

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

**ESTATÍSTICA ESPACIAL APLICADA À AGRICULTURA DE  
PRECISÃO**

**GUSTAVO HENRIQUE DALPOSSO**

**CASCADEL - PR**

**Janeiro – 2010**

**GUSTAVO HENRIQUE DALPOSSO**

**ESTATÍSTICA ESPACIAL APLICADA À AGRICULTURA DE  
PRECISÃO**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Agrícola, em cumprimento parcial aos requisitos para obtenção do título de Mestre em Engenharia Agrícola, pela UNIOESTE/Campus de Cascavel, área de concentração Engenharia de Sistemas Agroindustriais.

Orientador: Professor Dr. Miguel Angel Uribe  
Opazo

Co-orientador: Professor Dr. Erivelto Mercante

**CASCADEL**

**Janeiro - 2010**

Dalposso, Gustavo Henrique  
D149 Estatística espacial aplicada à agricultura de precisão. /  
Gustavo Henrique Dalposso. – Cascavel, 2010.  
66 f.

Orientador: Prof. Dr. Miguel Angel Uribe Opazo.  
Dissertação(Mestrado) – Universidade Estadual do Oeste do  
Paraná.

1. Estatística Aplicada - Agricultura. 2. Geoestatística . 3.  
Estatística Espacial. 4. Índice de Moran. I. Opazo, Miguel  
Angel Uribe. II. Título.


CDD – 519.5

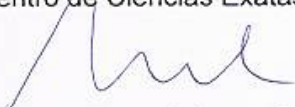
GUSTAVO HENRIQUE DALPOSSO

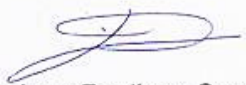
"Estatística espacial aplicada à agricultura de precisão"


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

Orientador:

  
Prof. Dr. Miguel Angel Uribe Opazo  
Centro de Ciências Exatas e Tecnológicas, UNIOESTE

  
Prof. Dr. Manuel Jesus Galea Rojas  
Universidade de Valparaiso, UV, Chile

  
Prof<sup>a</sup>. Dr<sup>a</sup>. Luciana Pagliosa Carvalho Guedes  
Coordenação do Curso Técnico em Agropecuária, UTFPR

  
Prof. Dr. Divair Christ  
Centro de Ciências Exatas e Tecnológicas, UNIOESTE

**DEDICO**

A Henrique Luis

## **AGRADECIMENTOS**

Ao meu orientador Dr. Miguel Angel Uribe Opazo que me conduziu neste trabalho e proporcionou o meu desenvolvimento no Programa de Pós-graduação e por ser um grande amigo e profissional.

Ao meu co-orientador Dr. Erivelto Mercante pela parceria firmada, pois esta além de render diversos trabalhos, rendeu uma grande amizade.

Ao meu amigo professor Jerry Adriani Johann, pelos trabalhos desenvolvidos no LEA, pois estes fizeram parte de minha formação.

A todos os pesquisadores do grupo GGEA, pela oportunidade oferecida.

Ao Programa de Pós-graduação em Engenharia Agrícola e a todos os professores do curso.

À CAPES pela bolsa concedida.

# SUMÁRIO

	LISTA DE TABELAS .....	viii
	LISTA DE FIGURAS .....	ix
	LISTA DE SÍMBOLOS .....	x
	RESUMO .....	xii
	ABSTRACT .....	xiii
<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>1</b>
<b>2</b>	<b>GEOESTATÍSTICA APLICADA À AGRICULTURA DE PRECISÃO .....</b>	<b>2</b>
2.1	INTRODUÇÃO .....	2
2.2	REVISÃO BIBLIOGRÁFICA .....	4
2.2.1	Evolução da geoestatística .....	4
2.2.2	Hipóteses e características importantes .....	4
2.2.3	Modelo geoestatístico .....	6
2.2.4	Parâmetros da estrutura de dependência espacial .....	8
2.2.5	Modelagem da estrutura de dependência espacial .....	10
2.2.6	Estimação de parâmetros .....	13
2.2.7	Estimador linear geoestatístico: Krigagem .....	15
2.2.8	Validação cruzada .....	16
2.2.9	Influência local .....	18
2.2.10	Acurácia em mapas temáticos .....	19
2.2.11	Modelagem cartográfica .....	24
2.3	MATERIAL E MÉTODOS .....	25
2.3.1	Material .....	25
2.3.2	Metodologia .....	25
2.4	RESULTADOS E DISCUSSÃO .....	29
2.5	CONCLUSÃO .....	42

2.6	REFERÊNCIAS .....	43
	<b>ESTATÍSTICA DE ÁREAS APLICADA À AGRÍCULTURA</b>	
3	<b>DE PRECISÃO .....</b>	<b>46</b>
3.1	INTRODUÇÃO .....	46
3.2	ANÁLISE DE DADOS DE ÁREAS .....	47
3.2.1	Análise exploratória .....	47
3.2.2	Elementos básicos .....	47
3.2.3	Indicador global de autocorrelação espacial .....	49
3.2.4	Diagrama de espalhamento de Moran .....	50
3.2.5	Mapa de espalhamento de Moran .....	51
3.2.6	Índice local de associação espacial ( <i>LISA</i> ) .....	52
3.2.7	Lisa map .....	52
3.2.8	Moran map .....	53
3.3	MATERIAL E MÉTODOS .....	54
3.3.1	Material .....	54
3.3.2	Metodologia .....	55
3.4	RESULTADOS E DISCUSSÃO .....	56
3.5	CONCLUSÃO .....	65
3.6	REFERÊNCIAS .....	66



## LISTA DE TABELAS

Tabela 1	Matriz dos erros genérica de ordem $k \times k$ .....	18
Tabela 2	Matriz de confusão da $k$ -ésima classe .....	20
Tabela 3	Métricas derivadas da matriz de confusão por classe .....	21
Tabela 4	Matriz de confusão total .....	21
Tabela 5	Métricas derivadas da matriz de confusão total .....	22
Tabela 6	Parâmetros do semivariograma omnidirecional .....	30
Tabela 7	Modelos ajustados e parâmetros obtidos do conjunto amostral completo da produtividade de trigo [ $t\ ha^{-1}$ ] .....	31
Tabela 8	Validação cruzada para a produtividade de trigo [ $t\ ha^{-1}$ ] relativa a todo o conjunto amostral .....	31
Tabela 9	Modelos ajustados e parâmetros do conjunto amostral da produtividade de trigo [ $t\ ha^{-1}$ ] sem as amostras influentes .....	33
Tabela 10	Validação cruzada para a produtividade de trigo [ $t\ ha^{-1}$ ] sem as amostras influentes .....	34
Tabela 11	Matriz dos erros dos <i>pixels</i> mapas de produtividade de trigo .....	36
Tabela 12	Produção estimada em cada classe da legenda .....	36
Tabela 13	Índices de acurácia de usuário ( $AU$ ) e produtor ( $AP$ ) .....	37
Tabela 14	Estatísticas do Índice Kappa ( $\hat{K}$ ) .....	37
Tabela 15	Elementos ( <i>pixels</i> ) da matriz de confusão da classe $C_k$ .....	37
Tabela 16	Métricas obtidas das matrizes de confusão de cada classe .....	38
Tabela 17	Matriz de confusão total da produtividade de trigo .....	38
Tabela 18	Índice $I$ de Moran dos índices NDVI e GVI .....	55
Tabela 19	Porcentagens de municípios em cada quadrante do gráfico de espalhamento de Moran dos índices NDVI e GVI .....	57
Tabela 20	Índice local de associação espacial (LISA) do índice NDVI .....	58
Tabela 21	Índice local de associação espacial (LISA) do índice GVI .....	59

## LISTA DE FIGURAS

Figura 1	Funções covariância e variograma .....	8
Figura 2	(A) Localização dos 50 pontos amostrais e (B) Divisão da área em talhões .....	23
Figura 3	Roteiro da análise geoestatística .....	26
Figura 4	Gráfico <i>boxcox</i> da produtividade de trigo .....	27
Figura 5	Gráfico espacial da produtividade de trigo .....	27
Figura 6	Gráfico <i>boxplot</i> da produtividade de soja .....	28
Figura 7	Gráfico das coordenada <i>versus</i> produtividade de trigo .....	28
Figura 8	Semivariogramas direcionais da produtividade de trigo .....	29
Figura 9	Semivariograma experimental da produtividade de trigo .....	29
Figura 10	Envelopes simulados .....	30
Figura 11	(A) Gráfico de diagnóstico $ L_{max} $ ; (B) Localização dos pontos influentes na área monitorada .....	32
Figura 12	Semivariograma experimental construído sem os pontos influentes ...	33
Figura 13	Semivariogramas ajustado: (A) completo e (B) sem os pontos influentes .....	34
Figura 14	(A) Mapa temático da produtividade de trigo utilizando 50 elementos amostrais na interpolação e (B) Mapa temático da produtividade de trigo utilizando 48 elementos amostrais na interpolação.....	35
Figura 15	Mapa da diferença de produtividade em módulo entre os dois mapas .	35
Figura 16	Representação dos tipos de contiguidade entre áreas. (A) Contiguidade Queen “rainha”, (B) Contiguidade Rook “torre” e (C) Contiguidade Bishop “bispo”.....	47
Figura 17	Construção do gráfico de dispersão do índice I de Moran .....	49
Figura 18	Legenda do mapa de espalhamento de Moran .....	50
Figura 19	<i>Shapefile</i> dos 36 municípios estudados na Região Oeste do Paraná .....	52
Figura 20	Roteiro de análise que será trabalhado com índices de vegetação .....	53
Figura 21	Evolução temporal do índice NDVI .....	54
Figura 22	Evolução temporal do índice GVI .....	55
Figura 23	Diagrama de espalhamento de Moran (Moran Scatterplot) para os índices NDVI e GVI em 23/11/2004, 25/12/2004 e 11/02/2005 .....	56
Figura 24	Mapa de significância LISA do índice de vegetação NDVI da soja ....	60
Figura 25	Mapa de significância LISA do índice de vegetação GVI da soja .....	61
Figura 26	LISA cluster map do índice NDVI da soja na safra 2004/2005.....	61

# LISTA DE SÍMBOLOS

## Capítulo 2

$Z(s)$	Valor do atributo na localização $s$
$Z(s+h)$	Valor do atributo na localização $s+h$
$C(h)$	Função de covariância
$\mu$	Média
$\mathfrak{R}^d$	Espaço euclidiano $d$ -dimensional ( $d \geq 1$ )
$\Sigma$	Matriz de covariância
$\mu(s)$	Modelo espacial linear
$\beta_i$	Parâmetros desconhecidos a serem estimados
$\varphi_1$	Efeito pepita ou erro de variância
$\varphi_2$	Contribuição ou variância de dispersão
$a$	Alcance ou raio de dependência espacial
$\varphi_3$	Função do alcance
$R$	Matriz que é função de $\varphi_3$
$I_n$	Matriz identidade de ordem $n$
$\gamma(h)$	Função semivariância
$C$	Patamar
$\theta$	Vetor de parâmetros.
$OLS$	Mínimos quadrados ordinários
$WLS1$	Mínimos quadrados ponderados
$ML$	Máxima verossimilhança
$RML$	Máxima verossimilhança restrita
$L(\theta)$	Função de verossimilhança
$l(\delta)$	Função log-verossimilhança
$EM$	Erro médio
$\hat{Z}(s_{(i)})$	Representa o valor predito por krigagem ordinária no ponto $s_i$ , sem considerar a observação $Z(s_i)$ .
$\overline{ER}$	Erro médio reduzido
$\sigma(s_{(i)})$	Desvio padrão da krigagem no ponto $s_i$ , sem considerar a observação $Z(s_i)$
$DP_{EM}$	Desvio padrão do erro médio
$S_{ER}$	Desvio padrão dos erros reduzidos
$EA$	Erro absoluto
$\omega$	Vetor de perturbação das respostas

$\omega_0$	Ponto de não perturbação
$LD(\omega)$	Afastamento da verossimilhança
$EG$	Exatidão global
$C_k$	Classe k do mapa temático
$AU$	Acurácia de usuário
$AP$	Acurácia de produtor
$\hat{K}$	Índice Kappa
$\sigma^2$	Variância
$T$	Índice Tau
$S$	Índice de sensibilidade
$E$	Índice de especificidade
$TFP_k$	Taxa de falso positivo
$TFN_k$	Taxa de falso negativo
$CCM$	Coefficiente de correlação de Matthews
$E$	Efeito pepita relativo

### Capítulo 3

$W$	Matriz de proximidade espacial
$Z$	Vetor dos desvios
$Wz$	Vetor de médias ponderadas
$I$	Índice de Moran Global
$\bar{x}$	Média dos atributos
$I_i$	Índice LISA
$\sigma^2$	variância da distribuição dos valores dos desvios

## RESUMO

As metodologias fornecidas pela estatística espacial são de grande importância para estudos envolvendo dados relacionados à agricultura, pois permitem conhecer a variabilidade espacial dos atributos estudados e identificar regiões que apresentam características semelhantes, o que permite realizar tratamentos localizados, maximizando as produtividades e minimizando os impactos causados pela aplicação de insumos em excesso.

Um dos ramos da estatística espacial é a geoestatística, que utiliza um conjunto de variáveis regionalizadas para modelar a estrutura de dependência espacial, possibilitando a elaboração de mapas temáticos. Atualmente os estudos geoestatísticos não terminam com a elaboração dos mapas, pois além de estimar o atributo monitorado em locais não amostrados se faz necessário investigar a qualidade destes mapas, investigando pontos influentes e utilizando medidas que permitam comparar mapas e realizar estimativas de áreas.

Outra forma de investigação é conhecida como estatística espacial de áreas, em que os objetos de análise são polígonos que representam talhões, bairros, municípios, estados entre outros. Neste tipo de análise, procura-se identificar autocorrelações espaciais em nível global e local, e a forma usual de apresentação dos resultados é feita utilizando mapas temáticos.

Neste trabalho utilizou-se a geoestatística para investigar a produtividade de trigo em uma área agrícola de 13,7 hectares no município de Salto do Lontra – Pr. Das 50 amostras coletadas, identificou-se duas como influentes e, com isso, optou-se por construir dois mapas temáticos e compará-los utilizando métricas derivadas da matriz dos erros. Os resultados mostraram que os mapas são diferentes e a retirada dos pontos influentes foi de fundamental importância para melhorar a qualidade do mapa temático, visto que a diferença entre a produtividade estimada e a produtividade real foi de apenas 40 quilos.

Para apresentar os recursos fornecidos pela estatística espacial de áreas comparou-se os índices de vegetação NDVI e GVI da produtividade de soja de 36 municípios da região Oeste do Paraná no ano agrícola 2004/2005. Os resultados permitiram identificar regiões com características semelhantes e que a soja é cultivada em períodos distintos na região.

**Palavras-chave:** matriz dos erros, índice I de Moran, matriz de contiguidade

# SPATIAL STATISTICS APPLIED TO PRECISION AGRICULTURE

## ABSTRACT

The methods provided by the spatial statistics are of great importance for studies involving data related to agriculture, for they allow one to know the space variability of the study and identify regions that have similar characteristics, which allows completely localized treatment, maximizing productivity and minimizing the impacts of excessive input application.

One of the branches of spatial statistics is geostatistics, which uses a set of regionalized variables to model the structure of spatial dependence, allowing the preparation of thematic maps. Currently, geostatistical studies do not end with the preparation of maps, but also estimates monitored the attribute in non-sampled locations. It is necessary to investigate the quality of these maps, investigating influential points and using measurements to compare maps and area estimations.

Another form of research is known as spatial statistics of areas where the objects of analysis are polygons representing blocks, neighborhoods, cities, states and others. This type of analysis seeks to identify spatial autocorrelation in global and local levels, and the usual form of reporting is through thematic maps.

In this work we used geostatistics to investigate the productivity of wheat in an agricultural area of 13.7 hectares in the municipality of Salto do Lontra – PR. Out of the 50 samples, two were identified as influential, and thus, we chose to build two thematic maps and to compare them using metrics derived from the matrix of errors. The results showed that the maps are different and the removal of influential points was essential to improve the quality of thematic map, since the difference between the estimated yield and actual yield was only 40 Kilos.

In order to display the resources provided by the spatial statistics of areas we compared to the vegetation rates NDVI and GVI's of soybean yield from 36 cities in Western Paraná in the agricultural year of 2004/2005. The results showed regions with similar characteristics and that soybeans grow at different times in the region.

**Keywords:** error matrix, Moran's I Index, contiguity matrix

## 1 INTRODUÇÃO

A constante evolução da informática, no que diz respeito ao desenvolvimento de softwares e hardwares, aliada ao crescimento da ciência, em relação à criação de novos métodos de pesquisa, está mudando a forma como o homem investiga a agricultura, e isso implica em melhores condições de trabalho para o profissional do campo, que em contrapartida, consegue oferecer produtos com melhor qualidade, proporcionando uma melhor condição de vida para todos, e principalmente, utilizando os recursos do meio ambiente racionalmente.

Esta prática de investigação, que visa estabelecer condições ideais as espécies cultivadas na agricultura, sejam elas químicas, físicas ou biológicas é conhecida como agricultura de precisão (AP). Segundo MOLIN (1997), a agricultura de precisão surge como uma nova demanda de informática na agricultura, pois implica na coleta e manipulação de uma grande quantidade de dados que só podem ser gerenciados por métodos computacionais. Todo o embasamento dessa nova tecnologia está na análise da variabilidade espacial dos fatores de produção, especialmente o solo. A partir dessa análise as decisões devem ser tomadas para que se faça então a aplicação dos insumos de uma forma localizada e com dosagens precisas.

Entre os diversos métodos utilizados para monitorar a agricultura, destaca-se a estatística espacial, termo este referente aos métodos estatísticos que incorporam nas análises informações sobre a localização, como as coordenadas geográficas do ponto em que a amostra foi coletada ou a localização do polígono que representa a região em estudo. Dentro da estatística espacial encontram-se diferentes abordagens para análise de dados georreferenciados, como por exemplo, a geoestatística e a estatística espacial de áreas.

A teoria das variáveis regionalizadas, ou geoestatística como é mais conhecida, é uma ciência que teve seu início em investigações realizadas em minas de ouro na África do Sul. Por meio de uma análise geoestatística é possível estimar dados em locais não amostrados analisando o comportamento espacial do fenômeno e minimizando o erro da estimação.

O conjunto de recursos proporcionado pela análise espacial em unidades de áreas é muito importante dentro do universo de possibilidades já disponibilizados em diversos Sistemas de Informação Geográfica. A apresentação dos resultados é geralmente feita através de mapas, que visam identificar regiões com características semelhantes, formando os conhecidos *clusters*. Nestes mapas, a variação do atributo no espaço não é representada por uma superfície contínua, como é o caso da geoestatística, e sim por variações abruptas de valores. Apesar de limitadas às características de cada área, muitas inferências sobre padrões ou comportamentos espaciais dos diversos atributos podem ser realizadas.

Neste contexto, o objetivo geral deste trabalho é utilizar recursos da estatística espacial para investigar variáveis vinculadas à agricultura.

## **2 GEOESTATÍSTICA APLICADA À AGRICULTURA DE PRECISÃO**

### **2.1 INTRODUÇÃO**

O conjunto de atividades agrícolas, compreendendo a produção de alimentos, fibras e energia, pode ser encarado como a alavanca da economia brasileira, e isto é de grande importância, visto que muitos países desenvolvidos têm excesso de produção agrícola, sendo esta uma das condições para se terem tornado ricos, pois o fato de serem auto-suficientes na produção agrícola tem sido determinante para os elevados níveis de prosperidade.

Um fato que de certa forma barra o crescimento da agricultura no Brasil é que muitas áreas agrícolas comerciais, por trabalharem com produção em larga escala, são tratadas como homogêneas, o que acaba trazendo prejuízos para o produtor e para o meio ambiente. Um exemplo é a aplicação de fertilizantes calculados com base em índices médios de fertilidade. Esta aplicação é feita de forma uniforme em toda a área e, como consequência, áreas com maior nível de fertilidade são adubadas em excesso e áreas mais pobres não são corrigidas. O mesmo ocorre na aplicação de agro químicos no combate de pragas e doenças, em que o tratamento uniforme gera problemas de poluição em áreas com baixos níveis de infestação.

Uma solução eficaz para contornar este problema é apresentada pela agricultura de precisão por meio da geoestatística, ramo da estatística espacial que permite estudar superfícies contínuas. A grande diferença em relação à estatística tradicional é que, por meio desta metodologia, além do valor do atributo no ponto em estudo, a sua localização também é incorporada na análise. Esse fato faz com que elementos amostrais próximos tenham valores mais semelhantes e sejam mais correlacionadas do que elementos amostrais mais distantes do ponto estudado. Desta maneira, é possível realizar estimações em locais não amostrados, permitindo criar mapas temáticos que descrevam o comportamento do atributo que está sendo investigado na área em estudo.

A geoestatística fornece uma solução para este tipo de problema através do interpolador krigagem, nome dado em homenagem ao engenheiro de minas Daniel Krige, pioneiro que, ao realizar trabalhos com dados de mineração, concluiu que a variação destes apresentava uma estruturação que dependia da distância de amostragem e, desta constatação, surgiram os conceitos básicos de geoestatística.



Atualmente a geoestatística encontra-se consolidada como ciência, porém, sua evolução é contínua e dia após dia surgem novas teorias, métodos, programas e índices que ajudam o pesquisador a melhor entender o comportamento do atributo na área estudada.

## **2.2 REVISÃO BIBLIOGRÁFICA**

### **2.2.1 Evolução da geoestatística**

A geoestatística tem por objetivo a caracterização da dispersão espacial e espaço-temporal das grandezas que definem a quantidade e a qualidade de recursos naturais, tais como florestas, recursos geológicos, hidrológicos, ecológicos ou outros fenômenos espaciais em que os atributos manifestam uma certa estrutura no espaço e/ou no tempo, como a contaminação de solos e aquíferos, temperatura e pluviometria de uma região. O seu corpo metodológico consiste basicamente num conjunto de instrumentos estatísticos que quantificam a continuidade espacial da grandeza em estudo, em modelos de interpolação espacial tendo por base a sua variabilidade estrutural (SOARES, 2000).

A teoria das variáveis regionalizadas surgiu em meados de 1960 ligada à escola francesa hoje conhecida como *Centre de Geoestatistique de Fontainebleau*, fundada por Georges Matheron. Ao estudar dados de concentração de ouro em 1951 na África do Sul, o engenheiro de minas Daniel Krige concluiu que não conseguia encontrar sentido nas variâncias se não levasse em consideração a distância entre as amostras. Baseado nessa observação, Matheron desenvolveu a teoria das variáveis regionalizadas, que contém os fundamentos da geoestatística.

Segundo LADIM (2006), as variáveis regionalizadas são constituídas por um duplo aspecto contraditório. Pela sua característica “aleatória” apresenta irregularidades e variação imprevisível de um ponto para outro e pela sua característica “estrutural” apresenta relações existentes entre os pontos no espaço motivadas pela sua gênese. Em outras palavras, é impossível estimar com exatidão o valor de um atributo - por exemplo, a produtividade num determinado ponto da área agrícola (aspecto aleatório) -, mas é provável que se encontre produtividade elevada perto de locais com produtividade alta (aspecto estrutural).

