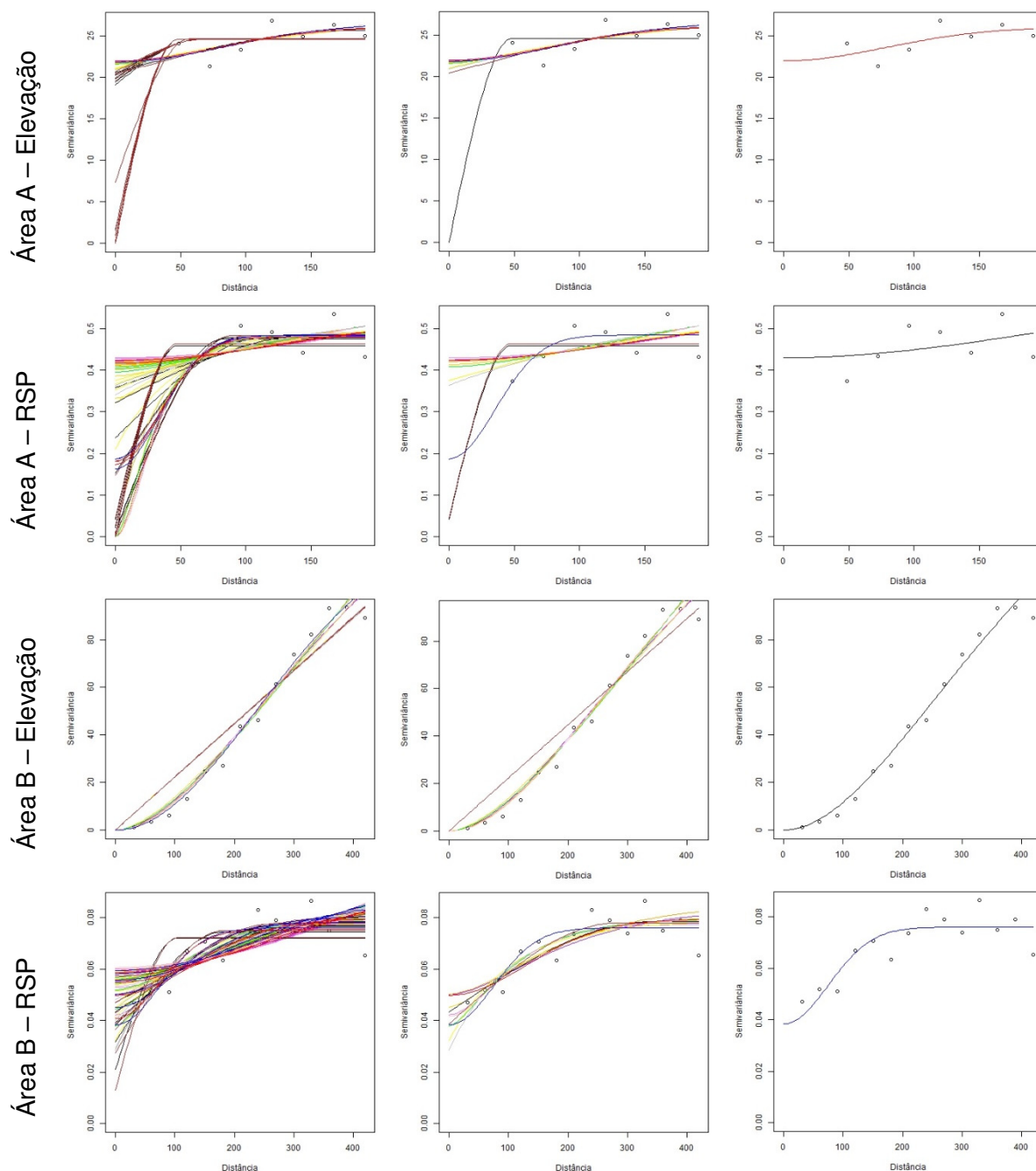


Figura 5 Semivariogramas e mapas temáticos de atributos da área B gerados com quatro classes, mensurados pelos métodos de interpolação: KRI - krigagem ordinária; IDW₁ – inverso da distância; IDW₂ – inverso da distância elevado ao quadrado.

5. 3. 2 Seleção do melhor ajuste do semivariograma (análise geoestatística)

Para cada variável selecionada para gerar as ZMs das áreas A e B, foram avaliadas as 300 diferentes combinações de ajustes para o semivariograma (Figura 6a) e identificado o melhor ajuste para cada modelo (Figura 6b) bem como o melhor modelo/ajuste geral (Figura 6c), os quais foram determinados com base na validação cruzada, a partir da utilização do ISI.



a) Geral (300 modelos) b) Melhores (12 modelos) c) Melhor modelo geral

Figura 6 Semivariogramas dos atributos elevação e RSP_0,1-0,2_2015 (RSP - resistência do solo a penetração) da área A, elevação e RSP_0,1-0,2_2013 da área B, ajustados por: todos os conjuntos de parâmetros e modelos testados (a); melhores parâmetros para cada um dos modelos testados (b); melhores parâmetros para o melhor modelo testado (c).

A distância mínima entre pontos amostrais foi de 25 m e a máxima 382 m para a área A, portanto, o *cutoff* utilizado foi de 191 m, caracterizando 8 lags com no mínimo 30 pares de pontos. Para a área B, a distância mínima entre pontos amostrais foi de 30 m e a máxima 838 m, com *cutoff* de 419 m, caracterizando 14 lags com no mínimo 30 pares de pontos.

O índice de seleção de interpolador (ISI, Figura 7) foi utilizado para selecionar o melhor modelo de ajuste do semivariograma. Na área A, o melhor modelo foi o gaussiano_ols (ISI = 0,036) para o atributo elevação e Matérn2_ols (parâmetro de forma $\kappa=2$) (ISI = 0,183) para RSP_0,1-0,2_2015 (Figura 7a). Para a área B (Figura 7b), o modelo Matérn2_wls (parâmetro de forma $\kappa=2$) (ISI = 0,760) para elevação e gaussiano_wls (ISI = 0,848) para RSP_0,1-0,2_2013, foram os melhores modelos ajustados.

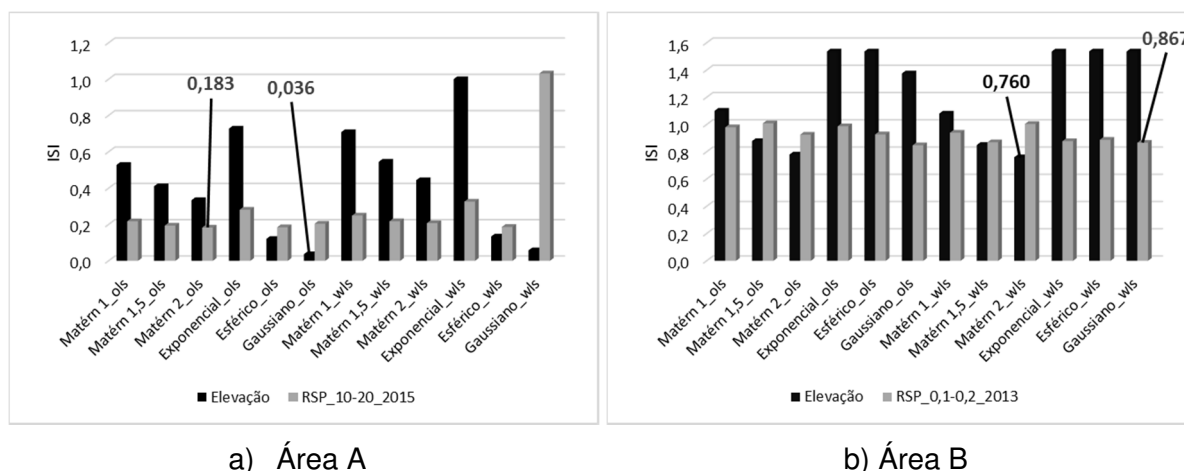


Figura 7 Índice de seleção de interpolador (ISI) para os 12 modelos de ajuste do semivariograma dos atributos elevação e resistência do solo à penetração (RSP) das áreas A (a) e B (b). Em destaque o melhor modelo de semivariograma correspondente ao menor ISI.

Outros modelos apresentaram ajustes e ISI semelhante aos escolhidos como melhores. Isso enfatiza (justifica) a qualidade dos procedimentos de geoestatística implementados que foram capazes de identificar, sem subjetivismo, os melhores ajustes para o semivariograma em cada um dos modelos analisados. E tal requisito é considerado importante para qualidade da interpolação por krigagem. Os parâmetros utilizados para identificar o melhor modelo de ajuste do semivariograma para cada um dos atributos escolhidos para gerar as ZMs são apresentados na Tabela 3. O IDE foi considerado como fraco para o atributo elevação na área A. Porém, na área B, o patamar (C_0+C_1) e a contribuição (C_1) obtiveram valores iguais, devido ao efeito pepita (C_0) resultar em valor igual a zero, acarretando em forte IDE, o que é recomendado pela literatura. O IDE foi considerado moderado para RSP em ambas as áreas (Tabela 3).

Tabela 3 Modelos e parâmetros de ajuste do semivariograma selecionados como melhores para interpolação por krigagem ordinária para os atributos elevação e resistência do solo à penetração (RSP) das áreas A e B

Áreas	Atributo	Modelo	C ₀	C ₁	C ₀ + C ₁	Alcance	IDE
A	Elevação	Gaussiano_OLS	21,9	4,1	26,0	110	84% (Fraca)
	RSP_0,1-0,2_2015	Matérn2_OLS	0,4	0,3	0,7	191	58% (Moderada)
B	Elevação	Matérn2_WLS	0,0	188,2	188,2	310	0% (Forte)
	RSP_0,1-0,2_2013	Gaussiano_WLS	0,05	0,03	0,1	200	58% (Moderada)

IDE – índice de dependência espacial; C₀ – efeito pepita; C₁ – contribuição; C₀+C₁ – patamar; RSP – resistência do solo à penetração; OLS – *ordinary least squares* (quadrados mínimos ordinários); WLS – *weight least squares* (quadrados mínimos ponderados).

5. 3. 3 Seleção do melhor expoente para o método de interpolação inverso da distância elevado a uma potência (IDW)

A escolha do melhor expoente para o interpolador IDW também foi determinada pelo índice ISI em que foram avaliados doze diferentes valores múltiplos de 0,5. Para os dados da área A, os expoentes que proporcionaram menor ISI foram obtidos quando foram utilizados os expoentes 1,0 e 0,5 para os atributos elevação e RSP_0,1-0,2_2015, respectivamente (Figura 8a). Para os dados da área B, o valor de ISI reduziu gradativamente com o acréscimo no expoente para ambos os atributos. O expoente 5,5 foi considerado o melhor para elevação e 3,5 para RSP, com ISI de 0,05 e 0,09, respectivamente (Figura 8b). Verificou-se, portanto, comportamento oposto de uma área com a outra, demonstrando que o melhor expoente pode variar de caso para caso, conforme identificado por Pasini et al. (2015) que encontraram valores diferentes como melhor expoente (4,0; 3,0; 6,0) para diferentes períodos de amostragem. Cecílio e Pruski (2003) registraram que o expoente 5,0 foi o melhor em seu estudo pluviométrico. Lu e Wong (2008) concluíram que o valor do melhor expoente varia conforme o padrão dos dados amostrais.

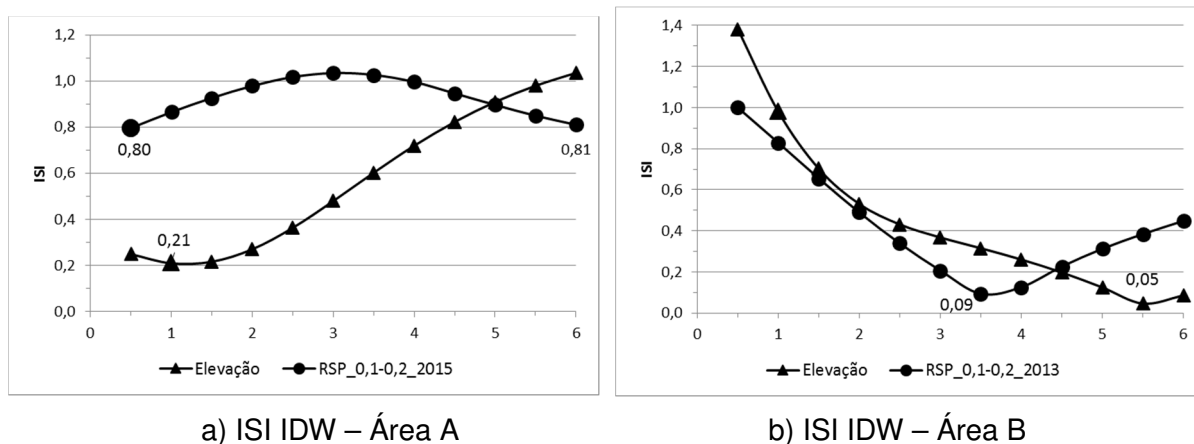


Figura 8 Gráfico de valores do índice de seleção de interpoladores (ISI) para escolha do melhor expoente para IDW (*inverse distance weighting* - inverso da distância elevado a uma potência) para os atributos selecionados para gerar zonas de manejo das áreas A (a) e B (b). Em destaque, o melhor expoente correspondente ao menor ISI.

5. 3. 4 Seleção do melhor de método de interpolação pelo uso do índice de seleção de interpoladores (ISI)

Quando foram confrontados os resultados obtidos pelos interpoladores krigagem (KRI), inverso da distância (IDW_1), inverso da distância elevado ao quadrado (IDW_2) e inverso da distância elevado ao melhor expoente (IDW_B) pelos índices EM, DP_{EM} e ISI, constatou-se que o ISI apresentou menor valor quando KRI foi utilizado para os dois atributos (Tabela 4) em ambas as áreas (A e B), com exceção da elevação na área A, onde o IDW_1 foi melhor. A justificativa é que a elevação na área A apresentou fraca dependência espacial (Tabela 3), assim, a krigagem deixou de ser uma boa alternativa. Estes resultados corroboram os encontrados por Bier (2015) em que KRI é melhor para a maioria dos atributos estudados, porém, não para todos. Na sequência, conforme constatado anteriormente, IDW_B foi considerado o melhor entre os IDWs, apresentando sempre o segundo melhor ISI para os atributos avaliados.

Tabela 4 Valores de erro médio (EM), desvio padrão do erro médio (DP_{EM}) e índice de seleção de interpolador (ISI) para cada método de interpolação utilizados para mensurar dados dos atributos selecionados para gerar zonas de manejo nas áreas A e B

Áreas	Atributos	Métodos	EM	DP_{EM}	ISI
A	Elevação	KRI_gaus_OLS	0,070	5,071	0,766
		IDW_1	0,066	5,014	<u>0,718</u>
		IDW_2	0,095	4,921	1,000
		$IDW_B = IDW_1$	0,066	5,014	<u>0,718</u>
	RSP_0,1-0,2_2015	KRI_matérn2_OLS	-0,001	0,711	<u>0,051</u>
		IDW_1	-0,020	0,720	0,900
		IDW_2	-0,023	0,724	1,017
		$IDW_B = IDW_{0.5}$	-0,018	0,726	0,829
B	Elevação	KRI_matérn2_WLS	-0,009	0,278	<u>0,312</u>
		IDW_1	0,029	1,490	1,813
		IDW_2	0,017	1,227	1,221
		$IDW_B = IDW_{5.5}$	0,002	1,045	0,563
	RSP_0,1-0,2_2013	KRI_gaus_WLS	0,000	0,244	<u>0,047</u>
		IDW_1	0,012	0,246	1,010
		IDW_2	0,007	0,251	0,595
		$IDW_B = IDW_{3.5}$	0,000	0,259	0,109

IDW – *inverse distance weighting* (inverso da distância elevado a uma potência); KRI – Krigagem; gaus – gaussiano; OLS – *ordinary least squares* (quadrados mínimos ordinários); WLS – *weight least squares* (quadrados mínimos ponderados); RSP – resistência do solo à penetração.

Analisando-se os mapas temáticos das variáveis interpoladas (Figura 9, área A e Figura 10, área B) verifica-se que os mapas gerados por KRI são mais suavizados, criando superfícies mais contínuas, com poucas ou sem manchas, e aristas suavizadas. Os mapas gerados por IDW_1 e IDW_B para o atributo elevação da área A são iguais, pois 1,0 foi considerado o melhor expoente, e o expoente 0,5 foi considerado o melhor para RSP.

