



Universidade Estadual da Paraíba
Centro de Ciências e Tecnologia
Departamento de Estatística

Alidiane Oliveira Farias

Modelagem geoestatística para os macronutrientes nitrogênio e fósforo sob o umbuzeiro num sistema de pastagem.

Campina Grande - PB

Dezembro 2013

Alidiane Oliveira Farias

Modelagem geoestatística para os macronutrientes nitrogênio e fósforo sob o umbuzeiro num sistema de pastagem.

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de Bacharel em Estatística.

Orientador:
Ricardo Alves de Olinda

Campina Grande - PB
Dezembro 2013

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL – UEPB

F224m Farias, Alidiane Oliveira.
Modelagem geoestatística para os macronutrientes nitrogênio e fósforo sob o umbuzeiro num sistema de pastagem [manuscrito] / Alidiane Oliveira Farias. – 2013.
30 f. : il. color.

Trabalho de Conclusão de Curso (Graduação em Estatística) – Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2013.

“Orientação: Prof. Dr. Ricardo Alves de Olinda, Departamento de Estatística”.

1. Semivariograma. 2. Geoestatística. 3. Krikagem. 4. Dependência espacial. I. Título.

21. ed. CDD 310

Alidiane Oliveira Farias

Modelagem geoestatística para os macronutrientes nitrogênio e fósforo sob o umbuzeiro num sistema de pastagem

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Aprovado em: 16 / 12 / 2013

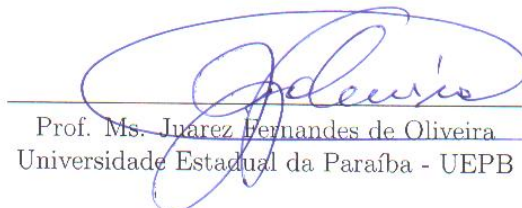
Banca Examinadora:



Prof. Dr. Ricardo Alves de Olinda
Orientador



Prof. Dr. João Gil De Luna
Universidade Estadual da Paraíba - UEPB



Prof. Ms. Juarez Fernandes de Oliveira
Universidade Estadual da Paraíba - UEPB

Dedicatória

Dedico esta vitória a Deus, razão da minha existência, aos meus pais, Alirio e Wilma, aos quais tenho muito respeito, e ao meu marido, Roberto, por ter acreditado em mim.

Agradecimentos

Agradeço primeiramente a Deus por ter me dado forças, nas noites que passei estudando e me dedicando até tarde, a este trabalho.

Aos meus pais, Alirio e Wilma, por sempre acreditarem em mim e sempre me incentivarem a ir em frente. Ao meu marido, Roberto, por estar sempre ao meu lado e acreditar que eu iria conseguir.

Ao meu orientador, Ricardo Olinda, pela sua paciência e pela competência e segurança que repassou um pouco do seu conhecimento para mim em prol da minha formação.

Aos amigos de classe que estiveram comigo suportando os dias maus, e compartilhando os dias bons nesses cinco anos, em especial a Francicleia, e a todos que começaram e por algum motivo não estão terminando, minhas saudades.

A todos que diretamente ou indiretamente contribuíram para que fosse possível a conclusão deste trabalho o meu mais sincero muito obrigado.

Resumo

Este trabalho teve por objetivo modelar por meio do enfoque Geoestatístico, a variabilidade espacial das propriedades físico-químicas do solo sob o umbuzeiro num sistema de pastagem. Os componentes analisados foram Nitrogênio(N) e Fósforo(P). Verificou-se a dependência espacial, por meio do semivariograma, e através do gráfico de envelope simulado, na sequência, ajustou-se diferentes funções de correlação, comparando-as por meio do critério Akaike(AIC). O índice de dependência espacial foi verificado utilizando-se a relação entre a variação estruturada e o patamar. Interpolações foram realizadas por meio da técnica da Krigagem. A precisão da Krigagem ficou num patamar aceitável o que implica na possibilidade de se obter boas informações sobre a região em estudos.

Palavras-chave: Semivariograma, krigagens, simulações.

Abstract

This study aimed to model by focusing Geostatistical, spatial variability of properties physicochemical soil under umbuzeiro a pasture system. The components were analyzed Nitrogen (N) and phosphorus (P). It was found spatial dependence, through the semivariogram, and through simulated envelope following graphic set different correlation functions, comparing them using the criteria Akaike (AIC). The index of spatial dependence was found using the relationship between the structured and the variation level. Interpolation was performed using the technique of Kriging. The accuracy of Kriging was an acceptable level which implies the possibility of obtaining good information about the area in studies.

Keywords: Semivariogram, kriging, simulations.

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 11
2	Fundamentação teórica	p. 13
2.1	Estacionaridade	p. 13
2.1.1	Hipótese de estacionaridade de primeira ordem	p. 13
2.1.2	Hipótese de Estacionariedade de Segunda Ordem	p. 14
2.2	Isotropia e anisotropia	p. 14
2.3	Variograma e semivariograma	p. 15
2.3.1	O semivariograma	p. 15
2.3.2	Modelos geoestatísticos gaussianos	p. 16
2.3.3	Funções de correlação para o semivariograma	p. 17
2.3.4	Critério de avaliação de modelos	p. 18
2.4	Interpolação de dados	p. 19
2.4.1	Krigagem	p. 19
3	Resultados e Discussão	p. 20
4	Conclusão	p. 28
	Referências	p. 29

Lista de Figuras

1	Gráfico descritivo do padrão espacial do umbuzeiro, variável nitrogênio.	p. 21
2	Transformação ótima de Box-Cox.	p. 22
3	Verificação de dependência espacial da variável nitrogênio da espécie umbuzeiro.	p. 22
4	Gráfico descritivo do padrão espacial do umbuzeiro, variável nitrogênio.	p. 22
5	Modelo ajustado ao semivariograma para as funções matérn $k=0,5$, matérn $k=1,0$, esférico, gaussiano e circular.	p. 23
6	Comparação dos valores observados com o valor de referência do teor médio de nitrogênio.	p. 24
7	Gráfico descritivo do padrão espacial do umbuzeiro, variável fósforo.	p. 25
8	Transformação ótima de Box-Cox.	p. 25
9	Verificação de dependência espacial através do envelope simulado.	p. 25
10	Gráfico descritivo do padrão espacial do umbuzeiro, variável fósforo, com os dados transformados.	p. 26
11	Modelo ajustado ao semivariograma para as funções matérn $k=0,5$, matérn $k=1,0$, esférico, gaussiano e circular.	p. 27
12	Comparação dos valores observados com o valor de referência do teor médio de fósforo.	p. 27

Lista de Tabelas

- 1 Estimativas dos parâmetros associado aos modelos por meio da máxima verossimilhança, assumindo-se a média da variável nitrogênio constante sobre a região de estudo. p. 23
- 2 Estimativas dos parâmetros associado aos modelos por meio da máxima verossimilhança, assumindo-se a média da variável fósforo constante sobre a região. p. 26

1 Introdução

Segundo Guimarães et.al.(2004), em várias áreas das ciências, as variáveis não apresentam um padrão de distribuição requerido pela estatística clássica. Os modelos da estatística clássica estão geralmente voltados para a verificação da distribuição de frequência dos dados, enquanto a geoestatística incorpora a interpretação da distribuição espacial, assim como a dependência espacial das amostras.