### **2.2.2 Hipóteses e características importantes**

Tem-se que uma variável regionalizada é uma função numérica com distribuição espacial, que varia de um ponto a outro com continuidade aparente, mas cujas variações não podem ser representadas por uma função simples, assim, para melhor compreensão dessa ferramenta, se faz necessário uma apresentação teórica de algumas características importantes.

- a) Localização

Os valores de uma variável regionalizada são dependentes de suas funções espaciais relativas dentro do campo geométrico e, além disso, estes valores são dependentes do tamanho da amostra, de sua forma e orientação (suporte amostral).

#### b) Continuidade

A continuidade espacial apresentada pela variável regionalizada entre amostras vizinhas reflete o grau de dependência ou independência espacial entre as amostras. Espera-se que as amostras mais próximas apresentem maior dependência do que as que estão mais distantes (CRESSIE, 1993). Dada uma característica quantitativa  $Z(s)$ , os diagramas de representação dos pares de pontos  $Z(s)$  vs  $Z(s+h)$ , são algumas das estatísticas que contém informação mais rica sobre a continuidade espacial de  $Z(s)$ . Quando a continuidade espacial não é identificada, diz-se que há presença de “efeito pepita puro”. Isto indica um tipo extremo de comportamento do variograma próximo à origem e reflete a variação espacial de um fenômeno de transição onde para um dado valor de patamar a amplitude terá um valor infinitesimalmente menor que as distâncias de observação (JOURNEL & HUIJBREGTS, 1978). Diante da presença do efeito pepita puro ressalta-se não se deve utilizar o método geoestatístico de interpolação.

#### c) Anisotropias

A anisotropia é uma característica muito frequente nos elementos da natureza, isto é, a variabilidade ou distribuição espacial de tais elementos ocorre mais intensamente numa direção e menos intensamente em outra direção. Quando a variabilidade espacial em uma área apresenta comportamento semelhante para distintas direções, dizemos que os dados são isotrópicos. A isotropia é de suma importância para a estimativa de locais não-amostrados. É utilizada na construção de mapas temáticos de variabilidade, pois, por meio da identificação de que a variável analisada é isotrópica, o estudo da dependência espacial pode ser feito por meio de um único semivariograma, chamado de omnidirecional (CARVALHO et al., 2008).

#### d) Estacionaridade

O conjunto de variáveis aleatórias  $Z(s_i)$ ,  $i = 1, \dots, n$ , constituem uma função aleatória da qual só se conhece uma realização  $z(s_i)$ , ou seja, o conjunto de dados amostrais.

Para estimar valores em locais não amostrados, deve-se introduzir as restrições de estacionaridade estatística. A existência de estacionaridade permite que o experimento

possa ser repetido mesmo que as amostras sejam coletadas em pontos diferentes, pois elas pertencem à mesma população com os mesmos momentos estatísticos (VIEIRA, 2000). Uma função aleatória é considerada estacionária quando o valor esperado para sua realização é o mesmo para todos os pontos da área em estudo, ou seja,  $E[Z(s_1)] = E[Z(s_2)] = \dots = E[Z(s_i)] = \mu$ .

Este parâmetro, que passa a ser independente da localização de  $s_i$ , pode ser estimado por uma média aritmética dos valores das realizações das variáveis aleatórias. Concretamente, a hipótese de estacionaridade da média, ou de primeira ordem como é conhecida, implica que  $\mu$  possa ser estimada pela média aritmética das observações. Julgar esta hipótese de estacionaridade como apropriada é julgar a média das amostras como representativa da área em estudo.

Para a análise geoestatística é necessário também a estacionaridade de segunda ordem, que implica que para cada par de uma variável aleatória, a função de covariância  $C(h) = Cov[Z(s), Z(s+h)]$  existe e seja dependente da distância  $h$ , implicando na ocorrência de variância finita. Usualmente, a aceitação de uma estacionaridade de segunda ordem não pode ser satisfeita. Necessita-se então de outro modelo estatístico, menos limitado, que é baseado na hipótese intrínseca, a qual considera apenas que a média dos valores  $Z(s)$  e a variância dos incrementos  $Z(s) - Z(s+h)$  ocorrem independentemente da localização na região, sendo função apenas do valor de  $h$ . Esta hipótese requer somente a hipótese de existência do semivariograma, sem a exigência da variância finita. Assim, a variância de  $Z(s)$  não é finita, mas a variância do primeiro incremento de  $Z$  é finita, e este incremento é fracamente estacionário (VIEIRA et al., 1983). Assumida a estacionaridade, por meio da hipótese intrínseca, e considerando que a associação das variáveis em pontos distintos é maior à medida que se reduz a distância entre eles, o passo seguinte é descrever e modelar estas relações entre distâncias e associação espacial.

### 2.2.3 Modelo geoestatístico

O fato dos modelos geoestatísticos incorporarem a incerteza na sua formalização não significa que o fenômeno em si tenha resultado de um processo aleatório, mas serve somente de base metodológica à inferência espacial ou estimação de grandezas em áreas não amostradas e à quantificação da incerteza associada ao estimador (SOARES, 2000).

MARDIA & MARSHALL (1984) consideram um processo estocástico gaussiano  $\{Z(s_i), s_i \in S\}$ , em que  $S \subset \mathfrak{R}^d$ , sendo  $\mathfrak{R}^d$  espaço euclidiano  $d$ -dimensional ( $d \geq 1$ ). Ou seja, o processo  $Z = (Z(s_1), \dots, Z(s_n))^T$ , em que  $s_i$  e  $s_u$  ( $i, u = 1, \dots, n$ ) são localizações espaciais conhecidas, têm distribuição normal  $n$ -variada com vetor de médias  $\mu \mathbf{1}$  e matriz de

covariância  $\Sigma$ , isto é,  $Z \sim N_n(\mu\mathbf{1}, \Sigma)$ , em que  $\mu$  é uma constante;  $\mathbf{1}$  é um vetor formado pelo algarismo 1 de ordem  $n \times 1$  e  $\Sigma$  é uma matriz definida positiva, de ordem  $n$ , dada por  $\Sigma = [Cov(Z(s_i), Z(s_u))]$ . Supondo-se que os dados são descritos pelo modelo da Equação (1):

$$Z(s_i) = \mu(s_i) + \varepsilon(s_i) \quad (1)$$

em que os termos determinístico  $\mu(s_i)$  e estocástico  $\varepsilon(s_i)$  podem depender da localização espacial em que  $Z(s_i)$  foi obtida. Assume-se que o erro estocástico  $\varepsilon$  tem média zero,  $E[\varepsilon(s_i)] = 0$ , e que a variação entre pontos no espaço é determinada por alguma função de covariância  $C(s_i, s_u) = Cov[\varepsilon(s_i), \varepsilon(s_u)]$ .

Assume-se que para algumas funções conhecidas de  $s$ , como  $x_1(s), x_2(s), \dots, x_p(s)$ , a média do processo estocástico é dada pelo modelo espacial linear apresentado na Equação (2):

$$\mu(s) = \sum_{u=1}^p x_u(s) \beta_u \quad (2)$$

em que  $\beta_1, \dots, \beta_p$  são parâmetros desconhecidos a serem estimados.

A função de covariância  $C(s_i, s_u)$  é também especificada por um vetor  $q$ -dimensional  $\varphi = (\varphi_1, \dots, \varphi_q)^T$ . Por simplicidade, pode-se utilizar as seguintes notações:  $Z(s_i) = z_i$ ,  $Z = (z_1, \dots, z_n)^T$ ,  $x_{iu} = x_u(s_i)$ ,  $x_i = (x_{i1}, \dots, x_{ip})^T$ , em que  $X$  é uma matriz  $n \times p$  com suas linhas  $x_i^T$ ,  $\beta = (\beta_1, \dots, \beta_p)^T$ ,  $\varepsilon_i = \varepsilon(s_i)$ ,  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ , com  $i = 1, \dots, n$  e  $u = 1, \dots, p$ .

Desta forma,  $\mu(s_i) = x_i^T \beta$ , e então,

$$z_i = x_i^T \beta + \varepsilon_i \quad (3)$$

De forma equivalente, em notação matricial, tem-se:

$$Z = X\beta + \varepsilon \quad (4)$$

Então,  $E(\varepsilon) = 0$  e a matriz de covariância de  $\varepsilon$  é  $\Sigma = [(\sigma_{iu})]$ , em que  $\sigma_{iu} = C(s_i, s_u)$ . Assume-se que  $\Sigma$  é não singular e que  $X$  tem colunas com posto completo. Considerando-se de maneira particular a forma paramétrica da matriz de covariância,

$$\Sigma = \varphi_1 I_n + \varphi_2 R \quad (5)$$

em que:

$\varphi_1$  : efeito pepita ou erro de variância;

$\varphi_2$  : contribuição ou variância de dispersão;

R: matriz que é função de  $\varphi_3$  onde  $\varphi_3$  é função do alcance ( $a$ ) do modelo;

$I_n$ : matriz identidade de ordem  $n$ .

A forma paramétrica da matriz de covariância, dada na Equação 5, ocorre para vários processos isotrópicos, nos quais a covariância  $C(s_i, s_u)$  é definida segundo a função de covariâncias  $C(h_{iu}) = \varphi_2 r_{iu}$ , em que  $h_{iu} = \|s_i - s_u\|$  é a distância euclidiana entre  $s_i$  e  $s_u$ .

Nas funções de covariâncias  $C(h_{iu})$ , a variância do processo estocástico  $Z$  é  $C(0) = \varphi_1 + \varphi_2$ . Assim, a função variograma pode ser definida como:

$$\gamma(h) = C(0) - C(h) \quad (6)$$

Pela Equação (6), pode-se obter a função covariância:

$$C(h) = C(0) - \gamma(h) \quad (7)$$

A covariância  $C(h)$  e o variograma  $\gamma(h)$  são equivalentes para caracterizar a dependência espacial.

#### 2.2.4 Parâmetros da estrutura de dependência espacial

Os parâmetros fornecidos pela estrutura de dependência espacial são de grande importância na análise geoestatística, pois são utilizados no processo de krigagem, que é o interpolador geoestatístico que cria o mapa temático do atributo estudado.

##### i) Efeito pepita ( $\varphi_1$ )

Teoricamente, o valor do variograma é nulo:  $\gamma(h) = 0$  para  $h = 0$ . Na prática há um valor mínimo de  $h$  entre amostras para o qual pode ser quantificado o valor de  $\gamma(h)$ . Quando esse valor de  $\gamma(h_{\min})$  é elevado, o que significa que há uma grande variabilidade à pequena escala, isto é, referente às menores distâncias entre amostras ou observações, pode acontecer de  $\gamma(h)$  não tender para 0 quando  $h$  tende para 0. Há na realidade uma

inflexão ou descontinuidade no andamento do variograma a uma escala não captada pelas amostras, isto é, entre  $h = 0$  e  $h_{min}$ . Nestas situações, o variograma é modelado por uma ordenada na origem igual a uma constante  $\varphi_1$  denominada por “efeito pepita” (SOARES, 2000).

ii) Contribuição ( $\varphi_2$ )

É também conhecida como variância de dispersão e representa as diferenças espaciais entre os valores de uma variável tomada em dois pontos separados por distâncias cada vez maiores.

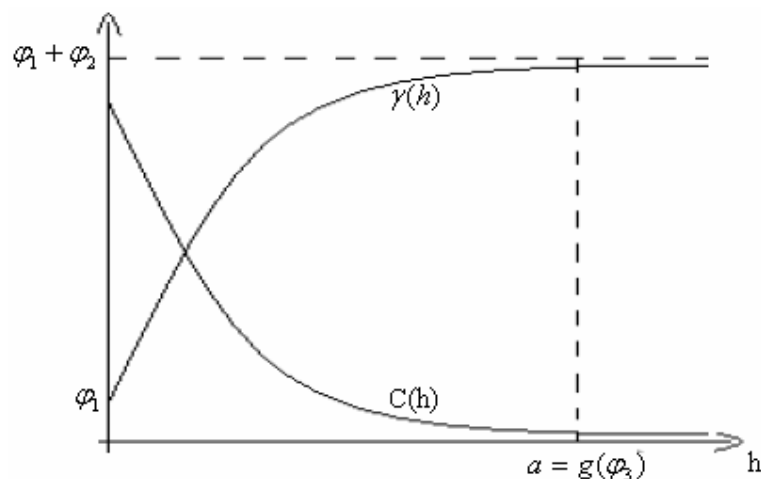
iii) Alcance ( $a = g(\varphi_3)$ )

É a distância a partir da qual os elementos amostrais passam a ser não correlacionados. Em outras palavras, o alcance reflete o grau de homogeneização entre os elementos amostrais, ou seja, quanto maior for o alcance maior será a homogeneidade entre as amostras. O alcance é a distância que separa o campo estruturado (amostras correlacionadas) do campo aleatório (amostras independentes).

iv) Patamar ( $C = \varphi_1 + \varphi_2$ )

O Patamar ( $\varphi_1 + \varphi_2$ ) corresponde ao ponto onde toda semivariância da amostra é de influência aleatória, correspondendo a variância total ( $s^2$ ) obtida pela estatística clássica (TRANGMAR et al., 1985).

A Figura 1 ilustra o andamento típico das funções variograma  $\gamma(h)$  e covariância  $C(h)$ . Como a função variograma é uma medida da variância das diferenças nos valores da variável regionalizada entre pontos separados por uma distância  $h$ , pontos mais próximos, por estarem correlacionados terão essa variância pequena, aumentando à medida que os pontos se distanciam.



**Figura 1** – Funções covariância e variograma.

O contrário acontece para a função covariância, que é grande para distâncias pequenas diminuindo à medida que a distância aumenta, pois esta função mede a correlação entre pontos separados por uma distância  $h$ .

## 2.2.5 Modelagem da estrutura de variabilidade espacial

Os semivariogramas e correlogramas amostrais podem não ser suficientes para descrever o padrão de variabilidade espacial do fenômeno estudado, visto que contém estimativas apenas para algumas distâncias que a malha amostral permite calcular. Um dos produtos finais da análise geoestatística é o mapa temático gerado por krigagem, e este tipo de interpolação exige medidas para outras distâncias.

Os modelos ajustados para o semivariograma experimental, apresentados por ISAACS & SRIVASTAVA (1989) e CRESSIE (1993), são divididos em modelos transitivos, que possuem patamar, e modelos não transitivos, que não possuem patamar. Modelos com patamar são aqueles em que a semivariância estabiliza-se em torno de um valor chamado patamar ( $C$ ) após certa distância, chamada alcance ( $a$ ). Nestes casos, observações separadas por distâncias maiores que  $a$  são consideradas não correlacionadas. Os modelos esféricos, exponencial, gaussiano, circular, Matérn e exponencial potência são exemplos de funções com patamar.

Modelos sem patamar são aqueles em que não há indicação de estabilização da semivariância. Este comportamento pode ser devido ao fato de que não foram obtidos elementos amostrais suficientes afastados para que o patamar fosse detectado, ou pode ser uma indicação de que a hipótese intrínseca não é válida.



a) Modelo esférico

Trata-se de um dos modelos mais usuais em geoestatística. Ele apresenta um crescimento rápido e atinge o patamar a 2/3 do alcance. A função semivariância deste modelo tem a forma:

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[ \frac{3}{2} \left( \frac{h}{\varphi_3} \right) - \frac{1}{2} \left( \frac{h}{\varphi_3} \right)^3 \right], & 0 < h \leq \varphi_3 \\ \varphi_1 + \varphi_2, & h > \varphi_3 \end{cases} \quad (8)$$

A função covariância é dada por:

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & h = 0 \\ \varphi_2 \left[ 1 - \frac{3}{2} \left( \frac{h}{\varphi_3} \right) + \frac{1}{2} \left( \frac{h}{\varphi_3} \right)^3 \right], & 0 < h < \varphi_3 \\ 0, & h > \varphi_3 \end{cases} \quad (9)$$

b) Modelo exponencial

Este modelo atinge o patamar assintoticamente, com o alcance prático definido como a distância na qual o valor do modelo é 95% do patamar (ISAAKS & SRIVASTAVA, 1989). A função semivariância deste modelo tem a forma:

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[ 1 - \exp\left(-\frac{h}{\varphi_3}\right) \right], & 0 < h \leq \varphi_3 \\ \varphi_1 + \varphi_2, & h > \varphi_3 \end{cases} \quad (10)$$

A função covariância é dada por:

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & h = 0 \\ \varphi_2 \left[ \exp\left(-\frac{h}{\varphi_3}\right) \right], & 0 < h \leq \varphi_3 \\ 0, & h > \varphi_3 \end{cases} \quad (11)$$

c) Modelo gaussiano

Os modelos esférico e exponencial apresentam um crescimento relativamente rápido junto da origem, denunciando um comportamento típico de fenômenos relativamente irregulares. Outros fenômenos, bem mais regulares e contínuos, denunciam um crescimento lento de  $\gamma(h)$  com um comportamento parabólico na origem. Este é o caso de variogramas ajustáveis por modelos gaussianos. A função semivariância deste modelo tem a forma:

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

A função covariância é dada por:

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

d) Modelo circular

O modelo circular apresenta validade apenas nos planos unidimensionais e bidimensionais, não podendo ser aplicado a planos tridimensionais, onde é aplicado o modelo esférico. A função semivariância deste modelo tem a forma:

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

A função covariância é dada por:

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & h = 0 \\ \varphi_2 \frac{2}{\pi} \left[ \cos^{-1} \left( \frac{h}{\varphi_3} \right) - \frac{h}{\varphi_3} \sqrt{1 - \frac{h^2}{\varphi_3^2}} \right], & 0 < h \leq \varphi_3 \end{cases} \quad (15)$$

e) Família Matérn

MATÉRN (1986) apresentou uma função chamada família Matérn. Esta função é definida, em termos de modelo teórico da função semivariância, e é dada pela Equação (16):

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[ (1 - 2^{k-1} \Gamma(k))^{-1} \left( \frac{h}{\varphi_3} \right)^k K_k \left( \frac{h}{\varphi_3} \right) \right], & h > 0 \end{cases} \quad (16)$$

em que:

$\varphi_1, \varphi_2, \varphi_3$  e  $k$  são parâmetros;

$K_k$  é a função de Bessel de terceiro tipo, de ordem  $k$  ;

A função covariância é dada por:

$$C(h) = \begin{cases} 0, & h = 0 \\ \varphi_2 \left[ (2^{k-1} \Gamma(k))^{-1} \left( \frac{h}{\varphi_3} \right)^k K_k \left( \frac{h}{\varphi_3} \right) \right], & h > 0 \end{cases} \quad (17)$$

### 2.2.6 Estimação de parâmetros

Escolher um modelo adequado implica em obter estimadores do vetor de parâmetro  $\theta = (\varphi_1, \varphi_2, \varphi_3)^T$  com métodos de estimação tais como: mínimos quadrados ordinários (OLS), mínimos quadrados ponderados (WLS1) (CRESSIE, 1985), máxima verossimilhança (ML) e máxima verossimilhança restrita (RML) (MARDIA & MASHALL, 1984).

O método de estimação de mínimos quadrados ordinários - OLS, consiste na obtenção de um vetor de parâmetros  $\hat{\theta}$  estimador de  $\theta = (\varphi_1, \varphi_2, \varphi_3)^T$ , que minimiza a Equação (18), onde  $k$  indica o número de lags que constituem o semivariograma experimental,  $\hat{\gamma}(h_j)$  é o valor estimado da semivariância experimental, que correspondente

ao  $j$ -ésimo *lag*, para  $j=1, 2, \dots, k$  e  $\gamma(h_j, \hat{\theta})$  representa o valor estimado que corresponde ao  $j$ -ésimo *lag* para  $j=1, 2, \dots, k$ ; obtido pelo modelo ajustado ao semivariograma experimental, e depende do vetor de parâmetros  $\theta$ .

$$\sum_{j=1}^k [\hat{\gamma}(h_j) - \gamma(h_j, \theta)]^2 \quad \text{ou} \quad \sum_{j=1}^k [\hat{C}(h_j) - C(h_j, \theta)]^2 \quad (18)$$

O método de estimação de mínimos quadrados ponderados - WLS1 define pesos diretamente proporcionais ao número de pares de pontos amostrais que contribuem para a semivariância estimada a cada *lag*. Segundo o método, o vetor de parâmetros  $\theta$  a ser escolhido é o que minimiza a Equação (19):

$$\sum_{j=1}^k N(h_j) [\hat{\gamma}(h_j) - \gamma(h_j, \theta)]^2 \quad \text{ou} \quad \sum_{j=1}^k N(h_j) [\hat{C}(h_j) - C(h_j, \theta)]^2 \quad (19)$$

A estimação de parâmetros por máxima verossimilhança - ML é um método analítico, muito utilizado na teoria estatística. A função de verossimilhança de  $n$  variáveis aleatórias  $Z_1, \dots, Z_n$  é definida como a densidade conjunta dada por  $f_{Z_1, \dots, Z_n}(z_1, \dots, z_n; \theta)$ , que deverá ser considerada como uma função do vetor de  $q$ -parâmetros desconhecidos  $\theta = (\theta_1, \dots, \theta_q) \in \Theta$  (espaço paramétrico). Seja  $(Z_1, \dots, Z_n)$  um vetor aleatório  $n$ -dimensional de uma população com função densidade de probabilidade conjunta  $f_{Z_1, \dots, Z_n}(z_1, \dots, z_n; \theta)$ , então a função de verossimilhança será:

$$L(\theta) = f_{Z_1, \dots, Z_n}(z_1, \dots, z_n; \theta) \quad (20)$$

A função de verossimilhança pode ser entendida como a intensidade de contribuições dos parâmetros na produção de uma dada amostra. Sob uma distribuição normal  $n$ -variada dos erros, tem-se  $Z \sim N_n(X\beta, \Sigma)$  e os parâmetros desconhecidos do modelo podem ser estimados maximizando-se a função de verossimilhança, como considerado por MARDIA & MARSHALL (1984). Por motivos de simplicidade de cálculos, utiliza-se o logaritmo da função verossimilhança, chamada de função log-verossimilhança, definida por:

$$l(\delta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma(\theta)| - \frac{1}{2} (Z - X\beta)^\top \Sigma(\theta)^{-1} (Z - X\beta) \quad (21)$$

A função de verossimilhança é, basicamente, uma função do vetor de parâmetros  $\theta \in \Theta$  então, a melhor estimativa para o vetor de parâmetros será aquela que maximiza a função de verossimilhança ou de logaritmo da função verossimilhança.

Outro método, freqüentemente utilizado para estimar os parâmetros da matriz de covariância  $\Sigma(\theta)$  é o método de máxima verossimilhança restrita – RML, que obtém estimadores menos viciados que os estimadores obtidos por ML. Nos ensaios de campo na agricultura e nas pesquisas com melhoramento animal, como método de seleção, a estimação de máxima verossimilhança restrita dos parâmetros de covariância espacial, tem sido preferida à de máxima verossimilhança (CRESSIE & LAHIRI, 1996). Neste caso, o logaritmo da função de log-verossimilhança restrita é dada pela equação (22):

$$l_R(\theta) = \left( \frac{n-p}{2} \right) \log(2\pi) + \frac{1}{2} \log(|X^T X|) - \frac{1}{2} \log|\Sigma| - \frac{1}{2} \log(|X^T \Sigma^{-1} X|) - \frac{1}{2} Z^T P Z, \quad (22)$$

em que ,

$$P = \Sigma^{-1}(I - A), \text{ com } A = X(X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1}.$$

Com o desenvolvimento de recursos computacionais os métodos ML e RML têm sido técnicas de estimação muito utilizadas, pois têm propriedades da teoria das grandes amostras que torna o seus resultados mais eficientes (DIGGLE & RIBEIRO JR., 2007).