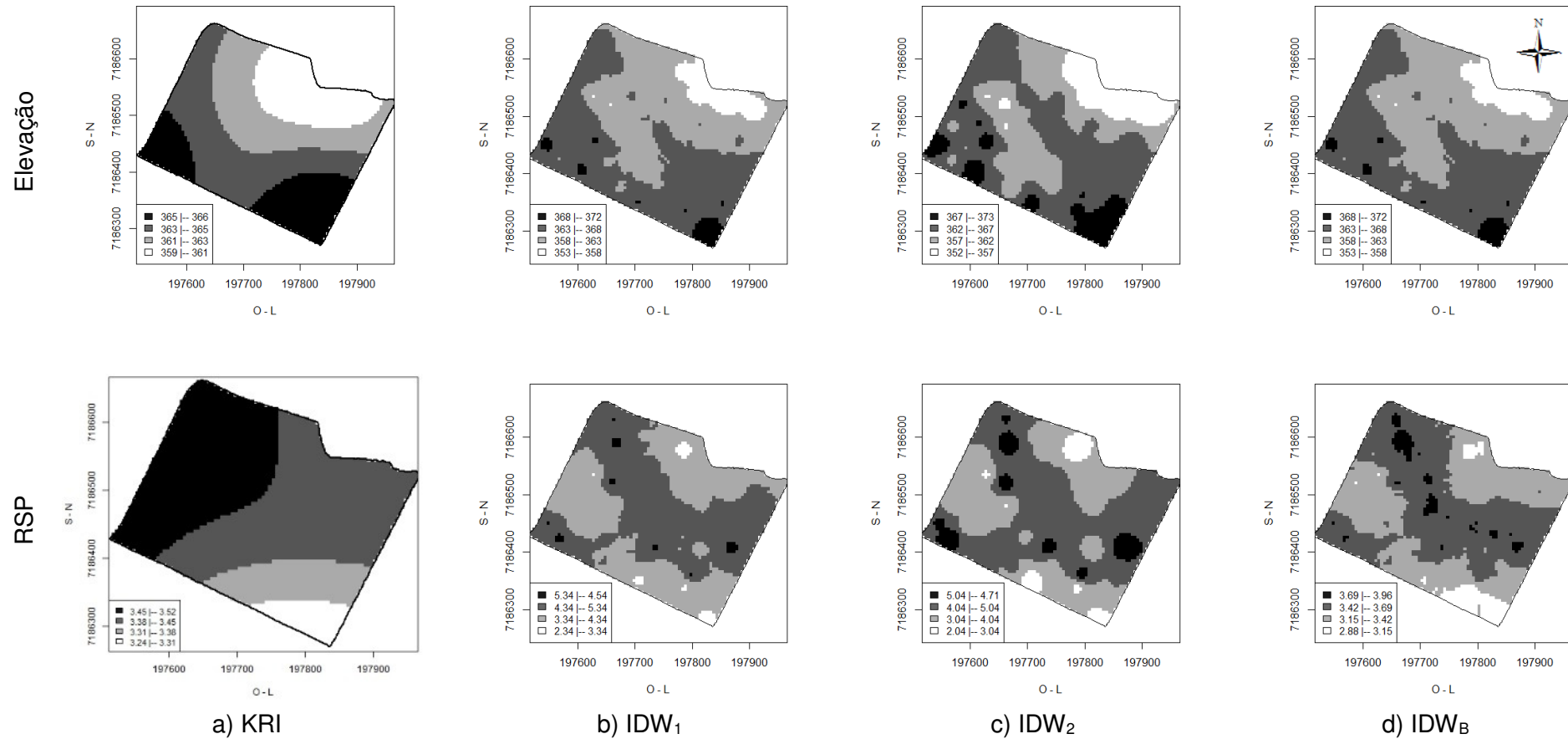


Figura 9 Mapas temáticos da área A dos atributos elevação e RSP_{0,1-0,2_2015} (RSP – resistência do solo à penetração) com quatro classes, mensurados pelos métodos de interpolação: KRI - krigagem ordinária (a); IDW₁ – inverso da distância (b); IDW₂ – inverso da distância elevado ao quadrado (c); IDW_{B(1,0, elevação; 0,5, RSP)} – inverso da distância elevado ao melhor expoente (d).

Os mapas para a área B do atributo elevação gerados pelos diferentes métodos de interpolação ficaram parecidos (Figura 10). Neste caso, o melhor expoente para IDW foi 5,5 (Figura 10d). Tal dado indica menor influência dos pontos mais distantes na mensuração dos valores, o que acarretou em mapas mais contínuos, semelhante ao da KRI (Figura 10a). O mapa que mais se diferenciou dos demais foi o gerado por IDW₁ (Figura 10b), menor expoente utilizado e portanto, o mais influenciado pelos pontos mais distantes na mensuração dos valores para locais não amostrados. O expoente 3,5 foi considerado o melhor para RSP, desta maneira, a influência dos pontos mais distantes foi pequena na mensuração dos valores, e já que é um atributo em que os valores amostrados variaram bastante, o mapa interpolado apresentou muitas manchas (Figura 10d), porém, é o mais parecido com KRI entre os IDWs.

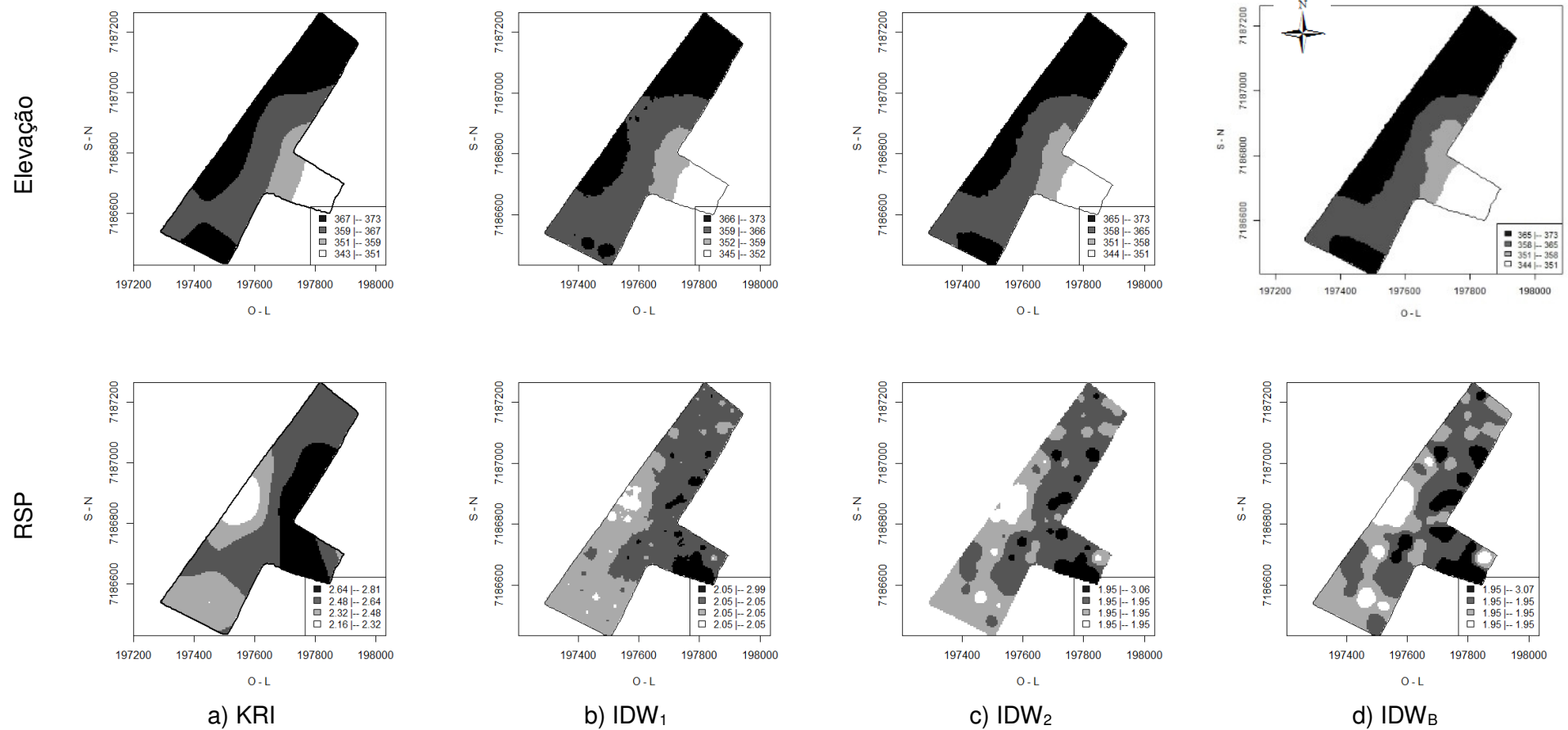


Figura 10 Mapas temáticos da área B dos atributos elevação e RSP_{0,1-0,2_2013} (RSP – resistência do solo à penetração) com quatro classes, mensurados pelos métodos de interpolação: KRI - krigagem ordinária (a); IDW₁ – inverso da distância (b); IDW₂ – inverso da distância elevado ao quadrado (c); IDW_{B(5,5, elevação; 3,5, RSP)} – inverso da distância elevado ao melhor expoente (d).

5. 3. 5 Definição, análise e avaliação das zonas de manejo (ZMs)

Os valores mensurados pelos quatro métodos de interpolação foram utilizados para subdividir as áreas em duas, três e quatro classes pelo agrupamento Fuzzy C-Means. As ZMs definidas para a área A (Figura 11), a partir dos dados interpolados por KRI, são visualmente diferentes dos demais métodos, principalmente quando divididas em quatro classes. Pode-se observar que os mapas mais semelhantes entre si são IDW_B (Figura 11d) e IDW_1 (Figura 11b), até pelo motivo do IDW_B ter sido utilizado com os expoentes 1,0 (elevação) e 0,5 (RSP_10-20_2015).

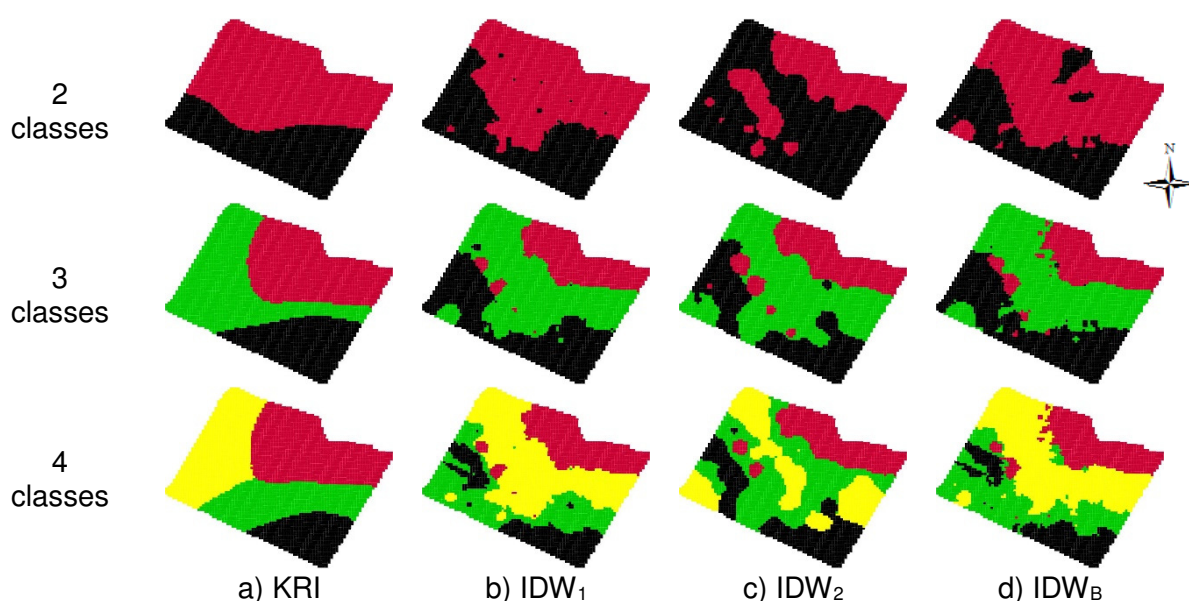


Figura 11 Zonas de Manejo definidas para a área A com atributos elevação e RSP_0,1-0,2_2015 (RSP – resistência do solo à penetração), com duas, três e quatro classes utilizando Fuzzy C-Means, mensurados pelos métodos de interpolação: KRI - krigagem ordinária (a); IDW_1 – inverso da distância (b); IDW_2 – inverso da distância elevado ao quadrado (c); $IDW_{B(1,0, \text{elevação}; 0,5, \text{RSP})}$ – inverso da distância elevado ao melhor expoente (d).

As ZMs da área B (Figura 12), definidas com duas classes a partir de dados interpolados por IDW_1 e IDW_2 , são semelhantes entre si. Da mesma maneira que por KRI e IDW_B , apesar de algumas manchas observadas nos mapas de IDW_B . Com três classes, independente do interpolador utilizado, os mapas são diferentes. E com quatro classes, os mapas apresentam semelhança entre si, porém com mais manchas nos gerados a partir dos dados interpolados por IDW (Figura 12).

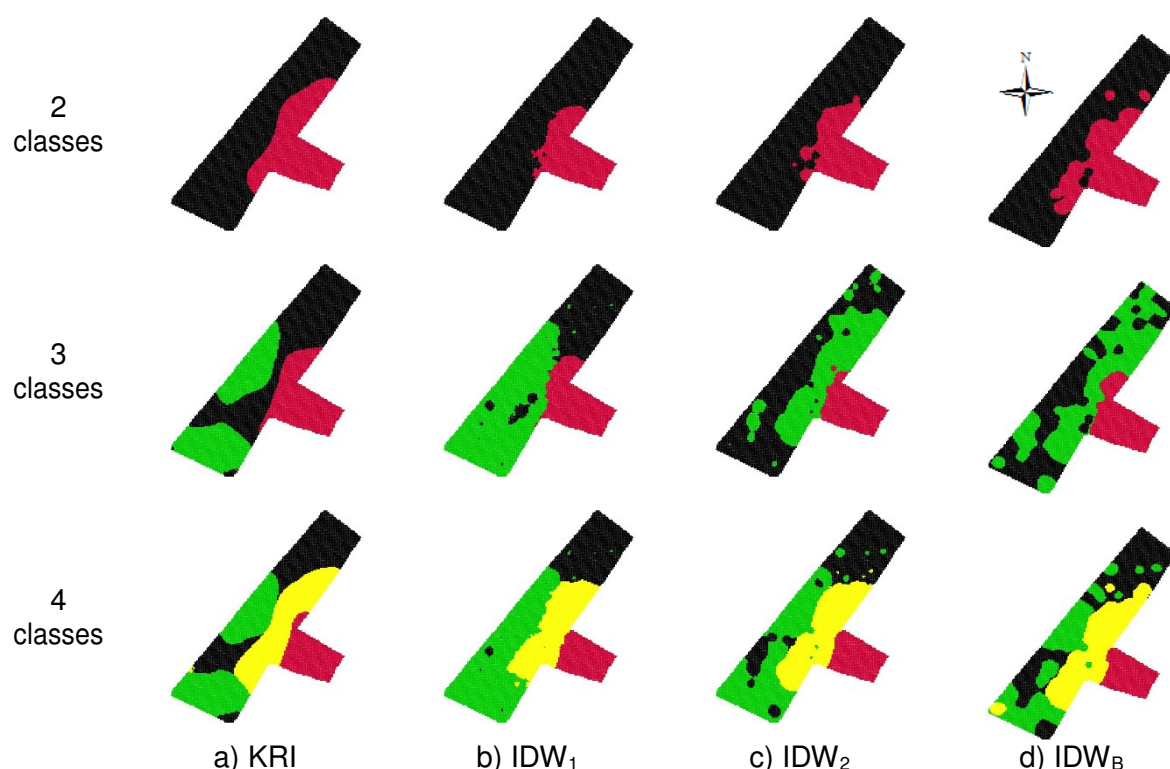


Figura 12 Zonas de Manejo definidas para a área B com atributos elevação e RSP_{0,1-0,2_2013} (RSP – resistência do solo à penetração), com duas, três e quatro classes utilizando Fuzzy C-Means, mensurados pelos métodos de interpolação: KRI - krigagem ordinária (a); IDW₁ – inverso da distância (b); IDW₂ – inverso da distância elevado ao quadrado (c); IDW_{B(5,5, elevação; 3,5, RSP)} – inverso da distância elevado ao melhor expoente (d).

A análise da ANOVA da variável produtividade média normalizada com os quatro métodos de interpolação e para 2, 3 e 4 classes de agrupamentos (Tabela 5) permitiu constatar que, para a área A, nenhum dos agrupamentos possuía todas as médias estatisticamente diferentes. Portanto, não é recomendável dividir a área A em diferentes ZMs, considerando-a como homogênea. Já para a área B, corroborando resultados apresentado por Schenatto et al. (2016), foi observado que a melhor maneira de agrupamento é subdividir a área em duas classes distintas, pois apresentaram resultados estatisticamente diferentes para todos os métodos de interpolação utilizados para mensurar valores dos atributos utilizados na geração das ZMs.

Tabela 5 Índices de avaliação calculados considerando zonas de manejo definidas com valores mensurados pelos métodos de interpolação krigagem ordinária (KRI), inverso da distância (IDW_1), inverso da distância elevado ao quadrado (IDW_2), inverso da distância elevado ao melhor expoente (IDW_B), para as áreas A e B.