Este aspecto da geoestatística está intimamente associado com a distribuição estatística dos dados no espaço. Assim, os métodos geoestatísticos fornecem um conjunto de ferramentas para entender a uma aparente aleatoriedade dos dados, mas com possível estruturação espacial, estabelecendo-se, desse modo, uma função de correlação espacial. Esta função representa a base da estimativa da variabilidade espacial.

A geoestatística teve sua origem na área de mineração (KRINGE, 1951). Atualmente, geoestatística é justamente um nome associado com uma classe de técnicas utilizadas para analisar e inferir valores de uma variável distribuída no espaço ou no tempo. Tais valores são implicitamente assumidos ser correlacionados com outros, e o estudo de tal correlação é denominada de análise estrutural ou modelagem do semivariograma.

O semivariograma é a principal ferramenta geoestatística para diagnosticar a presença de dependência entre as unidades amostradas. Os semivariogramas servem para caracterizar a estrutura de continuidade espacial da característica avaliada por exigirem hipóteses de estacionariedade menos restritivas(hipótese intrínseca). Após a análise estrutural, inferências em localizações não amostradas são realizadas utilizando-se a técnica da krigagem.

Os atributos físicos e químicos do solo influenciam diretamente e indiretamente o crescimento e o desenvolvimento das plantas. Por esse motivo a avaliação da variabilidade espacial destes atributos tem-se tornado importante ferramenta na determinação de estratégias de manejo do solo. O sistema silvopastoril é a combinação intencional de árvores, pastagem e gado numa mesma área ao mesmo tempo e manejados de forma integrada, com o objetivo de incrementar a produtividade por unidade de área. Nesses sistemas, ocorrem interações em todos os sentidos e em diferentes magnitudes, apresentam grande

potencial de benefícios econômicos e ambientais para os produtores e para a sociedade. São sistemas multifuncionais, onde existe a possibilidade de intensificar a produção pelo manejo integrado dos recursos naturais evitando sua degradação, além de recuperar sua capacidade produtiva. Portanto, uma proposta deste trabalho foi utilizar os estimadores de máxima verossimilhança, tomando-se como valores iniciais os parâmetros do semivariograma, para avaliar a estrutura de dependência espacial em variáveis de fertilidade do solo. O presente trabalho foi desenvolvido com o objetivo específico de explorar e avaliar a união de ferramentas da estatística espacial, com o propósito de analisar as relações entre os aspectos da distribuição espacial de carbono orgânico e da biomassa de herbáceas em sistema silvopastoril, a fim de explicar se a espécie arborea consegue manter e/ou preservar o nitrogênio e o fósforo no solo, verificando também o crescimento do capim na região de estudos.

Neste trabalho serão abordados aspectos básicos da metodologia geoestatística para a análise espacial dos macronutrientes, nitrogênio e fósforo sob o umbuzeiro num sistema de pastagem, com ênfase na análise do semivariograma como ferramenta de determinação da dependência espacial e a técnica da krigagem.

2 Fundamentação teórica

De acordo com Landim (2003), a estatística espacial pode ser definida como uma técnica que busca descrever os padrões existentes nos dados, considerando-se o local onde as informações são coletadas. A estatística espacial divide-se em três grandes áreas, segundo o tipo de dados analisados: processos pontuais, geoestatística e dados de área. Processos pontuais são fenômenos expressos por meio de ocorrências identificadas como pontos localizados no espaço, ou seja, considera-se que a própria localização do fenômeno é o evento, que se interessa estudar. Na geoestatística os dados se distribuem continuamente em uma região e as localizações onde são feitas as medidas são fixas e podem estar regularmente ou irregularmente distribuídas. Os dados de área não apresentam uma variação contínua no espaço, diferentemente da geoestatística.

Conforme Guimarães et.al.(2002), A geoestatística é uma área da estatística aplicada que se caracteriza pela modelagem da continuidade espacial de fenômenos naturais. A geoestatística trata de variáveis regionalizadas, dando-se a devida importância às relações espaciais existentes entre as amostras, a partir das quais efetua-se as avaliações quantificando-se, espacialmente, os erros cometidos nas avaliações.

2.1 Estacionaridade

Segundo Guimarães et.al. (2004), uma variável aleatória é estacionária se o desenvolvimento dessa variável no tempo ou no espaço ocorrer de maneira mais ou menos homogênea, com oscilações aleatórias contínuas em torno de um valor médio, em que, nem a amplitude média e nem as oscilações mudam bruscamente no tempo ou no espaço.

2.1.1 Hipótese de estacionaridade de primeira ordem

Sejam $Z(s_i)$ e $Z(s_i + \mathbf{h})$ dois valores de uma variável regionalizada obtidos dos pontos s_i e $s_i + \mathbf{h}$, separados por uma distância \mathbf{h} . Sendo assim, um campo aleatório é estacionário de primeira ordem se para qualquer deslocamento do vetor \mathbf{h} , a esperança matemática é constante em relação a origem (GUIMARÃES ET.AL, 2004), conforme mostrado na

relação a seguir

$$E[Z(s_i)] = E[Z(s_i) + \mathbf{h}] = \mu(s_i), \quad (2.1)$$

em que s_i é a média dos valores amostrais.

De acordo com a Equação (2.1), se for tomado um vetor \mathbf{h} de distância euclidiana entre dois pontos, o qual apresenta módulo e direção, para qualquer \mathbf{h} tem-se que:

$$E[Z(s_i) - Z(s_i + \mathbf{h})] = 0. \quad (2.2)$$

Considerando-se que a diferença entre as duas variáveis aleatórias $Z(s_i) - Z(s_i + \mathbf{h})$ é uma variável aleatória, isto corresponde a afirmar que o primeiro momento desta variável aleatória é igual a zero.

2.1.2 Hipótese de Estacionariedade de Segunda Ordem

De acordo com Guimarães et.al.(2004), um processo é estacionário de segunda ordem se o segundo momento em relação à origem é constante. Tem-se então que a variância é constante independente da origem no espaço ou no tempo, ou seja,

$$E[Z^2(s_i)] = \mu(s_i);$$

$$Var[Z(s_i)] = E[Z^2(s_i)] - E[Z(s_i)]^2 = \mu(s_i) - [\mu(s_i)]^2.$$

Consequentemente, define-se a covariância como sendo, a esperança do produto que ocorre em s_i e s_j , com $\mathbf{h} = s_j - s_i$, definida com

$$C(s_i, s_j) = E[Z(s_i)Z(s_j)] - \mu^2(s_i). \quad (2.3)$$

2.2 Isotropia e anisotropia

Quando o fenômeno em estudo revela diferentes padrões de dependência espacial, ou seja, apresenta uma variabilidade que não é a mesma em todas as direções, o fenômeno em estudo é chamado de anisotrópico. De acordo com Guedes et al. (2008), a anisotropia é encontrada por meio de semivariogramas, que, construídos em diferentes direções, revelam uma privilegiada direção com um distinto padrão de dependência espacial.