## 2.2.7 Estimador linear geoestatístico: Krigagem

O estimador geoestatístico é denominado por Krigagem, que é uma denominação que foi empregue pela primeira vez por Matheron, em homenagem aos trabalhos pioneiros de Krige. Krigagem Normal (*Ordinary Kriging*) é a denominação mais usual dos algoritmos de krigagem, uma família que cobre os estimadores não-estacionários (Krigagem Simples, Krigagem com modelo de deriva, também conhecida por krigagem universal, e krigagem com deriva externa), o estimador de correionalizações (Co-Krigagem), o estimador de funções de distribuição de probabilidades, o estimador de variáveis categóricas para a caracterização da morfologia de corpos em fenômenos espaciais (Krigagem da Indicatriz) e os estimadores não-lineares (Krigagem MultiGaussiana e Krigagem Disjuntiva).

A krigagem ordinária é interpretada como o valor interpolado de uma variável regionalizada Z, em um local  $s_0$ , podendo ser determinada pela Equação (23):

$$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i) \quad (23)$$

restrito a:

$$\sum_{i=1}^n \lambda_i = 1$$

em que:

$\hat{Z}(s_0)$  : Representa o valor estimado de  $Z$  em  $s_0$ ;

$n$ : Representa a quantidade de valores medidos  $Z(s_i)$ ;

$\lambda_i$  : São os pesos associados aos valores medidos.

A krigagem ordinária para  $n$  pontos amostrais constitui-se de  $n+1$  equações, com  $n+1$  incógnitas, podendo ser representada por meio de uma equação matricial dada em (24):

$$K \cdot \lambda = L \quad (24)$$

sendo:

$$K = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} & 1 \\ C_{21} & C_{22} & \dots & C_{2n} & 1 \\ \cdot & \cdot & \dots & \cdot & \cdot \\ C_{n1} & C_{n2} & \dots & C_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix}, \quad \lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdot \\ \lambda_n \\ \mu \end{bmatrix}, \quad L = \begin{bmatrix} C_{10} \\ C_{20} \\ \cdot \\ C_{n0} \\ 1 \end{bmatrix}$$

Nesta equação  $K$  é uma matriz de semivariâncias de ordem  $(n+1) \times (n+1)$  entre os pares de pontos amostrados;  $\lambda$  é o vetor de pesos;  $L$  é um vetor de semivariâncias da amostra em relação ao ponto a ser estimado  $s_0$ , e  $\mu$  representa o parâmetro de Lagrange. O vetor de parâmetros  $\lambda$  é obtido pelo produto do inverso da matriz  $K$  pelo vetor  $L$ .

No método da krigagem, os pesos são atribuídos de acordo com a variabilidade espacial expressa no semivariograma, no entanto, o que torna a krigagem um interpolador ótimo é a maneira como os pesos são distribuídos, não sendo tendencioso, tendo variância mínima e possibilitando que se conheça a variância da estimativa (WEBSTER & OLIVER, 1990).

### 2.2.8 Validação cruzada

A validação cruzada, ou *cross validation*, trata-se da adaptação da estatística paramétrica de validação cruzada a um conjunto de dados  $Z(s_\alpha)$ . O processo consiste em

retirar-se uma amostra  $Z(s_0)$  do conjunto de dados e estimar-se no local da amostra o valor  $[\hat{Z}(s_0)]$ , por exemplo, utilizando o estimador linear de krigagem. Do conjunto formado por valores reais e estimados  $[Z(s_\alpha), \hat{Z}(s_\alpha)]$ , calcula-se algumas estatísticas básicas, como médias e variância dos desvios e média dos módulos dos desvios, com objetivo de aferir a qualidade do modelo escolhido para o variograma.

CRESSIE (1993) apresenta o erro médio reduzido ( $\overline{ER}$ ), o desvio padrão dos erros médios ( $DP_{EM}$ ) e o desvio padrão dos erros reduzidos ( $S_{ER}$ ) como instrumento para avaliar modelos. O erro médio é definido como:

$$EM = \frac{1}{n} \sum_{i=1}^n (Z(s_i) - \hat{Z}(s_{(i)})) \quad (25)$$

em que:

$n$  : representa o número de dados;

$Z(s_i)$  : representa o valor observado no ponto  $s_i$ ;

$\hat{Z}(s_{(i)})$  : representa o valor predito por krigagem ordinária no ponto  $s_i$ , sem considerar a observação  $Z(s_i)$ .

Esse procedimento pode ser visto como um experimento no qual se imita o processo de estimação, ao supor que nunca se toma uma amostra naquela localização. Uma vez que a estimação é feita, pode-se compará-la ao valor da amostra que foi inicialmente removida do conjunto de dados amostrais. Este procedimento, método de “deixar um fora”, é repetido para todas as amostras disponíveis (FARACO et al. 2008).

Segundo McBRATNEY & WEBSTER (1986), o erro médio reduzido é definido como:

$$\overline{ER} = \frac{1}{n} \sum_{i=1}^n \frac{Z(s_i) - \hat{Z}(s_{(i)})}{\sigma(s_{(i)})} \quad (26)$$

em que:

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

O desvio padrão dos erros reduzidos é obtido a partir da seguinte expressão:

$$S_{ER} = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{|Z(s_i) - \hat{Z}(s_{(i)})|}{\sigma(s_{(i)})}} \quad (27)$$

FARACO et al. (2008) comentam que conhecendo-se o conjunto de valores medidos e preditos por krigagem ordinária  $Z(s_i)$  e  $\hat{Z}(s_{(i)})$ , respectivamente, é possível definir o erro absoluto na unidade da variável da unidade estudada, pela Equação (28):

$$EA = \sum_{i=1}^n |Z(s_i) - \hat{Z}(s_{(i)})| \quad (28)$$

Segundo McBRATNEY & WEBSTER (1986), aplicando-se a condição de não tendenciosidade, o valor populacional para o erro médio reduzido deve ser zero e do desvio padrão do erro reduzido deve ser igual a um. Assim, os valores de  $EM$  e  $\overline{ER}$  mais próximos de zero, o menor valor do  $DP$  e o valor de  $S_{ER}$  mais próximo de um são os critérios para a escolha do melhor modelo ajustado.

### 2.2.9 Influência local

Em muitas situações pode-se observar um conjunto de dados com observações aberrantes ou discrepantes que podem ser consideradas influentes, isto é, mudam algum tipo de decisão na construção de modelos geoestatísticos (BORSSOI et al., 2009). O método de influência local proposto por COOK (1986) avalia o efeito simultâneo de observações sobre os estimadores de máxima verossimilhança (MV) sem a necessidade de sua eliminação do conjunto de dados. CHRISTENSEN et al. (1993) estudam métodos de diagnósticos baseado em eliminação de casos em modelos lineares espaciais.

Para um conjunto de dados observados seja  $l(\theta)$  o logaritmo da função verossimilhança do modelo postulado, em que  $\theta = (\beta, \varphi_1, \varphi_2, \varphi_3)^T$  e seja  $\omega$  um vetor de perturbações pertencente a um espaço de perturbações  $\Omega$ . Seja  $l(\theta/\omega)$  o logaritmo da função verossimilhança correspondente ao modelo perturbado por  $\omega \in \Omega$ . Assume-se que existe um  $\omega_0 \in \Omega$  tal que  $l(\theta) = l(\theta/\omega_0)$ , para todo  $\theta$  e que  $l(\theta/\omega)$  é duas vezes diferenciável em  $(\theta^T, \omega^T)^T$ .

Considere-se o seguinte esquema de perturbação:  $Z_\omega = Z + \omega$ , com  $\omega = (\omega_1, \dots, \omega_n)^T$  vetor de perturbação das respostas e  $\omega_0 = (0, \dots, 0)^T$  o ponto de não perturbação. Com este esquema de perturbação pretende-se detectar possíveis *outliers* nos dados que afetam o estimador de máxima verossimilhança de  $\theta$ . Logo o logaritmo da função verossimilhança perturbada  $l(\theta/\omega)$ , para o modelo gaussiano está dado por:

$$l(\theta / \omega) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (Z_\omega - X\beta)^T \Sigma^{-1} (Z_\omega - X\beta)$$



A influência da perturbação  $\omega$  no estimador de MV do vetor de parâmetros  $\theta$  pode ser avaliada pelo afastamento da verossimilhança, definido por  $LD(\omega) = 2(l(\hat{\theta}) - l(\hat{\theta}_\omega))$ , em que,  $\hat{\theta}$ : estimador de MV de  $\theta$  do modelo postulado; e  $\hat{\theta}_\omega$ : estimador de MV de  $\theta$  do modelo perturbado. COOK (1986) propôs estudar o comportamento local de  $LD(\omega)$  em torno de  $\omega_0$ , utilizando a curva normal  $C_i$  de  $LD(\omega)$  em  $\omega_0$  na direção de algum vetor unitário  $l$ , definido  $C_i = 2||\Delta^T L^{-1} \Delta||$ , com  $||l|| = 1$ , em que,  $L$ : é a matriz de informação observada, avaliada em  $\theta = \hat{\theta}$ ;  $\Delta$ : é uma matriz de ordem  $(pq) \times n$  dada por  $\Delta = (\Delta\beta^T, \Delta\varphi^T)^T$ , avaliada em  $\theta = \hat{\theta}$  e em  $\omega = \omega_0$ , em que,

$$\Delta_\beta = X^T \Sigma^{-1} \text{ e } \Delta_\varphi = \frac{\partial^2 l(\theta / \omega)}{\partial \varphi \partial \omega^T}, \text{ com } \frac{\partial^2 l(\theta / \omega)}{\partial \varphi \partial \omega^T} = (Z_\omega - X\beta)^T \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_i} \Sigma^{-1},$$

$$L = \begin{pmatrix} L_{\beta\beta} & L_{\beta\varphi} \\ L_{\varphi\beta} & L_{\varphi\varphi} \end{pmatrix}, \text{ em que, } L_{\beta\beta} = -(X^T \Sigma^{-1} X); \quad L_{\beta\varphi} = \frac{\partial^2 l(\theta)}{\partial \beta \partial \varphi^T}, \text{ com}$$

$$\frac{\partial^2 l(\theta)}{\partial \beta \partial \varphi_j} = -X^T \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} \Sigma^{-1} \varepsilon, \quad j = 1, \dots, q; \quad L_{\varphi\beta} = L_{\beta\varphi}^T; \text{ e } L_{\varphi\varphi} = \frac{\partial^2 l(\theta)}{\partial \varphi \partial \varphi^T}, \text{ com}$$

$$\frac{\partial^2 l(\theta)}{\partial \varphi_i \partial \varphi_j} = \frac{1}{2} \text{tr} \left\{ \Sigma^{-1} \left( \frac{\partial \Sigma}{\partial \varphi_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} - \frac{\partial^2 \Sigma}{\partial \varphi_i \partial \varphi_j} \right) \right\} + \frac{1}{2} \varepsilon^T \Sigma^{-1} \left\{ \frac{\partial^2 \Sigma}{\partial \varphi_i \partial \varphi_j} - \frac{\partial \Sigma}{\partial \varphi_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} - \frac{\partial \Sigma}{\partial \varphi_j} \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_i} \right\} \Sigma^{-1} \varepsilon$$

Seja  $L_{max}$  o auto-vetor, normalizado, associado ao maior autovalor, em módulo, da matriz  $B = \Delta^T L^{-1} \Delta$ . O gráfico dos elementos  $|L_{max}|$  versus  $i$  (ordem dos dados) pode revelar qual o tipo de perturbação que tem a maior influência em  $LD(\omega)$ , na vizinhança de  $\omega_0$  (COOK, 1986; BORSSOI et al., 2009).

## 2.2.10 Acurácia em mapas temáticos

A elevada quantidade de recursos existentes para elaboração de informações acaba levando o pesquisador a confeccionar vários mapas temáticos, sendo necessário utilizar técnicas que possam compará-los, como a quantificação das áreas e a determinação da concordância entre seus *pixels*<sup>1</sup>, informações estas que são geradas pela matriz dos erros dos mapas. A Tabela 1 apresenta a forma genérica de uma matriz dos erros. Nesta matriz, os *pixels* do mapa real ou de referência são quantificados nas colunas enquanto que os *pixels* do mapa modelo são quantificados nas linhas. Cada elemento da matriz representa a quantidade de *pixels* pertencentes à classe  $i$  do mapa modelo e à classe  $j$  do mapa de referência.

<sup>1</sup> Aglutinação das palavras *Picture* e *Element*, e representa o menor elemento num dispositivo de exibição ao qual é possível atribuir-se um valor. Neste trabalho, os *pixels* serão definidos com os elementos da matriz dos valores krigados.

**Tabela 1:** Matriz dos erros genérica de ordem  $k \times k$ 

		Mapa de referência			Total por linha
		$C_1$	$C_2$	... $C_k$	$n_i$
Mapa modelo	$C_1$	$n_{11}$	$n_{12}$	... $n_{1k}$	$n_{1\cdot}$
	$C_2$	$n_{21}$	$n_{22}$	... $n_{2k}$	$n_{2\cdot}$
	·	·	·	·	·
	·	·	·	·	·
	$C_k$	$n_{k1}$	$n_{kk}$	... $n_{kk}$	$n_{k\cdot}$
Total por coluna		$n_{\cdot 1}$	$n_{\cdot 2}$	... $n_{\cdot k}$	$n$
		$n_j$			

$k$ : número de classes;  $n_{ij}$ : número de *pixels* classificados na classe  $i$  do mapa modelo e na classe  $j$  do mapa de referência.

A diagonal principal (quando  $i = j$ ) representa casos em que os *pixels* tiveram a mesma classificação nos dois mapas enquanto que os elementos fora da diagonal principal representam as classificações errôneas. Observa-se que se os dois mapas forem iguais, todos os elementos fora da diagonal principal serão nulos.

Uma métrica utilizada para avaliar a acurácia dos mapas construídos é a exatidão global (*EG*) do mapa modelo, que tem o objetivo de determinar o percentual de acerto da classificação realizada e pode ser extraída da matriz dos erros pela Equação (29).

$$EG = \frac{\sum_{i=1}^k n_{ii}}{n} \quad (29)$$

A exatidão global é uma estatística de avaliação utilizada para mensurar a similaridade entre algo real ou controlado e um modelo ajustado, porém, é conveniente apresentar esta medida associada a outras métricas que levem em consideração os demais valores contidos na matriz dos erros e não somente os valores da diagonal principal. Os índices de acurácia de usuário (*AU*) e acurácia de produtor (*AP*) são maneiras de representar a precisão de uma classe ou categoria individualmente. O índice de acurácia de usuário ( $AU_i$ ) representa o quociente entre o número de *pixels* classificados corretamente na classe  $i$  sobre o total de *pixels* classificados na classe  $i$  do mapa modelo. Este índice pode ser calculado com os elementos da matriz dos erros e é apresentado na Equação (30):

$$AU_i = \frac{n_{ii}}{n_{i\cdot}}; \quad i = 1, \dots, k \quad (30)$$

O índice de acurácia de produtor ( $AP_i$ ) é uma estatística que fornece a probabilidade de um pixel ser classificado na classe  $i$  se ele realmente pertence a classe  $i$ . Este índice é obtido pela Equação (31):

$$AP_i = \frac{n_{ii}}{n_{\cdot i}}; \quad i = 1, \dots, k \quad (31)$$

Como GONG & HOWARTH (1990) relatam, o índice Kappa (COHEN, 1960) é uma estatística utilizada para mensurar a exatidão das classificações temáticas. Este índice vem sendo recomendado como uma medida apropriada da exatidão por utilizar todos os elementos da matriz dos erros ao invés de apenas utilizar aqueles que se situam na diagonal principal da mesma, o que ocorre quando se calcula a exatidão global da classificação. O índice Kappa fornece uma medida de concordância entre os valores do mapa modelo e os valores do mapa de referência e segundo CONGALTON & GREEN (1999) pode ser calculado pela Equação (32):

$$\hat{K} = \frac{n \sum_{i=1}^k n_{ii} - \sum_{i=1}^k n_{i\cdot} n_{\cdot i}}{n^2 - \sum_{i=1}^k n_{i\cdot} n_{\cdot i}} \quad (32)$$

A variância do índice Kappa, utilizada na construção de intervalos de confiança e testes de hipóteses, pode ser calculada conforme a equação (33):

$$\sigma^2(\hat{K}) = \frac{1}{n} \left( \frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2 - \theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4 - 4\theta_2^2)}{(1-\theta_2)^4} \right) \quad (33)$$

em que:

$$\theta_1 = \frac{1}{n} \sum_{i=1}^k n_{ii}; \quad \theta_2 = \frac{1}{n^2} \sum_{i=1}^k n_{i\cdot} n_{\cdot i}; \quad \theta_3 = \frac{1}{n^2} \sum_{i=1}^k n_{ii} (n_{i\cdot} + n_{\cdot i}); \quad \theta_4 = \frac{1}{n^3} \sum_{i=1}^k \sum_{j=1}^k n_{ij} (n_{i\cdot} + n_{\cdot j})^2.$$

Utilizando o índice Kappa juntamente com sua variância, é possível construir um intervalo de confiança de  $(1 - \alpha)\%$  utilizando a Equação (34):

$$IC(K; (1 - \alpha)\%) = \hat{K} \pm z_{\alpha/2} \sqrt{\sigma^2(\hat{K})} \quad (34)$$

Além dos índices de exatidão global e Kappa, que são muito utilizados, MA & REDMOND (1995) apresentaram o índice Tau ( $T$ ), que fornece uma medida quantitativa relativamente precisa e intuitiva sobre a acurácia da classificação. O índice Tau é similar ao índice Kappa, sendo calculado conforme a equação (35), em que  $p_k$  indica as probabilidades *a priori* para cada classe.

$$T = \frac{\sum_{i=1}^k n_{ii}}{n} - p_k \quad (35)$$

Quando as probabilidades *a priori* para as classes forem iguais, temos  $p_i = 1/k$ , onde  $k$  representa o número de classes da matriz dos erros.

Uma outra forma de comparar mapas temáticos pode ser utilizada mediante a criação da matriz de confusão por classe (FIELDING & BELL, 1997), matriz esta que pode ser obtida com os elementos da matriz dos erros. A Tabela 2 apresenta a forma genérica da matriz de confusão da  $k$ -ésima classe.

**Tabela 2:** Matriz de confusão da  $k$ -ésima classe

		Mapa de referência	
		Verdadeiro (V)	Falso (F)
Mapa modelo	Verdadeiro (V)	$a_k$	$b_k$
	Falso (F)	$c_k$	$d_k$

Em que:

$a_k = n_{kk}$  : quantidade de *pixels* da classe  $k$  do mapa real que foram classificados corretamente como pertencentes a classe  $k$  no mapa modelo.

$b_k = n_{\bullet k} - n_{kk}$  : quantidade de *pixels* que não pertencem a classe  $k$  do mapa real e foram classificados como pertencentes a classe  $k$  no mapa modelo.

$c_k = n_{\bullet k} - n_{kk}$  : quantidade de *pixels* que pertencem a classe  $k$  do mapa real e que pertencem a uma classe diferente no mapa modelo.

$d_k = n - (a_k + b_k + c_k)$  : quantidade de *pixels* que não pertencem a classe  $k$  no mapa real e foram classificados como não pertencente a classe  $k$  no mapa modelo.

A Tabela 3 apresenta algumas métricas, derivadas da matriz de confusão por classe, e utilizadas na comparação dos mapas temáticos. O índice de sensibilidade ( $S_k$ ) é uma medida que indica a probabilidade de um *pixel* no mapa modelo ser classificado como pertencente à classe  $k$  se realmente ele pertence à classe  $k$  no mapa real, sendo assim uma medida equivalente à acurácia de produtor (FIELDING & BELL 1997).

**Tabela 3:** Métricas derivadas da matriz de confusão por classe

Métrica	Equação
Índice de sensibilidade ( $S_k$ )	$a_k/(a_k+c_k)$
Índice de especificidade ( $E_k$ )	$d_k/(b_k+d_k)$
Taxa de falso positivo ( $TFP_k$ ) (Erro de comissão)	$b_k/(b_k+d_k)$
Taxa de falso negativo ( $TFN_k$ ) (Erro de omissão)	$c_k/(a_k+c_k)$

O índice de especificidade ( $E_k$ ) indica a probabilidade de um *pixel* não pertencente à classe  $k$  do mapa real ser classificado como não pertencente à classe  $k$  no mapa modelo (LURZ et al. 2001). Os erros de comissão também são conhecidos como erros de superestimativa, ou seja, indicam a proporção de *pixels* que não pertencem à classe  $k$  no mapa real que são classificados como pertencentes à classe  $k$  no mapa modelo. Os erros de omissão são considerados graves, pois indicam a proporção de *pixels* que pertencem à classe  $k$  do mapa real e foram classificados em outras classes no mapa modelo. Com base nessas métricas, é possível comparar individualmente as classes do modelo com as classes do mapa real. Para uma comparação global dos mapas, JENNESS & WYNNE (2005) apresentam a matriz de confusão total, indicada na Tabela 4.

**Tabela 4:** Matriz de confusão total

		Mapa de referência	
		Verdadeiro (V)	Falso (F)
Mapa Modelo	Verdadeiro (V)	$a = \sum_{i=1}^k n_{ii}$	$b = \sum_{i=1}^k \sum_{j \neq i}^k n_{ij}$
	Falso (F)	$c = \sum_{j=1}^k \sum_{i \neq j}^k n_{ij}$	$d = \sum_{i=1}^k \sum_{i \neq j}^k \sum_{j \neq i}^k n_{ij}$

A Tabela 5 apresenta algumas métricas derivadas da matriz de confusão total. Os índices de sensibilidade total ( $S$ ) e especificidade total ( $E$ ) são análogos às suas versões por classe, com a diferença de que agora apresentam medidas para o mapa completo. O coeficiente de correlação de Matthews ( $CCM$ ) é uma versão discreta do coeficiente de correlação de Pearson, e seus valores se encontram no intervalo  $[-1;1]$ , sendo que o valor 1 equivale a uma predição perfeita, 0 equivale a uma predição aleatória e -1 significa uma predição inversa.