Áreas	Classes	Métodos	ANOVA				VR(%)	ICVI	SI
			1	2	3	4			
A	2	KRI	A	a			-4,2	1,27	98,5
		IDW_1	A	a			0,4	0,91	96,7
		IDW_2	A	a			3,2	0,67	95,4
		IDW_B	A	a			-2,8	1,26	95,3
	3	KRI	A	a	a		4,9	0,76	97,2
		IDW_1	A	a	a		5,7	0,74	94,6
		IDW_2	A	b	ab		10,2	0,67	93,9
		IDW_B	A	a	a		3,3	0,79	93,4
	4	KRI	A	a	a	a	3,7	0,71	96,9
		IDW_1	A	a	a	a	2,3	0,92	92,0
		IDW_2	A	b	ab	ab	10,6	0,67	91,0
		IDW_B	A	a	a	a	9,3	0,64	90,8
B	2	KRI	A	b			13,9	0,53	98,8
		IDW_1	A	b			10,4	0,59	98,7
		IDW_2	A	b			9,5	0,62	98,6
		IDW_B	A	b			5,4	0,87	97,7
	3	KRI	A	b	a		17,3	0,62	97,6
		IDW_1	A	b	a		2,3	0,83	97,4
		IDW_2	Ab	a	b		7,6	0,83	95,6
		IDW_B	A	b	ab		6,4	0,88	94,5
	4	KRI	A	b	c	ac	14,2	0,62	97,1
		IDW_1	A	b	ac	abc	6,5	0,61	97,0
		IDW_2	A	b	ac	abc	3,0	0,87	95,7
		IDW_B	A	b	ac	abc	6,2	0,86	94,9

IDW_B área A = 1,0 e 0,5; área B = 5,5 e 3,5; VR – *variance reduction* (redução da variância); ICVI – *improved cluster validation index* (índice de validação de cluster); SI – *smooth index* (índice de suavidade).

Após a identificação do número ideal de classes para a divisão de cada área, foram analisados os índices VR e ICVI (Tabela 5). Considerando a divisão da área B em duas classes, pode-se observar que os melhores valores para os índices foram obtidos pelo método de interpolação KRI com maior VR (13,9%) e menor ICVI (0,53). Estes resultados corroboram os obtidos por meio do ISI (Tabela 4) na identificação do melhor método de interpolação para dados dos atributos selecionados para gerar as ZMs. Também foi utilizado o SI para medir a suavidade dos limites das ZMs e como era de se esperar, foi maior para KRI.

A fim de identificar se as ZMs apresentaram diferenças significativas quando avaliadas por outros atributos diferentes da produtividade, foi realizada a análise de variância (ANOVA) para dados de características químicas do solo (C, pH, Ca, Mg, K, Cu, Zn, Fe, Mn), RSP, textura do solo (argila, silte e areia) e elevação, para os agrupamentos de duas classes para ambas áreas (Tabela 6). Apesar da avaliação das ZMs da área A em relação à produtividade

não ter apresentado resultados estatisticamente diferentes, foram realizadas as avaliações com outros atributos a fim de confirmar a homogeneidade da área.

Além da produtividade, outros oito atributos (elevação, argila, RSP_0,1-0,2_2013, RSP_0,0-0,1_2014, RSP_0,0-0,1_2015, Cu, K e Mn) apresentaram diferença significativa entre as duas classes em que a área B foi subdividida, independente do método de interpolação utilizado (Tabela 6). Outros três atributos (areia, silte e Zn) também apresentaram diferença significativa pela utilização de três dos quatro métodos de interpolação testados.

Tabela 6 Análise da variância (ANOVA) para produtividade e atributos do solo das áreas A e B, considerando duas classes definidas com valores mensurados pelos métodos de interpolação krigagem ordinária (KRI), inverso da distância (IDW₁), inverso da distância elevado ao quadrado (IDW₂) e inverso da distância elevado ao melhor expoente (IDW_B)

Áreas	A				B			
Atributos	KRI	IDW ₁	IDW ₂	IDW _B	KRI	IDW ₁	IDW ₂	IDW _B
Produtividade normalizada					*	*	*	*
Elevação	*	*	*	*	*	*	*	*
Argila					*	*	*	*
RSP_0,1-0,2_2013			*		*	*	*	*
RSP_0,0-0,1_2014					*	*	*	*
RSP_0,0-0,1_2015				*	*	*	*	*
Cu		*	*		*	*	*	*
K					*	*	*	*
Mn					*	*	*	*
Areia				*	*	*	*	
Silte						*	*	*
Zn					*	*	*	
Mg					*			
RSP_0,0-0,1_2013		*	*					
RSP_0,2-0,3_2013								
RSP_0,1-0,2_2014				*				
RSP_0,2-0,3_2014								
RSP_0,1-0,2_2015		*	*	*				
RSP_0,2-0,3_2015				*				
C								
Ca								
Fe		*						
P		*	*					
pH	*							

*: existe diferença média entre as classes, considerando um nível de 5% de significância; RSP – resistência do solo à penetração.

Porém, para a área A, apenas os atributos elevação e RSP_0,1-0,2_2015, que foram os selecionados para gerar as ZMs, apresentaram diferença significativa a partir da ANOVA

por todos os métodos de interpolação utilizados. Outros dez atributos foram identificados como estatisticamente diferentes por um (RSP_0,1-0,2_2013, RSP_0,1-0,2_2014, RSP_0,1-0,2_2015, RSP_0,2-0,3_2015, areia, Fe, e pH) ou dois (RSP_0,0-0,1_2013, Cu, e P) métodos de interpolação.

Apreciou-se a dispersão entre as variáveis elevação e RSP_0,1-0,2_2013 (Figura 13) e observou-se que os dados interpolados por KRI apresentam envoltório mais suave que os outros métodos. Por isso, estão mais compactados e próximos aos centroides (destacados em preto) em comparação aos demais métodos IDWs.

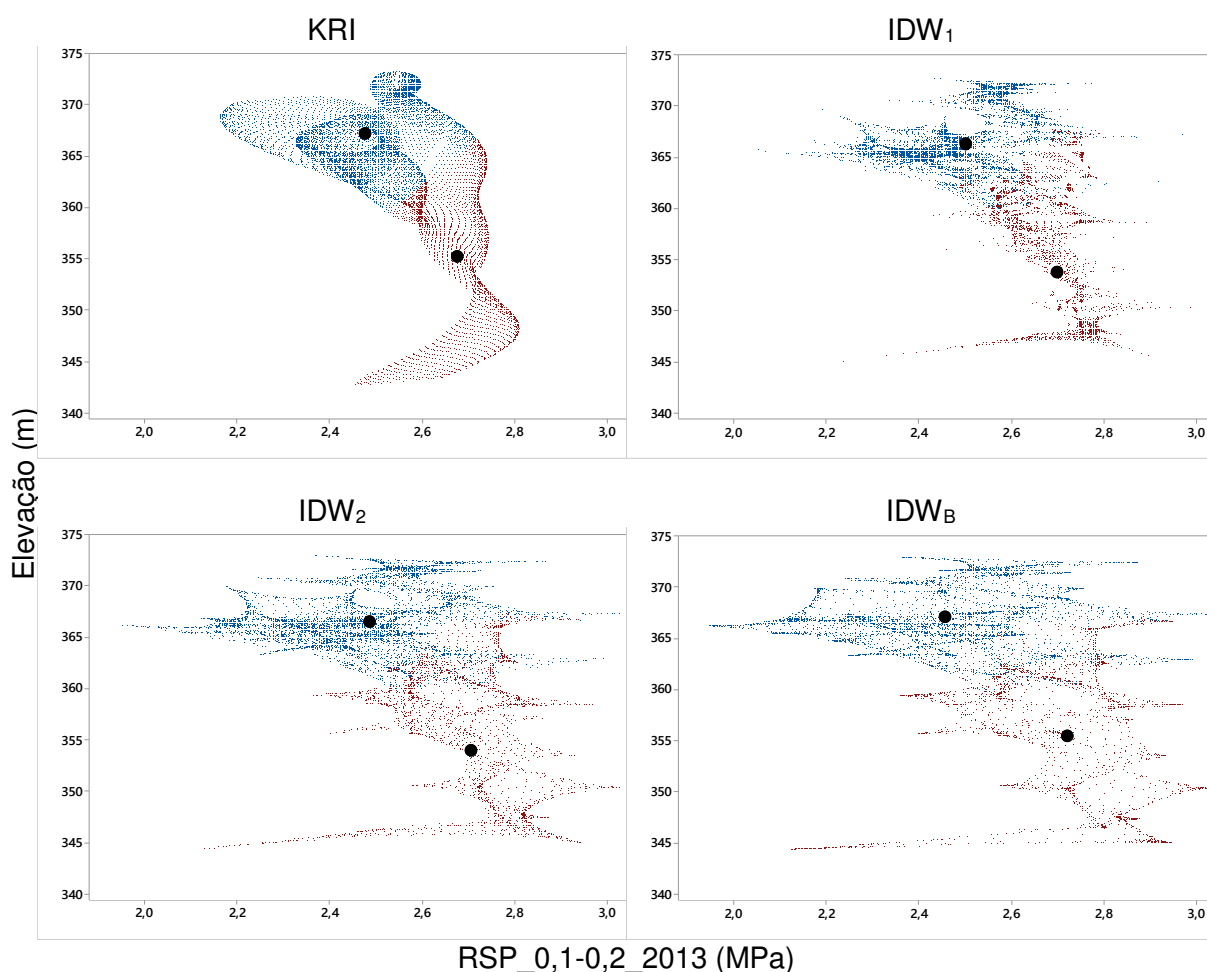


Figura 13 Dispersão dos pontos interpolados por KRI - krigagem ordinária (a); IDW₁ – inverso da distância (b); IDW₂ – inverso da distância elevado ao quadrado (c); IDW_{B(5,5, elevação; 3,5, RSP)} – inverso da distância elevado ao melhor expoente (d), da área B subdividida em duas classes. Os centroides são destacados em preto.

Em muitos casos, há inconsistência em se determinar qual o melhor método de interpolação para a geração de dados a serem utilizados para gerar ZMs (FRIDGEN et al., 2004). Desta maneira, é importante utilizar diferentes métodos de interpolação e comparar seus resultados. Como a krigagem é um método complexo, é importante utilizar o aplicativo proposto (Apêndices A e B) para identificar o melhor modelo de ajuste para o semivariograma, pois automatiza, agiliza e facilita a utilização da krigagem ordinária.

5.4 Conclusão

As funções computacionais implementadas foram eficientes e capazes de identificar o melhor ajuste para o semivariograma a ser utilizado na krigagem ordinária, bem como o melhor expoente a ser utilizado pelo método de interpolação inverso da distância elevado a uma potência (IDW).

O índice de seleção de interpoladores (ISI), utilizado em ambos os casos, demonstrou ser um índice com sensibilidade capaz de realizar as comparações necessárias para se definir o melhor método de interpolação.

No estudo de caso realizado, quando os dados apresentaram pelo menos moderada dependência espacial, o interpolador krigagem se mostrou o mais adequado, além de corroborar com a literatura. Caso contrário, o interpolador IDW₁, que neste caso específico foi considerado o melhor expoente, foi o mais adequado. Foi possível ainda afirmar que as zonas de manejo mais adequadas foram definidas com dados interpolados por krigagem, e considerado o melhor método de interpolação.

5.5 Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), à Fundação Araucária (FA), à Universidade Estadual do Oeste do Paraná (UNIOESTE) e à Universidade Tecnológica Federal do Paraná (UTFPR), pelo aporte financeiro e aos proprietários pela concessão das áreas experimentais.

5.6 Referências bibliográficas

- BAZZI, C. L.; SOUZA, E. G.; URIBE-OPAZO, M. A.; NÓBREGA, L. H. P.; ROCHA, D. M. Management zones definition using soil chemical and physical attributes in a soybean area. **Engenharia Agrícola**, v. 34, n. 5, p. 952–964, 2013.
- BIER V. A. **Construção e avaliação de mapas**. 2015. 82 f. Tese (Doutorado em Engenharia Agrícola) - Universidade Estadual do Oeste do Paraná, Cascavel, 2015.
- BORSSOI, J. A.; URIBE-OPAZO, M. A.; GALEA, M. Técnicas de diagnóstico de influência local na análise espacial da produtividade da soja. **Engenharia Agrícola**, v. 31, n. 2, p. 376–387, 2011.
- CAMPOS, M. C. C.; SOARES, M. D. R.; SANTOS, L. A. C.; OLIVEIRA, I. A.; AQUINO, R. E., BERGAMIN, A. C. Variabilidade espacial dos atributos físicos em um Argissolo Vermelho sob floresta. **Comunicata Scientiae**, v. 4, n. 2, p. 168–178, 2013.
- CECÍLIO, R. A.; PRUSKI, F. F. Interpolação dos parâmetros da equação de chuvas intensas com uso do inverso de potências da distância. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 7, n. 3, p. 501-504, 2003.

COLAÇO, A. F.; POVH, F. P.; MOLIN, J. P.; ROMANELLI, T. L. Energy assessment for variable rate nitrogen application. **Agric. Eng. Int: CIGR Journal**, v. 14, n. 3, p. 85-90, 2012.

CÓRDOBA, M. A.; BRUNO, C. I.; COSTA, J. L.; PERALTA, N. R.; BALZARINI, M. G. Protocol for multivariate homogeneous zone delineation in precision agriculture. **Biosystems Engineering**, v. 143, p. 95-107, 2016.

COUTO, E. G.; SCARAMUZZA, J. F.; MARASCHINI, L. Influência dos métodos de interpolação dos dados nos mapas usados na agricultura de precisão. **Simpósio Internacional de Agricultura de Precisão**, 2002.

CRESSIE, N. A. C. **Statistics for spatial data**. Revised ed. New York: Wiley-Interscience Publication, 1993. 900 p.

CZAPLEWSKI, R. L.; REICH, R. M. Expected value and variance of Moran's bivariate spatial autocorrelation statistic for a permutation test. **USDA Forest Service**, v. 309, p. 13, 1993.

DEMATTÊ, J. A. M.; DEMATTÊ, J. L. I.; ALVES, E. R.; BARBOSA, R. N.; MORELLI, J. L. Precision agriculture for sugarcane management: a strategy applied for brazilian conditions. **Acta Scientiarum. Agronomy**, v. 36, n. 1, p. 111, 2014.

DIGGLE, P. J.; RIBEIRO JR, P. J. **Model-based geostatistics**. 1º ed. New York: Springer, 2007. 228 p.

DOBERMANN, A.; PING, J. L.; ADAMCHUK, V. I.; SIMBAHAN, G. C.; FERGUSON, R. B. Classification of Crop Yield Variability in Irrigated Production Fields. **Agronomy Journal**, v. 95, n. 5, p. 1105–1120, 2003.

DOERGE, T. A. Management Zone Concepts. **Site-Specific Management Guidelines**, p. 1-4, 2000.

DRUCK, S.; CARVALHO, M. S.; CÂMARA, G.; MONTEIRO, A. M. V. **Análise Espacial de Dados Geográficos**. 1. ed. Brasília: Embrapa, 2004. 190 p.

FARACO, M. A.; URIBE-OPAZO, M. A.; SILVA, E. A. A. DA; JOHANN, J. A.; BORSSOI, J. A. Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. **Revista Brasileira Ciência do Solo**, v. 32, n. 1, p. 463-476, 2008.

FERGUSON, R. B.; HERGERT, G. W. Soil Sampling for Precision Agriculture. **Precision Agriculture**, p. 1-4, 2009.

FRANZEN, D. W.; HOPKINS, D. H.; SWEENEY, M. D.; ULMER, M. K.; HALVORSON, A. D. of Site-Specific Nitrogen Management. **Agronomy Journal**, v. 94, p. 381-389, 2002.

FRIDGEN, J. J.; KITCHEN, N. R.; SUDDUTH, K. A.; DRUMMOND, S. T.; WIEBOLD, W. J.; FRAISSE, C. W. Management zone analyst (MZA): software for management zone delineation. **Agronomy Journal**, v. 96, p. 100-108, 2004.

GAVIOLI, A.; SOUZA, E. G.; BAZZI, C. L.; GUEDES, L. P. C.; SCHENATTO, K. Optimization of management zone delineation by using spatial principal components. **Computers and Electronics in Agriculture**, v. 127, p. 302-310, 2016.

GRZEGOZEWSKI, D. M.; URIBE-OPAZO, M. A.; BASTIANI, F. DE; GALEA, M. Local influence when fitting Gaussian spatial linear models : an agriculture application. **Ciencia e Investigacion Agraria**, v. 40, n. 3, p. 523-535, 2013.

GUASTAFERRO, F.; CASTRIGNANÒ, A.; BENEDETTO, D.; SOLLITTO, D.; TROCCOLI, A.; CAFARELLI, B. A comparison of different algorithms for the delineation of management zones. **Precision Agriculture**, v. 11, n. 6, p. 600-620, 2010.

ISAAKS, E. H.; SRIVASTAVA, R. M. **Applied Geostatistics**. 1^a ed. New York: Oxford, 1989. 561 p.

KESTRING, F. B. F.; GUEDES, L. P. C.; BASTIANI, F. DE; URIBE-OPAZO, M. A. Comparação de mapas temáticos de diferentes grades amostrais para a produtividade da soja. **Engenharia Agrícola**, v. 35, n. 4, p. 733-743, 2015.