Segundo Guedes L.P.C.(2008), os principais tipos de anisotropia estudados na geostatística são: geométrica, zonal e combinada. A anisotropia geométrica é aquela em que

existe uma direção com maior continuidade espacial, isto é, maior valor de alcance ϕ no semivariograma experimental em determinada direção; a anisotropia zonal, quando existe uma direção com maior valor de patamar ($\sigma^2 + \tau^2$) nos semivariogramas experimentais em relação às demais direções; e a anisotropia combinada, quando houver determinadas direções com diferentes valores de alcance e patamar nos semivariogramas experimentais.

2.3 Variograma e semivariograma

Segundo Guerra (1988), o variograma é utilizado para verificar a presença de correlação espacial entre as unidades amostrais. Existem três tipos de variogramas: experimental ou observado (é o variograma obtido a partir do conjunto de amostras derivadas da amostragem realizada), verdadeiro (real mas sempre desconhecido) e teórico (utilizado para o ajuste do modelo). O variograma permite representar quantitativamente a variação de um fenômeno regionalizado no espaço, que é fundamental para as técnicas de krigagem.

2.3.1 O semivariograma

De acordo com Guimarães et.al. (2002), o semivariograma é definido como

$$\gamma(\mathbf{h}) = \frac{1}{2} \text{Var}[Z(s_i) - Z(s_i + \mathbf{h})]. \quad (2.4)$$

Note que $\text{Var}[Z(s_i) - Z(s_i + \mathbf{h})]$ é a variância dos dados separados por uma distância (\mathbf{h}) , mas, na expressão acima, esta variância está sendo dividida por dois, então se utiliza o prefixo 'semi' para distinguir da variância e daí vem o nome semivariância para $\gamma(\mathbf{h})$ e semivariograma para o gráfico de $\gamma(\mathbf{h})$ em função de (\mathbf{h}) . A função de semivariância $\gamma(\mathbf{h})$ é definida como sendo a esperança matemática do quadrado da diferença entre os valores de pontos no espaço, separados por um vetor de distância euclidiana (\mathbf{h}) , conforme a seguinte equação

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

e pode ser estimado por

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2} \frac{\sum_{i=1}^n [Z(s_i) - Z(s_i + \mathbf{h})]^2}{n(\mathbf{h})}, \quad (2.6)$$

em que $n(\mathbf{h})$ é o número de pares de valores medidos $Z(s_i), Z(s_i + \mathbf{h})$, separados por um vetor (\mathbf{h}) .

Segundo Guimarães et.al. (2002), analisando-se a expressão da função semivariância, pode-se imaginar que quanto mais próximos estiverem os pontos amostrados, maior será a semelhança entre eles e, portanto, menor a semivariância; e quanto mais distantes estiverem os pontos amostrados, menor será a semelhança e, conseqüentemente, maior a dispersão (variância). O gráfico de $\gamma(\mathbf{h})$ versus (\mathbf{h}) representa o semivariograma, que permite obter a estimativa do valor de semivariância para as diferentes combinações de pares de pontos e assim analisar o grau de dependência espacial da variável estudada e definir os parâmetros necessários para a estimativa de suas características em locais não amostrados (SOUZA, 1999).

2.3.2 Modelos geoestatísticos gaussianos

Segundo Diggle e Ribeiro Jr. (2007), considera-se que em alguma região S exista um campo aleatório gaussiano Z latente, ou seja, o processo existe mas não é observável, é necessário fazer uma amostragem de n localizações espaciais dentro da região S e observar valores de atributo de interesse nas localizações amostradas. Sendo assim, existe um vetor $\mathbf{Y}(\mathbf{s})$ de $n \times 1$ valores observados em $\mathbf{s} = s_i; i = (1, \dots, n)$, sob o pressuposto de gaussianidade, pode ser modelado como

$$\mathbf{Y}(\mathbf{s}) = \mu(\mathbf{x}) + Z(\mathbf{s}) + \boldsymbol{\varepsilon}, \quad (2.7)$$

em que $\mu(\mathbf{x}) = \mathbf{X}\boldsymbol{\beta}$ sendo \mathbf{X} uma matriz $n \times q$ contendo $q - 1$ possíveis covariáveis, $\boldsymbol{\beta}$ um vetor $q \times 1$ de parâmetros associados a \mathbf{X} , $Z(\mathbf{s})$ um campo aleatório gaussiano que possui vetor de médias $n \times 1$ nulo e matriz de covariâncias $\boldsymbol{\Sigma}$ de dimensão $n \times n$, sendo que cada elemento $\Sigma_{i,j}$ é igual a $Cov(Z(s_i); Z(s_j))$, para todo s_i e s_j pertencentes a \mathbf{s} , e $\boldsymbol{\varepsilon}$ um vetor $n \times 1$ de ruídos brancos, que por suposição são independentes e identicamente distribuídos com distribuição de probabilidade normal com média zero e desvio padrão σ .

Segundo Diggle e Ribeiro Jr. (2007), por simplicidade matemática, normalmente utiliza-se o logaritmo da função de verossimilhança para fazer a estimação, que associada a (1) tem a seguinte forma

$$l(\boldsymbol{\theta}; \mathbf{Y}(s)) = -0.5(n \ln(2\pi) + \ln(|\boldsymbol{\Sigma}_Y|) + (\mathbf{Y}(s) - \mathbf{X}\boldsymbol{\beta})^t \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y}(s) - \mathbf{X}\boldsymbol{\beta})). \quad (2.8)$$

No contexto de geoestatística, pode-se adotar a utilização da reparametrização $\nu = \tau/\sigma$, a qual facilita a estimação de $\mu(s_i)$. O vetor de parâmetros a ser estimado passa a ser $\boldsymbol{\theta}^* = (\beta, \sigma^2, \phi, \nu^2)'$ e $\boldsymbol{\Sigma}_Y$ pode ser escrita como $\sigma^2 \mathbf{V}$, em que, \mathbf{V} é uma matriz $n \times n$ que depende apenas de ν e ϕ . Sendo que existem formas analíticas apenas para os estimadores

de β e σ^2 , além das formas fechadas para os estimadores, é possível encontrar a matriz de informação de Fisher observada para os mesmos e portanto a matriz de covariâncias associada a $(\hat{\beta}, \hat{\sigma}^2)$ (DIGGLE; RIBEIRO Jr., 2007).

2.3.3 Funções de correlação para o semivariograma

Conforme Mello (2004), a partir do semivariograma é possível ajustar uma função matemática que expressa a estrutura de dependência espacial da característica avaliada, é necessário ajustá-lo a um modelo teórico que deve fornecer soluções estáveis para o estimador (krigagem). Isto quer dizer que as covariâncias têm de ser definidas positivamente. A condição de não negatividade do modelo, limita o conjunto de funções usadas na modelagem do semivariograma experimental.