**Tabela 5 –** Métricas derivadas da matriz de confusão total

Métrica	Equação
$S$	$a/(a+c)$
$E$	$d/(b+d)$
$CCM$	$((a \cdot d) - (b \cdot c)) / ((a+b) \cdot (a+c) \cdot (d+b) \cdot (d+c))^{(1/2)}$

### 2.2.11 Modelagem cartográfica

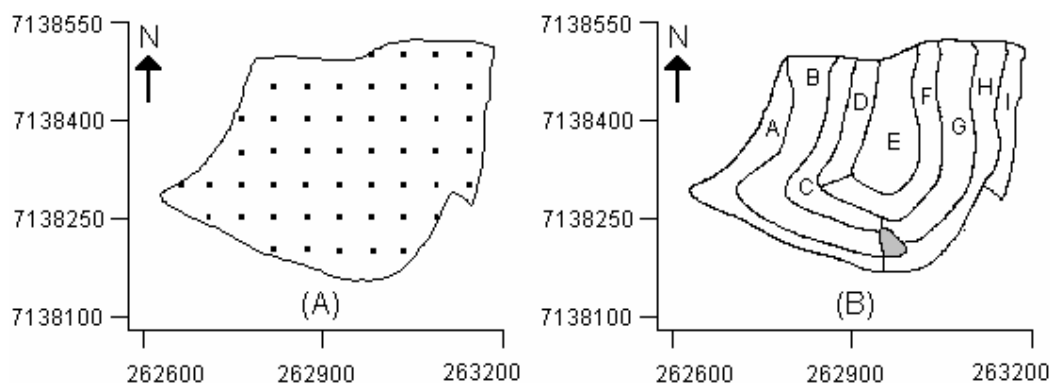
A modelagem cartográfica é um tipo de técnica de análise espacial, que consiste numa coleção de mapas sobre determinados atributos, onde cada mapa é uma variável sujeita a operações matemáticas tradicionais. A modelagem cartográfica apresenta-se hoje como uma importante técnica de análise espacial geográfica. Ela aparece, como definição, num contexto de utilização de tecnologias de Geoprocessamento, no início da década de 1990, com TOMLIN (1990), em que ele introduz um conceito novo - álgebra de mapas -, atualmente muito explorado por pesquisadores e desenvolvido em alguns softwares de Geoprocessamento.

Segundo BERRY (1993), uma linguagem de modelagem cartográfica usa uma sequência de funções primitivas para realizar uma análise complexa de mapas. Nesse sentido, ela é semelhante à álgebra tradicional, na qual operadores primitivos (adição, subtração, exponenciação) são logicamente sequenciados com variáveis para que se forme uma equação. No entanto, na álgebra cartográfica, mapas inteiros representam as variáveis. A unidade básica de processamento é o *pixel*, que pode ser processado independentemente, integrado numa vizinhança ou integrado em região de elementos com o mesmo atributo.

## 2.3 MATERIAL E MÉTODOS

### 2.3.1 Material

A análise de um conjunto de dados reais foi efetuada com o propósito de ilustrar a aplicação das técnicas geoestatísticas. A área experimental compreende uma área de produção de grãos com 13,7 hectares de extensão, localizada no município de Salto do Lontra – PR, que foi cultivado com trigo variedade IAPAR 78. Na área monitorada, foi definida uma grade amostral regular de 50 x 50 m totalizando 50 posições georreferenciadas (Figura 2 (A)). Na Figura 2 (B) observa-se que esta área também foi dividida em nove talhões denominados de “A” a “I”, para facilitar a colheita mecânica e com isso poder determinar a produção real. A pequena área em cinza na Figura 2 (B) representa uma região de pedras que não é monitorada no experimento.



**Figura 2** – (A) Localização dos 50 pontos amostrais e (B) Divisão da área em talhões.

Nos pontos de amostragem, foram coletadas as parcelas de trigo. Estas amostras foram trilhadas de forma manual, pesadas e posteriormente transformadas em  $t\ ha^{-1}$ .

A análise geoestatística foi realizada utilizando os pacotes geoR (RIBEIRO JR & DIGGLE, 2001) e SplanCs (ROWLINGSON & DIGGLE, 1993) do software R.

### 2.3.2 Metodologia

Uma análise geoestatística é composta de diversos procedimentos que devem ser utilizados com bastante cautela, pois não se tratam de uma simples aplicação de fórmulas para obter resultados. Para uma correta investigação o pesquisador deve sempre estar refazendo e comparando suas análises, e isto faz com que o roteiro a ser seguido, bem como os resultados intermediários e finais não sejam obtidos de forma única.

A seguir são descritos os passos da análise geoestatística. Inicia-se com uma análise exploratória que vem seguida de diversos procedimentos até a confecção de mapas temáticos que detalhem o comportamento do atributo estudado.

#### a) Análise descritiva geral

Em uma análise descritiva geral, a única informação investigada é o conjunto amostral do atributo estudado - assim, a posição do atributo não é considerada. O objetivo fundamental é a identificação inicial do comportamento dos dados, sem nenhuma pretensão de inferência. O uso de medidas de posição, tendência central, variação, assimetria e curtose, associadas a informações gráficas como o histograma, *boxplot* e ramo-e-folhas permite uma primeira visão dos aspectos gerais do conjunto de dados.

#### b) Análise descritiva espacial

Embora seja mais detalhada que a análise descritiva geral, uma vez que leva em consideração a posição da amostra, a intenção é tão somente ter um indicativo do comportamento dos dados agora associando-os às suas posições na região amostrada.

Um gráfico que é rico em informação é o *postplot*, que consiste em apresentar a malha de pontos coletados separados por cores/símbolos identificando os quartis da distribuição dos dados. A partir deste gráfico é possível identificar se existem regiões com concentração de altos ou baixos valores amostrais. Uma outra opção de análise é um gráfico tridimensional que mostra a superfície definida pela amostra. A análise é idêntica à feita com o gráfico *postplot* e dá uma idéia visual inicial do comportamento do atributo na região estudada.

#### c) Análise variográfica

Para que a interpolação por krigagem possa ser realizada com eficácia, são necessárias estimativas de semivariâncias (covariâncias) para distâncias além das fornecidas pela malha amostral. Um modelo que descreva a semivariância (covariância) em função da distância é então ajustado a partir de um conjunto discreto de pontos, estimativas da semivariância (covariância) para os diferentes “lags”. No presente trabalho foram avaliados os modelos exponencial, esférico e gaussiano utilizando os métodos de estimação de mínimos quadrados ordinários (OLS), mínimos quadrados ponderados, máxima verossimilhança (ML) e máxima verossimilhança restrita (RML).



#### d) Validação cruzada

A idéia central avaliada pela validação cruzada é que se o fenômeno foi satisfatoriamente modelado, é possível reproduzir, com boa aproximação, informações da amostra. Desta forma, o procedimento consiste em retirar um ponto do conjunto original das amostras coletadas e estimá-lo utilizando o restante das amostras. Após realizar a estimativa de todos os pontos, é possível obter medidas que ajudem a escolher o modelo que melhor descreva a variabilidade espacial do atributo estudado.

#### e) Krigagem ordinária

Na maioria das pesquisas relacionadas à geoestatística, o profissional está interessado em estimar valores em locais não amostrados, seja por um interesse local ou pela intenção de obter um detalhamento da área que vai além do permitido pela amostra. Nestes casos, o procedimento adotado é a interpolação.

A literatura apresenta uma vasta coleção de interpoladores, desde os conhecidos modelos deterministas, como polígonos de influência, triangulação e inverso do quadrado das distâncias, até a proposta de predição geoestatística, conhecida como krigagem. Neste trabalho será utilizado a krigagem ordinária, interpolador mais eficaz que os deterministas por incorporar a estrutura de dependência espacial do atributo estudado.

#### f) Análise de diagnóstico

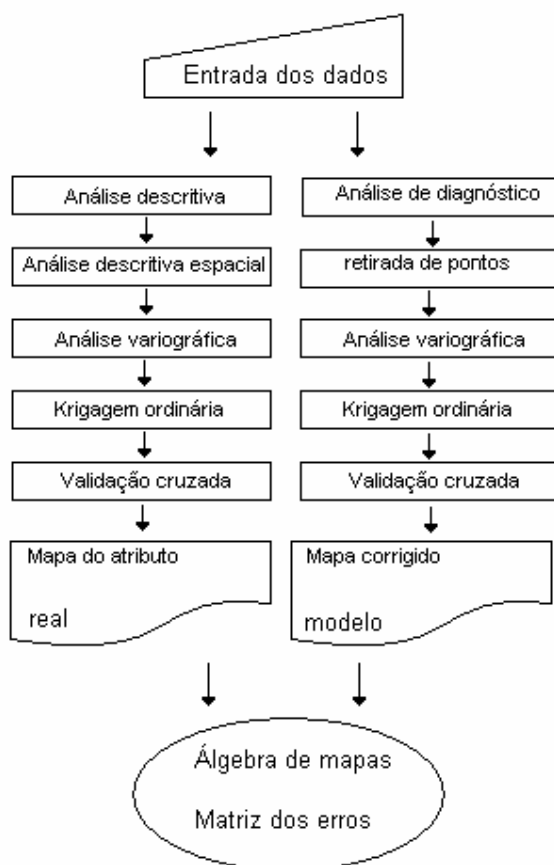
Com o objetivo de investigar a existência de pontos influentes no conjunto amostral do atributo estudado, será analisado o gráfico de influência local  $|L_{\max}|_{x_i}$  (ordem dos dados), que propõe avaliar a influência conjunta das observações sob pequenas perturbações no modelo, identificando assim, pontos que possam causar alguma interferência ou provoquem grandes desvios nos resultados do ajuste.

#### g) Comparação de mapas

Para comparar mapas obtidos por diferentes técnicas, serão utilizadas as métricas derivadas da matriz dos erros e também será criando um mapa de diferenças utilizando a álgebra de mapas.

A Figura 3 apresenta o roteiro desta análise. Observa-se que dois mapas serão gerados. O primeiro obtido pelo ajuste de um modelo teórico, selecionado por validação cruzada, utilizando todos os pontos do conjunto de dados amostrais, e o segundo, utilizando

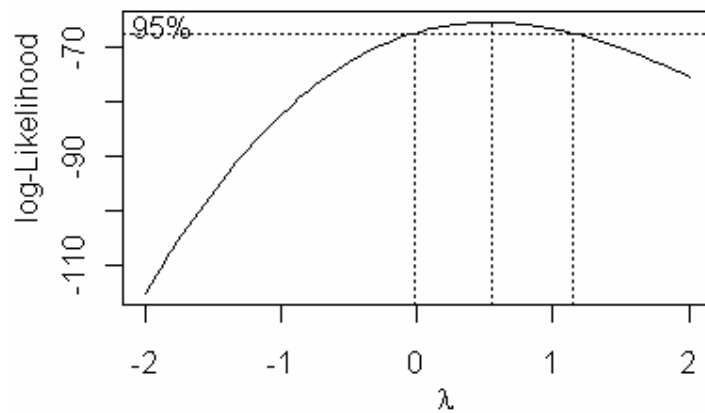
a análise de diagnósticos para retirar pontos detectados como influentes. Para finalizar o estudo, os mapas serão comparados utilizando a subtração em módulo de um mapa pelo outro, que evidencia regiões que foram diferentes em ambos os mapas e também será construída a matriz dos erros, considerando o mapa com todos os pontos como o de referência e o mapa sem os pontos influentes como modelo. Com esta matriz, será possível calcular as métricas de acurácia.



**Figura 3** – Roteiro da análise geoestatística.

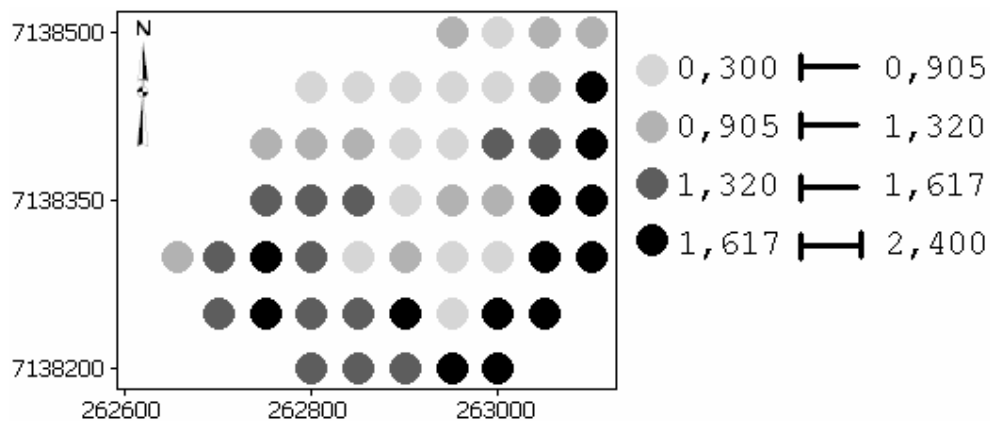
## 2.4 RESULTADOS E DISCUSSÃO

A análise exploratória dos elementos amostrados apresentou uma produtividade média de trigo de  $1,31 \text{ t ha}^{-1}$ , sendo a produção mínima de  $0,30 \text{ t ha}^{-1}$  e máxima de  $2,40 \text{ t ha}^{-1}$  com desvio padrão de  $0,54 \text{ t ha}^{-1}$  e coeficiente de variação (CV) de 41,40%. Apesar desta alta heterogeneidade entre os valores amostrais ( $CV > 30\%$ ), pode-se considerar que o conjunto apresenta uma distribuição normal de probabilidades, pois pelo teste de normalidade de Anderson-Darling obteve-se um nível descritivo (p-valor) de 0,102 o que indica normalidade dos dados com nível de 5% de significância. Analisando o gráfico *boxcox* da produtividade de trigo na Figura 4, observa-se que não há necessidade de transformar os dados, visto que o  $\lambda = 1$  encontra-se no intervalo de 95 % de confiança.



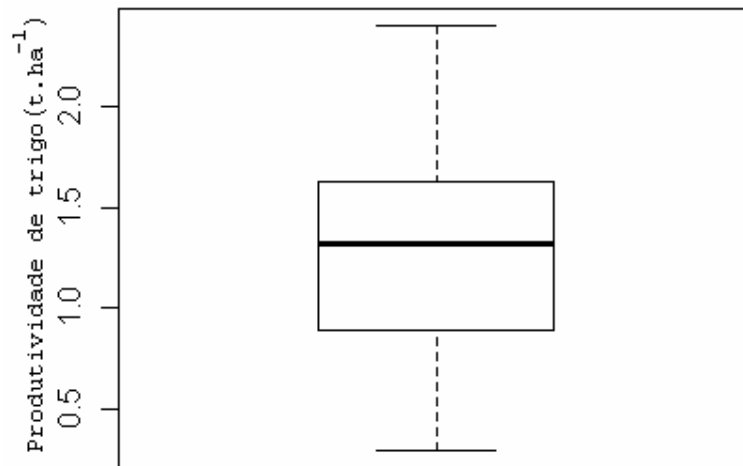
**Figura 4** – Gráfico *boxcox* da produtividade de trigo.

A Figura 5 apresenta o gráfico espacial da área em estudo, representando a produtividade de trigo classificada por quartis.



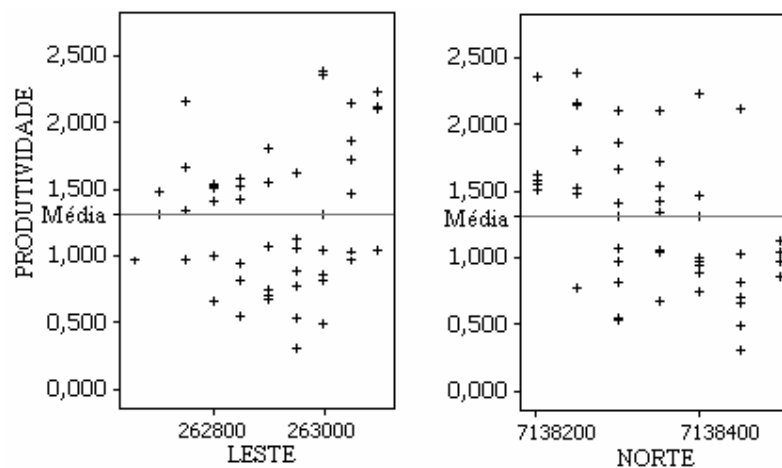
**Figura 5** - Gráfico espacial da produtividade de trigo.

Destaca-se nesse gráfico a região norte da área por apresentar um aglomerado com produtividade baixa e a região leste por apresentar um aglomerado com maiores valores de produtividade. A Figura 6 apresenta o gráfico *boxplot* da produtividade de soja.



**Figura 6** – Gráfico *boxplot* da produtividade de trigo.

Observando o gráfico da Figura 6 observa-se que o conjunto de dados é simétrico e também destaca-se a ausência de pontos discrepantes no conjunto de dados amostrais. Com finalidade de verificar a existência de tendências, heterogeneidades de variâncias e comportamentos isolados foram construídos os gráficos da Figura 7. Para cada linha e coluna foram assinalados os valores encontrados.

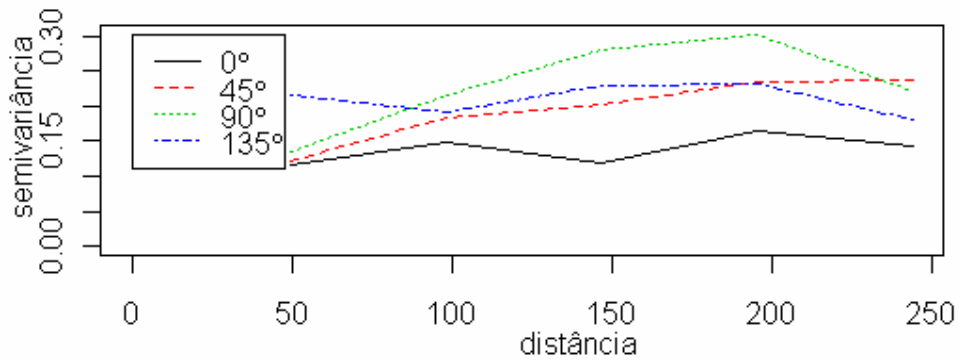


**Figura 7** – Gráfico das coordenada *versus* produtividade de trigo.

Nos dois gráficos existem pequenas flutuações dos valores em relação à média. Porém, isto não caracteriza problemas visto que em ambas as direções as dispersões são semelhantes. As análises exploratórias indicam que o comportamento do conjunto de dados

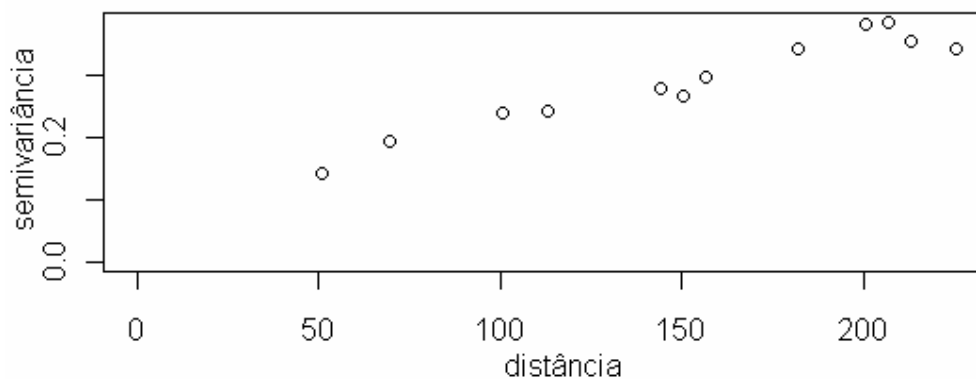
de produtividade de trigo não mostra problemas que afrontem as hipóteses de estacionaridade.

Para realizar o estudo variográfico, é necessário conhecer a distância máxima entre os pontos amostrais, com intuito de definir o ponto de corte. Para construir o semivariograma utilizando o conjunto amostral a distância máxima entre duas amostras é de 487,8781 m e como vamos padronizar um corte de 50% (CLARK, 1979), a distância a ser considerada será de 244 m. Para identificar a existência de anisotropia, construiu-se os semivariogramas direcionais da Figura 8.



**Figura 8** – Semivariogramas direcionais da produtividade de trigo.

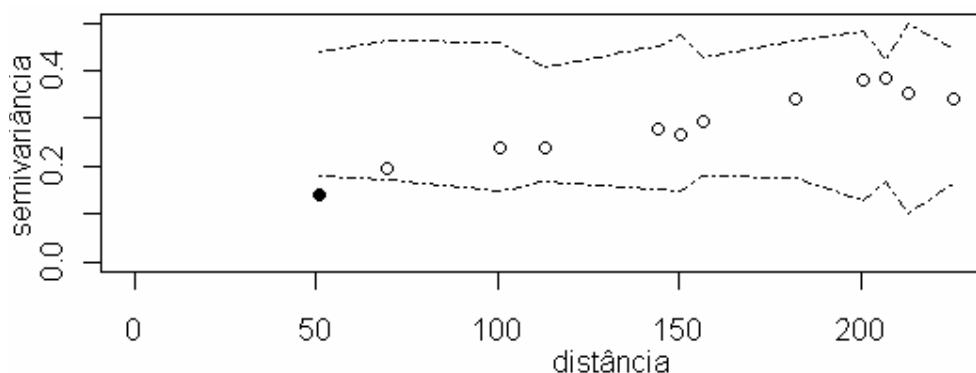
Observa-se na Figura 8 que os semivariogramas direcionais apresentam um comportamento semelhante para as direções monitoradas, indicando que os dados são isotrópicos e assim, podemos considerar um único semivariograma, conhecido como semivariograma omnidirecional, apresentado na Figura 9.



**Figura 9** – Semivariograma experimental da produtividade de trigo.

Observa-se que a primeira semivariância encontra-se na distância de 50 m, indicando que não se tem conhecimento do fenômeno a menores distâncias. Este semivariograma foi gerado utilizando o estimador de Matheron e um ângulo de tolerância de 22,5°. Uma forma de investigar a dependência espacial é proporcionada pela construção de

envelopes simulados, apresentados na Figura 10. Na análise dos envelopes simulados deve haver ao menos um ponto do semivariograma fora do envelope. Se isso ocorrer rejeita-se a hipótese nula, de que não há dependência espacial.



**Figura 10** – Envelopes simulados.

Nesta simulação encontrou-se um ponto fora dos envelopes, conforme é destacado na Figura 9. Este fato sugere a existência de correlação espacial da produtividade de trigo na área monitorada. Destaca-se que foi estipulado um mínimo de 30 pares de pontos para cada semivariância plotada nos gráficos das Figuras 8,9 e 10. A Tabela 6 apresenta informações mais detalhadas do semivariograma experimental da produtividade de trigo, como a distância em que foi computada a semivariância e o número de pares utilizados.

**Tabela 6** – Parâmetros do semivariograma omnidirecional

$h$	pares	$\gamma(h)$
50,846	82	0,141
69,538	75	0,194
100,692	66	0,240
113,154	123	0,241
144,308	53	0,277
150,538	53	0,265
156,769	97	0,295
181,692	82	0,342
200,385	38	0,382
206,615	71	0,383
212,846	32	0,354
225,308	60	0,342

Observa-se na Tabela 6 que a maior quantidade de pares ocorreu a uma distância de 113,15 m, com 123 pares. A menor quantidade de pares ocorreu a uma distância de 212,84 m, com 32 pares - porém, superior ao limite definido *a priori* de 30 pares de pontos.

A Tabela 7 apresenta as estimativas dos parâmetros referentes aos ajustes dos modelos teóricos de semivariograma, juntamente com os respectivos métodos de

estimação. Observa-se que o maior raio de dependência espacial é obtido com o modelo exponencial, com aproximadamente 389 m nos métodos WLS1, ML e RML.