LU, G. Y.; WONG, D. W. An adaptive inverse-distance weighting spatial interpolation technique. **Computers & Geosciences**, v. 34, n. 9, p. 1044-1055, 2008.

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

MAZZINI, P. L. F.; SCHETTINI, C. A. F. Avaliação de metodologias de interpolação espacial aplicadas a dados hidrográficos costeiros quase-sinóticos. **Braz. J. Aquat. Sci. Technol.**, v. 13, n. 1, p. 53-64, 2009.

MELLO, J. M.; BATISTA, J. L. F.; RIBEIRO JR, P. J.; OLIVEIRA, M. S.. Ajuste e seleção de modelos espaciais de semivariograma visando à estimativa volumétrica de Eucalyptus grandis. **Scientia Forestalis**, v. 1, n. 69, p. 25-37, 2005.

MINASNY, B.; MCBRATNEY, A. B. The Matérn function as a general model for soil variograms. **Geoderma**, v. 128, n. 1, p. 192-207, 2005.

MOLIN, J. P.; RABELLO, L. M. Estudos sobre a mensuração da condutividade elétrica do solo. **Engenharia Agrícola**, v. 31, n. 1, p. 90-101, 2011.

MONDO, H. V. V.; GOMES JUNIOR, F. G.; PINTO, T. L. F.; MARCHI, J. L., MOTOMIYA, A. V. A.; MOLIN, J. P.; CICERO, S. M. Spatial variability of soil fertility and its relationship with seed physiological potential in a soybean production area. **Revista Brasileira de Sementes**, v. 34, n. 2, p. 193-201, 2012.

ODA-SOUZA, M.; BARBIN, D.; RIBEIRO JR, P. J.; STAPE, J. L. Aplicação de métodos geoestatísticos para identificação de dependência espacial na análise de dados de um ensaio de espaçamento florestal em delineamento sistemático tipo leque. **Revista Árvore**, v. 32, n. 3, p. 499-509, 2008.

ODEH, I. O. A.; MCBRATNEY, A. B.; CHITTLEBOROUGH, D. J. Soil pattern recognition with fuzzy c-means: Application to classification and soil-landform interrelationships. **Soil Sci. Soc. Am. J.**, v. 56, p. 505-516, 1992.

PASINI, M. P. B.; LÚCIO, A. D.; FRONZA, D.; WEBER, L. D. S. Krigagem ordinária e inverso da distância ponderada aplicados na espacialização da população da mosca-do-figo. **Revista Brasileira de Ciências Agrárias**, v. 10, n. 3, p. 452-459, 2015.

PEDROSO, M.; TAYLOR, J.; TISSEYRE, B.; CHARNOMORDIC, B.; GUILLAUME, S. A segmentation algorithm for the delineation of agricultural management zones. **Computers and Electronics in Agriculture**, v. 70, n. 1, p. 199-208, 2010.

PIMENTEL-GOMES, F. **Curso de estatística experimental**. 15^a ed. Piracicaba: FEALQ, 2009. 451 p.

REZA, S. K.; SARKAR, D.; DARUAH, U.; DAS, T. H. Evaluation and comparison of ordinary kriging and inverse distance weighting methods for prediction of spatial variability of some chemical parameters of Dhalai district, Tripura. **Agropedology**, v. 20, n. 1, p. 38-48, 2010.

ROCHA, M. M. DA; LOURENÇO, D. A.; LEITE, C. B. B. Aplicação de Krigagem com Correção do Efeito de Suavização em Dados de Potenciometria da Cidade de Pereira Barreto - SP. **Geologia USP Série Científica**, v. 7, n. 2, p. 37-48, 2007.

RUSPINI, E. H. A new approach to clustering. **Inf. Control.**, v. 15, p. 22-32, 1969.

SCHENATTO, K.; GODOY DE SOUZA, E.; BAZZI, C. L.; BETZEK, N. M.; GAVIOLI, A. Two approach comparison to define crop management zones (MZs). **African Journal of Agricultural Research**, v. 11, n. 38, p. 3680-3689, 2016.

SHANKEN, J.; ZHOU, G. Estimating and testing beta pricing models: Alternative methods and their performance in simulations. **Journal of Financial Economics**, v. 84, n. 1, p. 40-86, 2007.

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.

SOUZA, E. G.; BAZZI, C. L.; KHOSLA, R.; URIBE-OPAZO, M. A.; REICH, R. M. Interpolation type and data computation of crop yield maps is important for precision crop production. **Journal of Plant Nutrition**, v. 39, n. 4, p. 531-538, 2016.

TOMCZAK, M. Spatial Interpolation and its Uncertainty Using Automated Anisotropic Inverse Distance Weighting (IDW) - Cross-Validation / Jackknife Approach. **Journal of Geographic Information and Decision Analysis**, v. 2, n. 2, p. 18-30, 1998.

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

VALENTE, D. S. M.; QUEIROZ, D. M.; PINTO, F. D. A. D. C.; SANTOS, N. T.; SANTOS, F. L. Definition of management zones in coffee production fields based on apparent soil electrical conductivity. **Scientia Agricola**, v. 69, n. 3, p. 173-179, 2012.

WANDERLEY, H. S.; AMORIM, R. F. C.; CARVALHO, F. O. Interpolação espacial da precipitação no estado de Alagoas utilizando técnicas geoestatísticas. **Revista Campo Digital**, v. 8, n. 1, p. 34-42, 2013.

WOLLENHAUPT, N. C.; WOLKOWSKI, R. P.; CLAYTON, M. K. Mapping soil test phosphorus and potassium for variable-rate fertilizer application. **Journal of Production Agriculture**, v. 7, n. 4, p. 441-448, 1994.

XIANG, L.; YU-CHUN, P.; ZHONG-QIANG, G.; CHIN-JIANG, Z. Delineation and Scale Effect of Precision Agriculture Management Zones Using Yield Monitor Data Over Four Years. **Agricultural Sciences in China**, v. 6, n. 2, p. 180-188, 2007.

YAMAMOTO, J. K.; LANDIM, P. M. B. **Geoestatística - conceitos e aplicações**. 1ª ed. São Paulo: Oficina de Textos, 2013. 215 p.

6. ARTIGO 2 - MÉTODOS DE RETIFICAÇÃO PARA OTIMIZAÇÃO DE ZONAS DE MANEJO³

RESUMO: A utilização de zonas de manejo (ZMs) é uma abordagem para a agricultura de precisão em que se faz uso de máquinas convencionais, ao invés de máquinas de aplicação localizada. Independentemente do método utilizado para delimitar as ZMs, geralmente surgem pequenas manchas ou pixels isolados. Desta maneira, com o intuito de suavizar e melhorar a continuidade, foi implementada uma função computacional de retificação de ZMs. Esta função realiza a análise de 8 (máscara 3×3) ou 24 (máscara 5×5) pixels vizinhos utilizando as estatísticas mediana e moda para avaliar cada pixel no mapa e identificar se esse deve ser transferido para outra ZM. Depois de interpolados, dados amostrais de três áreas experimentais foram usados para criar agrupamentos por algoritmo Fuzzy C-Means e gerar mapas com duas, três, quatro e cinco classes. A função de retificação foi aplicada cinco vezes para cada mapa para a eliminação de pixels isolados e praticamente todas as manchas e proporcionar suavização nos limites entre classes, principalmente quando utilizados os métodos com máscara 5×5 . A suavização mais intensa foi identificada na primeira retificação, e aumentou conforme o número de classes foi incrementado. Os resultados demonstram que essas técnicas são uma forma eficaz de aumentar a continuidade e suavidade de ZMs, e são adequadas para aplicação em agricultura de precisão.

PALAVRAS-CHAVE: agricultura de precisão; índice de suavização; unidade de manejo.

ADJUSTMENT METHODS FOR MANAGEMENT ZONES OPTIMIZATION

ABSTRACT: The use of management zones (MZs) is an approach to precision agriculture that involves conventional machinery instead of site-specific machinery. Regardless of the method used to delimit these zones, patches or isolated pixels generally appear. In order to smooth the MZs and improve their continuity, a computational adjustment function was implemented. This function analyzes 8 (3×3 mask) or 24 (5×5 mask) close pixels using the statistical median and mode to evaluate whether each pixel in the map should be reassigned to a different MZ. Sample data from three experimental fields were used after interpolation. This occurs to create clusters by means of Fuzzy C-Means algorithm, and generate maps with two, three, four, and five classes. The adjustment function was applied five times to each map in order to eliminated isolated pixels as well as almost all patches, in addition to smoothing the ranges between classes, especially when using the 5×5 mask methods. The most intense smoothing was identified in the first adjustment and increased as the number of classes was increased. The results have shown that these techniques are an effective way to increase the continuity and MZ smoothness. They are also appropriate to be applied in precision agriculture.

KEYWORDS: management units; precision agriculture; smoothness index.

³ Artigo enviado em Janeiro de 2017 ao periódico internacional Computers and electronics in agriculture, com classificação A2 no QUALIS/CAPES na área de Ciências Agrárias.

6.1 Introdução

A variabilidade dos nutrientes no solo interfere diretamente na produtividade da lavoura e, segundo Mallarino e Wittry (2004), pode estar relacionada a fatores como variações no clima, topografia, matéria orgânica, vegetação e práticas de manejo do solo. Esta influência ocorre em diferentes escalas, logo dificulta o manejo e reduz a eficácia da operação se aplicado em escala uniforme (MOHAMMADI, 2002). Por isso, diferentes requisitos de fertilização podem ser necessários para a mesma área. Devido à interferência que gera no rendimento da cultura, é essencial se estudar e se conhecer esta variabilidade bem como a de outros componentes da produção agrícola, a fim de identificar técnicas de manejo adequadas (MZUKU et al., 2005; RODRIGUES et al., 2016).

O conhecimento da variabilidade espacial é obtido por monitoramento e medições para possibilitar o planejamento de eventuais correções de deficiências, principalmente quando se deseja realizar manejo localizado a fim de melhorar a qualidade do solo e consequentemente aumentar a produção (DAVIDSON, 2014; MZUKU et al., 2005). A informação sobre a variabilidade e distribuição espacial da produtividade pode ser usada no manejo localizado (agricultura de precisão), visando aumentar a eficiência do uso de fertilizantes e sustentabilidade ambiental, mas muitas vezes a um custo elevado (KHOSLA et al., 2002).

Normalmente, amostras de solo são analisadas para determinar os níveis de nutrientes no solo. A amostragem deve ser suficientemente densa para permitir a determinação da variabilidade de nutrientes no solo, de modo que os fertilizantes possam ser utilizados de forma rentável e ambientalmente responsável (FERGUSON; HERGERT, 2009; FRANZEN et al., 2002). Para determinar a densidade de amostras do solo de uma área, é necessário considerar a disponibilidade de tempo e o orçamento disponível para a amostragem.

Zona de manejo (ZM) é uma sub-região de um campo que expressa a combinação relativamente homogênea de fatores limitantes à produção para os quais uma taxa fixa de fertilizantes pode ser aplicada (BOBRYK et al., 2016; DOERGE, 2000; MORAL; TERRÓN; SILVA, 2010; MOSHIA et al., 2014). Assim, é possível melhorar a eficácia da utilização de nutrientes, manter ou aumentar o rendimento e reduzir potencialmente a sobrecarga de nutrientes para o meio ambiente (KHOSLA et al., 2002; MOSHIA et al., 2014).

É possível coletar uma amostra composta por zona para que se reduza o número de pontos amostrais utilizados na amostragem, (WOLLENHAUPT; WOLKOWSKI; CLAYTON, 1994), em que subamostras são coletadas em torno do ponto georreferenciado, e se garanta a avaliação correta dos nutrientes na área. A partir desta abordagem de amostragem, é possível a redução de custos de laboratório, além de se manter o nível de confiabilidade (FERGUSON; HERGERT, 2009; MALLARINO; WITTRY, 2004). Muitos estudos relacionados com a densidade de amostragem foram realizados (DEMATTÊ et al., 2014; DOERGE, 2000; FERGUSON; HERGERT, 2009; FRANZEN et al., 2002; JOURNELL; HUIJBREGTS, 1978;

WOLLENHAUPT; WOLKOWSKI; CLAYTON, 1994), resultando em densidade mínima sugerida de uma amostra ha^{-1} (FERGUSON; HERGERT, 2009) e 2,5 amostras ha^{-1} (DOERGE, 2000; JOURNELL; HUIJBREGTS, 1978), que devem ser compostas por, pelo menos, oito amostras individuais (WOLLENHAUPT; WOLKOWSKI; CLAYTON, 1994).

Dados amostrais de diversos atributos podem ser utilizados para definir ZMs, porém, é vantajoso utilizar um conjunto de dados multivariados de atributos que não variam sensivelmente ao longo do tempo (topografia, propriedades físicas do solo) que estejam correlacionados com a produtividade e assim produzam ZMs mais estáveis (DOERGE, 2000).

É importante diferenciar os termos utilizados no presente contexto. De acordo com Pedroso et al. (2010), ZM é uma área espacialmente contínua à qual pode ser aplicado um tratamento único, ao passo que uma classe de manejo é a área sobre a qual determinado tratamento pode ser aplicado, e esse pode ser composto por mais de uma ZM.

Várias abordagens foram desenvolvidas para definir ZMs e, de acordo com a técnica utilizada, podem ser classificadas como empírica ou de agrupamento (CÓRDOBA et al., 2016; GUASTAFERRO et al., 2010). A abordagem de agrupamento é a mais típica e tem sido usada por muitos pesquisadores com resultados satisfatórios (ARNO et al., 2011; MORAL; TERRÓN; SILVA, 2010; SALEH; BELAL, 2014; TAGARAKIS et al., 2013). No entanto, independentemente do processo utilizado para gerar as ZMs, pequenas manchas geralmente surgem dentro de uma classe. Muitas soluções foram propostas a fim de se amenizar este problema. Lowrance (2014) desenvolveu o software EZZone para suavizar ZMs pela fusão de pequenos polígonos com polígonos maiores de diferentes zonas. Outros pesquisadores tais como, Pramanik et al. (2013), propuseram a incorporação de manchas a zonas maiores pela atribuição de peso para pixels, isto é, a partir da análise da assimetria de pixels vizinhos. Para a eliminação de células isoladas ou manchas, Xiang et al. (2007) utilizaram filtragem pela maioria de pixels (moda). De acordo com Córdoba et al. (2016) e Dobermann et al. (2003), é recomendado o uso de filtros espaciais aplicados aos resultados para melhorar a continuidade do zoneamento.

O objetivo do presente estudo foi desenvolver e aplicar técnicas computacionais baseadas nas estatísticas moda e mediana, a fim de se retificar ZMs e torná-las mais contínuas, portanto, mais viáveis do ponto de vista operacional.

6.2 Materiais e métodos

Os dados foram coletados com base em grade amostral irregular em três áreas localizadas na zona rural dos municípios de Serranópolis do Iguaçu (Área A: 25°24'28" S, 54°00'17" O; Área B: 25°26'49" S, 54°04'59" O) e Céu Azul (Área C: 25°06'32" S, 53°49'55" O) na região Oeste do estado de Paraná, Brasil. As áreas têm sido cultivadas sob o sistema de plantio direto há mais de 10 anos com rotação de culturas de soja (*Glycine max* L.) e milho

(*Zea mays* L.) para fins comerciais. As Áreas A, B e C medem aproximadamente 9,9 ha, 19,8 ha, e 15,5 ha, respectivamente. Foram definidos 42 (4,2 amostras ha⁻¹), 58 (2,9 amostras ha⁻¹) e 40 (2,6 amostras ha⁻¹) pontos amostrais, respectivamente (Figura 1). Esta densidade de pontos de amostrais é suficiente para identificar a variabilidade de atributos, uma vez que é maior do que 2,5 amostras ha⁻¹ (DOERGE, 2000; JOURNELL; HUIJBREGTS, 1978). O mínimo de 40 amostras compostas foram utilizadas a fim de satisfazer a sugestão de Journal e Huijbregts (1978) para a utilização de pelo menos 30 pares de pontos para cada semivariância na definição do semivariograma.

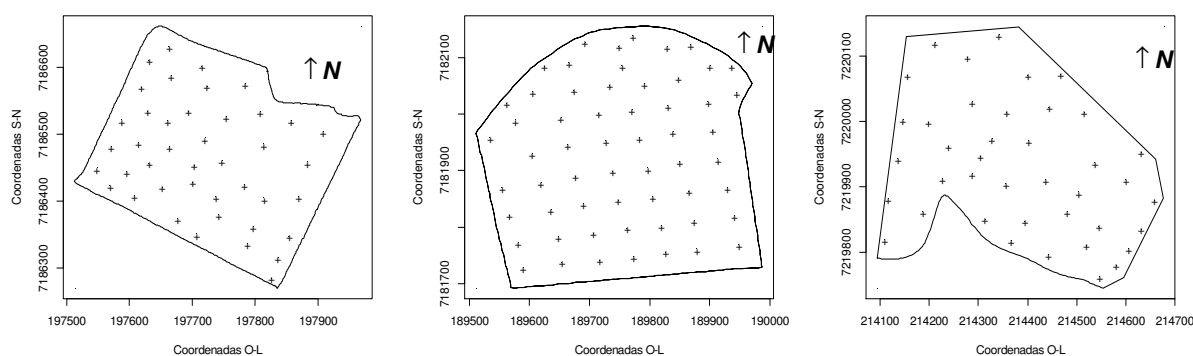


Figura 1 Localização dos pontos amostrais e contorno das áreas.