Função de correlação gaussiana

Segundo Kruger (2005), o modelo gaussiano é um modelo transitivo, muitas vezes usado para modelar fenômenos extremamente contínuos.

$$\rho(\mathbf{h}) = \tau^2 + \sigma^2 \left[1 - \exp \left[-3 \left(\frac{\mathbf{h}}{\phi} \right)^2 \right] \right] \quad 0 \leq \mathbf{h} \leq \pi.$$

Função de correlação exponencial

$$\rho(\mathbf{h}) = \tau^2 + \sigma^2 \left[1 - \exp \left[-3 \left(\frac{\mathbf{h}}{\phi} \right) \right] \right] \quad 0 \leq \mathbf{h} \leq \pi.$$

Segundo Kruger (2005), a função de correlação exponencial frequentemente produz um ajuste qualitativamente razoável à estrutura de correlação dos dados, mas as previsões baseadas nesse modelo tendem a ser pouco robustas a pequenas variações sobre o modelo escolhido. Tanto na função exponencial quanto no modelo de gaussiano, π é a distância máxima na qual o semivariograma é definido e nestes modelos o patamar $\sigma^2 + \tau^2$ é atingido apenas assintoticamente. O parâmetro ϕ é determinado visualmente como a distância após a qual o semivariograma se estabiliza.

Função de correlação Esférica

$$\rho(\mathbf{h}) = \begin{cases} \tau^2 + \sigma^2 \left[\frac{3}{2} \left(\frac{\mathbf{h}}{\phi} \right) - \frac{1}{2} \left(\frac{\mathbf{h}}{\phi} \right)^3 \right] & \text{se } 0 \leq \mathbf{h} \leq \phi; \\ \tau^2 + \sigma^2 & \text{se } \mathbf{h} > \phi. \end{cases}$$

Segundo Kruger (2005), função de correlação esférica depende apenas do parâmetro ϕ , não há flexibilidade para a forma. Adicionalmente, a lógica de sua aplicação é questionável, em problemas no espaço bidimensional.

Função de correlação Circular

$$\rho(\mathbf{h}) = \begin{cases} \frac{2}{\pi} (\text{sen}^{-1}(\mathbf{h} + \sqrt{1 - \mathbf{h}^2})); \\ \rho(\mathbf{h}) = 1 \end{cases}, \quad \mathbf{h} > 1.$$

Função de correlação Matérn

Segundo Kruger (2005), a função de correlação Matérn, proposta por Bertil Matérn em 1986 é provavelmente a melhor escolha para função de correlação de uso geral por sua flexibilidade e simplicidade (apenas dois parâmetros) e possui a seguinte forma

$$\rho(\mathbf{h}) = \frac{2^{2-\nu}}{\Gamma(\nu)} \left(\frac{\mathbf{h}}{\phi}\right)^\nu K_\nu\left(\frac{\mathbf{h}}{\phi}\right), \quad \mathbf{h} > 0,$$

em que $K_\nu(\cdot)$ é a função *Bessel* modificada de terceiro grau de ordem $\nu > 0$ e $\Gamma(\cdot)$ é a função *Gamma*.

2.3.4 Critério de avaliação de modelos

O princípio do Critério de informação de Akaike é selecionar uma combinação de variáveis exploratórias a modelos para a função de correlação que minimize o valor do *AIC*. É importante observar que, em muitas situações clássicas, tais como regressão linear ou em modelos de séries temporais, o *AIC* não é uma condição consistente para a seleção de modelos. Ou seja, como o crescente aumento do tamanho das amostras, há uma alta probabilidade de que um modelo selecionado pelo *AIC* não corresponda ao verdadeiro modelo (OLINDA, 2008). O critério de informação de Akaike é expresso por

$$AIC = -2 \times (L(\boldsymbol{\theta}; y)) + 2p,$$

em que, $L(\boldsymbol{\theta}; y)$ é o logaritmo da função de verossimilhança maximizado e p é o número de parâmetros. Segundo este critério, o melhor modelo é o que possui menor valor de *AIC*.

2.4 Interpolação de dados

ZIMBACK (2003), define a técnica da construção dos mapas de isolinhas, onde são geradas estimativas de dados de pontos não amostrados a partir de pontos amostrados, denomina-se **interpolação de dados**. Existem diferentes métodos de interpolação de dados, entretanto, a maioria desses métodos não fornecem o algoritmo dos erros associados aos resultados obtidos o que, efetivamente, é fornecido pelo método geoestatístico da Krigagem, ou seja, um modelo contínuo de variação espacial.

2.4.1 Krigagem

Para a aplicação da krigagem assume-se, que sejam conhecidas as realizações $z(s_1)$, $z(s_2)$, ..., $z(s_n)$ da variável $Z(s)$, nos locais s_1, s_2, \dots, s_n ; que o semivariograma da variável já tenha sido determinado; e que o interesse seja estimar um valor \hat{s} na posição s_0 . A Krigagem Simples é a mais comum das estimações usadas na geoestatística. A Krigagem Ordinária, que é a variação mais utilizada da Krigagem simples, descrita por Trangmar et al. (1985), como o valor interpolado $\hat{Z}(s_0)$ de uma variável regionalizada Z , num local s_0 pode ser determinada por

$$\hat{Z}(s_0) = \sum_{i=1}^n w_i Z(s_i),$$

em que, n é o número de amostras de $Z(s_i)$ envolvidas na estimativa de $\hat{Z}(s_0)$, e w_i são os pesos associados a cada valor medido em $Z(s_i)$.

3 Resultados e Discussão

Dar-se início aos resultados discutindo-se as análises obtidas para a variável nitrogênio em espécies de umbuzeiro. O estudo foi realizado em 1996, em uma fazenda de gado de corte pertencente à Agropecuária Jaçanã, situada na Cidade de Custódia-PE. ($8^{\circ} 13.70$ ao Sul e $37^{\circ}44.70$ a Oeste). A precipitação média de chuva no local é 740 milímetros e a temperatura média anual é de $26^{\circ} C$ (dados a partir dos registros da fazenda). A vegetação dentro da área da Agropecuária, inicialmente consistia em uma vegetação, de caatinga semidecídua seca, a qual foi desmatada por tratores em 1984, e em seguida foram retirados os entulhos como tocos e pedras que porventura ficaram no local dentre os 5.000 ha (MENEZES; SALCEDO, 1999).

Em uma área da fazenda, com cerca de 2.000 ha, foi preservada a espécie nativa da caatinga, *Spondias tuberosa* Arruda (Umbuzeiro), com uma distância média entre as árvores variando cerca de 30m a 40m. A idade das árvores preservadas neste local foram estimadas em pelo menos 50 anos de idade (Dados levantados junto aos trabalhadores da fazenda). Segundo Tiessen et al. (2003), as espécies arbóreas nativas representam situações diferentes em termos de diferenças de potencial no solo e características herbáceas em cada sistema, as árvores nativas já estavam presentes quando a pastagem foi estabelecida. A lotação de animais, nos pastos da fazenda situava-se, na época do estudo, em torno de 0; 17% animais por ha, bem abaixo da densidade normal em pastagens de capim buffel na região. Segundo Menezes e Salcedo (1999), dentro das áreas de estudo, foi selecionado *S. tuberosa*(Umbuzeiro). Deu-se preferência por árvores maduras isoladas uma das outras. Para cada árvore, estabeleceram-se 81 pontos de amostragem regularmente distribuídos, onde a espécie arborea sempre está no centro da área.

