



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

Francicleia Araújo Bezerra

**Análise da variabilidade espacial de
componentes inorgânicos do solo sob árvores
de algaroba.**

Campina Grande - PB

Dezembro 2013

Francicleia Araújo Bezerra

Análise da variabilidade espacial de componentes inorgânicos do solo sob árvores de algaroba.

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

B574a Bezerra, Francicleia Araújo.
Análise da variabilidade espacial de componentes inorgânicos do solo sob árvores de algaroba [manuscrito] / Francicleia Araújo Bezerra. – 2013.
31 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. Semivariância. 2. Dependência espacial. 3. Algaroba. I. Título.

21. ed. CDD 310

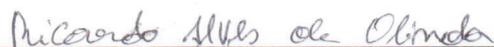
Francicleia Araújo Bezerra

Análise da variabilidade espacial de componentes inorgânicos do solo sob árvores de algaroba

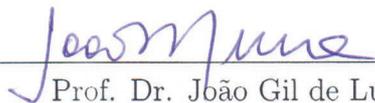
Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de Bacharel em Estatística.

Aprovado em: 16 / 12 / 2013

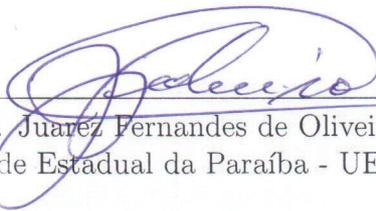
Banca Examinadora:



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 este trabalho aos meus pais, Francisco e Maria de Lourdes, que sempre me apoiaram e incentivaram durante toda a graduação, ajudando-me assim, a superar todas as dificuldades.

Agradecimentos

A Deus, por ter me dado saúde e sabedoria para superar as dificuldades encontradas durante esta caminhada. Ao professor Ricardo Alves de Olinda, pela confiança e paciência durante o período em que estivemos juntos na graduação e orientação deste trabalho. Também por ter me aceito como orientanda apesar do pouco tempo que dispunha e a quantidade elevada de orientandos que já tinha. Aos amigos: Ariely Aretusa, por ter me ajudado a superar as dificuldades encontradas durante a digitação deste trabalho; Sara Verusca, que sempre me ajudou quando eu precisei, mesmo tendo, muitas vezes, deixado de fazer suas tarefas; Alidiane, Flávia e Cleber Xavier que sempre estiveram comigo durante toda a graduação, apoiando e ajudando a superar os obstáculos. A minha família que sempre esteve ao meu lado, incentivando e apoiando durante toda a graduação. Sem eles nada disso seria possível.

Resumo

Este Trabalho tem por finalidade verificar a dependência espacial dos nutrientes nitrogênio e fósforo no solo sob árvores de algaroba, ou seja, verificar como se comporta espacialmente os teores de nitrogênio e fósforo no solo sob arvores de algaroba. Utilizou-se de semivariogramas e krigagem para a realização das análises. Observou-se a existência de dependência espacial por meio do envelope simulado e variogramas empíricos. Verificou-se o índice de dependência espacial por meio da relação entre a variação estruturada e o patamar. Entre as diferentes funções de correlação estudadas, o modelo circular foi o que melhor se ajustou aos dados para o nitrogênio, e o modelo esférico foi o que melhor se ajustou aos dados para o fósforo. Utilizou-se como critério de seleção o método de Akaike(AIC). Os resultados obtidos demonstraram que há indícios de dependência espacial do nitrogênio e do fósforo em função da presença dos indivíduos da algaroba. A variável nitrogênio apresenta dependência espacial, descrita pelo modelo circular com alcance de aproximadamente 7,884m e a variável fósforo apresenta dependência espacial, descrita pelo modelo esférico com alcance de aproximadamente 8,074m.

Palavras-chave: Semivariância, interpolação, variogramas.

Abstract

This work aims to check the spatial dependence of the nutrients nitrogen and phosphorus in soil in connection with mesquite trees, ie, check how it behaves spatially the levels of nitrogen and phosphorus in soil under mesquite trees. We used semivariograms and kriging for the analyzes. Observed the existence of spatial dependence through the simulated envelope and empirical variogram. There was the index of spatial dependence through relação between structured and level variation. Among the different functions correlation study, the circular model was the best fit to the data for the nutrient nitrogen, and the spherical model was the best fit to the data for the nutrient phosphorus. Was used as a selection criterion the method of Akaike(AIC). The results showed that there is evidence of spatial dependence of nitrogen and phosphorus as a function of presence of individuals of *Prosopis juliflora*. The nitrogen has variable spatial dependence described the circular model with variable range of approximately 7,884 m and phosphorus presents spatial dependence described by spherical with a range of approximately 8,074m model.

Keywords: semivariance, interpolation, variograms.