A localização dos pontos amostrais foi realizada por um aparelho receptor GPS da marca Trimble, modelo Juno SB e a elevação obtida pela utilização de estação total eletrônica Topcon GPT-7505. No entorno de cada ponto definido na grade amostral (a distância máxima de 3 m), foram realizadas quatro medições da resistência do solo à penetração (RSP) com a utilização de um penetrômetro eletrônico da marca Falker PenetroLOG. Calcularam-se posteriormente as médias das medições para representação do ponto amostral nas profundidades de 0-0,1; 0,1-0,2; e 0,2-0,3 m. Nos mesmos locais, foram coletadas oito subamostras de solo na profundidade de 0-0,2 m em um raio de 3 metros do ponto determinado na grade (adaptado de WOLLENHAUPT et. al., 1994). Posteriormente as amostras foram encaminhadas ao laboratório para análise e obtenção de dados de textura (argila, silte e areia), e características químicas do solo (C, pH, Ca, Mg, K, Cu, Zn, Fe, Mn). Porém, somente atributos estáveis, isto é, aqueles recomendados para estudar a definição de ZMs (DOERGE, 2000), foram analisados (Tabela 1).

Tabela 1 Atributos coletados em cada ponto amostral

Atributos/ Ano	Área A	Área B	Área C
Produtividade de Soja (2012)		X	X
Produtividade de Milho (2012)		X	
RSP 0,0–0,1 m, 0,1–0,2 m, 0,2–0,3 m (2012)	X	X	
Produtividade de Soja (2013)	X		X
Produtividade de Milho (2013)	X		
RSP 0,0–0,1 m, 0,1–0,2 m, 0,2–0,3 m (2013)	X		X
Produtividade de Soja (2014)	X		X
Produtividade de Milho (2014)	X		
RSP 0,0–0,1 m, 0,1–0,2 m, 0,2–0,3 m (2014)	X		X
Produtividade de Soja (2015)	X		
Argila	X	X	X
Silte	X	X	X
Areia	X	X	X
Elevação	X	X	X

RSP: resistência do solo à penetração.

Todos os dados de produtividade foram normalizados pela amplitude (Equação 1) para a eliminação dos efeitos das variações sazonais e da cultura utilizada, a fim de transformar o valor P_{ij} obtido em cada ponto amostral em um valor normalizado ($P_{ij_Amplitude}$) da seguinte maneira:

$$P_{ij_Amplitude} = \frac{P_{ij} - \tilde{x}_j}{A_j}, \quad \text{Eq. (1)}$$

em que, $P_{ij_Amplitude}$ – produtividade normalizada pela amplitude no ponto i no ano j ; P_{ij} – produtividade no ponto i no ano j ; $\tilde{x}_j$ – mediana da produtividade no ano j ; A_j – amplitude das amostras no ano j .

O programa “Software para definição de unidades de manejo (SDUM)” (BAZZI et al., 2013) foi utilizado para realizar a análise estatística dos dados amostrais, para interpolar os dados por inverso da distância e para definir e avaliar ZMs. O tamanho do pixel foi definido como 5 × 5 m a fim de proporcionar maior densidade de pontos por área e, por conseguinte, gerar ZMs com contorno mais suave.

O cálculo da estatística de autocorrelação bivariada espacial de Moran foi utilizado para avaliar as correlações espaciais entre os atributos analisados (CZAPLEWSKI; REICH, 1993). A matriz de correlação espacial permite a identificação dos atributos que influenciam a produtividade e se cada variável está correlacionada espacialmente com ela mesma (autocorrelação espacial). Em seguida, as variáveis a serem utilizadas no processo de agrupamento foram escolhidas usando a abordagem proposta por Bazzi et al. (2013): 1) eliminação de variáveis com dependência espacial não significativa; 2) eliminação de variáveis que não têm correlação com a variável dependente; e 3) eliminação das variáveis redundantes (as que são correlacionadas umas com as outras), dando preferência à

manutenção de variáveis que têm maior correlação com a variável dependente. Uma vez que as variáveis (atributos) foram selecionadas, mapas com duas, três, quatro e cinco classes foram criados usando o algoritmo de agrupamento Fuzzy C-Means (RUSPINI, 1969) com o valor de ponderação recomendado de 1,3 (FRIDGEN et al., 2004; ODEH; MCBRATNEY; CHITTLEBOROUGH, 1992), e o erro mínimo possível (isto é, o critério de parada do algoritmo) que foi definido como 0,0001.

Filtros de mediana e moda foram utilizados para a suavização das ZMs, com o objetivo de aumentar a continuidade das ZMs e reduzir sua fragmentação. Uma função computacional foi implementada diretamente no banco de dados PostgreSQL usando a linguagem procedural PL/pgSQL a fim de se analisar o comportamento de cada pixel em relação aos seus vizinhos usando as estatísticas mediana e moda. Para cada pixel do mapa, foram analisados 8 ou 24 pixels vizinhos mais próximos (usando máscara 3×3 ou máscara 5×5 , respectivamente). A grade consistiu da combinação de três pixels na direção horizontal por três pixels na direção vertical (máscara 3×3) ou cinco pixels na direção horizontal por cinco pixels na direção vertical (máscara 5×5) (Figura 2). Assim, foram implementados os seguintes modelos de classificação: mediana 3×3 , mediana 5×5 , moda 3×3 , e moda 5×5 .

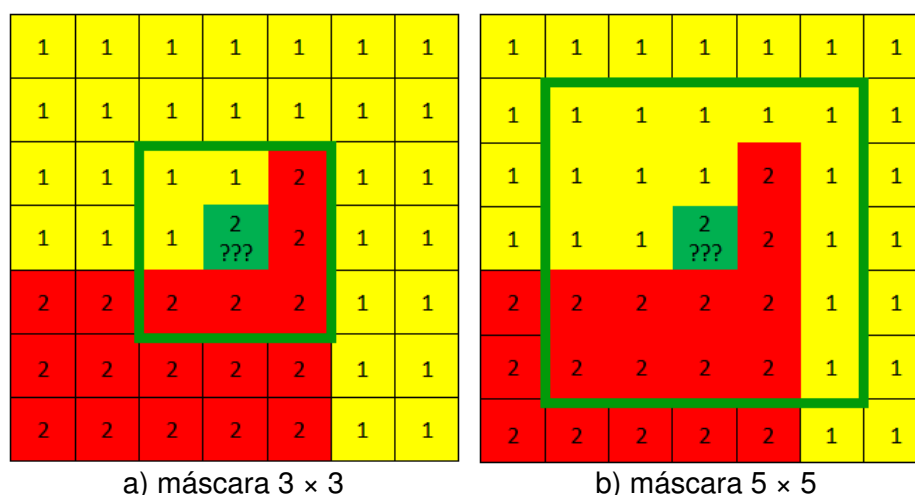


Figura 2 Definição da máscara de retificação: a) 8 pixels vizinhos (máscara 3×3); b) 24 pixels vizinhos (máscara 5×5).

Quatro séries de cinco mapas retificados foram geradas para cada um dos quatro mapas originais (duas, três, quatro e cinco classes). Cada série usou um dos quatro métodos de classificação (mediana 3×3 , mediana 5×5 , moda 3×3 , e moda 5×5), com um total de 80 mapas retificados para cada área analisada.

A medida em que as ZMs foram retificadas, a suavização foi determinada pelo índice de suavidade (*smooth index*, SI) (GAVIOLI et al., 2016), que representa o resultado da frequência de alterações de classes no mapa, nas direções horizontal, vertical e diagonais (Equação 2), pixel a pixel.

$$SI = 100 - \left(\frac{\sum_{i=1}^k C_{H_i}}{4P_H} + \frac{\sum_{j=1}^k C_{V_j}}{4P_V} + \frac{\sum_{l=1}^k C_{DR_l}}{4P_{DR}} + \frac{\sum_{m=1}^k C_{DL_m}}{4P_{DL}} \right) \times 100, \quad \text{Eq. (2)}$$

em que, C_{H_i} – número de alterações na linha i (horizontal); C_{V_j} – número de alterações na coluna j (vertical); C_{DR_l} – número de alterações na diagonal l (diagonal direita - DR); C_{DL_m} – número de alterações na diagonal m (diagonal esquerda - DL); k – número máximo de pixels em uma linha, coluna ou diagonal; P_H – possibilidade de alteração de pixels na horizontal; P_V – possibilidade de alteração de pixels na vertical; P_{DR} – possibilidade de alteração de pixels na diagonal direita (DR); e P_{DL} – possibilidade de alteração de pixels na diagonal esquerda (DL).

O grau de semelhança entre o mapa original e os mapas retificados foi avaliado pelo coeficiente Kappa (COHEN, 1960), que é frequentemente utilizado como índice para avaliar a acurácia espacial entre dois mapas com base na classificação proposta por Landis e Koch (1977): muito forte quando o índice varia de 0,81-1,00; forte: 0,61-0,80; moderada: 0,41-0,60; e fraca: 0,21-0,40. Além disso, admite-se que não existe concordância se o índice é menor ou igual a 0,20.

6.3 Resultados e discussão

Os resultados da matriz de correlação espacial obtidos utilizando o SDUM demonstraram que os seguintes atributos tiveram autocorrelação: elevação, argila, silte, resistência do solo à penetração (RSP) 0,0-0,1 m em 2013, RSP 0,0-0,1 m em 2014 e RSP 0,2-0,3 m em 2012 na Área A; elevação, areia, argila, silte e RSP 0,2-0,3 m em 2012 na Área B; e elevação, argila, silte, RSP 0,0-0,1 m em 2013 e RSP 0,2-0,3 m em 2014 na Área C. De acordo com a metodologia proposta por Bazzi et al. (2013), os seguintes atributos foram eliminados por não apresentar correlação com a produtividade média normalizada: argila, silte e RSP 0,0-0,1 m em 2014 na Área A; e RSP 0,2-0,3 m em 2014 na Área C. Além disso, os seguintes atributos foram eliminados por apresentarem correlação com a elevação (a correlação de elevação com a produtividade foi maior): RSP 0,2-0,3 m em 2012 na Área A; areia, argila e silte na Área B; e silte e argila na Área C. Assim, foram selecionados os seguintes atributos para definir as ZMs: elevação e RSP 0,0-0,1 m em 2013 nas Áreas A e C, e elevação e RSP 0,2-0,3 m em 2012 na Área B.

Em geral, a qualidade dos mapas gerados piorou à medida que o número de classes aumentou (Figura 3), corroborando os resultados apresentados por Lowrance (2014). Na Área A, pode-se observar que muitas zonas de uma mesma classe foram geradas, especialmente em mapas com quatro ou cinco classes, resultando em mapas com má aparência visual e

pouca continuidade (Figura 3). Manchas e irregularidades nas bordas foram observadas mesmo no mapa com duas classes, que aumentaram no mapa com três classes. Nas Áreas B e C, independentemente do número de classes, as ZMs apresentaram apenas algumas manchas e irregularidades nas bordas.

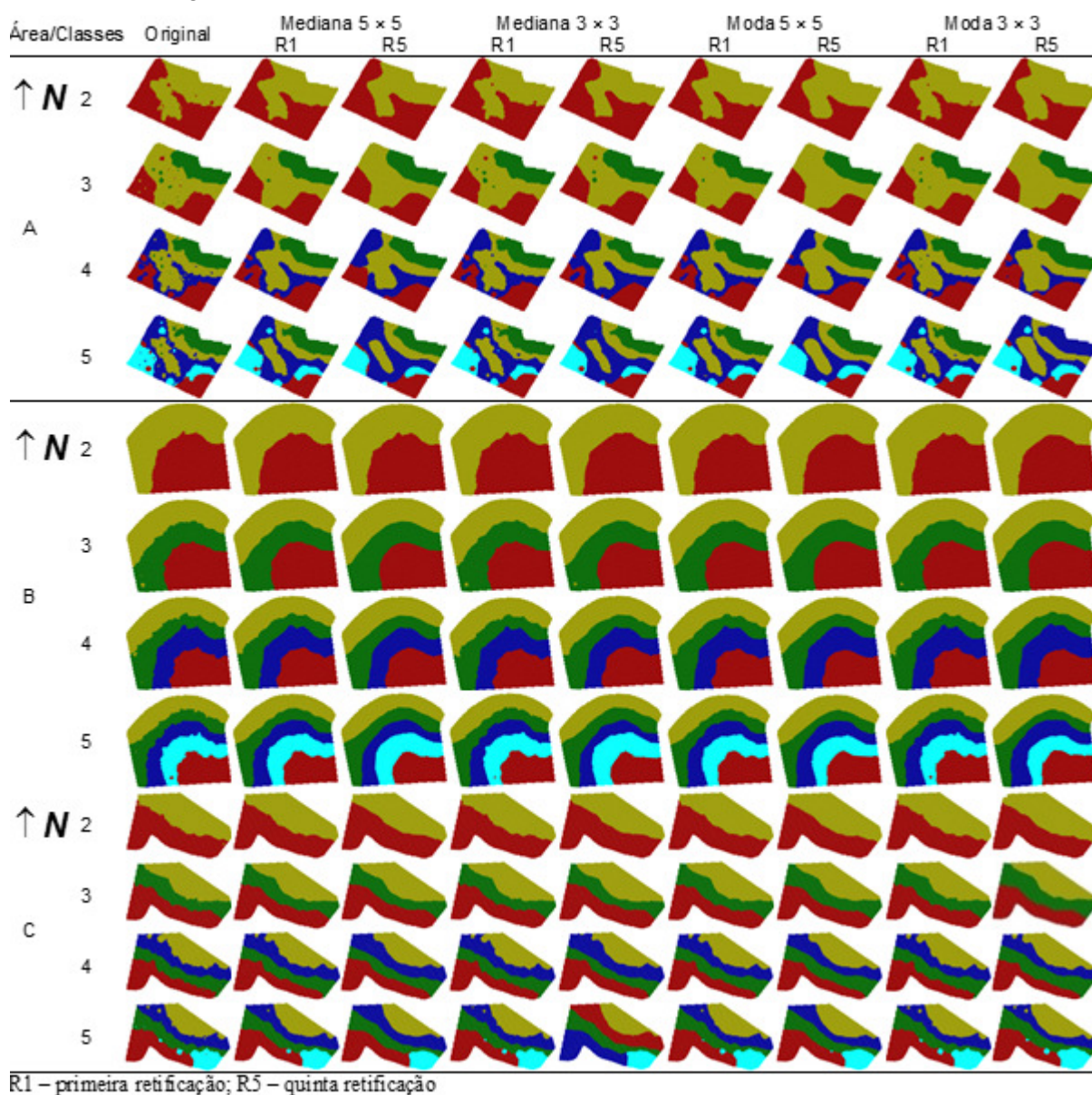


Figura 3 Mapas temáticos das zonas de manejo originais, após a primeira e após a quinta retificação, considerando o método de retificação e o número de classes, para as Áreas A, B, e C.

A suavização dos mapas na primeira retificação é visualmente identificada (Figura 3) pela utilização de todos os quatro métodos definidos para este estudo (mediana 3 × 3, mediana 5 × 5, moda 3 × 3, e moda 5 × 5). No entanto, houve melhora progressiva ao longo das demais retificações, culminando em mapas contínuos após a quinta retificação. Os mapas originais da Área A demonstraram os piores resultados das três áreas e apresentaram muitas manchas e baixa continuidade. Mesmo assim, quando a área foi dividida em duas classes, todos os métodos foram capazes de eliminar as manchas e pixels isolados. No entanto,

quando a área foi dividida em três classes, o método mediana 3×3 não foi capaz de eliminar todas as manchas, mesmo após a quinta retificação. Os mapas originais com quatro classes tinham muitas manchas e pixels isolados. No entanto, após a quinta retificação, todos os métodos foram capazes de definir mais apropriadamente as ZMs, especialmente com a utilização dos métodos com máscara 5×5 . Os mapas originais também mostraram muitas manchas e pixels isolados quando divididos em cinco classes. Melhoraram após a quinta retificação, em especial com o método moda 5×5 , embora algumas manchas e pixels isolados permaneceram, independentemente do método utilizado.

Não foi possível identificar qual o método mais adequado para retificação na Área B, pois as ZMs foram bem definidas nos mapas originais. Quando a Área C foi dividida em quatro e cinco classes, algumas diferenças foram encontradas, e os métodos usando a máscara 5×5 foram capazes de remover todas as manchas e pixels isolados ao longo das cinco retificações.