A coleta de dados foi realizada por pesquisadores e estudantes do Departamento de Energia Nuclear da Universidade Federal Rural de Pernambuco-UFRPE. As amostras de solo foram coletadas na profundidade de 0-15cm, com as devidas coordenadas x e y , as amostras foram secas ao ar e passadas numa peneira de 2mm, analisou-se quanto ao total de nitrogênio(N). Do mesmo modo, as amostras de fósforo(P), foram tomadas em cada posição da grade ($0,7cm \times 0,7cm$), toda a biomassa viva em pé foi cortada ao nível do solo e colocadas em sacos de papel, as amostras foram secas em estufa durante 48 horas

a 60°C (MENEZES; SALCEDO, 1999).

Pode-se observar por meio da Figura 1 os quartis da variável nitrogênio, isto é, “+” 1º quartil, “Δ” 2º quartil, “o” 3º quartil, “×” 4º quartil. O gráfico do canto superior direito e do canto inferior esquerdo, apresentam a dispersão dos dados em torno da média geral.

Na sequência, por meio da Figura 2, utilizando-se a transformação ótima do Box-Cox, pode-se observar a necessidade de transformação para a variável em estudo, tendo em vista que o valor 1(um) não está contido no intervalo que maximiza o logaritmo da função de máxima verossimilhança para o valor de λ , ao nível de significância de 0,05. Neste caso, para a variável nitrogênio, foi encontrado o valor de $\lambda = -0,06$.

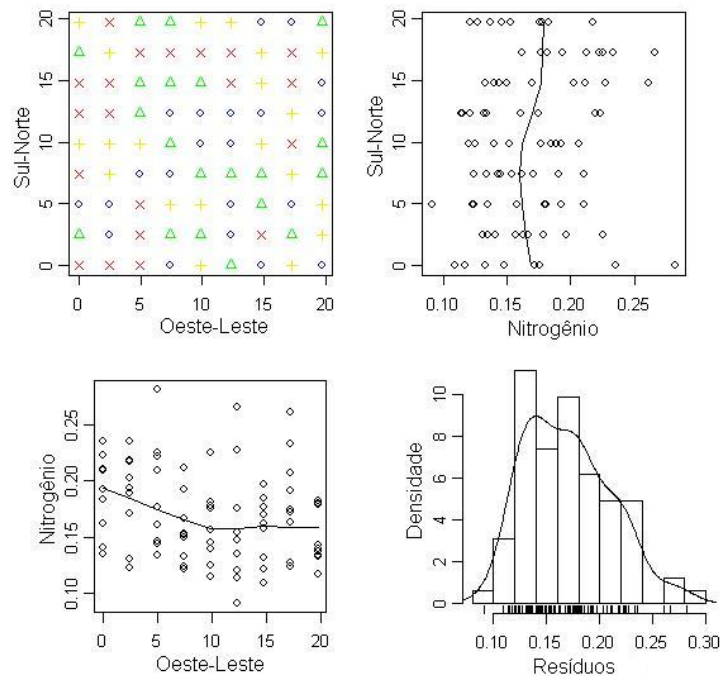


Figura 1: Gráfico descritivo do padrão espacial do umbuzeiro, variável nitrogênio.

A existência de dependência espacial pode ser verificada por meio da Figura 6, utilizando-se envelopes simulados e variogramas empíricos, nessa pesquisa foram utilizadas 1.000 simulações. Conforme Diggle e (RIBEIRO Jr. 2007), se existir pelo menos um ponto fora do envelope simulado, há evidências para a existência de dependência espacial do nitrogênio no solo. A dependência espacial observada para os valores do nitrogênio mostra que as análises da estatística clássica, em que as observações são consideradas aleatórias e independentes espacialmente, devem ser substituídas por análises espaciais, que levam em consideração as relações entre observações vizinhas.

Após a transformação dos dados do teor de nitrogênio no solo, percebe-se que o gráfico

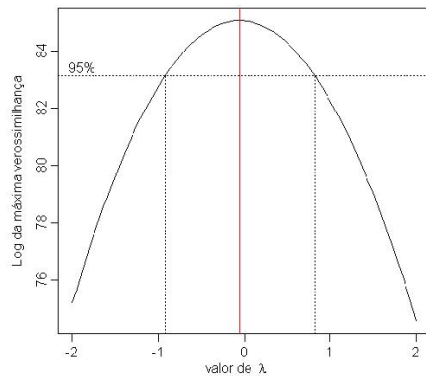


Figura 2: Transformação ótima de Box-Cox.

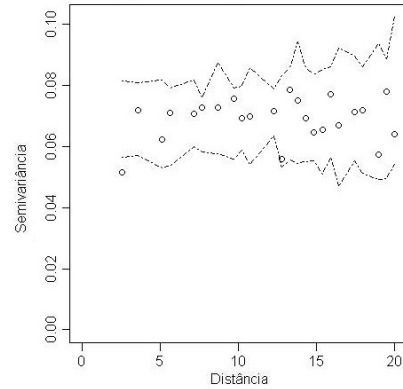


Figura 3: Verificação de dependência espacial da variável nitrogênio da espécie umbuzeiro.

do canto inferior direito da Figura 3 condiz com a hipótese de que os resíduos aproximam-se da distribuição normal. Pode-se observar por meio da Tabela 1 as estimativas dos parâmetros dos modelos ajustados por meio da máxima verossimilhança, considerando-se a média constante sob a região de estudo.

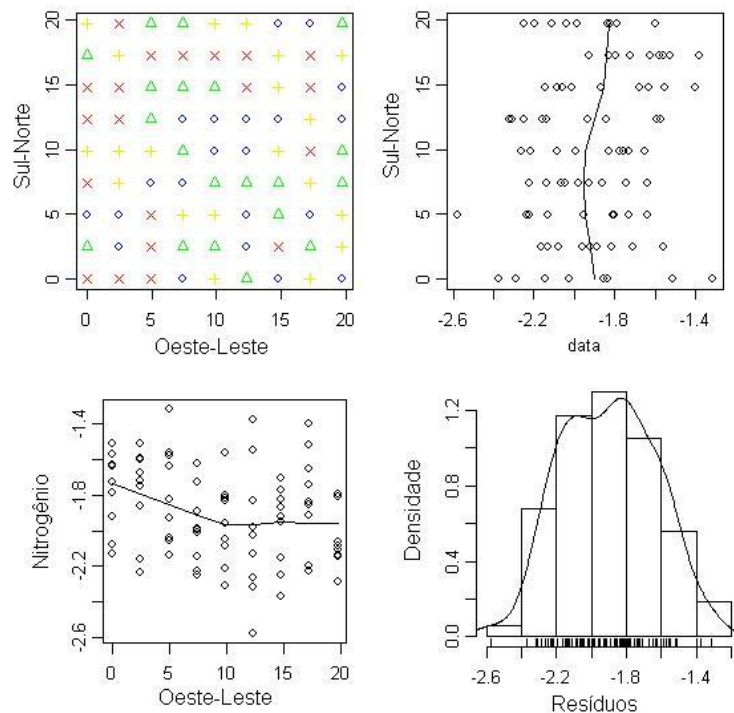


Figura 4: Gráfico descritivo do padrão espacial do umbuzeiro, variável nitrogênio.