**Tabela 7** - Modelos ajustados e parâmetros obtidos do conjunto amostral completo da produtividade de trigo [ $t\ ha^{-1}$ ]

Modelo	Método	$a$	$\varphi_1$	$\varphi_2$	$\varphi_1 + \varphi_2$	$E(\%)$
Exponencial	OLS	359,4886	0,0000	0,4229	0,4229	0,00
	WLS1	389,4457	0,0000	0,4373	0,4373	0,00
	ML	389,4450	0,0166	0,3518	0,3684	4,51
	RML	389,4452	0,0086	0,3848	0,3934	2,19
Esférico	OLS	310,1955	0,0672	0,3392	0,4064	16,54
	WLS1	311,3712	0,0669	0,3405	0,4074	16,42
	ML	137,6487	0,0117	0,2611	0,2728	4,29
	RML	141,7012	0,0105	0,2759	0,2864	3,67
Gaussiano	OLS	276,7686	0,1300	0,2797	0,4097	31,73
	WLS1	273,0037	0,1284	0,2787	0,4071	31,54
	ML	269,1884	0,1051	0,2961	0,4012	26,20
	RML	302,6388	0,1056	0,4164	0,522	20,23

OLS: mínimos quadrados ordinários, WLS1: mínimos quadrados ponderados, ML: máxima verossimilhança, RML: máxima verossimilhança restrita,  $a$ : ancance,  $\varphi_1$ : efeito pepita,  $\varphi_2$ : contribuição,  $\varphi_1 + \varphi_2$ : patamar e  $E$ : efeito pepita relativo 100. ( $\varphi_1 / (\varphi_1 + \varphi_2)$ )

O efeito pepita relativo ( $E$ ) na Tabela 7 é utilizado para mensurar o grau de dependência espacial segundo a classificação proposta por CAMBARDELLA et al. (2004). Observa-se que os ajustes utilizando modelos exponenciais apresentam os menores valores para o efeito pepita relativo, indicando uma forte dependência espacial. Dos quatro ajustes realizados com o modelo gaussiano, 3 deles apresentam coeficiente de efeito pepita relativo superior a 25 %, indicando uma moderada dependência espacial.

Como tem-se vários modelos e métodos, é necessário utilizar técnicas que permitam escolher o melhor ajuste, como por exemplo, a validação cruzada, apresentada na Tabela 8.

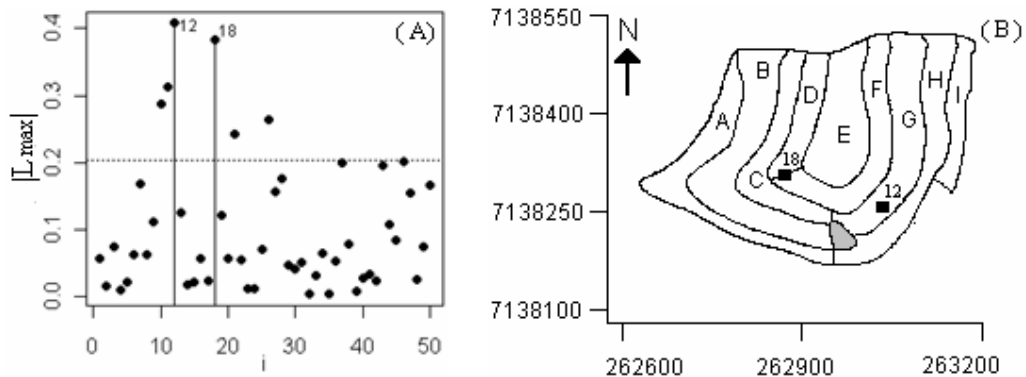
**Tabela 8** - Validação cruzada para a produtividade de trigo [ $t\ ha^{-1}$ ] relativa a todo o conjunto amostral

Modelo	Método	$EM$	$\overline{ER}$	$DP_{EM}$	$S_{ER}$	$EA$
Exponencial	OLS	-0,002074	-0,002220	0,3564242	0,9521044	12,78630
	WLS1	-0,002310	-0,0025460	0,3563375	0,9716647	12,78598
	ML	-0,001719	-0,001899	0,3581258	0,9950070	12,84087
	RML	-0,002006	-0,002238	0,3571692	0,9924911	12,79904
Esférico	OLS	-0,002417	-0,002660	0,3674936	0,9379480	13,43437
	WLS1	-0,002439	-0,002689	0,3675304	0,9391278	13,43475
	ML	0,0005531	0,0010620	0,3620052	0,9890566	13,10038
	RML	0,0002486	0,0006794	0,3630399	0,9865988	13,20557
Gaussiano	OLS	-0,001874	-0,001957	0,3687667	0,9190851	13,54737
	WLS1	-0,001864	-0,001948	0,3683358	0,9220606	13,52683
	ML	-0,002367	-0,002709	0,3652186	1,0011898	13,41738
	RML	-0,003308	-0,003882	0,3645614	1,0007248	13,42031

$EM$ : erro médio,  $\overline{ER}$ : erro médio reduzido,  $DP_{EM}$ : desvio padrão do erro médio,  $S_{ER}$ : desvio padrão dos erros reduzidos e  $EA$ : erro absoluto.

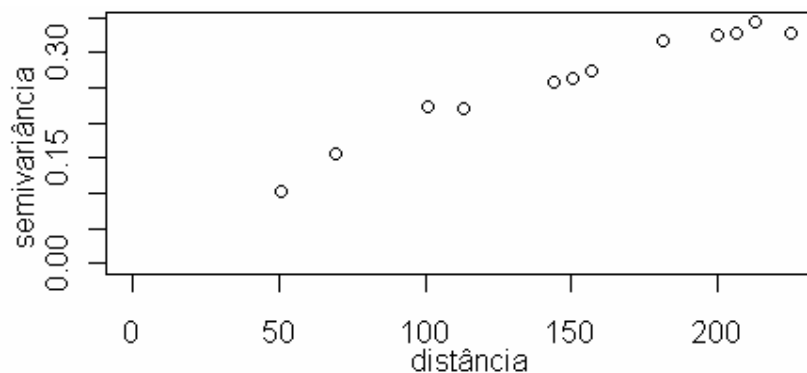
Segundo o critério de validação cruzada o melhor modelo é o que fornece os valores de EM e  $\overline{ER}$  mais próximos de zero, o valor  $DP_{EM}$  menor e o valor de  $S_{ER}$  mais próximo de um. Optou-se por escolher o modelo exponencial - ML como o que melhor se ajustou, visto que das cinco estatísticas investigadas ele apresenta um melhor desempenho em três.

A Figura 11 (A) apresenta o gráfico de diagnóstico dos auto-vetores  $|L_{max}|$ , utilizando o modelo espacial exponencial – ML.



**Figura 11** - (A) Gráfico de diagnóstico  $|L_{max}|$ ; (B) Localização dos pontos influentes na área monitorada.

Verifica-se que as observações 12 e 18 foram consideradas influentes e podem mudar algum tipo de decisão na construção de modelos geoestatísticos e/ou na construção dos mapas temáticos. A Figura 11 (B) apresenta a localização dos pontos influentes na área monitorada. A amostra n° 12 refere-se a uma produtividade de  $2,40 \text{ t ha}^{-1}$  e a amostra n° 18 representa um elemento amostral com produtividade de  $0,55 \text{ t ha}^{-1}$ . Desta forma, optou-se por fazer uma nova análise variográfica retirando os dois pontos considerados influentes e verificando de que maneira eles afetam a análise espacial. A Figura 12 apresenta o semivariograma experimental da produtividade de trigo gerado sem os dois pontos influentes. Destaca-se um comportamento semelhante quando comparado com o semivariograma que foi construído com todos os pontos amostrais na Figura 9.



**Figura 12** – Semivariograma experimental construído sem os pontos influentes.



A Tabela 9 apresenta os parâmetros ajustados utilizando os métodos de estimação de mínimos quadrados ordinários (OLS), mínimos quadrados ponderados (WLS1), máxima verossimilhança (ML) e máxima verossimilhança restrita (RML) sem os pontos influentes.

**Tabela 9** - Modelos ajustados e parâmetros do conjunto amostral da produtividade de trigo [ $t\ ha^{-1}$ ] sem as amostras influentes

Modelo	Método	$a$	$\varphi_1$	$\varphi_2$	$\varphi_1 + \varphi_2$	$E(\%)$
Exponencial	OLS	545,3916	0,0000	0,4836	0,4836	0,00
	WLS1	549,0667	0,0000	0,4847	0,4847	0,00
	ML	562,9345	0,0000	0,4178	0,4178	0,00
	RML	389,4452	0,0000	0,317	0,317	0,00
Esférico	OLS	246,9547	0,0144	0,3262	0,3406	4,23
	WLS1	233,5746	0,0000	0,3332	0,3332	0,00
	ML	140,2367	0,0000	0,2173	0,2173	0,00
	RML	142,0542	0,0000	0,2243	0,2243	0,00
Gaussiano	OLS	225,0066	0,0786	0,2700	0,3486	22,55
	WLS1	225,0065	0,0783	0,2695	0,3478	22,51
	ML	121,3996	0,0143	0,2477	0,2620	5,46
	RML	125,6102	0,0165	0,2644	0,2809	5,87

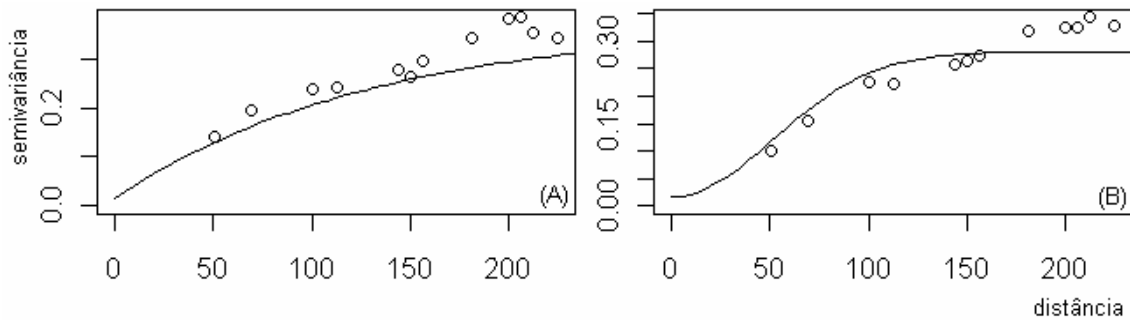
Observa-se que os ajustes realizados utilizando o modelo exponencial com os métodos OLS, WLS1 e ML apresentaram um raio de dependência espacial superior à distância máxima que é de 487, 8781 m e portanto, estes modelos serão desconsiderados da análise, pois apesar de estarem matematicamente corretos, eles não condizem com a realidade. A Tabela 10 apresenta as estatísticas da validação cruzada referentes ao conjunto de dados sem os pontos influentes. Novamente, procura-se por valores de  $EM$  e  $\overline{ER}$  mais próximos de zero, menores valores para o  $DP_{EM}$ , valores de  $S_{ER}$  mais próximo de um e o menor erro absoluto. Analisando a Tabela 10, optou-se por escolher o gaussiano – RML como o modelo que melhor se ajusta ao semivariograma experimental gerado com o conjunto de dados composto de 48 elementos amostrais.

**Tabela 10** - Validação cruzada para a produtividade de trigo [ $t\ ha^{-1}$ ] sem as amostras influentes

Modelo	Método	$EM$	$\overline{ER}$	$DP_{EM}$	$S_{ER}$	$EA$
Exponencial	RML	0,0009800	0,0017480	0,3053484	0,9413290	10,76085
	OLS	0,0010440	0,0018390	0,2974865	0,8996713	10,42330
Esférico	WLS1	0,0011600	0,0021600	0,2918369	0,9390484	10,26458
	ML	0,0037200	0,0057110	0,3017940	0,9260413	10,74068
	RML	0,0036180	0,0055120	0,3026852	0,9203176	10,77810
Gaussiano	OLS	0,0007296	0,0015100	0,3222009	0,9596583	11,35211
	WLS1	0,0007271	0,0015100	0,3221772	0,9614916	11,35204
	ML	0,0026980	0,0048850	0,2720037	1,0011951	9,813561
	RML	0,0019660	0,0036590	0,2723834	0,9993793	9,73965

EXP: exponencial, ESF: esférico, GAUSS: gaussiano,  $EM$ : erro médio,  $\overline{ER}$ : erro médio reduzido,  $DP_{EM}$ : desvio padrão do erro médio,  $S_{ER}$ : desvio padrão dos erros reduzidos e  $EA$ : erro absoluto.

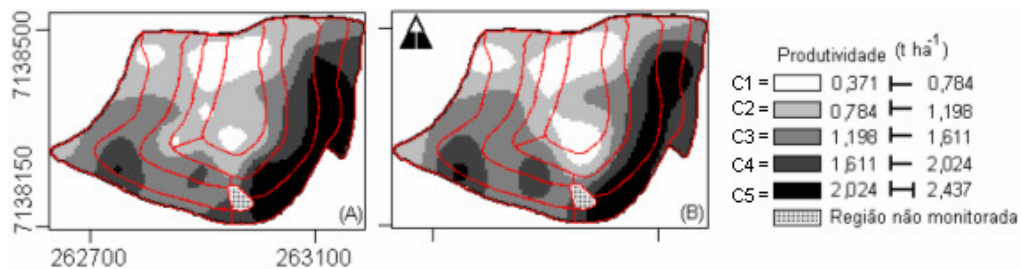
A Figura 13 (A) apresenta o semivariograma experimental do conjunto completo de dados juntamente com o modelo ajustado (exponencial – ML), e a Figura 13 (B) apresenta o semivariograma experimental gerado sem os pontos influentes juntamente com o modelo ajustado (gaussiano – RML). Conclui-se que os pontos influentes alteram a escolha do modelo e os valores dos parâmetros.



**Figura 13** – Semivariogramas ajustado: (A) completo e (B) sem os pontos influentes.

A Figura 14 (A) apresenta o mapa da produtividade de trigo gerado por krigagem ordinária utilizando os parâmetros ( $\phi_1 = 0,0166$ ;  $\phi_2 = 0,3518$ ;  $a = 389,4450$ ) fornecidos pelo modelo exponencial – ML, obtidos do estudo realizado com todo o conjunto amostral e a Figura 14 (B) apresenta o mapa gerado utilizando os parâmetros ( $\phi_1 = 0,0165$ ;  $\phi_2 = 0,2644$ ;  $a = 125,6102$ ) fornecidos pelo modelo gaussiano – RML, obtidos pelo estudo dos dados sem as amostras influentes. Para construção destes mapas, foi criado um *grid* de interpolação formado por 5423 *pixels*.

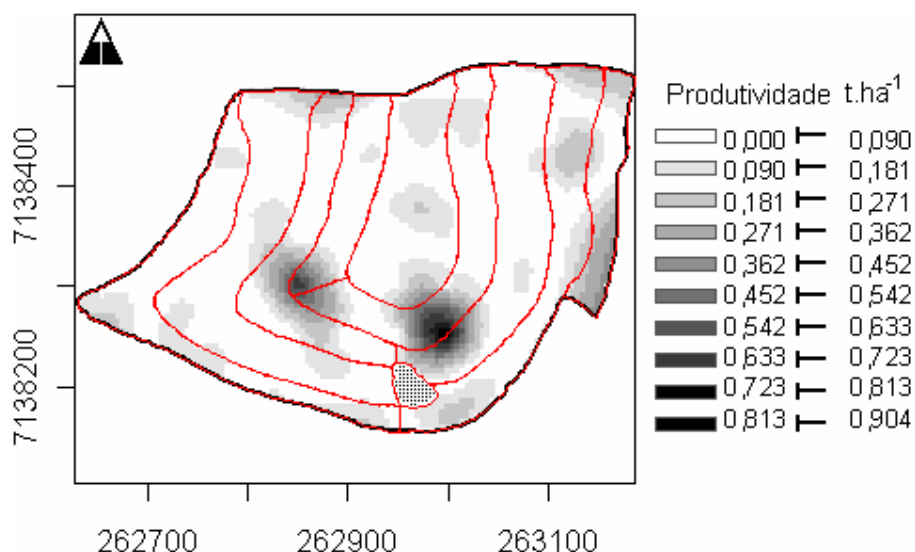
Observa-se que os mapas apresentam um comportamento semelhante. Porém, é possível identificar diferenças, não somente nos locais dos pontos influentes como também em outras regiões.



**Figura 14** - (A) Mapa temático da produtividade de trigo utilizando 50 elementos amostrais na interpolação e (B) Mapa temático da produtividade de trigo utilizando 48 elementos amostrais na interpolação.

Destaca-se que os mapas foram construídos utilizando 5 classes, sendo este número definido de tal forma que possibilite tanto uma identificação visual das áreas de

produtividade quanto uma praticidade em realizar aplicações localizadas de insumos, pois uma grande quantidade de intervalos implica em áreas muito pequenas em cada classe, o que dificulta a demarcação no campo. A diferença entre os mapas é melhor visualizada na Figura 15, que apresenta o mapa temático da diferença de produtividade em módulo entre os dois mapas.



**Figura 15** - Mapa da diferença de produtividade em módulo entre os dois mapas

Observa-se na Figura 15 que as regiões de maior diferença são exatamente as regiões dos pontos influentes. Porém, diversas outras regiões apresentam diferenças, o que mostra que a retirada dos pontos influentes afetou o mapa como um todo.

Para melhor comparar estes mapas, é conveniente quantificar seus 5423 *pixels* em uma matriz, conhecida como matriz dos erros, definindo o mapa gerado com todos os pontos como o de referência, visto que ele é elaborado por métodos geoestatísticos mais conhecidos e o mapa gerado sem os pontos influentes como o mapa modelo, devido à implantação dos estudos de diagnósticos em influência local. A Tabela 11 apresenta a matriz dos erros que compara os dois mapas.

**Tabela 11** - Matriz dos erros dos *pixels* dos mapas de produtividade de trigo  
Mapa Padrão

Mapa Modelo	C1	C2	C3	C4	C5	TOTAL
C1	485	272	9	0	0	766
C2	45	1202	124	18	0	1389
C3	7	206	1218	99	19	1549
C4	0	0	135	616	241	992
C5	0	0	0	135	592	727
TOTAL	537	1680	1486	868	852	5423

C1: [0,371; 0784), C2:[0,784;1,198), C3:[1,198; 1,611), C4:[1,611; 2,024) e C5:[2,024; 2,437).

Na matriz dos erros, os *pixels* do mapa padrão ou de referência são quantificados nas colunas e os *pixels* do mapa modelo são quantificados nas linhas. Assim, tem-se que dos 5423 *pixels*, 537 estão na primeira classe do mapa padrão e 766 estão na primeira classe do mapa modelo.

Esta primeira informação permite realizar uma estimativa da produção de trigo na área para ser comparada com a produção de trigo real, colhida e pesada pelo produtor, que foi de 18,49 toneladas. A área total monitorada é de 13,551 hectares e esta corresponde a 5423 *pixels*. Realizando uma regra de três, podemos verificar que os 537 *pixels* da primeira classe do mapa gerado com todos os pontos representam uma área de 1,3419 hectares. Multiplicando esta área pelo ponto médio da primeira classe (0,5778), obtemos a estimativa de 0,78 toneladas. Seguindo o mesmo procedimento para as outras classes, obtemos as estimativas da Tabela 12.

**Tabela 12** – Produção (t) estimada em cada classe

MAPA	C1	C2	C3	C4	C5	Total
Referência	0,78	4,16	5,21	3,94	4,75	18,84
Modelo	1,11	3,44	5,43	4,50	4,05	18,54

Observa-se na Tabela 12 que a produção total estimada pelo mapa padrão foi de 18,84 toneladas, 350 quilos superior à produção real, que foi de 18,49 toneladas e a produção total estimada utilizando o mapa modelo foi de 18,54 toneladas, 50 quilos superior à produção real. Diante desta observação, destaca-se que a retirada dos pontos influentes foi de grande importância na análise, pois permitiu estimar a produção com maior precisão.

A diagonal principal da matriz dos erros da Tabela 11 indica a quantidade de *pixels* classificados identicamente nos dois mapas. Estes elementos permitem calcular a exatidão global (EG), uma das primeiras medidas de acurácia que pode ser utilizada na comparação dos mapas. Da Tabela 11 tem-se que EG = 0,76 ou seja, a exatidão global foi de 76 %. Este valor também indica que os mapas temáticos apresentam diferenças, pois de acordo com ANDERSON et al. (1976), o nível mínimo de precisão aceitável é de 85 %. A acurácia pode ser investigada para cada classe individualmente por meio da análise dos índices de acurácia de usuário e de produtor, apresentados na Tabela 13.

**Tabela 13** – Índices de acurácia de usuário (AU) e produtor (AP)

	C1:[0,371;0784)	C2:[0,784;1,198)	C3:[1,198; 1,611)	C4:[1,611;2,024)	C5:[2,024;2,437)
AU <sub>i</sub>	0,6432	0,8654	0,7863	0,6210	0,8143
AP <sub>i</sub>	0,9032	0,7155	0,8197	0,7097	0,6948

Observa-se pela Tabela 8 que a classe C4 do mapa modelo apresentou um índice de acurácia de usuário de 0,6210 o que indica que 37,9 % dos *pixels* que pertencem a esta

classe no mapa modelo pertencem a outras classes no mapa de referência. Observando o índice de acurácia de produtor da classe C<sub>4</sub>, tem-se que a probabilidade de um *pixel* ser classificado nesta classe do mapa modelo se ele realmente pertence a esta classe no mapa de referência é de 0,7097. A Tabela 14 apresenta o índice Kappa com sua respectiva variância. Analisando a Tabela 14, observa-se que o valor de Kappa é 0,69, o que, segundo classificação de LADIS & KOCK (1997), indica uma semelhança muito boa.

**Tabela 14** - Estatísticas do Índice Kappa ( $\hat{K}$ )

$\hat{K}$	0,69
$\sigma^2(\hat{K})$	0,00084
IC(K, 95%)	[0,63; 0,75]

O índice Tau obtido da matriz dos erros considerando a igualdade das probabilidades *a priori* para cada classe foi 0,70, que apesar de ser superior ao índice Kappa, também indica uma semelhança muito boa entre os mapas temáticos da produtividade de trigo. A Tabela 15 apresenta os elementos das matrizes de confusão da classe C<sub>k</sub>, que são utilizados para calcular métricas adicionais, proporcionando uma investigação mais detalhada.

**Tabela 15** – Elementos (*pixels*) da matriz de confusão da classe C<sub>k</sub>

C <sub>k</sub>	a <sub>k</sub>	b <sub>k</sub>	c <sub>k</sub>	d <sub>k</sub>
C <sub>1</sub>	485	281	52	4605
C <sub>2</sub>	1202	187	478	3556
C <sub>3</sub>	1218	331	268	3606
C <sub>4</sub>	616	376	252	4179
C <sub>5</sub>	592	135	260	4436

Como os elementos a<sub>k</sub> e d<sub>k</sub> representam predições corretas e os elementos b<sub>k</sub> e c<sub>k</sub> representam predições errôneas, destaca-se a classe C<sub>1</sub> com apenas 4% de *pixels* classificados erroneamente e a classe C<sub>4</sub> com 11% de *pixels* classificados erroneamente. A Tabela 16 apresenta as métricas obtidas das matrizes de confusão de cada classe C<sub>k</sub>.