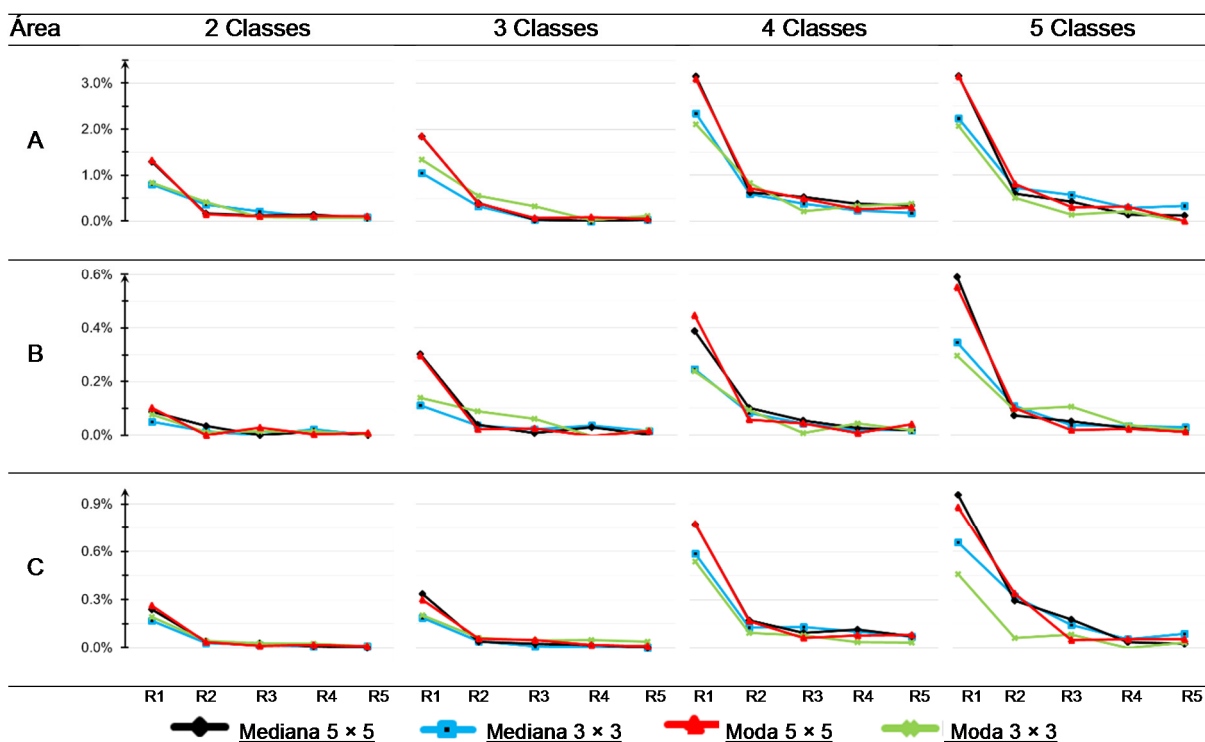
Corroborando os resultados apresentados por Córdoba et al. (2016), neste estudo, foi possível identificar que, na maioria dos casos, os melhores resultados foram obtidos após cinco retificações pelo método com máscara 5×5 , uma vez que apresentaram mapas mais suavizados e consequentemente com melhores condições operacionais. Independente da área analisada, o custo (tempo) computacional entre os métodos da mediana e moda foram equivalentes, porém, com máscara 5×5 , o tempo foi seis vezes maior em relação à máscara 3×3 .

A suavização dos mapas (ZMs) foi determinada utilizando o SI. Em todos os casos, houve aumento do SI (Figura 4) após cada retificação. O incremento foi mais elevado após a primeira retificação e aumentou com o número de classes. Isto demonstra que a aplicação de múltiplas retificações foi mais importante nos mapas com quatro e cinco classes. No entanto, em todos os casos, as retificações restantes foram benéficas, corroborando Lowrance (2014) que assumiu que o uso de funções de retificação apresentaram bons resultados.

A utilização de máscara 5×5 permite a análise de até 24 pixels vizinhos, a fim de determinar se o pixel em questão deve ser transferido para outra classe. Teoricamente, isto proporciona melhor resultado do que quando utilizada máscara 3×3 , o que é confirmado na Figura 4. De modo geral, a partir dos gráficos de todas as áreas estudadas (Figura 4), pode-se observar que, na primeira retificação, a porcentagem de alteração pixels é maior quando a máscara 5×5 é utilizada. No entanto, em retificações posteriores isso não ocorreu e a porcentagem de alteração de pixels pela utilização de métodos com máscara 5×5 foi próxima ao observado com os métodos usando máscara 3×3 .

Mesmo que em escalas diferentes, os gráficos para as três áreas mostrados na Figura 4 são estruturalmente semelhantes, e tendem ao mesmo padrão quando mapas com o mesmo número de classes são analisados. Houve maior variação do SI na primeira retificação, e o índice aumentou proporcionalmente com o número de retificações. Também é

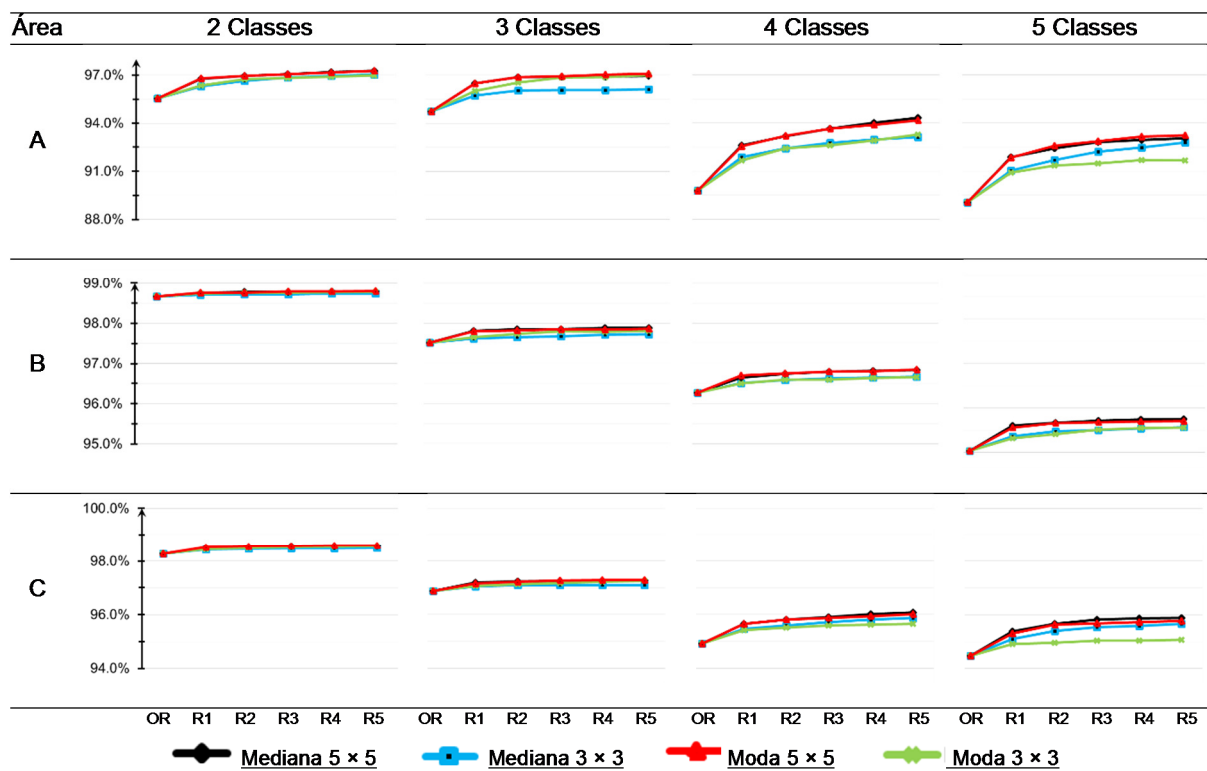
possível identificar que os gráficos dos mapas com duas e três classes sofreram poucas mudanças após a segunda retificação. Assim, indicam que se tornaram mais contínuos, porém ainda propensos a pequenas, mas importantes mudanças nas retificações restantes. Os gráficos dos mapas com quatro e cinco classes sofreram alterações até a quinta retificação. Os métodos de retificação com máscara 5×5 proporcionaram maior suavização total, o que se tornou mais evidente com maior número de classes.



R1 – retificação 1; R2 – retificação 2; R3 – retificação 3; R4 – retificação 4; R5 – retificação 5.

Figura 4 Acréscimo do índice de suavidade após cada retificação, considerando cada método de retificação e o número de classes para as Áreas A, B, e C.

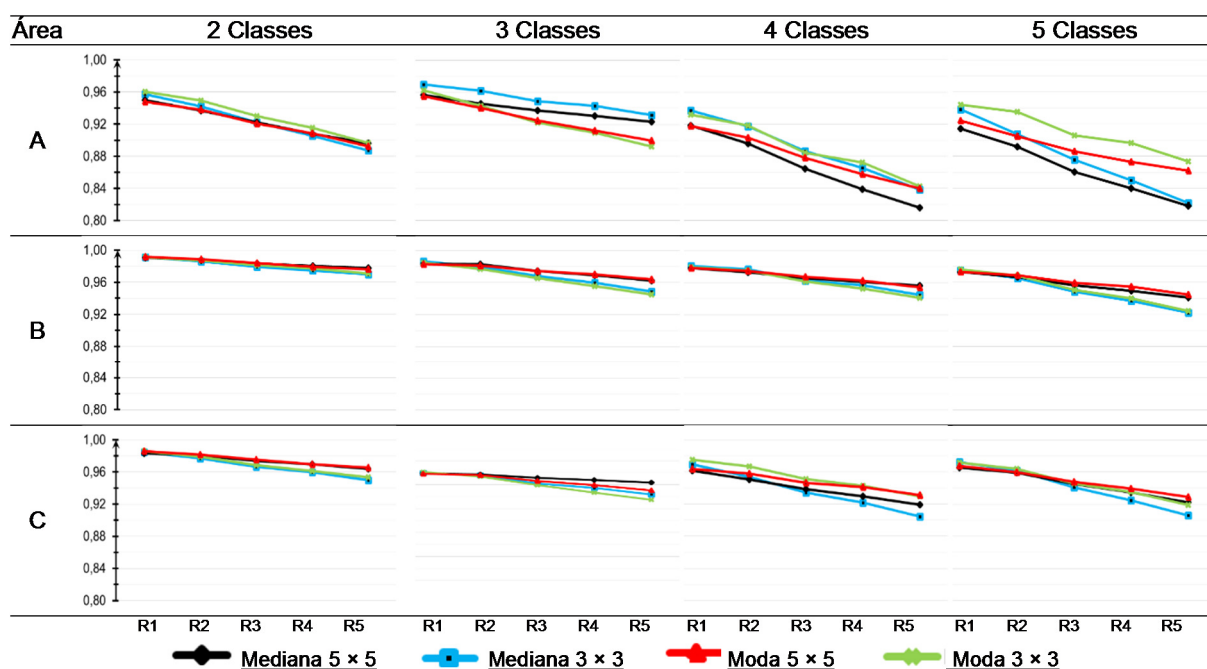
Na análise do índice de suavidade (SI - equação 2), após cada retificação em função do método de retificação e o número de classes das Áreas A, B, e C, verificou-se que, em todos os casos, os melhores resultados foram obtidos por métodos com máscara 5×5 (Figura 5). Com duas classes, o SI variou ligeiramente com o número de retificações, independentemente do método utilizado. Na maioria dos casos, quando houve diferenças entre os métodos de retificação, a moda 3×3 apresentou os piores resultados, com exceção da Área A com três classes, para os quais o método mediana 3×3 foi o pior. As maiores diferenças foram observadas na Área A.



OR – original; R1 – retificação 1; R2 – retificação 2; R3 – retificação 3; R4 – retificação 4; R5 – retificação 5.

Figura 5 Índice de suavidade (SI) após cada retificação, considerando cada método de retificação e o número de classes para as Áreas A, B, e C.

O coeficiente Kappa foi utilizado para avaliar a concordância entre os mapas originais e retificados. Foram obtidas concordâncias muito fortes ($\text{kappa} \geq 0,8$) entre os mapas em todos os casos (Figura 6). Apesar do fato de os métodos com máscara 5×5 terem apresentado maior aumento do SI, a concordância com o mapa original também foi a mais elevada. Nos mapas da Área A (para a qual os mapas originais são menos suaves), o coeficiente Kappa apresentou maior variação (0,96-0,82). Houve praticamente concordância total na primeira retificação das Áreas B e C, mas, em alguns casos, após as demais retificações, o coeficiente Kappa reduziu para cerca de 0,90.



R1 – retificação 1; R2 – retificação 2; R3 – retificação 3; R4 – retificação 4; R5 – retificação 5.

Figura 6 Índice Kappa do mapa original e mapa retificado, considerando cada método de retificação e o número de classes, para as Áreas A, B, e C.

6.4 Conclusão

O índice de suavidade apresentou maior variação na primeira retificação bem como o aumento do número de classes. Com as aplicações sucessivas da função de retificação, progressivamente, as bordas foram suavizadas, as manchas reduzidas e consequentemente, a qualidade das ZMs melhorada. Após a quinta retificação, na maioria dos casos, as ZMs tornaram-se contínuas. Os melhores resultados foram obtidos pela utilização da máscara 5 × 5, independentemente do método estatístico (mediana ou moda). Portanto, a função computacional implementada é capaz de proporcionar ajustes que facilitam a aplicação da agricultura de precisão.

6.5 Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), à Fundação Araucária (FA), à Universidade Estadual do Oeste do Paraná (UNIOESTE) e à Universidade Tecnológica Federal do Paraná (UTFPR), pelo aporte financeiro e aos proprietários pela concessão das áreas experimentais.

6.6 Referências bibliográficas

- ARNO, J.; MARTINEZ-CASASNOVAS, J. A.; RIBES-DASI, M.; ROSELL, J. R. Clustering of grape yield maps to delineate site-specific management zones. **Spanish Journal of Agricultural Research**, v. 9, n. 3, p. 721–729, 2011.
- BAZZI, C. L.; SOUZA, E. G.; URIBE-OPAZO, M. A.; NÓBREGA, L. H. P.; ROCHA, D. M. Management zones definition using soil chemical and physical attributes in a soybean area. **Engenharia Agrícola**, v. 34, n. 5, p. 952–964, 2013.
- BOBRYK, C. W.; MYERS, D. B.; KITCHEN, N. R.; SHANAHAN, J. F.; SUDDUTH, K. A.; DRUMMOND, S. T.; GOMEZ RABOTEAUX, N. N. Validating a digital soil map with corn yield data for precision agriculture decision support. **Agronomy Journal**, v. 108, n. 3, p. 957-965, 2016.
- BUTTAUFUOCO, G.; CASTRIGNANO, A.; COLECCHIA, A. S.; RICCA, N. Delineation of management zones using soil properties and a multivariate geostatistical approach. **Italian Journal of Agronomy**, v. 5, n. 4, p. 323-332, 2010.
- COHEN, J. A coefficient of agreement for nominals scales. **Journal of Education and Measurement**, v. 20, n. 1, p. 37-46, 1960.
- CÓRDOBA, M. A.; BRUNO, C. I.; COSTA, J. L.; PERALTA, N. R.; BALZARINI, M. G. Protocol for multivariate homogeneous zone delineation in precision agriculture. **Biosystems Engineering**, v. 143, p. 95-107, 2016.
- CZAPLEWSKI, R. L.; REICH, R. M. Expected value and variance of Moran's bivariate spatial autocorrelation statistic for a permutation test. **USDA Forest Service**, v. 309, p. 13, 1993.
- DAVIDSON, D. Evaluating the quality of your soil. **Crops & soils magazine**, p. 4-13, 2014.
- DEMATTÊ, J. A. M.; DEMATTÊ, J. L. I.; ALVES, E. R.; BARBOSA, R. N.; MORELLI, J. L. Precision agriculture for sugarcane management: a strategy applied for brazilian conditions. **Acta Scientiarum. Agronomy**, v. 36, n. 1, p. 111, 2014.
- DOBERMANN, A.; PING, J. L.; ADAMCHUK, V. I.; SIMBAHAN, G. C.; FERGUSON, R. B. Classification of Crop Yield Variability in Irrigated Production Fields. **Agronomy Journal**, v. 95, n. 5, p. 1105-1120, 2003.
- DOERGE, T. A. Management Zone Concepts. **Site-Specific Management Guidelines**, p. 1-4, 2000.
- FERGUSON, R. B.; HERGERT, G. W. Soil Sampling for Precision Agriculture. **Precision Agriculture**, p. 1-4, 2009.
- FRANZEN, D. W.; HOPKINS, D. H.; SWEENEY, M. D.; ULMER, M. K.; HALVORSON, A. D. of Site-Specific Nitrogen Management. **Agronomy Journal**, v. 94, p. 381-389, 2002.
- FRIDGEN, J. J.; KITCHEN, N. R.; SUDDUTH, K. A.; DRUMMOND, S. T.; WIEBOLD, W. J.; FRAISSE, C. W. Management zone analyst (MZA): software for management zone delineation. **Agronomy Journal**, v. 96, p. 100-108, 2004.
- GAVIOLI, A.; SOUZA, E. G.; BAZZI, C. L.; GUEDES, L. P. C.; SCHENATTO, K. Optimization of management zone delineation by using spatial principal components. **Computers and Electronics in Agriculture**, v. 127, p. 302-310, 2016.