O próximo passo é selecionar o melhor modelo que se ajusta ao semivariograma. Comparando-se as funções de correlação ajustadas ao semivariograma (Tabela 1), para a variável nitrogênio onde está inserida a espécie umbuzeiro, por meio do critério de AIC,

Tabela 1: Estimativas dos parâmetros associado aos modelos por meio da máxima verossimilhança, assumindo-se a média da variável nitrogênio constante sobre a região de estudo.

Função de correlação	$\hat{\beta}$	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$	AIC	IDE
matérn kappa = 0,5	-1,9057	0,0000	0,0692	1,6948	-295,5	0%
matérn kappa = 1,0	-1,9070	0,0000	0,0695	1,1620	-296,4	0%
esférico	-1,9061	0,0000	0,0672	4,3013	-298,7	0%
gaussiano	-1,9071	0,0000	0,0696	2,1537	-298,7	0%
circular	-1,9055	0,0000	0,0673	3,8130	-299,6	0%

a que melhor se ajustou foi a função de correlação circular.

Para verificar o grau de dependência espacial, entre as características de fertilidade do solo com relação ao nitrogênio, utilizou-se da relação efeito pepita τ^2 e o patamar $\sigma^2 + \tau^2$. Observa-se que o nitrogênio apresenta uma forte dependência espacial, que pode ser descrita pelo modelo circular com alcance de aproximadamente $4m$, ou seja, amostras de nitrogênio de solo selecionadas à distâncias inferiores estão correlacionadas entre si. A análise variográfica apresentada na Figura 5, apresenta evidências de que todas as características estudadas nos procedimentos até aqui descritos, apresentam estruturas espaciais bem definidas, ou seja, existe uma função estrutural com semivariância de comportamento modelável.

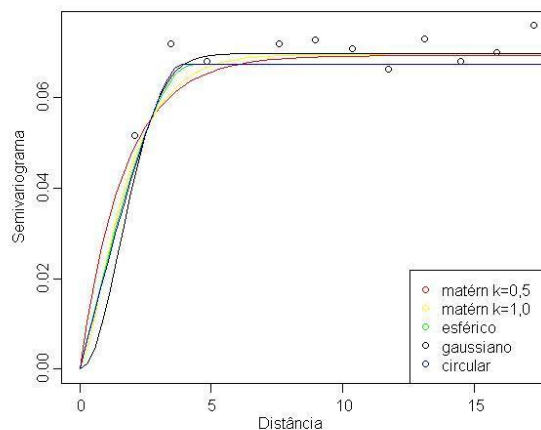


Figura 5: Modelo ajustado ao semivariograma para as funções matérn $k=0,5$, matérn $k=1,0$, esférico, gaussiano e circular.

Com o modelo estabelecido, para o semivariograma, foram calculados as previsões espaciais para o índice de nitrogênio no solo, levando-se em consideração os valores observados. Observa-se por meio da Figura 6 o mapa preditivo do campo aleatório para o efeito do nitrogênio sobre o macronutriente nitrogênio num sistema de pastagem. Levando-

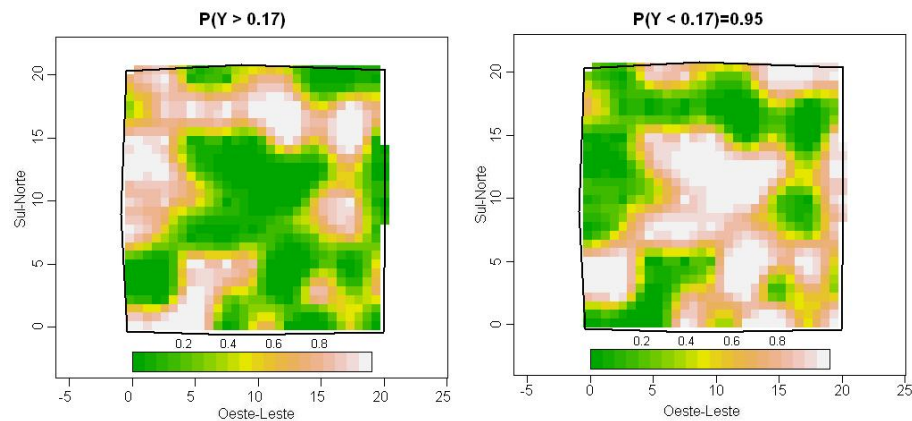


Figura 6: Comparação dos valores observados com o valor de referência do teor médio de nitrogênio.

se em consideração o valor do teor do nitrogênio ($P > 0,17$), pode-se observar uma maior concentração de nitrogênio a medida que afasta-se da copa da árvore do umbuzeiro.

Da mesma forma com a variável nitrogênio, segue-se as análises para a variável fósforo. Pode-se observar por meio da figura 7 os quartis da variável fósforo, ou seja, “+” 1º quartil, “ Δ ” 2º quartil, “o” 3º quartil, “ \times ” 4º quartil. O gráfico do canto superior direito e do canto inferior esquerdo, apresentam a dispersão dos dados em torno da média geral.

Na sequência, por meio da Figura 8, utilizando-se a transformação ótima do Box-Cox pode-se observar a necessidade de transformação para a variável em estudo, tendo em vista que o valor 1 (um) não está contido no intervalo que maximiza o logaritmo da função de verossimilhança para o valor de λ , nível de significância ao de 0,05. Neste caso, para a variável fósforo, foi encontrado o valor de $\lambda = -0,67$.

A existência de dependência espacial pode ser verificada por meio da Figura 9 utilizando-se envelopes simulados e variogramas empíricos, nessa pesquisa foram utilizadas 1.000 simulações. Conforme (DIGGLE e RIBEIRO Jr. 2007), se existir pelo menos um ponto fora do envelope simulado, há evidências para a existência de dependência espacial do fósforo no solo. A dependência espacial observada para os valores do fósforo evidencia que as análises estatística clássica, em que as observações são consideradas aleatórias e independentes espacialmente, devem ser substituídas por análises espaciais, que levam em consideração as relações entre observações vizinhas.