**Tabela 16** – Métricas obtidas das matrizes de confusão de cada classe

Classes	S <sub>k</sub>	E <sub>k</sub>	TFP <sub>k</sub>	TFN <sub>k</sub>
C <sub>1</sub> : [0,371;0,784)	0,9032	0,9425	0,0575	0,0968
C <sub>2</sub> : [0,784;1,198)	0,7155	0,9500	0,0500	0,2845
C <sub>3</sub> : [1,198;1,611)	0,8197	0,9159	0,0841	0,1803
C <sub>4</sub> : [1,611;2,024)	0,7097	0,9175	0,0825	0,2903
C <sub>5</sub> : [2,024;2,437]	0,6948	0,9705	0,0295	0,3052

S<sub>k</sub>: Sensibilidade, E<sub>k</sub>: Especificidade, TFP<sub>k</sub>: Taxa de falso positivo e TFN<sub>k</sub>: Taxa de falso negativo

Destaca-se a classe  $C_5$  com o menor índice de sensibilidade e o maior índice de especificidade, assim, a probabilidade de um *pixel* ser classificado na classe  $C_5$  do mapa modelo se este *pixel* pertence a classe  $C_5$  do mapa de referência é a menor se comparada com as outras classes e a probabilidade de um *pixel* do mapa modelo não apresentar produtividade de trigo no intervalo [2,024; 2,437] dado que o *pixel* não apresenta produtividade de trigo no mesmo intervalo no mapa de referência é maior se comparada com as demais classes. As taxas de falso positivo TFPk representam os erros de superestimativa e as taxas de falso negativo TFNk representam os erros de omissão. Observa-se que o maior erro de superestimativa ocorre na classe  $C_3$  e o maior erro de omissão ocorre na classe  $C_5$ . A Tabela 17 apresenta a matriz de confusão total da produtividade de trigo, utilizada para calcular métricas globais de comparação.

**Tabela 2** - Matriz de confusão total da produtividade de trigo

		Mapa gerado com todos os pontos	
		verdadeiro	falso
Mapa gerado sem os pontos influentes	verdadeiro	a = 4113	b = 1310
	falso	c = 1310	d = 20382

Utilizando os valores da Tabela 17 obtém-se  $S = 0,76$ ,  $E = 0,94$  e  $CCM = 0,70$ . Uma característica importante do índice de sensibilidade total ( $S$ ) é que ele equivale ao índice de exatidão global ( $EG$ ). Assim, a capacidade geral do mapa modelo em classificar *pixels* corretamente é de 0,76. O índice de especificidade total ( $E$ ) indica que a capacidade do mapa modelo em evitar classificações incorretas é de 0,94. O coeficiente de correlação de Matthews ( $CCM$ ) obtido foi 0,70 e como este índice fornece o valor 1 se os mapas forem idênticos, verifica-se que o mapa gerado sem os pontos influentes apresenta diferenças quando comparado com o mapa gerado com todos os pontos.

É importante destacar que o aumento do número de classes tende a diminuir a similaridade entre os mapas, por exemplo, utilizando 10 classes obtém-se  $EG = 0,61$ ,  $\hat{K} = 0,55$  e  $T = 0,56$ . Ao utilizarmos 15 classes, estes valores tornam-se  $EG = 0,50$ ,  $\hat{K} = 0,45$  e  $T = 0,46$  e utilizando 20 classes tem-se  $EG = 0,40$ ,  $\hat{K} = 0,36$  e  $T = 0,37$ . Assim, é importante definir um número de classes que esteja de acordo com o objetivo da análise, pois um número muito elevado tende a dificultar o manejo localizado, sendo assim inconveniente para o produtor.

## **CONCLUSÃO**

A análise geoestatística permitiu identificar com precisão o comportamento da produtividade de trigo na região monitorada e com isso, a estimação da produção de trigo ficou muito próxima a produção real. Destaca-se a validação cruzada como excelente técnica para escolher o modelo que melhor se ajusta ao semivariograma experimental.

Destaca-se que a precisão da estimação da produtividade de trigo deu-se pelo fato de ter sido retirado dois elementos amostrais detectados como influentes. Assim, conclui-se que a metodologia dos diagnósticos é de fundamental importância em estudos de geoestatística.

Os resultados mostraram que as métricas derivadas da matriz dos erros são uma excelente opção para realizar comparação entre mapas temáticos, pois fornecem valores globais e ainda permitem fazer comparações por classes.

## REFERÊNCIAS

ANDERSON, J. R. et al. A land use and land cover classification system for use with remote sensor data. **Geological Survey Professional Paper 964**. Washington: USGS. 1976. 41 p.

BERRY, J. K. Cartographic modeling: the analytical capabilities of GIS. In: Goodchild, M; Parks, B. O.; STEYAERT, L. T. **Environmental modelling with GIS**, New York, Oxford University Press. 1993. 488 p.

BORSSOI, J. A. ; URIBE-OPAZO, M. A ; GALEA, M. Diagnostic techniques applied in geostatistics for agricultural data analysis. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 33, n.6, p. 1561-1570, 2009.

CAMBARDELLA, C. A.; MOORMAN, T. B.; NOVAK, J. M.; PARKIN, T. B.; KARLEN, D. L.; TURCO, R. F.; KONOPKA, A. E. Field-scale variability of soil properties in central Iowa soils. **Soil Science Society America Journal**, Madison, v.58, n.5, p.1501-1511, 1994.

CARVALHO, L. P.; URIBE-OPAZO, M. A.; JOHANN, J. A.; SOUZA, E. G. Estudo da anisotropia na variabilidade espacial em propriedades químicas do solo. **Revista Brasileira de Ciência do Solo**, v.32, n.6, p.2217-2226, 2008.

CHRISTENSEN, R.; JOHNSON, W.; PEARSON, L. Covariance function diagnostics for spatial linear models. **International Association for Mathematical Geology**, New York, v. 25, n.2, p.145-160, 1993.

CLARK, I. **Practical geostatistics**. England: Applied Science Publishers, 1979. 129 p.

COHEN, J. A. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, v.20, p.37-46, 1960.

CONGALTON, R. G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. New York: Lewis Publishers, 136p., 1999. 136 p.

COOK, R. D. Assessment of local influence (with discussion). **Journal of the Royal Statistical Society**, Series B, London, v.48, p.133-169, 1986.

CRESSIE, N. Fitting variogram models by weighted least squares. **Mathem. Geol.**, n.17, p. 563-586, 1985.

CRESSIE, N. **Statistics for spatial data**. New York: Jonh Wiley. 900p., 1993.

CRESSIE, N.; LAHIRI, S.N. Asymptotics for REML estimation of spatial covariance parameters. **Journal of Statistical Planning and Inference**, p.327-341, 1996.

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



FARACO, M. A. et al . Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. **Rev. Bras. Ciênc. Solo**, Viçosa, v.32, n.2, p. 463-476, 2008.

FIELDING, A.H.; BELL, J.F. A review of methods for the assessment of prediction errors in conservation presence/absence models. **Environmental Conservation**, n. 24. p. 38-49, 1997.

GONG, P., HOWART, P.J. An assessment of some factors influencing multispectral land-cover classification. **Photogrammetric Engineering and Remote Sensing**. Bethesda, 56(5), p.597-603, 1990.

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

JENNESS, J; WYNNE, J. J. **Cohen's Kappa and classification table metrics 2.0**: an ArcView 3x extension for accuracy assessment of spatially explicit models. U.S. Geological Survey Open-File Report OF 2005-1363. U. S. Geological Survey, Southwest Biological Science Center, Flagstaff, AZ. 2005. 86 p.

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

LANDIM, P.M.B. Sobre Geoestatística e mapas. **Terra Didática**, 2(1), p. 19-33, 2006.

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

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

MA, Z., REDMOND, R.L. Tau coefficients for accuracy assessment of classification of remote sensing data. **Photogrammetric Engineering and Remote Sensing**, Bethesda, 61(4), p. 453-439, 1995.

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

MATÉRN, B. Spatial Variation. **Lecture notes in statistics**. 2 ed. New York: Springer. 1986. 151 p.

McBRATNEY, A.; WEBSTER, R. Choosing functions for semi-variograms os soil properties and fitting them to sample estimates. J. **Soil Sci.**, n.37, p.617-639, 1986.

MOLIN, J. P. Agricultura de precisão. Parte 2: diagnóstico, aplicação localizada e considerações econômicas. **Engenharia Agrícola**, v.17, n.2, p.108-121, 1997.

RIBEIRO JR., P.J.; DIGGLE, P.J. geoR: A package for geostatistical analysis. **R-NEWS**, v.1, n.2 , ISSN 1609-3631. 2001.

ROWLINSON, B.S.; DIGGLE, P. J. Splancs: Spatial Point Pattern Analysis Code in S-Plus. **Technical Report** 92/63, Lancaster University, U.K, 1992.

SOARES, A. **Geoestatística para as Ciências da Terra e do Ambiente**. Coleção Ensino da Ciência e Tecnologia, nº 9, Lisboa, IST Press. 2000. 206 p.

TOMLIN, C. D. **Geographic Information Systems and Cartographic Modelling**. Prentice-Hall, Inc. New Jersey. 1990. 249 p.

TRANGMAR, B.B.; YOST, R.S.; WADE, M.K.; UEHARA, G. Applications of geostatistics to spatial studies of soil properties. **Advances in Agronomy**, v.38, p.45-94, 1985.

VIEIRA, S.R.; HATFIELD, J.L.; NIELSEN, D.R.; BIGGAR, J.W. Geostatistical theory and application to variability of some agronomical properties. **Hilgardia**, v.51, n.3, p.1-75, 1983.

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

WEBSTER, R.; OLIVER, M.A. **Statistical Methods in Soil and Land Resource Survey**. Oxford University Press, Oxford, 1990. 316 p.

### 3 ESTATÍSTICA ESPACIAL DE ÁREAS APLICADA À AGRICULTURA DE PRECISÃO

#### 3.1 INTRODUÇÃO

O avanço da informática proporciona, a cada dia, o acesso a novas tecnologias que podem ser utilizadas na agricultura, resultando em grande fonte de informação para todos. Uma destas tecnologias consiste na criação e exploração de mapas que descrevam o comportamento de um atributo em determinada região, como, por exemplo, os bairros de uma cidade, os municípios de um estado ou até mesmo os estados de um país. Os Sistemas de Informação Geográfica ou SIGs como são conhecidos, realizam tais tarefas, pois são *softwares* que permitem a criação, manipulação e extração de informações de bases de dados geográficos.

A utilização de técnicas estatísticas que fazem uso explícito da localização dos fenômenos é um recurso que vem ganhando destaque na comunidade científica pela diversidade de suas aplicações. Um dos casos pioneiros em que se incorporou a localização espacial em uma investigação aconteceu em 1854, em Londres, em que o médico britânico John Snow, ao marcar em uma mapa a localização dos casos de cólera e das fontes de água, conseguiu identificar o local contaminado e acabar com a epidemia.

Porém, mesmo com a crescente disponibilidade de bases de dados georreferenciados e a expansão de tecnologias eficientes e de baixo custo afirmam CÂMARA et al. (2004), grande parte dos usuários de SIG's limita-se ao uso do mesmo apenas para visualização de mapas, ignorando que é possível ir muito além, já que a combinação de técnicas estatísticas adaptadas a dados espaciais com funções de visualização formam um conjunto de técnicas de análise com grande potencial para estudos urbanos e regionais.

Segundo SERRANO & VALCARCE (2000) quando se trabalha com dados de natureza espacial, pode aparecer um efeito denominado autocorrelação espacial. A dependência ou autocorrelação espacial surge sempre que o valor de uma variável em um lugar do espaço está relacionado com seu valor em outro ou outros lugares do espaço. Este fenômeno pode ser entendido como uma situação em que observações próximas no espaço possuam valores similares. Neste caso, o desafio da análise é medir o grau de associação espacial entre observações de uma ou mais variáveis.

Quando são utilizadas técnicas de estatística espacial para análise de dados de áreas, essas são apresentadas em forma de índices que medem a associação espacial em nível global (Índice de *Moran*) e nível local (índice *LISA*), gráficos de espalhamento (*Moran Scatterplot*), que permitem estudar o índice global em um diagrama de dispersão e mapas da distribuição dos índices.

## 3.2 ANÁLISE DE DADOS DE ÁREAS

### 3.2.1 Análise exploratória

A forma mais simples e intuitiva de análise exploratória é a visualização de valores extremos nos mapas. Vale ressaltar que o uso de diferentes pontos de corte da variável induz a visualização de diferentes aspectos. Os *softwares* de análise espacial dispõem usualmente de três métodos de corte de variável: intervalos iguais, percentis e desvios-padrão.

No caso de intervalos iguais, em que os valores máximo e mínimo são divididos pelo número de classes, se a variável tem uma distribuição muito concentrada de um lado, caracterizando uma assimetria, esse corte deixa apenas um número pequeno de áreas nas classes da perna mais longa da distribuição; como resultado, a maior parte das áreas será alocada a uma ou duas cores. O uso de percentis para definição de classes obriga a alocação dos polígonos em quantidades iguais pelas cores; isto pode mascarar diferenças significativas em valores extremos e dificultar a identificação de áreas críticas. Finalmente, o uso de desvios-padrão, no qual a distribuição da variável é apresentada em gradações de cores diferentes para valores acima e abaixo da média, faz a suposição da normalidade da distribuição da variável.

Em resumo, é parte importante da análise exploratória experimentar diferentes pontos de corte da variável na visualização dos mapas (CÂMARA et al., 2004). Os procedimentos consistem em adaptações de consagradas técnicas da estatística, como histograma, quartis, desvio padrão e pontos discrepantes, sendo que o detalhe é que esses recursos agora são representados em mapas temáticos, indicando as localizações.

### 3.2.2 Elementos básicos

Em todas as técnicas de estatística espacial de dados de área empregadas estão presentes três elementos básicos: a matriz de proximidade espacial ( $W$ ), o vetor de desvios ( $Z$ ) e o vetor de médias ponderadas ( $Wz$ ).

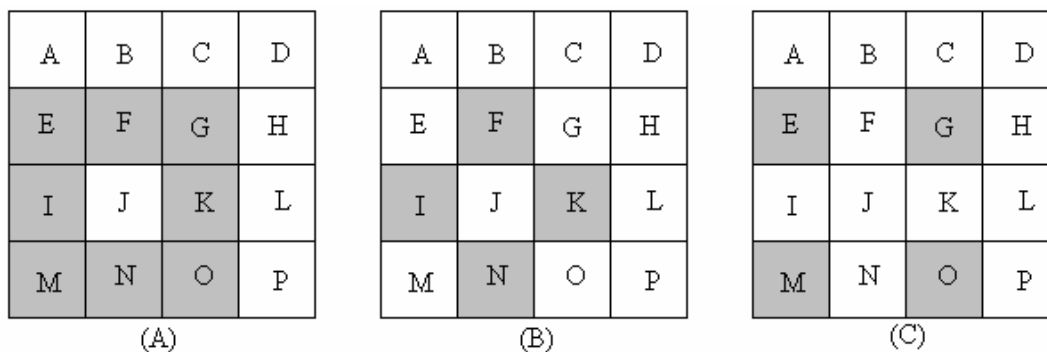
#### a) Matriz de proximidade espacial

Para estimar a variabilidade espacial de dados de área, uma ferramenta básica é a matriz de proximidade espacial, também chamada matriz de vizinhança. A matriz de proximidade espacial é uma ferramenta geral e bastante útil para descrever o arranjo

especial dos objetos (BAILEY & GATRELL, 1995). Dado um conjunto de  $n$  áreas  $\{A_1, A_2, \dots, A_n\}$ , constrói-se a matriz  $W$  de ordem  $n \times n$  onde cada um dos elementos  $w_{ij}$  representa uma medida de proximidade entre  $A_i$  e  $A_j$ . Essa medida de proximidade pode ser calculada a partir de um dos seguintes critérios:

- i)  $w_{ij} = 1$ , se o centróide de  $A_i$  está a uma determinada distância de  $A_j$ ; caso contrário  $w_{ij} = 0$ ; para  $i \neq j = 1, \dots, n$ .
- ii)  $w_{ij} = 1$ , se  $A_i$  compartilha um lado comum com  $A_j$ , caso contrário  $w_{ij} = 0$ ; para  $i \neq j = 1, \dots, n$ .
- iii)  $w_{ij} = l_{ij}/l_i$ , onde  $l_{ij}$  é o comprimento da fronteira entre  $A_i$  e  $A_j$  e  $l_i$  é o perímetro de  $A_i$ ; para  $i \neq j = 1, \dots, n$ .

Ao utilizar matrizes de proximidade espacial que considerem a contiguidade, é importante convencionar as formas de vizinhança. Para tanto, os critérios são baseados em movimentos de algumas peças do jogo de xadrez, como a rainha, torre e bispo. Por exemplo, a Figura 16 apresenta os critérios de vizinhança para identificar os vizinhos da área J. Na Figura 16 (A) o critério adotado é o rainha “Queen”, o seja, todas as áreas que tiverem interseção não nula com a área J serão vizinhas da área J. Na Figura 16 (B) é apresentado o critério torre “Rook”, em que os vizinhos serão apenas aqueles que tiverem um lado em comum. Para o critério bispo “Bishop”, apenas as áreas que se localizam nas diagonais são consideradas vizinhas (Figura 16 (C)).



**Figura 16** – Representação dos tipos de contiguidade entre áreas. (A) Contiguidade *Queen* “rainha”, (B) Contiguidade *Rook* “torre” e (C) Contiguidade *Bishop* “bispo”.

A idéia da matriz de proximidade espacial pode ser generalizada para vizinhos dos vizinhos, e isso caracteriza a ordem da matriz. Por exemplo, considerando o critério de vizinhança *Rook* e uma matriz de ordem 2, os vizinhos de F, I, N e K também seriam considerados vizinhos de J. Como a matriz de proximidade é utilizada em cálculos de indicadores na fase de análise exploratória, é muito útil normalizar suas linhas, para que a

soma dos pesos de cada linha seja igual a 1. Isto simplifica os cálculos de índices de autocorrelação espacial.

#### b) Vetor dos desvios

Sabe-se que a cada área  $i$  está associado um número real ( $x_i$ ), que representa o valor do atributo na área  $i$ . Para o cálculo do vetor de desvios  $Z$  é calculada, primeiramente, a média ( $\mu$ ) dos valores dos atributos, considerando as  $n$  áreas. Cada elemento  $i$  de  $Z$ ,  $z_i$ , é obtido subtraindo-se o valor da média, do valor do atributo, ou seja,  $z_i = x_i - \mu$ .

#### c) Vetor de médias ponderadas

O terceiro elemento básico, o vetor de médias ponderadas ( $Wz$ ), é obtido pela multiplicação do vetor transposto dos desvios, pela matriz de proximidade espacial com linhas normalizadas, onde cada elemento de uma linha  $i$  qualquer, originariamente com valor 1, é dividido pelo número de elementos não-nulos da mesma linha. Desta forma, como resultado, cada elemento  $wz_i$ , contém um valor correspondente à média dos desvios dos vizinhos da área  $i$ , caracterizando uma média móvel espacial.

Com a definição do vetor de médias ponderadas podemos estabelecer outro mapa em uma análise exploratória, simplesmente calculando a média móvel espacial dos atributos estudados. Segundo CÂMARA et al. (2004) o cálculo de uma média móvel espacial é uma forma de explorar a variação da tendência espacial dos dados, pois a operação tende a produzir uma superfície com menor flutuação que os dados originais.

### 3.2.3 Indicador global de autocorrelação espacial

Um teste usualmente utilizado para indicar a possibilidade de autocorrelação espacial global é o teste  $I$  de Moran, assim batizado em homenagem ao estatístico australiano Patrick Alfred Pierce Moran que o definiu no artigo MORAN (1950). Segundo ANSELIN (2005), o índice  $I$  de Moran é uma medida cujo resultado indica se os dados se distribuem ou não de forma aleatória no espaço. A estatística  $I$  de Moran calcula uma medida de autocorrelação espacial global cujo significado pode ser interpretado da seguinte forma: caso esta autocorrelação seja positiva, então tem-se que, se uma dada região apresentar valores elevados ou (baixos) para uma variável qualquer, então seu entorno deve apresentar valores elevados ou (baixos), o que caracterizaria uma espécie de *cluster*

no espaço. Caso esta estatística seja negativa, então pode-se esperar que uma região qualquer que apresente um valor elevado ou (baixo) para uma dada variável, deverá estar cercada de regiões cujos valores para esta mesma variável devem ser baixos ou (elevados). O índice global de Moran é indicado na Equação (36):

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (36)$$

O mesmo cálculo feito para matrizes de proximidade de maior ordem permite estimar a função de autocorrelação para cada ordem de vizinhança (ou “lag”). Para testar a significância do índice I de Moran, utiliza-se uma distribuição de referência aleatorizando os valores do atributo estudado nas áreas e calcula-se um novo valor do índice para cada permutação feita, permitindo a realização do seguinte teste de hipóteses:  $H_0: \rho = 0$ , que representa a hipótese nula de ausência de correlação espacial *versus* a hipótese alternativa  $H_1: \rho \neq 0$ . Se o nível descritivo (p-valor) for menor que o nível de significância estipulado (usualmente 0,05 de probabilidade) rejeita-se a hipótese nula de ausência de correlação espacial. Ao contrário, se o nível descritivo for maior que o nível de significância, não se rejeita a hipótese nula.

O índice I de Moran também pode ser definido em função do vetor de desvios e do vetor de médias ponderadas, ou seja:

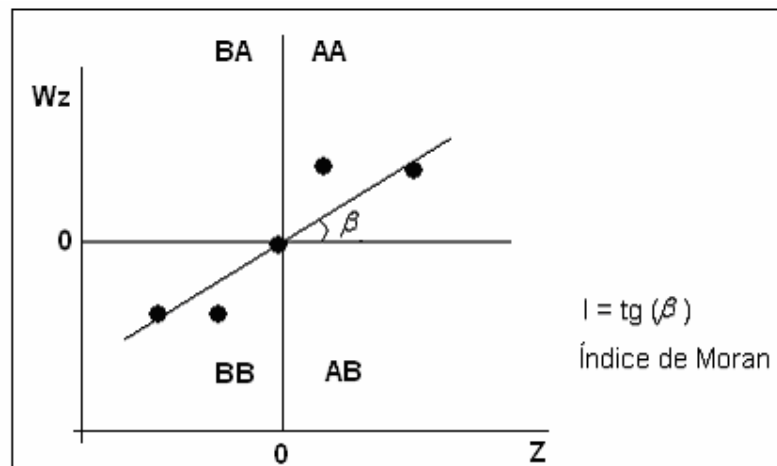
$$I = \frac{Z^t \cdot Wz}{Z^t \cdot Z} \quad (37)$$

### 3.2.4 Diagrama de espalhamento de Moran

Segundo ANSELIN (1996), o gráfico de dispersão de Moran é uma das formas de interpretar o índice I de Moran. Como este índice indica a inclinação da reta de regressão de  $Wz$  (média dos vizinhos dos desvios  $z$  normalizada) em  $Z$  (desvios normalizados) (NETER e WASSERMAN, 1974), pode-se construir um dispositivo gráfico para visualizar a associação espacial entre cada elemento  $z_i$  com a média dos valores dos seus vizinhos  $wz_i$ . A Figura 17 mostra como o gráfico de dispersão do índice I de Moran é construído, gráfico este rotineiramente chamado de *Moran Scatterplot*.