GUASTAFERRO, F.; CASTRIGNANÒ, A.; BENEDETTO, D.; SOLLITTO, D.; TROCCOLI, A.; CAFARELLI, B. A comparison of different algorithms for the delineation of management zones. **Precision Agriculture**, v. 11, n. 6, p. 600-620, 2010.

JOURNELL, A. G.; HUIJBREGTS, C. J. **Mining geostatistics**. London, 1978. 600 p.

KHOSLA, R.; FLEMING, K.; DELGADO, J. A.; SHAVER, T. M.; WESTFALL, D. G. Use of site-specific management zones to improve nitrogen management for precision agriculture. **Journal of Soil and Water Conservation**, v. 57, n. 6, p. 513-518, 2002.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1, p. 159-174, 1977.

LOWRANCE, C. **Open source hardware and software in agriculture: an autonomous sap flow measurement wireless network & a user friendly management zone delineation tool**. 2014. 121 f. Thesis (Master's degree in crop and soil sciences) - University of Georgia, 2014.

MALLARINO, A. P.; WITTRY, D. J. Efficacy of grid and zone soil sampling approaches for site-specific assessment of phosphorus, potassium, pH, and organic matter. **Precision Agriculture**, v. 5, n. 2, p. 131-144, 2004.

MOHAMMADI, J. Spatial variability of soil fertility, wheat yield and weed density in a one-hectare field in Shahre Kord. **Journal of Agricultural Science and Technology**, v. 4, n. 1/2, p. 83-92, 2002.

MORAL, F. J.; TERRÓN, J. M.; SILVA, J. R. M. DA. Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. **Soil and Tillage Research**, v. 106, n. 2, p. 335-343, 2010.

MOSHIA, M. E.; KHOSLA, R.; LONGCHAMPS, L.; REICH, R.; DAVIS, J. G.; WESTFALL, D. G. Precision manure management across site-specific management zones: Grain yield and economic analysis. **Agronomy Journal**, v. 106, n. 6, p. 2146-2156, 2014.

MZUKU, M.; KHOSLA, R.; REICH, R.; INMAN, D.; SMITH, F.; MACDONALD, L. Spatial Variability of Measured Soil Properties across Site-Specific Management Zones. **Soil Science Society of America Journal**, v. 69, n. 5, p. 1572-1579, 2005.

ODEH, I. O. A.; MCBRATNEY, A. B.; CHITTLEBOROUGH, D. J. Soil pattern recognition with fuzzy c-means: Application to classification and soil-landform interrelationships. **Soil Sci. Soc. Am. J.**, v. 56, p. 505-516, 1992.

PEDROSO, M.; TAYLOR, J.; TISSEYRE, B.; CHARNOMORDIC, B.; GUILLAUME, S. A segmentation algorithm for the delineation of agricultural management zones. **Computers and Electronics in Agriculture**, v. 70, n. 1, p. 199-208, 2010.

PRAMANIK, S.; PRUSTY, S.; BHATTACHARJEE, D.; BHUNRE, P. K. A Region-to-pixel based Multi-sensor Image Fusion. **Procedia Technology**, v. 10, p. 654-662, 2013.

RODRIGUES, K. M.; HURTADO, S. M. C.; DECHEN, S. C. F.; VIEIRA, S. R. Spatial variability in soil fertility and particle size and their effects on sugarcane yield. **Sugar Tech.**, v. 18, n. 1, p. 39-48, 2016.

RUSPINI, E. H. A new approach to clustering. **Inf. Control.**, v. 15, p. 22-32, 1969.

SALEH, A.; BELAL, A. A. Delineation of site-specific management zones by fuzzy clustering of soil and topographic attributes: A case study of East Nile Delta, Egypt. **IOP Conference**

Series: Earth and Environmental Science, v. 18, p. 1-6, 2014.

TAGARAKIS, A.; LIAKOS, V.; FOUNTAS, S.; KOUNDOURAS, S.; GEMTOS, T. A. Management zones delineation using fuzzy clustering techniques in grapevines. **Precision Agriculture**, v. 14, n. 1, p. 18-39, 2013.

WOLLENHAUPT, N. C.; WOLKOWSKI, R. P.; CLAYTON, M. K. Mapping soil test phosphorus and potassium for variable-rate fertilizer application. **Journal of Production Agriculture**, v. 7, n. 4, p. 441-448, 1994.

XIANG, L.; YU-CHUN, P.; ZHONG-QIANG, G.; CHIN-JIANG, Z. Delineation and scale effect of precision agriculture management zones using yield monitor data over four years. **Agricultural Sciences in China**, v. 6, n. 2, p. 180-188, 2007.

7. CONSIDERAÇÕES FINAIS

A automação de máquinas e equipamentos, aliada à utilização de aplicativos computacionais tais como: sistemas de armazenamento de dados, sistemas de análise estatística, sistemas de informações gerenciais, sistemas de informações geográficas, entre outros, fazem parte da realidade atual, inclusive em atividades agrícolas. A praticidade incorporada pela utilização de softwares frequentemente facilita operações consideradas complexas. Desta maneira, rotinas computacionais foram implementadas, validadas e aplicadas neste estudo, com o objetivo de proporcionar eficácia e facilitar a utilização de métodos de interpolação.

Costumeiramente, os valores 1 e 2 são utilizados como potência para o método de interpolação inverso da distância. Porém, como constatado por diversos autores, nem sempre estes valores são os apropriados, pois a definição do melhor expoente para o inverso da distância depende de qual é o atributo e dos valores dos dados coletados. Da mesma maneira, o melhor ajuste para o semivariograma é essencial para identificar valores mais apropriados para os parâmetros utilizados no processo de interpolação por krigagem ordinária. Assim, destaca-se a importância das rotinas implementadas que foram capazes de identificar automaticamente o melhor expoente para inverso da distância, bem como o melhor modelo e valores dos parâmetros de ajuste do semivariograma, e conseqüentemente, mensurar dados com maior precisão.

Porém, independentemente do método de interpolação e parâmetros utilizados para mensurar dados, pequenas áreas representadas por pixels isolados ou manchas geralmente são observadas nos mapas temáticos de zonas de manejo, o que dificulta a aplicação de insumos e fertilizantes à taxa constante em determinadas áreas. Assim, as rotinas implementadas com o objetivo de retificar as zonas de manejo foram consideradas úteis e suficientemente capazes para identificar e realizar alterações capazes de proporcionar continuidade às zonas de manejo, facilitando seu manejo.

Estudos futuros poderão identificar se os resultados obtidos pela retificação de zonas de manejo com a utilização de outras máscaras proporcionam melhores resultados, considerando inclusive o tempo computacional. No que se refere aos métodos de interpolação, outros modelos e métodos de ajuste do semivariograma poderão ser incorporados às rotinas implementadas a fim de que se aumentem as possibilidades para mensurar dados com maior precisão.

APÊNDICES

Apêndice A – Rotina de geoestatística

```

require(geoR)
require(splancs)
require(classInt)
require(sp)
library(stats)
require(MASS)
library(e1071)
require(gstat)
library(ade4)
library(spdep)
library(RPostgreSQL)
library(data.table)
require(gstat)
require(nortest)

##### variáveis #####

ISI = TRUE          #identifica que será utilizado o índice ISI para escolha do melhor modelo

#variáveis para semivariograma
v_lambda = 1        # 1 = dados NÃO transformados
auto_lags= TRUE      # parametro que define automaticamente o nro de lags
nro_lags = 11        # parametro semivariograma KO que estipula o nro de lags
estimador = "classical" #parametro semivariograma KO = Matheron
cutoff = 50          #parametro semivariograma KO-% da distancia maxima entre os pontos
nro_pares = 30        #parametro semivariograma KO
nro_intervalos_alc= 5 #parametro Alcance parametros do semivariograma (KO)
nro_intervalos_contr= 5 #parametro Contribuição parametros do semivariograma (KO)

#parâmetros do expand.grid para criar vals da matriz de contribuição/alcance
min_seq_contr = 0    # por padrão = 0 e receberá o valor de min_var, ou usuário informa valor
min_seq_alc = 0      # por padrão = 0 e e receberá o valor de min_dist_var mais adiante, ou
usuário informa valor

t_cor_linha_ols = data.table(rbind("GOLD", "PURPLE", "VIOLET", "YELLOW", "BLACK",
    "GREEN", "ORANGE", "PINK", "GRAY", "BROWN", "BLUE", "RED", "MAGENTA"))
t_modelos = data.table(rbind("matern", "matern", "matern", "exp", "sph", "matern", "matern",
    "matern", "exp", "sph", "gaus", "gaus"))
t_kappa = data.table(rbind(1.0, 1.5, 2.0, 0.5, 0.5, 1.0, 1.5, 2.0, 0.5, 0.5, 0.5, 0.5))
t_metodo = data.table(rbind("ols", "ols", "ols", "ols", "ols", "wl", "wl", "wl", "wl", "wl", "wl", "ols"))
nro_modelo=12
vlr_kappa=0          #parametro semivariograma KO

#Constante para gráfico Semivariograma
legenda_x_semiv = "Distância"
legenda_y_semiv = "Semivariância"
titulo_semiv = "Semivariograma experimental"
#Constante do gráfico Mapa de dispersão dos pontos amostrais
titulo_pamostrais = "Mapa de dispersão dos pontos amostrais"
fonte_pamostrais = 3
leg_x_pamostrais = "O - L"

```

```

leg_y_pamostrais = "S - N"
titulo_BoxPlot = "Gráfico Boxplot"
titulo_PostPlot = "Gráfico Postplot"

#parametros que identificam o tamanho do pixel do mapa final
tam_pixel_x = 5 #parametro dos mapas
tam_pixel_y = 5 #parametros dos mapas

classes = 4    #número de classe no mapa (intervalos) - só no R

#####
# Estabelece conexão com o PoststgreSQL usando RPostgreSQL
drv <- dbDriver("PostgreSQL")
projeto = "db_projeto_ischenatto_a"      #área
atributo = "tb_produtividade_media_soja" #amostra
local = 29182

# Configuração completa da conexão com o banco de dados
con <- dbConnect(drv, dbname=projeto,host="localhost",port=5432,
  user="postgres",password="xxxx")
#leitura dos dados da tabela de atributos do banco de dados
atrib = paste("select st_x(st_transform(the_geom,", local, ")),
  st_y(st_transform(the_geom, ", local, ")),amo_medida from ", atributo)
frame_dados <- dbGetQuery(con,atrib)
dados <- as.geodata(frame_dados)
names(dados)
min (dados$data)
#Análise exploratória dos dados
summary(dados$data)
mean (dados$data)
min (dados$data)
var(dados$data)
sd(dados$data)
CV = sd(dados$data)*100/mean(dados$data)
CV
skewness(dados$data)
kurtosis(dados$data)
length(dados$data)
plot(dados)
hist(dados$data)
boxplot(dados$data,main=titulo_BoxPlot)
points(dados,pt.div="quartile",col=c("yellow","green","red","blue"),main=titulo_PostPlot,
xlab=leg_x_pamostrais, ylab=leg_y_pamostrais)
max_dist <- max(dist(dados$coords))
min_dist <- min(dist(dados$coords))
vlr_cutoff <- max_dist*cutoff/100
if (auto_lags==TRUE){
  nro_lags = round(vlr_cutoff/min_dist)    ## menor distancia das variancias
}
nro_lags
dados.var <- variog(dados,coords=dados$coords, data=dados$data,
  uvec=seq(min_dist,vlr_cutoff,l=nro_lags), lambda=v_lambda,
  estimator.type=estimador, max.dist=vlr_cutoff, pairs.min=nro_pares)
dados.var
plot(dados.var, xlab = legenda_x_semiv, ylab = legenda_y_semiv, main = titulo_semiv)
# Informações do semivariograma experimental

```

```

distancia <- dados.var$u
semivariancia <- dados.var$v
pares <- dados.var$n
tabela <- cbind(distancia,semivariancia,pares)
tabela
min_dist_var = min(distancia) ## menor distancia das variancias
min_dist_var
## para ajustar os valores para o semivariograma
min_var = min(semivariancia) # menor variância
max_var = max(semivariancia) # maior variância
min_var
max_var
vlr_cutoff
if (min_seq_alc==0){
  min_seq_alc = vlr_cutoff/4
}
if (min_seq_contr==0){
  min_seq_contr = min_var
}
min_seq_alc
vals <- expand.grid(seq(min_seq_contr,max_var, l=nro_intervalos_contr),
seq(min_seq_alc, vlr_cutoff, l=nro_intervalos_alc))
semiv_geral= paste0("semiv_geral_",atributo,".jpg")
semiv_geral
jpeg (filename = semiv_geral, quality=1600)
plot(dados.var,xlab='Distância',ylab='Semivariância',
      main= paste ("Semivariograma ajustado -",atributo) )
vals
cont = nro_intervalos_contr * nro_intervalos_alc
cont
#cria matriz para armazenar informações do ice
matriz_ice<-matrix(nrow=0,ncol=7,
  dimnames = list(c(),c("modelo","metodo","min_ice", "melhor_contrib",
    "melhor_alcance", "melhor_vlr_kappa", "gid" )))
vetor_ice = c() ### vetor para armazenar o menor ice de cada molelo
t_modelos
t_kappa
j=0
j
nro_modelo
metodo="wl"
while (j<nro_modelo)
{
  #cria matriz para armazenar informações da validação cruzada
  matriz_vc<-matrix(nrow=0,ncol=9,
    dimnames = list(c(),c("Modelo", "EM", "EMR", "DP_EM", "DP_EMR",
      "DP_EMR_1", "EA","Metodo", "SDAE")))
  #cria vetores para armazenar informações da validação cruzada
  vetor_em = c()
  vetor_emr = c()
  vetor_dp_em = c()
  vetor_dp_emr = c()
  vetor_dp_emr_1 = c()
  vetor_ea = c()
  vetor_modelo = c()
  vetor_metodo =c()

```

```

ice = c()
A = c()
B = c()
vetor_contr = c()
vetor_alcance = c()
vetor_vlr_kappa = c()
vetor_sdae=c()
j=j+1
j
modelo = t_modelos$V1[j]
vlr_kappa = as.numeric(t_kappa$V1[j])
cor_linha_ols = t_cor_linha_ols$V1[j]
metodo = t_metodo$V1[j]
modelo
vlr_kappa
metodo
i=0
i
cont
t_cont = 0
while (i<cont)
{
    i= i+1
    i
    contrib = as.numeric(vals$Var1[i])
    alcance = as.numeric(vals$Var2[i])
    contrib
    alcance
    if (modelo=="matern"){
        if (metodo=="ols"){
            variograma.ols=variofit(dados.var,ini=c(contrib,alcance),
            weights= "equal",cov.model= modelo, kappa= vlr_kappa,
            max.dist=vlr_cutoff)
        } else {
            variograma.ols=variofit(dados.var,ini=c(contrib,alcance),
            cov.model= modelo, kappa= vlr_kappa,
            max.dist=vlr_cutoff)
        }
    } else {
        if (metodo=="ols"){
            variograma.ols=variofit(dados.var,ini=c(contrib,alcance),
            weights= "equal",cov.model= modelo,
            max.dist=vlr_cutoff)
        } else {
            variograma.ols=variofit(dados.var,ini=c(contrib,alcance),
            cov.model= modelo, max.dist=vlr_cutoff)
        }
    }
    lines(variograma.ols,col=cor_linha_ols)
    variograma.ols
    #armazena informações da validação cruzada em variáveis
    vc=xvalid(dados,model=variograma.ols,micro.scale=0)
    vc
    emr=1
    dp_emr =1
    dp_emr_1 = 0

```

```

if ((mean (vc$std.error) != "NaN"))
{
    emr = mean (vc$std.error) #erro médio reduzido
    dp_emr = round(sd (vc$std.error),digits=20)
    dp_emr_1 = ((sd (vc$std.error))-1)
}
vc$error
em = round(mean (vc$error),digits=20) #erro médio
##### DPem calculado
nro_amstras = length(vc$error) #conta o número de elementos
n = 0
somatorio = 0
while (n < nro_amstras)
{
    n = n+1
    somatorio = somatorio + (vc$error[n]*vc$error[n])
}
media_em2 = somatorio / nro_amstras #média dos EM2
dp_em = sqrt(media_em2) #raiz quadrada EM2
media_em2
dp_em
ea=round(sum(abs(vc$predicted-vc$data)),digits=20) #erro absoluto

##### # ISI
vc
vc$data
vc$predicted
em
dif = (vc$data - vc$predicted)^2
dif
media_dif = mean(dif)
media_dif
sdae = sqrt(media_dif)
sdae

#popula vetores com informações da validação cruzada
vetor_em <- rbind(vetor_em,c(em))
vetor_emr = rbind(vetor_emr,c(emr))
vetor_dp_em = rbind(vetor_dp_em,c(dp_em))
vetor_dp_emr = rbind(vetor_dp_emr,c(dp_emr))
vetor_dp_emr_1 = rbind(vetor_dp_emr_1,c(dp_emr_1))
vetor_ea = rbind(vetor_ea,c(ea))
vetor_modelo = rbind(vetor_modelo,c(modelo))
vetor_metodo = rbind(vetor_metodo,c(metodo))
vetor_contr = rbind(vetor_contr,c(contrib))
vetor_alcance = rbind(vetor_alcance,c(alcance))
vetor_vlr_kappa = rbind(vetor_vlr_kappa,c(vlr_kappa))
vetor_sdae = rbind(vetor_sdae,c(sdae))

#popula matriz com informações da validação cruzada
matriz_vc<-rbind(matriz_vc,
    c(modelo,em,emr,dp_em,dp_emr,dp_emr_1,ea,metodo,sdae))
matriz_vc
t_cont = t_cont +1 ###variável para armazenar o tamanho de table_ice
}
matriz_vc