Após a transformação dos dados do teor de fósforo no solo, percebe-se que o gráfico do canto inferior direito da Figura 10 condiz com a hipótese de que os resíduos aproximam-se da distribuição normal. Pode-se observar por meio da Tabela 2 as estimativas os parâmetros dos modelos ajustados por meio da máxima verossimilhança, considerando-se

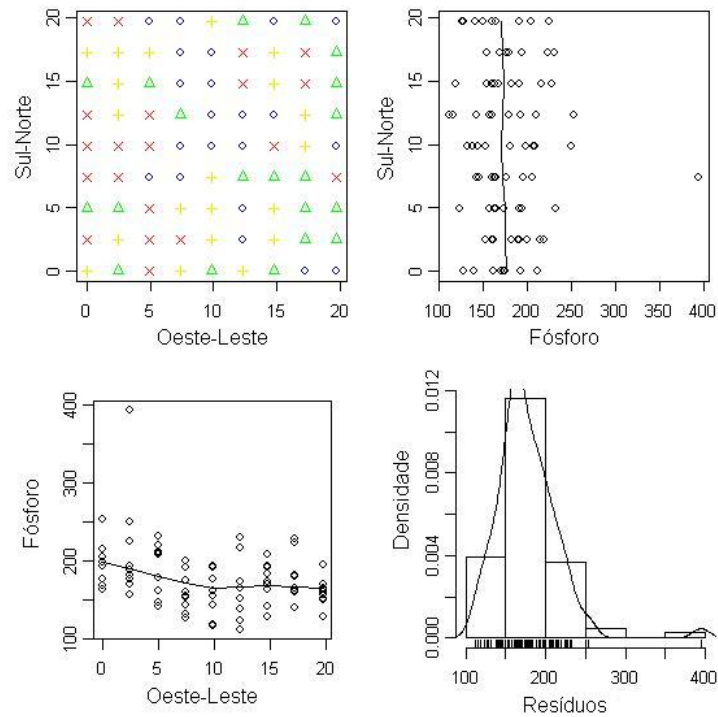


Figura 7: Gráfico descritivo do padrão espacial do umbuzeiro, variável fósforo.

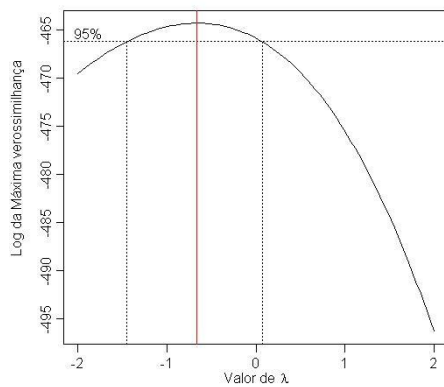


Figura 8: Transformação ótima de Box-Cox.

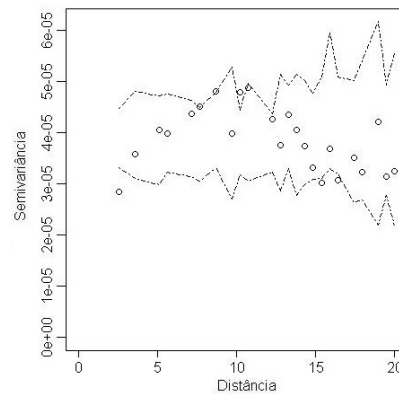


Figura 9: Verificação de dependência espacial através do envelope simulado.

a média constante sob a região de estudo.

O próximo passo é selecionar o melhor modelo que se ajusta ao semivariograma. Comparando-se as funções de correlação ajustadas ao semivariograma (Tabela 2), para a variável fósforo onde está inserida a espécie umbuzeiro, por meio do critério de AIC, a que melhor se ajustou foi a função de gaussiana.

Para verificar o grau de dependência espacial, entre as características de fertilidade

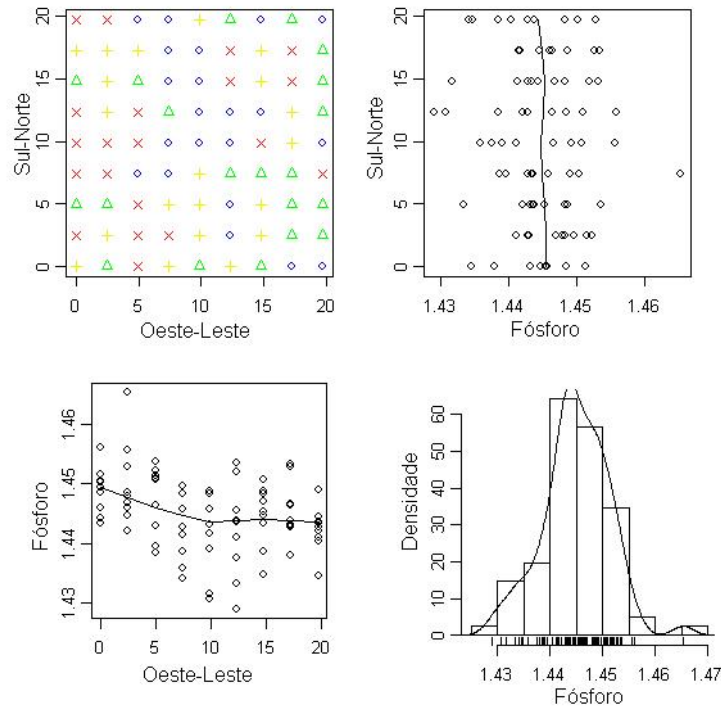


Figura 10: Gráfico descritivo do padrão espacial do umbuzeiro, variável fósforo, com os dados transformados.

Tabela 2: Estimativas dos parâmetros associado aos modelos por meio da máxima verossimilhança, assumindo-se a média da variável fósforo constante sobre a região.

função de correlação	$\hat{\beta}$	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$	AIC	IDE
matérn kappa = 0,5	1,445	0,000	0,000	1,919	798,9	0%
matérn kappa = 1,0	1,445	0,000	0,000	1,235	798,5	0%
esférico	1,445	0,000	0,000	4,721	798,4	0%
gaussiano	1,445	0,000	0,000	2,361	798,4	0%
circular	1,445	0,000	0,000	4,359	798,4	0%

do solo com relação ao fósforo, utilizou-se da relação efeito pepita τ^2 e o patamar $\sigma^2 + \tau^2$. Nota-se que o fósforo apresenta uma forte dependência espacial, que pode ser descrita pelo modelo gaussiano com alcance de 2,361m, ou seja, os teores de fósforo no solo obtidas de amostras a uma distância inferior a 2,361m, correlacionadas entre si. A análise variográfica apresentada na Figura 11, mostrou que todas as características estudadas nos procedimentos até aqui descritos, apresentam estruturas espaciais bem definidas, ou seja, existe uma função estrutural com semivariância de comportamento modelável.