O gráfico de dispersão de Moran apresentado na Figura 17 é dividido em quadrantes. Esses quadrantes correspondem a quatro padrões de associação espacial entre as regiões e seus vizinhos. O primeiro quadrante (localizado na parte superior direita) mostra as regiões que apresentam altos valores para a variável em análise cercadas por

regiões que apresentam valores semelhantes. A classificação atribuída a esse quadrante é alto-alto (AA). O segundo quadrante (localizado na parte superior esquerda) mostra as regiões com valores baixos cercados por vizinhos que apresentam valores altos.







**Figura 17** – Construção do gráfico de dispersão do índice I de Moran.

Esse quadrante é classificado como baixo-alto (BA). O terceiro quadrante (localizado no canto inferior esquerdo) é constituído pelas regiões com valores baixos para as variáveis em análise cercados por vizinhos que também apresentam baixos valores. Esse quadrante é identificado como baixo-baixo (BB). O quarto quadrante (localizado no canto inferior direito) é formado pelas regiões com altos valores para as variáveis em análise cercadas por regiões com baixos valores. Tal quadrante classifica-se como alto-baixo (AB). As regiões que estão localizadas nos quadrantes AA e BB apresentam autocorrelação espacial positiva, ou seja, essas regiões formam *clusters* de valores similares. Por outro lado, os quadrantes BA e AB apresentam autocorrelação espacial negativa, indicando regiões em que seu comportamento é diferente de seus vizinhos.

### 3.2.5 Mapa de espalhamento de Moran

Para uma melhor compreensão do diagrama de espalhamento de Moran, é conveniente apresentá-lo associado com um mapa temático bidimensional, no qual cada polígono da região estudada é apresentado segundo seu quadrante no diagrama de espalhamento de Moran. Este mapa é conhecido como *Box Map* e a sua legenda é construída segundo a Figura 18.



	Alto - Alto	AA
	Baixo - Baixo	BB
	Alto - Baixo	AB
	Baixo - Alto	BA

**Figura 18** – Legenda do mapa de espalhamento de Moran.

### 3.2.6 Índice local de associação espacial (*LISA*)

Enquanto o indicador global de correlação espacial *I* de Moran fornece um único valor como medida da associação espacial para todo o conjunto de dados, os indicadores locais produzem um valor específico para cada objeto, permitindo assim, a identificação de agrupamentos de objetos com valores de atributos semelhantes (*clusters*) e objetos anômalos (*outliers*). Segundo ANSELIN (1995), um *LISA* tem que atender a dois objetivos:

- i) permitir a identificação de padrões de associação espacial significativos;
- ii) ser uma decomposição do índice global de associação espacial.

O *LISA* utilizado neste trabalho é o índice local de Moran. Uma das formas de representação deste índice, em função dos elementos básicos, é exposta na Equação 38:

$$I_i = z_i \cdot Wz_i / \sigma^2 \quad (38)$$

em que:

$I_i$  : índice local para o objeto  $i$ ;  $i = 1, \dots, n$ .

$z_i$  : valor do desvio do objeto  $i$ ;  $i = 1, \dots, n$ .

$Wz_i$  : valor médio dos desvios dos objetos vizinhos de  $i$ ;

$\sigma^2$  : variância da distribuição dos valores dos desvios.

### 3.2.7 Lisa map

A significância estatística do uso do índice de Moran local é computada de forma similar ao caso do índice global. Para cada área, calcula-se o índice local e, depois permuta-se aleatoriamente o valor das demais áreas, até obter-se uma pseudo-distribuição para a

qual possamos computar os parâmetros de significância. Uma vez determinada a significância estatística do índice local de Moran, é útil gerar um mapa indicando as regiões que apresentam correlação local significativamente diferente do resto do dados. Estas regiões podem ser vistas como "bolsões" de não-estacionariedade, pois são áreas com dinâmica espacial própria e que merecem análise detalhada.

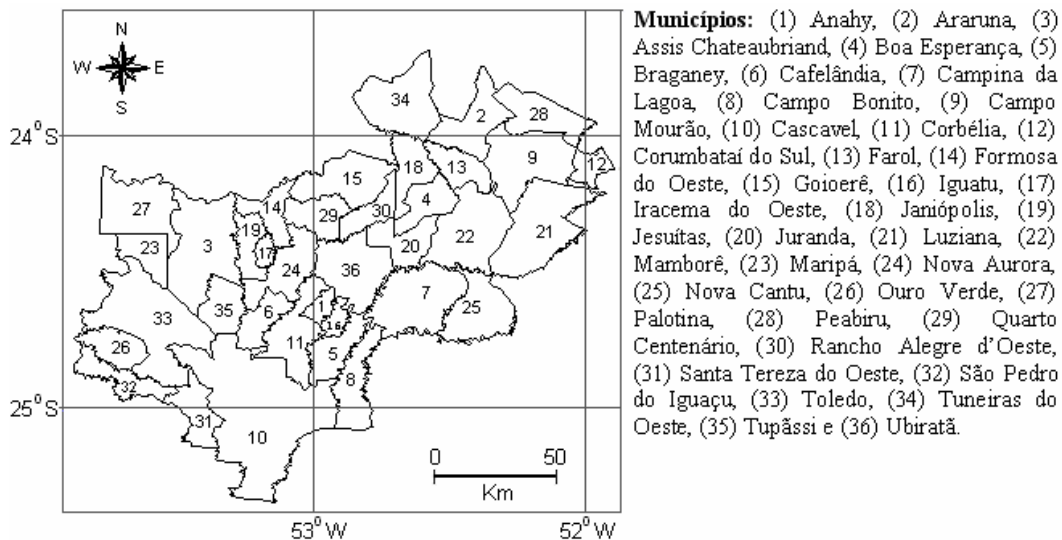
### **3.2.8 Moran map**

O Moran map é uma mapa gerado pela combinação do diagrama de espalhamento de Moran com o Lisa map. Desta forma, as regiões detectadas como significativas no Lisa map são classificadas conforme sua localização no diagrama de espalhamento de Moran.

### 3.3 MATERIAL E MÉTODOS

#### 3.3.1 Material

A análise de um conjunto de dados reais será efetuada com o propósito de ilustrar a aplicação das técnicas de estatística espacial de áreas. Para criação do banco de dados espaciais, será utilizado um *shapefile* dos municípios da região Oeste do Estado do Paraná (Figura 19).



**Figura 19** – *Shapefile* dos 36 municípios estudados na Região Oeste do Paraná.

Pretende-se estudar a autocorrelação espacial de índices de vegetação da soja no ano agrícola de 2004/2005. Para tanto, será utilizado um conjunto de índices NDVI e GVI de reflectância de superfície da safra 2004/2005 obtidos por MERCANTE (2007) nas datas de 23/11/2004, 09/12/2004, 25/12/2004, 26/01/2005 e 11/02/2005.

Entre os diversos índices de vegetação existentes, o índice de vegetação por diferença normalizada (NDVI - Normalized Difference Vegetation Index) (ROUSE et al., 1973), obtido a partir da razão entre a subtração e soma das refletâncias das bandas da região do infravermelho próximo e do vermelho do espectro eletromagnético, é o que vem sendo mais largamente utilizado (KLERING et. al, 2007).

Em estudos de monitoramento, frequentemente é buscado o estabelecimento de relações entre o padrão de evolução do crescimento das plantas ao longo do tempo com o padrão de evolução temporal do NDVI. A utilidade do índice NDVI foi demonstrada em diversos campos, tais como, agricultura de precisão, estudo da fenologia da vegetação, em classificações espectrais e avaliações primárias de produtividades (SUGAWARA, 2001).

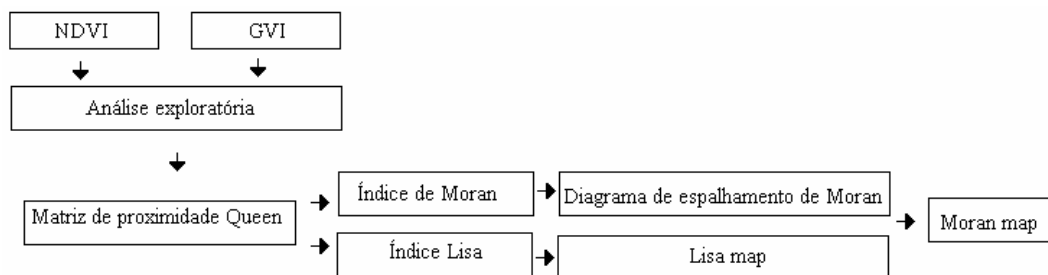
Segundo SONG & WOODCOCK (2003), a transformação *Tasseled Cap* é um dos índices de vegetação mais utilizado para as imagens geradas pelos satélites do projeto

Landsat. Nessa transformação, as resultantes são três imagens-síntese denominadas *Brightness*, *Greenness* e *Wetness* (CRIST & CICONE, 1984). A imagem “*greenness*” ou índice de vegetação GVI (*green vegetation index*) apresenta altos valores em alvos com alta densidade de vegetação verde, sendo obtida de acordo com a Equação (39), em que T(*i*) representa a banda(*i*) do sensor Thematic Mapper do satélite Landsat 5.

$$GVI = -0.2848(T1) -0.2435(T2) -0.5436(T3) +0.7243(T4) +0.0840(T5) -0.1800(T7) \quad (39)$$

### 3.3.2 Metodologia

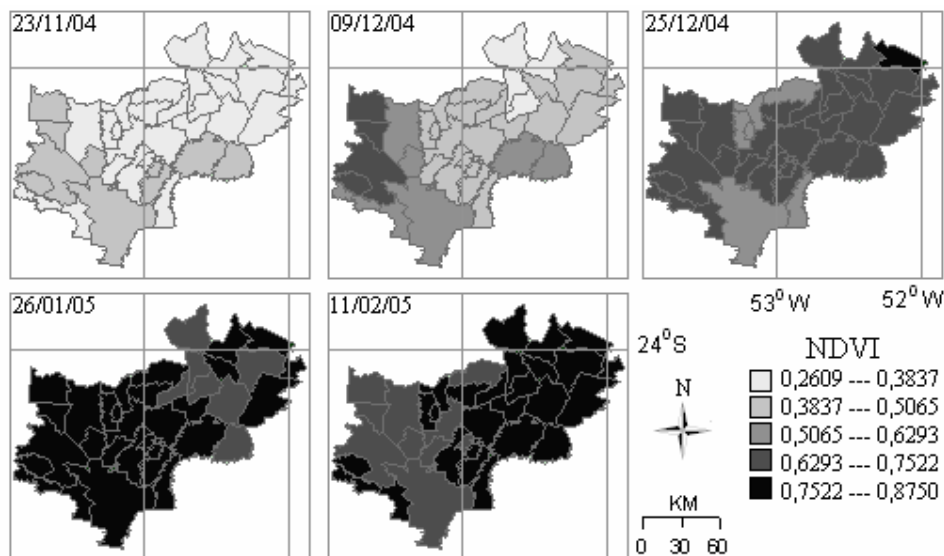
A Figura 20 mostra o roteiro de análise que será trabalhado com índices de vegetação. Será feita uma análise exploratória dos índices de vegetação e posteriormente, será criada a matriz *queen* de proximidade espacial referente ao shapefile da Região Oeste do Paraná. Serão calculados índices de correlação espacial global e local e confeccionados diagramas de espalhamento de Moran, *Lisa Map's* e *Moran Map's*.



**Figura 20** – Roteiro de análise que será trabalhado com índices de vegetação.

### 3.4 RESULTADOS E DISCUSSÃO

A Figura 21 apresenta a evolução temporal do índice NDVI. Os intervalos foram definidos agrupando os valores de NDVI de todas as datas para obter o máximo (0,8750) o mínimo (0,2609) e a amplitude geral (0,61403) dos índices. Posteriormente, dividiu-se a amplitude geral pelo número de classes desejado (5) para obter o intervalo de classe.

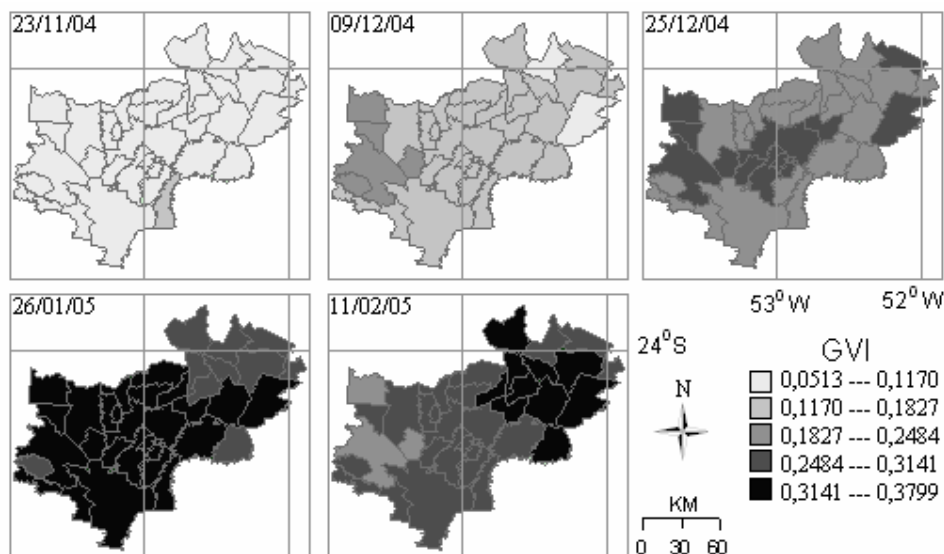


**Figura 21** – Evolução temporal do índice NDVI.

Observa-se que a evolução temporal do índice NDVI esteve relacionada com as diferentes fases do ciclo fenológico da cultura, pois os valores mais baixos dos índices são observados em 23/11/04, onde a cultura já está em desenvolvimento. Porém, em período inicial e os valores mais altos se encontram entre 26/01/05 e 11/02/05, período esse de maior desenvolvimento das plantas. É possível observar uma diferença nas datas de plantio, visto que alguns municípios têm seus ciclos mais tardios que os outros. A Figura 22 apresenta a evolução temporal do índice GVI, em que as classes foram calculadas de maneira análoga ao índice NDVI. A evolução temporal do índice GVI apresenta um comportamento semelhante à evolução do índice NDVI, no que diz respeito à relação com o ciclo fenológico da cultura da soja, pois nas datas iniciais os valores são baixos e nas finais os valores são altos. Destaca-se certa insensibilidade do índice em 23/11/04, pois, como 97,22 % dos municípios encontram-se no primeiro intervalo da legenda, não se pode identificar quais municípios apresentam um plantio antecipado.

A Tabela 20 apresenta o índice I de Moran dos índices de vegetação NDVI e GVI com seu respectivo nível descritivo (p-valor) referente à hipótese de falta de correlação espacial. Nesta análise está sendo considerada que os dados apresentam uma distribuição normal. Para cada data foram geradas distribuições empíricas, por meio de 99 permutações

escolhidas independentes e com igual probabilidade dentre todas as possíveis permutações dos índices entre os municípios estudados.



**Figura 22** – Evolução temporal do índice GVI.

Considerando o NDVI observa-se que, para todas as datas em estudo, o índice I de Moran obtido foi maior que zero, assim, existe uma correlação positiva indicando que a região estudada apresenta municípios com altos e/ou baixos índices NDVI cercados por municípios com a mesma situação.

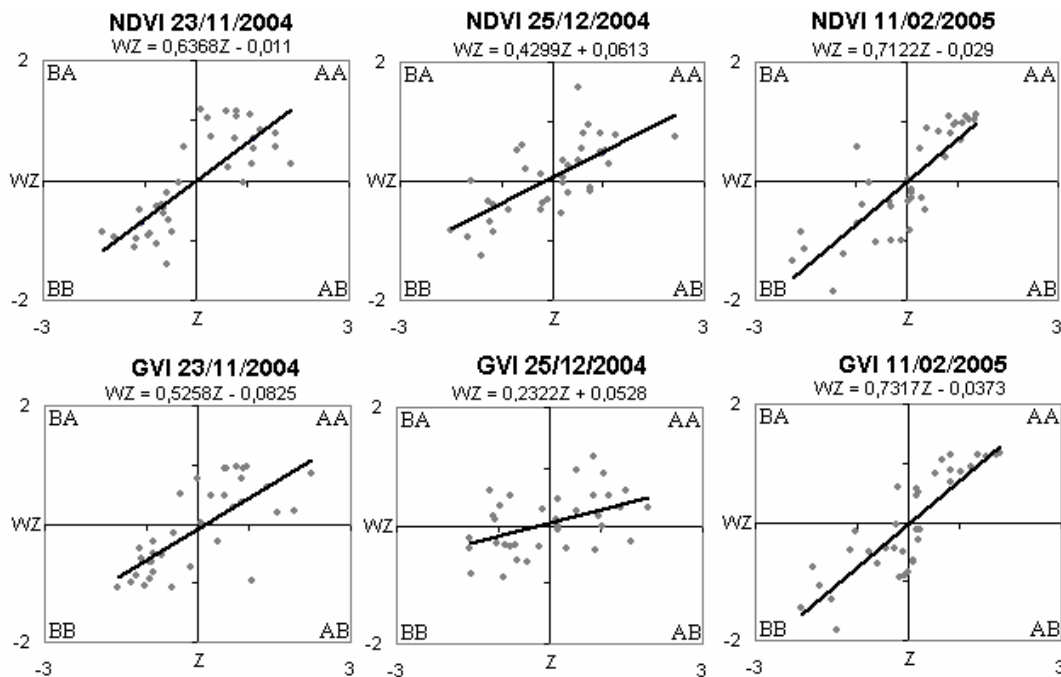
**Tabela 18** – Índice I de Moran dos índices NDVI e GVI

Índice	Data	I	p-valor
NDVI	23/11/04	0,6368	0,0010*
	09/12/04	0,7315	0,0010*
	25/12/04	0,4299	0,0010*
	26/01/05	0,5694	0,0010*
	11/02/05	0,7122	0,0010*
GVI	23/11/04	0,5258	0,0010*
	09/12/04	0,6814	0,0010*
	25/12/04	0,2322	0,0140*
	26/01/05	0,6904	0,0010*
	11/02/05	0,7317	0,0010*

\*Rejeita-se a hipótese de falta de correlação espacial ao nível de 5% de probabilidades.

Considerando o índice GVI, verifica-se que em todas as datas os índices calculados foram positivos e significativos, evidenciando a presença de aglomerados de valores semelhantes. Os níveis descritivos (p-valores) referentes a todas as datas em estudo são inferiores aos níveis de significâncias estipulados (0,05 de probabilidade), mostrando que as

correlações espaciais positivas dos índices NDVI e GVI, nas datas monitoradas, são significativas. Comparando o índice I de Moran dos índices de vegetação NDVI e GVI, destaca-se que o comportamento de ambos é semelhante, porém, nas três primeiras datas monitoradas os índices calculados com o NDVI são superiores aos calculados com o GVI, fato esse que se inverte nas duas últimas datas, em que o índice I de Moran do GVI é mais elevado. A Figura 23 apresenta os diagramas de espalhamento de Moran dos índices NDVI e GVI em 23/11/2004, 25/12/2004 e 11/02/2005. Primeiramente, destaca-se o coeficiente angular da reta de regressão de  $Wz$  versus  $Z$ , que representa o índice I de Moran.



**Figura 23** – Diagrama de espalhamento de Moran (Moran Scatterplot) para os índices NDVI e GVI em 23/11/2004, 25/12/2004 e 11/02/2005.

Nos diagramas de espalhamento de Moran (Figura 23), cada ponto representa um município da região estudada - assim, temos uma maneira visual de investigar o índice I de Moran. Considerando a data 23/11/2004 observa-se que a maioria dos municípios encontra-se nos quadrantes AA e BB, tanto para o índice NDVI quanto para o GVI, o que evidencia a correlação espacial positiva. Considerando os diagramas em 25/12/04, verifica-se que os municípios, que anteriormente se localizavam nos quadrantes AA e BB, começam a se distribuir nos demais quadrantes, evidenciando um período de transição. Este comportamento é devido ao fato que nas regiões em que os índices são altos, alguns municípios começam a apresentar índices menores, em virtude de estarem passando do pico vegetativo. Em contrapartida, municípios que iniciaram o plantio posteriormente, começam a apresentar valores elevados, em virtude de estarem no estágio de maior

biomassa. Para uma investigação mais detalhada, quantificou-se a percentagem de municípios alocados em cada quadrante, o que é apresentado na Tabela 19.

**Tabela 19** – Percentagens de municípios em cada quadrante do gráfico de espalhamento de Moran dos índices NDVI e GVI

Índice	DATA	AA	AB	BB	BA
NDVI	23/11/04	44,44 %	02,78 %	50,00 %	02,78 %
	09/12/04	30,56 %	05,56 %	55,56 %	08,33 %
	25/12/04	41,67 %	13,89 %	30,56 %	13,89 %
	26/01/05	38,89 %	05,56 %	38,89 %	16,67 %
	11/02/05	41,67 %	19,44 %	36,11 %	02,78 %
GVI	23/11/04	41,67 %	05,56 %	47,22 %	05,56 %
	09/12/04	33,33 %	08,33 %	52,78 %	05,56 %
	25/12/04	36,31 %	16,67 %	30,56 %	16,67 %
	26/01/05	50,00 %	05,56 %	33,33 %	11,11 %
	11/02/05	38,80 %	13,80 %	44,44 %	02,78 %

Conforme observado, pontos que se enquadram nos quadrantes AB e BA do gráfico de espalhamento de Moran representam municípios com associação espacial negativa, ou seja, são entendidos como municípios com alto (baixo) índice de vegetação cercados por municípios com baixo (alto) índice de vegetação. Observando o NDVI é possível constatar que a data que apresenta a maior percentagem de municípios alocados nos quadrantes AB e BA é 25/12/04 com 27,78 % dos municípios. Para o GVI o mesmo ocorre, porém, a percentagem de municípios alocados nos quadrantes que representam correlação inversa é 33,34 %. Assim, verifica-se que o índice GVI foi mais eficaz na detecção do período de transição. Considerando a última data monitorada (11/02/2005), observa-se que, para o índice NDVI a percentagem de municípios alocados no quadrante AA é superior à percentagem de municípios alocados no quadrante BB. Desta maneira existe maior quantidade de municípios com índice NDVI alto cercados por municípios com índice NDVI altos, do que municípios com índice NDVI baixos cercados por municípios com índice NDVI baixos. Este comportamento se inverte quando consideramos o índice GVI, pois para a mesma data, a percentagem de municípios alocados em AA é menor do que a percentagem de municípios alocados em BB. Este comportamento pode ser explicado pelo fato do índice GVI utilizar de mais informações que são oriundos de maior quantidade de bandas TM, como por exemplo, a banda 5 (1,55 – 1,75  $\mu m$ ) que é sensível ao teor de umidade das plantas e a banda 7 (2,08 – 2,35  $\mu m$ ) que é sensível à morfologia do terreno.

Sabe-se que o índice I de Moran é uma medida de associação espacial para toda a região em estudo, e como se investigam 36 municípios, é provável que ocorram diferentes regimes de associação espacial. Assim, é necessário complementar a análise com a utilização de um índice que produza um valor específico para cada município, o que vai permitir a identificação de agrupamentos. A Tabela 20 apresenta os índices locais de



associação espacial (LISA) do índice NDVI para todas as datas monitoradas acompanhados de sua classificação quanto à significância estatística.