```

```

if (ISI==TRUE)
{
  max_abs_em = max (abs(vetor_em))
  min_sdae = min (vetor_sdae)
  max_sdae = max (vetor_sdae)
  max_abs_em
  min_sdae
  max_sdae
  A = (abs(vetor_em))/max_abs_em
  B = ((vetor_sdae) - min_sdae)/ max_sdae
}
A
B
ice = round(A + B, digits=20)
ice
min_ice=min(ice)
min_ice
table_ice <- data.table(cbind(ice, vetor_contr, vetor_alcance, vetor_modelo,
                             vetor_metodo, vetor_vlr_kappa, vetor_em, vetor_dp_em))
table_ice
i=0
while (i<t_cont)
{
  i= i+1
  if (table_ice$V1[i] == min_ice) {
    melhor_contrib = as.numeric(table_ice$V2[i])
    melhor_alcance = as.numeric(table_ice$V3[i])
    melhor_modelo = table_ice$V4[i]
    melhor_metodo = table_ice$V5[i]
    melhor_vlr_kappa = as.numeric(table_ice$V6[i])
    melhor_em = as.numeric(table_ice$V7[i])
    melhor_dp_em = as.numeric(table_ice$V8[i])
    i=t_cont
  }
}
table_ice
melhor_contrib
melhor_alcance
melhor_modelo
melhor_metodo
melhor_vlr_kappa
melhor_em
melhor_dp_em

#popula matriz com informações do melhor ICE de cada modelo
matriz_ice<-rbind(matriz_ice, c(modelo, metodo, min_ice, melhor_contrib,
                                melhor_alcance, melhor_vlr_kappa, j))
matriz_ice
vetor_ice = rbind(vetor_ice,c(min_ice)) ## vetor com menor ice de cada modelo
vetor_ice
}
dev.off()
vetor_em_melhor = c()
vetor_dp_em_melhor = c()
matriz_isi<-matrix(nrow=0,ncol=3,

```

```

dimnames = list(c(),c("em", "dpem", "isi"))
isi = c("isi")
A = c()
B = c()
isi
matriz_ice
matriz_isi
#### gráfico com o melhor semivariograma de cada modelo/metodo ###
matriz_legenda<-matrix(nrow=0,ncol=3,
dimnames = list(c(),c("modelo","metodo","vlr_kappa")))
semiv_melhores= paste0("semiv_melhores_",atributo,".jpg")
jpeg (filename = semiv_melhores, quality=1600)
plot(dados.var,xlab='Distância',ylab='Semivariância',
      main= paste ("Semiariograma ajustado -",atributo) )

i=0
i
while (i<nro_modelo)
{
  i= i+1
  i
  modelo = matriz_ice[i,1]
  metodo = matriz_ice[i,2]
  contrib = as.numeric(matriz_ice[i,4])
  alcance = as.numeric(matriz_ice[i,5])
  vlr_kappa = as.numeric(matriz_ice[i,6])
  cor_linha_ols = t_cor_linha_ols$V1[i]
  if (modelo=="matern"){
    if (metodo=="ols"){
      variograma.ols<-variofit(dados.var,ini=c(contrib,alcance),
                              weights= "equal",cov.model= modelo,
                              kappa= vlr_kappa, max.dist=vlr_cutoff)
    } else {
      variograma.ols<-variofit(dados.var,ini=c(contrib,alcance),
                              cov.model= modelo,
                              kappa= vlr_kappa, max.dist=vlr_cutoff)
    }
  } else {
    if (metodo=="ols"){
      variograma.ols<-variofit(dados.var,ini=c(contrib,alcance),
                              weights= "equal",cov.model= modelo,
                              max.dist=vlr_cutoff)
    } else {
      variograma.ols<-variofit(dados.var,ini=c(contrib,alcance),
                              cov.model= modelo, max.dist=vlr_cutoff)
    }
  }

  lines(variograma.ols,col=cor_linha_ols)
  vc=xvalid(dados,model=variograma.ols,micro.scale=0)
  em_melhor = round(mean (vc$error),digits=20) #erro médio
  dif = (vc$data - vc$predicted)^2
  media_dif = mean(dif)
  sdae = sqrt(media_dif)
  vetor_em_melhor <- rbind(vetor_em_melhor,c(em_melhor))
  vetor_dp_em_melhor = rbind(vetor_dp_em_melhor,c(sdae))
}

```

```

vetor_em_melhor
if (ISI==TRUE)
{
  max_abs_em_melhor = max (abs(vetor_em_melhor))
  min_abs_sdae = min (vetor_dp_em_melhor)
  max_abs_sdae = max (vetor_dp_em_melhor)
  A = (abs(vetor_em_melhor))/max_abs_em_melhor
  B = (vetor_dp_em_melhor - min_abs_sdae)/ max_abs_sdae
}
isi = round(A + B, digits=20)
matriz_isi<-cbind(vetor_em_melhor,vetor_dp_em_melhor,isi)
matriz_isi_melhor = cbind (matriz_ice,matriz_isi)
dev.off()
nome_tab = paste0 ("tb_ice_",atributo)
nome_arq = paste0 (atributo,".txt")
write.table(matriz_isi_melhor,nome_arq)

# Grava as coordenadas e valores interpolados no banco de dados em "tb_utm"
dbWriteTable(con, nome_tab, as.data.table(matriz_isi_melhor), overwrite = T)

# Encerra a conexão com o banco de dados
dbDisconnect(con)

```

Apêndice B – Rotina de interpolação de dados por krigagem ordinária

```

require(geoR)
require(splancs)
require(classInt)
require(sp)
library(stats)
require(MASS)
library(e1071)
require(gstat)
library(ade4)
library(spdep)
library(RPostgreSQL)
library(data.table)
require(gstat)

#variáveis para semivariograma
v_lambda = 1 # 1 = dados NÃO transformados
cor_linha = "BLACK"
auto_lags=TRUE # parametro que define automaticamente o nro de lags
nro_lags = 11 #parametro semivariograma KO que estipula o nro de lags
arbitrariamente
estimador = "classical" #parametro semivariograma KO = Matheron
cutoff = 50 #parametro semivariograma KO - % da distancia maxima entre os pontos
nro_pares = 30 #parametro semivariograma KO
#parametros que identificam o tamanho do pixel do mapa final
tam_pixel_x = 5 #parametro dos mapas
tam_pixel_y = 5 #parametros dos mapas
classes = 4 #número de classe no mapa (intervalos) - só no R
#Constante para gráfico Semivariograma
legenda_x_semiv = "Distância"
legenda_y_semiv = "Semivariância"
titulo_semiv = "Semivariograma experimental"
leg_x_pamostrais = "O - L"
leg_y_pamostrais = "S - N"

#####
# Estabelece conexão com o PoststgreSQL usando RPostgreSQL
drv <- dbDriver("PostgreSQL")
projeto = "db_projeto_ischenatto_a" #área
atributo = "tb_elevacao" #amostra
local = 29182
# Configuração completa da conexão com o banco de dados
con <- dbConnect(drv, dbname=projeto,host="localhost",port=5432,
  user="postgres",password="xxxx")
borda <- dbGetQuery(con,"select x, y from tb_borda") #em utm
#leitura dos dados da tabela de atributos do banco de dados
atrib = paste("select st_x(st_transform(the_geom,", local, ")),
  st_y(st_transform(the_geom, ", local, ")),amo_medida from ", atributo)
frame_dados <- dbGetQuery(con,atrib)
dados <- as.geodata(frame_dados)

```

```
#####mapa temático
plot(borda,xlab=leg_x_pamostrais, ylab=leg_y_pamostrais)
polygon(borda)
points(dados, add=T)
apply(borda,2,range) #Mostra o mínimo e máximo das coordenadas
menor_X <- min(borda[,1]) #identifica o menor valor de X, primeira coluna
menor_Y <- min(borda[,2]) #identifica o menor valor de Y, segunda coluna
maior_X <- max(borda[,1]) #identifica o maior valor de X, primeira coluna
maior_Y <- max(borda[,2]) #identifica o maior valor de Y, segunda coluna
gr<-expand.grid(x=seq(menor_X,maior_X,by=tam_pixel_x), y=seq(menor_Y,maior_Y,
by=tam_pixel_y))
plot(gr)
points(dados, pt.div="equal") #monta o grid de interpolação
gi<- polygrid(gr,bor=borda)
length(gi$x)
points(gi, pch="+", col=2) #o novo grid considerando apenas a região limitada

##### krigagem #####
max_dist <- max(dist(dados$coords))
min_dist <- min(dist(dados$coords))
vlr_cutoff <- max_dist*cutoff/100
f (auto_lags==TRUE){
  nro_lags = round(vlr_cutoff/min_dist)      ## menor distancia das variancias
}

##### modelo escolhido #####
tb_matriz_ice = paste0 ("tb_ice_",atributo)
query_modelo = sprintf("select * from %s order by isi desc", tb_matriz_ice)
frame_modelo <- dbGetQuery(con,query_modelo)
query_cont = sprintf("select count (*) from %s", tb_matriz_ice)
v_cont <- dbGetQuery(con,query_cont)
i=0

while (i<v_cont)
{
  i= i+1
  melhor_contrib_geral = as.numeric(frame_modelo[i,5])
  melhor_alcance_geral = as.numeric(frame_modelo[i,6])
  melhor_modelo_geral = frame_modelo[i,2]
  melhor_metodo_geral = frame_modelo[i,3]
  melhor_vlr_kappa_geral = as.numeric(frame_modelo[i,7])
  id_modelo = as.numeric(frame_modelo[i,8])      #parametro do modelo escolhido
  if (melhor_modelo_geral == "matern") {
    if (melhor_vlr_kappa_geral == 1.5)
      nome= paste0 (atributo,"_",melhor_modelo_geral,"_1_5_",
        melhor_metodo_geral)
    else
      nome= paste0 (atributo,"_",melhor_modelo_geral,"_",
        melhor_vlr_kappa_geral,"_",melhor_metodo_geral)
  } else {
    nome= paste0 (atributo,"_",melhor_modelo_geral,"_",melhor_metodo_geral)
  }
  dados.var <- variog(dados,coords=dados$coords, data=dados$data,
    uvec=seq(min_dist,vlr_cutoff,l=nro_lags), lambda=v_lambda,
    estimator.type=estimador, max.dist=vlr_cutoff, pairs.min=nro_pares)
  plot(dados.var, xlab = legenda_x_semiv, ylab = legenda_y_semiv, main = titulo_semiv)
```

```

#faz o ajuste novamente com o melhor modelo
if (melhor_modelo_geral == "matern"){
  if (melhor_metodo_geral=="ols"){
    variograma.ols<-variofit(dados.var,
      ini=c(melhor_contrib_geral,melhor_alcance_geral),
      weights= "equal",cov.model= "matern",
      kappa=melhor_vlr_kappa_geral, max.dist=vlr_cutoff)
  } else {
    variograma.ols<-variofit(dados.var,
      ini=c(melhor_contrib_geral,melhor_alcance_geral),
      cov.model= "matern", kappa=melhor_vlr_kappa_geral,
      max.dist=vlr_cutoff)
  }
} else {
  if (melhor_metodo_geral=="ols"){
    variograma.ols<-variofit(dados.var,
      ini=c(melhor_contrib_geral,melhor_alcance_geral),
      weights= "equal",cov.model= melhor_modelo_geral,
      max.dist=vlr_cutoff)
  } else {
    variograma.ols<-variofit(dados.var,
      ini=c(melhor_contrib_geral,melhor_alcance_geral),
      cov.model= melhor_modelo_geral, max.dist=vlr_cutoff)
  }
}
nome_semiv = paste0 ("semiv_",nome, ".jpg")
jpeg (filename = nome_semiv, quality=1600)
plot(dados.var,xlab = legenda_x_semiv, ylab = legenda_y_semiv,
  main = paste ("Semivariograma ajustado pelo modelo",melhor_modelo_geral,
  melhor_vlr_kappa_geral,melhor_metodo_geral,atributo))
lines(variograma.ols,col=cor_linha,pch=3)
dev.off()
kC=krige.control(obj=variograma.ols,lambda=v_lambda,micro.scale=0) #krigagem
dados.kC=krige.conv(dados, loc=gi, krige=kC) #Faz o mapa da variável dados por KRI
length(dados.kC$pred)
max(dados.kC$pred)
min(dados.kC$pred)
valores<-dados.kC$predict
range(dados.kC$predict)
nome_mapa = paste0 ("mapa_",nome, ".jpg")
jpeg (filename = nome_mapa, quality=1600)
image(dados.kC, loc=gr, border=borda, col=gray(seq(1,0,l=classes)),
  xlab=leg_x_pamostrais, ylab=leg_y_pamostrais,
  zlim=range(dados.kC$predict))
zlim=range(dados.kC$predict)
max_medida = round (max(valores),digits=2)
min_medida = round (min(valores),digits=2)
intervalo = round ((max_medida - min_medida) / 4, digits=2)
intervalo1 = min_medida + intervalo
intervalo2 = intervalo1 + intervalo
intervalo3 = intervalo2 + intervalo
c1 = paste0 (min_medida," |-- ",intervalo1)
c2 = paste0 (intervalo1," |-- ",intervalo2)
c3 = paste0 (intervalo2," |-- ",intervalo3)
c4 = paste0 (intervalo3," |-- ",max_medida)

```

```

        legend("bottomleft", fill=gray(c(0.1, 0.4, 0.7, 1.0)),
              c(c4,c3,c2,c1),cex=0.9)
    dev.off()
}

nome_env = paste0 ("semiv_env_",nome, ".jpg")
jpeg (filename = nome_env, quality=1600)
    dados.env<-variog.mc.env(dados, obj.v=dados.var,nsim = 99)
    plot(dados.var,env=dados.env,xlab="Distância", ylab="Semivariância")
    lines(variograma.ols,col=cor_linha,pch=3)
dev.off()

# Concatena colunas: coordenadas em UTM dos pontos e valores interpolados
lng = gi$x
lat = gi$y
medida = valores
interpol <- data.table(cbind(lng, lat, medida))
dbWriteTable(con, "tb_utm", interpol, overwrite = T)
nome_mapa = paste0 ("inter_ko_",nome)
nome_tab = paste0 ("tb_",nome_mapa)

##### select para incluir parametros no insert
#leitura dos dados da tabela de atributos do banco de dados
sql_cod_amostra = paste("select distinct amo_codamostra from ", atributo)
codamostra= dbGetQuery(con,sql_cod_amostra)
cod_amostra = codamostra$amo_codamostra
sql_amostra = sprintf("select amo_codtipoatributo, amo_codarea from tb_amostra where
gid=%i", cod_amostra)
amostra <- dbGetQuery(con,sql_amostra)
tipo_atrib = amostra$amo_codtipoatributo
cod_area = amostra$amo_codarea

## cria tabela com dados interpolados ##
queryINTER = sprintf ("select f_cria_mapa_R('%s',%i,%i,'%s',%i,%i,%i,'%s','%s',%.1f,%i,%i)",
    nome_mapa,cod_area,tipo_atrib,nome_tab,cod_amostra,tam_pixel_x,tam_pixel_y,
    'POLYGON','KO',melhor_vlr_kappa_geral,0,0)
queryINTER
dbGetQuery(con, queryINTER)

vc=xvalid(dados,model=variograma.ols)
em = round(mean (vc$error),digits=20) #erro médio
nro_amostras = length(vc$error) #conta o número de elementos
n = 0
somatorio = 0

while (n < nro_amostras)
{
    n = n+1
    somatorio = somatorio + (vc$error[n]*vc$error[n])
}

media_em2 = somatorio / nro_amostras #média dos erros médios ao quadrado
dp_em = sqrt(media_em2) #raiz quadrada da média dos erros médios ao quadrado

#####criar e gravar EM e DPEM na tabela ISI #
nome_tab = paste0 ("tb_isi_",atributo)

```

```
nome_isi = paste0 ("ko_",melhor_modelo_geral,"_",melhor_vlr_kappa_geral,  
  "_",melhor_metodo_geral)  
dados_isi <- data.table(cbind(em,dp_em,nome_isi))  
dbWriteTable(con, nome_tab, as.data.table(dados_isi), overwrite = T)  
  
# Encerra a conexão com o banco de dados  
dbDisconnect(con)
```