Com o modelo estabelecido, para o semivariograma, foram calculadas as previsões espaciais para o índice de fósforo no solo, levando-se em consideração os valores observados. Observa-se por meio da Figura 12, o mapa preditivo do campo aleatório para o efeito do fósforo sobre o macronutrientes fósforo num sistema de pastagem. Levando-se em consi-

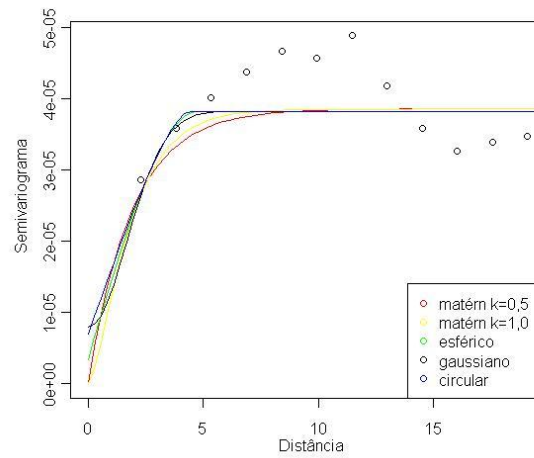


Figura 11: Modelo ajustado ao semivariograma para as funções matérn $k=0,5$, matérn $k=1,0$, esférico, gaussiano e circular.

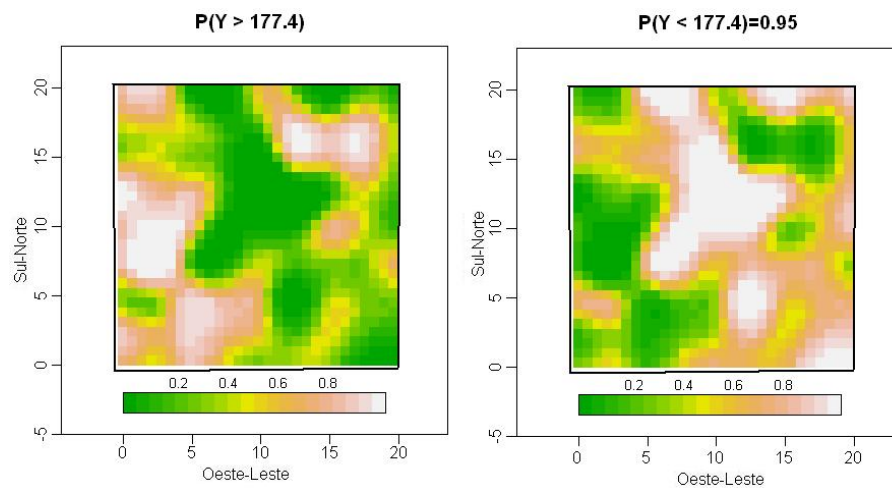


Figura 12: Comparação dos valores observados com o valor de referência do teor médio de fósforo

deração o valor do teor do fósforo ($P > 17,4$), pode-se observar uma maior concentração de fósforo a medida que se afasta-se da copa da árvore do umbuzeiro.

4 Conclusão

A partir do semivariograma é possível verificar e modelar a dependência espacial através de funções de correlação que expressam a estrutura de dependência espacial das características avaliadas. As funções de correlação utilizadas nesse trabalho foram: gaussiana, matérn, esférica e circular.

O modelo que melhor se ajustou aos dados foi o modelo circular, para a variável nitrogênio, e para a variável fósforo foi o modelo gaussiano, o qual foi selecionado pelo método de seleção de *AIC*, método este que seleciona como melhor modelo, aquele que apresentar o menor valor para a estatística de *AIC*. Observou-se que houve uma forte dependência espacial, para as duas variáveis estudadas, (nitrogênio e fósforo), o que justifica-se também pelo baixo valor do efeito pepita.

A técnica da krigagem é considerada uma boa metodologia de interpolação de dados. Ela utiliza dado tabular e sua posição geográfica para calcular as interpolações. A precisão da krigagem ficou em um patamar aceitável o que implica em boas informações sobre a região em estudo.

Referências

- DIGGLE, P. J.; RIBEIRO JÚNIOR, P. J. **Model Based Geostatistics**. SINAPE: Caxambu. Anais. São Paulo: Associação Brasileira de Estatística, 2000. 140p.
- DIGGLE, P. J.; RIBEIRO Jr., P. J. **Model-Based geostatistics**. New York: Springer, 2007. 230p.
- DRUCK, S.; CARVALHO, M. S.; CÂMARA, G.; MONTEIRO, A. M. V. **Análise Espacial de Dados Geográficos**. Brasília: EMBRAPA 2004. 89p.
- GUEDES, L. P. C.; URIBE, A. M.; JOHANN, J. A.; SOUZA, E. G. **Anisotropia no Estudo da Variabilidade Espacial de Algumas Variáveis Químicas do Solo**. 2008. 158p.
- GUERRA, P. A. G. **Geoestatística operacional**. Brasília, MME/DNPM, 1988. 145p.
- GENÚ, A. M.; **Geoestatística Multivariada**. 2004. 350p.
- GUIMARÃES, A.; OLIVEIRA. **Modelagem Geológica e Geoestatística de Reservatório da Bacia de Campos**, RJ, 2002. 198p.
- GUIMARÃES, E. CARVALHO.; **Geoestatística Básica e Aplicada**, Universidade Federal de Uberlândia Faculdade de Matemática, Núcleo de Estudos Estatísticos e Biométricos, MG, 2004. 120p.
- KRIGE, D.G. A statistical approach to some basic mine valuation problems on the Witwatersrand. **Journal of the Chemistry, Metal. and Mining Society of South Africa**, v.52, p.119-139, 1951.
- KRUGER, M. C. **Análise geoestatística de dados meteorológicos do estado do paran  utilizando um software livre**, 2005. 96p.
- LANDIM, P. M. B. **An lise estat stica de dados geol gicos**. [S.l.]: 2 edi o revista e ampliada. S o Paulo, 2003. 89p.
- MATHERON, G. Principles of geostatistics. **Economic Geology**, v.58, p.1246-1266, 1963.
- MENEZES, R. S. C.; SALCEDO, I. H. Influence of tree species on the herbaceous understory and soil chemical characteristics in a silvopastoral system in semi-arid northeastern brazil. **Revista Brasileira de Ci ncias do Solo**, v.1, p1-19, 1999.
- MELLO, J. M. de. **Geoestat stica aplicada ao invent rio florestal**. 2004. Dispon vel em: <http://www.teses.usp.br/teses/disponiveis/11/11150/tde-06122004-100612/pt-br.php>.

OLINDA, R. A. **Métodos para análise de Independência entre marcas e pontos em Processos Pontuais Marcados**. Dissertação (Mestrado em estatística e experimentação agrônômica), ULFA, 2008. 104p.

SOUZA, L.C. de. **Variabilidade espacial da salinidade de um solo aluvial no semi-árido paraibano**. Campina Grande, Dissertação (Mestrado) - Universidade Federal da Paraíba. 1999. 77p.

TIESSEN, H.; MENEZES, R.; SALCEDO, I.; B.WICK. Organic matter transformations and soil fertility in a treed pasture in semiarid ne brazil. **Plant and Soil**, v.1, p.1-22, 2003.

ZIMBACK, C.R.L. **Análise espacial de atributos químicos de solos para fins de mapeamento da fertilidade**. Tese de Livre-Docência (Livre-Docência em Levantamento do solo e ftopedologia), FCA/UNESP, 2001. 114p.