**Tabela 20** – Índice local de associação espacial (LISA) do índice NDVI

Municípios	23/11/04	09/12/04	25/12/04	26/01/05	11/02/05
Anahy	0,85 ns	0,14 ns	0,97 ns	0,93 *	-0,02 ns
Araruna	1,57 *	1,14 *	0,74 ns	0,67 *	0,91 **
Assis Chateaubriand	-0,14 *	1,06 **	0,06 ns	-0,01 ns	1,53 **
Boa esperança	1,51 **	0,94 *	0,45 ns	1,13 **	1,39 **
Braganey	0,79 *	0,00 ns	-0,01 ns	1,15 **	-0,03 ns
Cafelândia	0,22 ns	0,02 ns	-0,08 ns	1,39 *	0,62 *
Campina da Lagoa	0,28 ns	-0,11 ns	0,08 ns	0,00 ns	0,16 ns
Campo Bonito	0,12 *	0,03 ns	0,43 ns	0,98 ns	-0,02 ns
Campo Mourão	0,87 **	0,42 *	0,68 **	0,56 ns	1,07 **
Cascavel	1,07 **	0,57 *	-0,01 ns	0,93 **	0,72 *
Corbélia	0,53 *	0,05 ns	0,18 ns	1,36 **	0,13 ns
Corumbataí do Sul	0,40 ns	0,03 ns	0,84 **	0,27 ns	0,36 ns
Farol	1,35 **	0,78 **	0,32 ns	0,77 **	1,30 **
Formosa do Oeste	0,35 *	0,02 ns	1,58 **	0,00 ns	-0,18 ns
Goioerê	0,78 *	0,56 ns	0,86 ns	0,64 *	-0,57 ns
Iguatu	0,25 **	0,29 ns	0,50 *	1,24 **	0,02 ns
Iracema do Oeste	0,54 ns	0,09 ns	1,73 *	0,11 ns	-0,09 ns
Janiópolis	1,15 **	1,34 **	0,01 ns	1,42 **	0,73 *
Jesuítas	0,34 ns	-0,03 ns	1,66 *	0,09 ns	0,12 ns
Juranda	0,29 ns	0,43 ns	0,54 ns	0,48 *	0,69 **
Luiziana	0,13 ns	0,20 ns	0,08 ns	0,36 *	1,54 **
Mamborê	0,28 ns	0,26 ns	0,45 ns	0,75 **	1,28 **
Maripá	0,85 *	2,91 **	-0,02 ns	0,00 ns	2,70 **
Nova Aurora	0,02 ns	-0,01 ns	0,03 ns	0,24 *	0,35 ns
Nova Cantu	-0,05 ns	-0,16 ns	-0,36 ns	0,70 ns	0,96 *
Ouro Verde do Oeste	1,17 ns	2,76 *	0,11 ns	-0,26 ns	0,10 ns
Palotina	0,52 ns	2,61 *	-0,06 ns	0,00 ns	3,03 *
Peabiru	0,82 *	0,31 ns	1,77 ns	-0,42 *	0,54 ns
Quarto Centenário	0,33 ns	0,06 ns	0,40 ns	0,22 ns	-0,01 ns
Rancho Alegre d'Oeste	0,79 *	0,12 *	-0,11 ns	0,92 *	0,32 ns
Santa Tereza do Oeste	0,91 *	1,72 **	1,01 *	0,85 ns	0,32 *
São Pedro do Iguaçu	0,68 *	1,59 **	0,47 ns	0,24 ns	-0,06 ns
Toledo	1,22 **	2,68 **	-0,11 ns	-0,20 ns	1,80 **
Tuneiras do Oeste	0,77 **	1,59 **	-0,35 ns	1,92 *	0,91 **
Tupãssi	0,59 ns	1,13 *	-0,14 ns	0,53 ns	2,32 **
Ubiratã	0,14 ns	0,05 ns	0,32 ns	-0,02 ns	0,00 ns

(ns) não significativo; (\*) significativo com 5% de probabilidade e (\*\*) significativo com 1% de probabilidade

Segundo Anselin (1995), a interpretação do índice local  $I_i$  é semelhante à interpretação do índice global, pois valores positivos de  $I_i$  indicam uma agregação espacial de valores similares enquanto que valores negativos desta estatística, indicam uma agregação espacial de valores dissimilares. Observando a Tabela 20, destaca-se a data 25/12/04 com apenas 7 índices classificados como significativos, o que pode ser explicado pelo fato da soja se encontrar em diferentes fases do seu desenvolvimento na região e este ser um período de transição - visto que nas regiões em que inicialmente as plantas apresentavam as folhas em desenvolvimento, as vagens começam a apresentar a cor de vagem madura e, nas regiões que iniciaram um plantio posterior, as plantas começam a

apresentar um pleno enchimento das vagens. A Tabela 21 apresenta os índices locais de associação espacial  $I_i$  referentes aos índices de vegetação GVI em todas as datas monitoradas.

**Tabela 21** – Índice local de associação espacial (LISA) do índice GVI

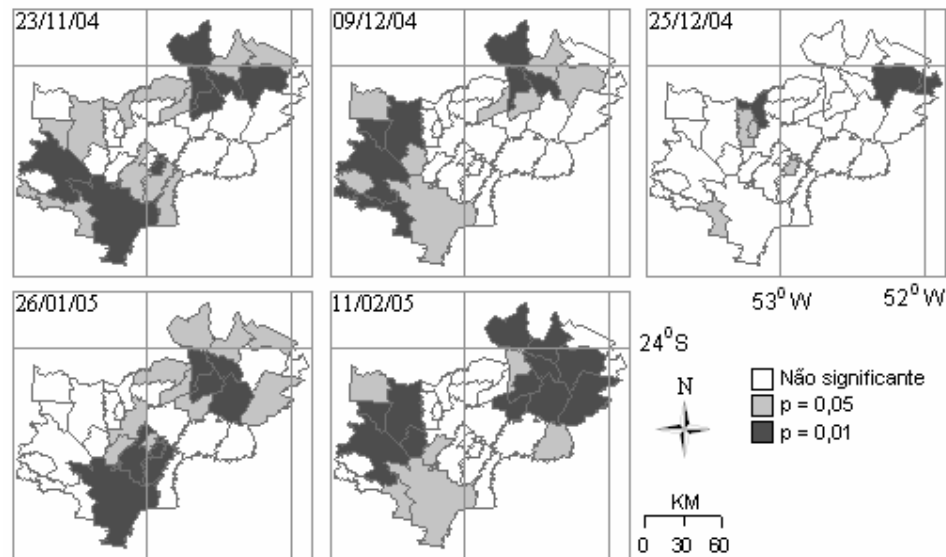
Municípios	23/11/04	09/12/04	25/12/04	26/01/05	11/02/05
Anahy	0,31 ns	0,11 ns	1,06 *	0,26 ns	0,20 ns
Araruna	0,87 *	1,27 *	0,23 ns	3,18 **	0,43 *
Assis Chateaubriand	-0,18 ns	1,11 **	0,08 ns	0,50 *	1,96 **
Boa esperança	1,67 **	0,44 ns	-0,13 ns	0,28 ns	2,09 **
Braganey	0,66 *	-0,04 ns	-0,04 ns	0,55 *	-0,06 ns
Cafelândia	0,12 ns	0,06 ns	0,46 ns	1,11 *	0,15 *
Campina da Lagoa	0,20 ns	-0,06 ns	0,02 ns	-0,04 ns	0,08 ns
Campo Bonito	1,90 *	0,10 ns	0,33 ns	0,41 ns	-0,02 ns
Campo Mourão	0,66 *	0,64 **	-0,01 ns	1,51 **	0,89 **
Cascavel	0,92 **	0,42 ns	-0,33 ns	0,97 **	0,39 ns
Corbélia	0,25 ns	0,01 ns	0,60 ns	0,96 **	0,05 ns
Corumbataí do Sul	-1,05 ns	-0,08 ns	-0,41 ns	2,00 *	-0,12 ns
Farol	1,31 **	0,43 **	0,36 ns	1,22 **	1,54 **
Formosa do Oeste	0,09 *	0,08 ns	0,64 ns	0,02 ns	0,12 ns
Goioerê	0,71 *	0,28 ns	1,29 *	0,16 ns	0,10 ns
Iguatu	-0,01 ns	0,11 ns	0,51 **	0,49 ns	0,15 ns
Iracema do Oeste	0,48 ns	0,26 ns	0,38 ns	0,50 ns	-0,08 ns
Janiópolis	1,11 **	0,48 *	0,28 *	0,64 ns	1,16 **
Jesuítas	0,36 ns	-0,02 ns	0,29 ns	0,73 ns	0,42 ns
Juranda	0,62 *	0,23 ns	0,15 ns	-0,05 ns	0,74 *
Luiziana	0,38 ns	0,65 ns	-0,45 ns	0,44 ns	1,96 **
Mamborê	0,59 *	0,37 *	0,01 ns	0,28 *	1,69 **
Maripá	0,74 *	4,38 **	0,97 ns	0,37 ns	2,55 **
Nova Aurora	0,07 ns	0,00 ns	0,14 ns	0,45 **	0,54 *
Nova Cantu	-0,12 ns	0,10 ns	-0,72 ns	0,46 ns	0,95 *
Ouro Verde do Oeste	0,84 ns	1,47 ns	0,00 ns	-0,29 ns	0,06 ns
Palotina	0,40 ns	4,51 *	0,94 ns	0,59 ns	3,05 *
Peabiru	0,27 ns	0,50 *	-0,37 ns	2,09 **	0,07 ns
Quarto Centenário	0,45 ns	0,05 ns	0,07 ns	-0,04 ns	-0,03 ns
Rancho Alegre d'Oeste	1,04 **	0,04 ns	-0,19 ns	-0,06 ns	0,56 *
Santa Tereza do Oeste	0,55 *	0,71 *	0,30 ns	0,86 ns	0,01 ns
São Pedro do Iguaçú	0,50 ns	0,77 *	-0,12 ns	0,17 ns	-0,06 ns
Toledo	0,82 *	2,36 **	-0,02 ns	-0,11 **	1,45 **
Tuneiras do Oeste	0,56 *	1,02 *	0,79 *	2,74 **	0,70 *
Tupãssi	0,30 ns	1,14 *	0,46 ns	0,77 *	1,89 **
Ubiratã	0,00 ns	-0,07 ns	0,59 ns	0,03 ns	0,01 ns

(ns) não significativo; (\*) significativo com 5% de probabilidade e (\*\*) significativo com 1% de probabilidade

Destaca-se também o fato da data 25/12/04 apresentar uma grande quantidade de índices não significativos, pois apenas os municípios de Anahy (1), Goioerê (15), Iguatu (16), Janiópolis (18) e Tuneiras do Oeste (34) apresentaram índices de correlação espacial significativos ( $p$ -valor < 0,05).

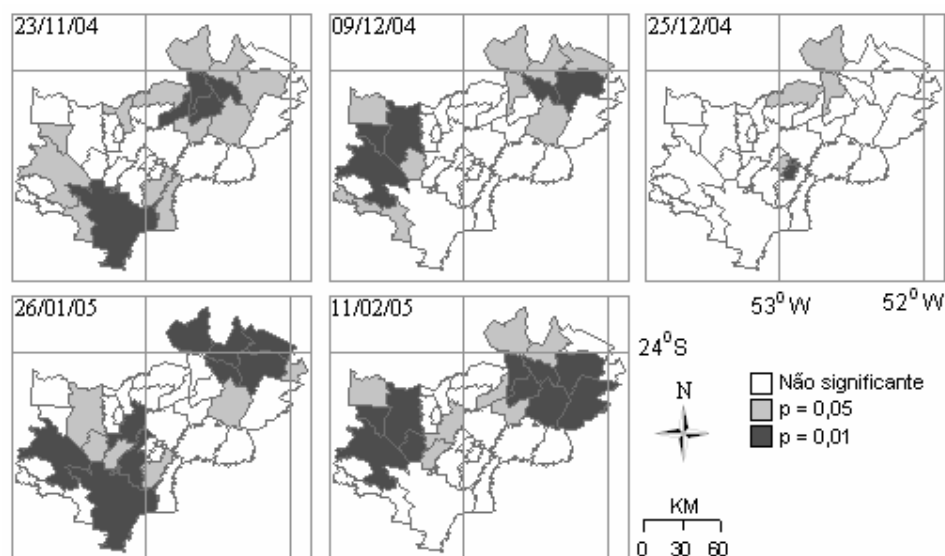
Para um melhor entendimento do índice local de associação espacial  $I_i$ , é conveniente apresentá-lo em um mapa temático que indique os municípios que apresentam associação espacial significativa, pois segundo Druck et al. (2004), estas regiões podem ser vistas como "bolsões" de não-estacionariedade, pois são áreas com dinâmica espacial

própria e que merecem análise detalhada. A Figura 24 apresenta o mapa de significância LISA do índice de vegetação NDVI da soja na safra 2004/2005 na região monitorada. A legenda deste mapa é formada por 3 classes, sendo que a primeira indica os municípios em que o índice local  $I_i$  não é significativo, a segunda indica os municípios em que o índice local  $I_i$  é significativo ao nível de 5 % de probabilidade e a última classe indica os municípios em que o índice local  $I_i$  é significativo ao nível de 1 % de probabilidade.



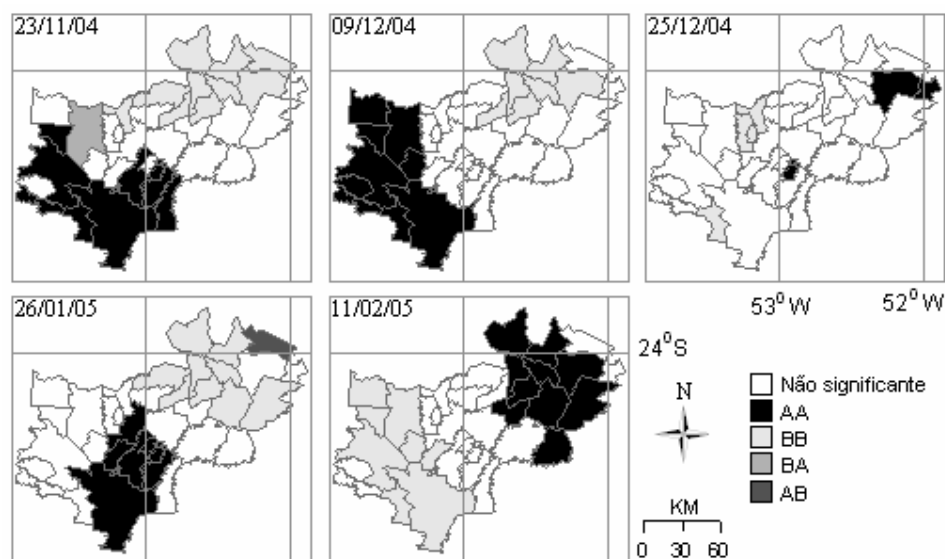
**Figura 24** – Mapa de significância LISA do índice de vegetação NDVI da soja.

Observando a Figura 24 é possível identificar o comportamento do índice de vegetação NDVI durante o período analisado. Em 23/11/2004, destaca-se a presença de um agrupamento de municípios que se inicia em torno de Cascavel (10) e se estende pela região noroeste até atingir a região nordeste. Este agrupamento indica os municípios que apresentam correlação espacial significativa. Observe que este agrupamento vai diminuindo até quase desaparecer em 25/12/04, período em que o desenvolvimento da cultura está em transição. Na última data monitorada (11/02/05), fica evidente a existência de dois grupos de municípios onde o índice de vegetação NDVI é correlacionado espacialmente. A Figura 25 apresenta o Mapa de significância LISA do índice GVI da soja na safra 2004/2005.



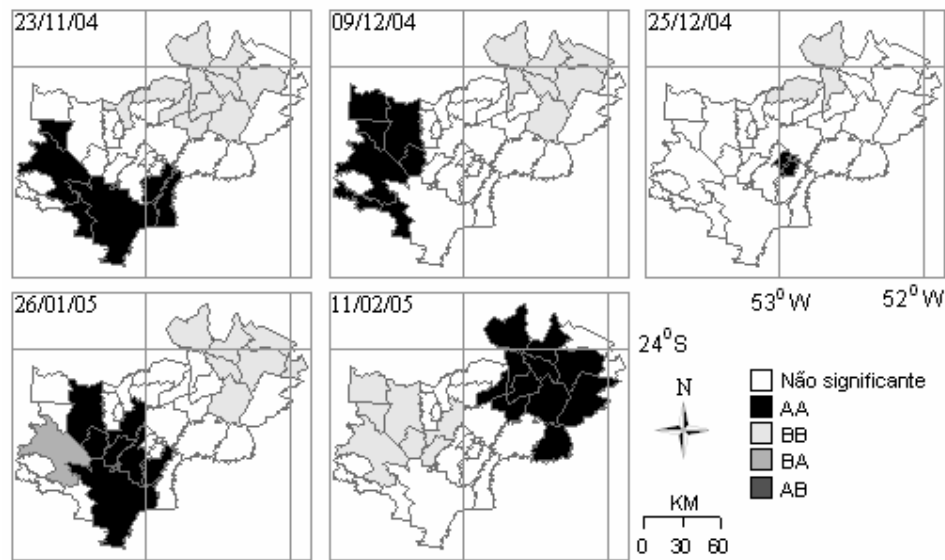
**Figura 25** – Mapa de significância LISA do índice de vegetação GVI da soja.

Destaca-se também, na Figura 25, o baixo número de municípios com índices significantes em 25/12/04 e a presença de dois agrupamentos de municípios com índices significantes nas demais datas monitoradas. Na Figura 26 destaca-se o mapa de significância LISA do índice NDVI agora classificado conforme a legenda do diagrama de espalhamento de Moran. Este mapa é conhecido como LISA *cluster map* e permite identificar qual tipo de correlação existe entre os municípios identificados como significantes.



**Figura 26** – LISA *cluster map* do índice NDVI da soja na safra 2004/2005.

Observando a data 23/11/04 na Figura 26 verifica-se que o aglomerado de municípios que apresenta correlação espacial significativa é formado por dois grupos com características diferentes. O primeiro grupo, localizado na região de Cascavel (10) e Toledo (33), indica os municípios em que o índice de vegetação NDVI é alto cercado por municípios onde o índice também é alto, e o outro grupo, na região de Araruna (2) e Tuneiras do Oeste (34), indica os municípios em que o índice de vegetação NDVI é baixo cercado por municípios com índice baixo. Este comportamento se inverte na última data monitorada, o que deixa claro que a soja não é cultivada no mesmo período na região Oeste do Paraná, pois enquanto na região de Cascavel (10) o desenvolvimento da cultura encontra-se no estágio de maturação, a soja na região Norte ainda está no estágio de enchimento das vagens, atingindo o seu pico vegetativo. A Figura 27 apresenta o LISA cluster map elaborado com base no índice GVI.



**Figura 27** – LISA *cluster map* do índice GVI da soja na safra 2004/2005.

O comportamento do índice de vegetação GVI é semelhante ao comportamento do índice de vegetação NDVI, pois observando a Figura 27 destacam-se as duas regiões em que o plantio da soja ocorre em diferentes datas. Em ambos os mapas das Figuras 26 e 27 destaca-se a pouca aparição de municípios classificados nas categorias referentes às correlações inversas, porém, quando isso ocorre, é interessante investigar com mais detalhes. Por exemplo, considerando a Figura 26 em 26/01/05, verifica-se que apesar da região norte apresentar um aglomerado de municípios com valores baixos de NDVI, o município de Peabiru (28) apresenta um valor elevado, indicando que neste município a soja encontra-se em um estágio mais avançado. Um análise semelhante é feita considerando a Figura 27 na data 26/01/05, pois apesar da região de Cascavel (10) apresentar altos valores

de GVI, verifica-se que a cidade de Toledo (33) apresenta um valor baixo, indicando que neste município a soja encontra-se em um estágio menos desenvolvido que nos demais.

### **3.5 CONCLUSÃO**

Os resultados mostraram que a estatística espacial de áreas é uma boa alternativa para estudar o comportamento espectral da cultura da soja, pois ela permite identificar regiões com características semelhantes, o que por sua vez serve de informação para os participantes da cadeia produtiva da soja, permitindo que produtores, cooperativas, consumidores e demais membros da cadeia produtiva da soja fiquem mais informados e possam traçar melhor seus planos para o fortalecimento da cultura.

## REFERÊNCIAS

ANSELIN, L. Local Indicators of Spatial Association--LISA. **Geographical Analysis** vol. 27, p. 93-115, 1995.

ANSELIN, L. The Moran scatterplot as ESDA tool to assess local instability in spatial association. In: Fisher, M.; Scholten, H. J.; Unwin, D. **Spatial Analytical Perspectives on GIS**. London: Taylor & Francis, p. 111-126, 1996.

ANSELIN, L. **Exploring Spatial Data with GeoDa: a Workbook**, Urbana-Champaign, 2005. 244 p.

BAILEY, T. C.; GATRELL, A. C. **Interactive Spatial Data Analysis**. Essex: Longman. 1995. 432 p.

CÂMARA, G.; MONTEIRO, A. M.; DRUK, S.; CARVALHO, M. **Análise espacial de dados geográficos**. Planaltina, DF: Embrapa Cerrados. 2004. 209 p.

CRIST E.P.; CICONE C.C. A Physically-Based transformation of Thematic Mapper data – The TM Tasseled Cap. **Transactions on Geosciences and Remote Sensing**, v.GE-22, n.3. p. 256 – 263. 1984.

KLERING, E. V.; CUSTÓDIO, M. S.; FONTANA D. C.; BERLATO, M. A. **Relação entre os perfis temporais de NDVI/MODIS da cultura do arroz irrigado, a insolação e o rendimento na região da Campanha do Rio Grande do Sul**. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 8, Florianópolis, 2007.

MERCANTE, E. **Dinâmica Espectral da Cultura da Soja ao Longo do Ciclo Vegetativo e sua Relação com a Produtividade na Região Oeste do Paraná**. Tese de Doutorado. FEAGRI UNICAMP. Campinas. 2007. 221 p.

MORAN, P.A.P. Notes on continuous stochastic phenomena. **Biometrika**, v.37, p. 17–23, 1950.

NETER, J.; WASSERMAN, W. **Applied Linear Statistical Models: regression, analysis of variance and experimental designs**. Illinois: Richard D.Irwin, Inc. 1974. 842 p.

ROUSE, J. W.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W. **Monitoring vegetation systems in the great plains with ERTS**. In: Earth Resources Technology Satellite – 1 Symposium, 3., Washington, D. C., 1973. Proceedings. Washington, D. C.: NASA. Goddard Space Flight Center, v. 1, p. 309-317. (NASA SP-351).1973.

SERRANO, R. M.; VALCARCE, E. V. **Técnicas econométricas para el tratamiento de datos espaciales: la econometría espacial**. 1 ed. Barcelona: Ediciones de la Universitat de Barcelona. 2000. 158 p.

SONG, C.; WOODCOCK, C. E. Monitoring Forest Succession with Multitemporal Landsat Images: Factors of Uncertainty. **Transactions on Geosciences and Remote Sensing**, v. 41, n.11, nov., p. 2557 – 2567. 2003.

**SUGAWARA, L. M. Avaliação de modelo agrometeorológico e imagens NOAA/AVHRR no acompanhamento e estimativa de produtividade da soja no estado do Paraná.** Dissertação (Mestrado em Sensoriamento Remoto) INPE, São José dos Campos, 2001. 181p.