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 11
2	Fundamentação Teórica	p. 12
2.1	Tipos de dados em análise espacial	p. 12
2.1.1	Teoria das variáveis regionalizadas	p. 13
2.2	Estacionaridade	p. 13
2.2.1	Hipótese de estacionaridade de primeira ordem	p. 13
2.2.2	Hipótese de estacionariedade de segunda ordem	p. 14
2.3	Variograma e semivariograma	p. 15
2.4	Modelos geoestatísticos gaussianos	p. 15
2.4.1	Funções de correlação para o semivariograma	p. 16
2.4.2	Critério de avaliação de modelos	p. 17
2.5	Interpolação de dados	p. 18
3	Material e métodos	p. 20
3.1	Região de estudo	p. 20
4	Resultados e Discussão	p. 21
5	Conclusão	p. 29

Lista de Figuras

1	Gráfico descritivo do padrão espacial do nitrogênio no solo sob árvores de algaroba, com os dados originais.	p. 21
2	Gráfico descritivo do padrão espacial do nitrogênio no solo sob árvores de algaroba, com os dados transformados.	p. 22
3	Gráfico de possíveis transformações de variáveis por meio da transformação ótima de Box-Cox (Esquerdo) e a verificação de dependência espacial por meio do envelope simulado (direito) da variável nitrogênio da espécie algaroba.	p. 23
4	Modelo ajustado ao semivariograma conforme o método da máxima verossimilhança para a variável nitrogênio da espécie algaroba.	p. 24
5	Comparação dos valores observados com o valor de referência para solos com baixo teor de nitrogênio, na espécie algaroba	p. 25
6	Gráfico descritivo do padrão espacial da algaroba, variável fósforo, com os dados originais.	p. 26
7	Gráfico descritivo do padrão espacial da algaroba, variável fósforo, com os dados originais.	p. 26
8	Modelo ajustado ao semivariograma conforme o método da máxima verossimilhança para a variável fósforo da espécie algaroba.	p. 27
9	Comparação dos valores observados com o valor de referência para solos com baixo teor de fósforo, na espécie algaroba	p. 28

Lista de Tabelas

- 1 Estimativa dos parâmetros associado aos modelos por meio da máxima verossimilhança, assumindo a média da variável nitrogênio constante sobre a região de estudo. p.24
- 2 Estimativa dos parâmetros associado aos modelos por meio da máxima verossimilhança, assumindo a média da variável fósforo constante sobre a região de estudo. p.27

1 Introdução

No Brasil, os estudos com a utilização de Estatística Espacial tiveram início na década de 80 e desde então os trabalhos com a utilização desta ferramenta tem ocorrido em diferentes regiões do país, devido à necessidade de se estudar a dependência espacial de determinados atributos. Os procedimentos usados na estatística espacial compreendem a distribuição espacial de dados oriundos de fenômenos ocorridos no espaço. Constitui hoje um grande desafio para a elucidação de questões centrais em diversas áreas do conhecimento.

Conceitos como dependência espacial e auto-correlação espacial são fundamentais para a análise espacial. Entende-se por dependência espacial o fato de que a maioria das ocorrências naturais ou sociais apresentam entre si uma relação que depende da distância. A ideia é verificar a variabilidade da dependência espacial, a partir de observações entre os valores de uma amostra e de seus vizinhos (SANTOS, 2011). A ênfase da Análise Espacial é mensurar propriedades e relacionamentos, levando-se em conta a localização espacial do fenômeno em estudo de forma explícita.

De acordo com Druck et al.(2004), a forma mais utilizada para classificar e descrever o problema da análise espacial considera três tipos de dados: processos pontuais, superfícies contínuas e áreas com contagens. A análise de cada tipo de dados tem suas próprias particularidades. Este trabalho concentra-se na análise dos atributos cuja descrição se enquadra em superfícies contínuas, ou seja, geoestatística. Na agronomia, por exemplo, o objetivo pode ser analisar uma região para fins de zoneamento agrícola, ou mapear o teor de nutrientes no solo. Outras áreas que também se deparam com tais problemas são epidemiologia, entomologia, construção civil entre outras.

Diante do exposto este trabalho tem por objetivo verificar a dependência espacial dos atributos (nitrogênio e fósforo) na algaroba na região de estudo; ajustar as funções de correlação aos dados por meio de semivariograma experimental; prever a partir dos dados amostrais, o grau de dependência das variáveis nitrogênio e fósforo.

2 Fundamentação Teórica

A estatística espacial está inserida em um contexto mais amplo, a análise espacial de dados. A análise espacial tem como objetivo mensurar propriedades e relacionamentos, considerando-se a localização espacial do fenômeno em estudo, ou seja, permite estudar, explorar e modelar fenômenos geográficos (DRUCK et al. 2004; GOODCHILD e HAINING, 2004). Segundo Druck et al. (2004), a análise espacial é composta por um conjunto de procedimentos encadeados cuja finalidade é a escolha de um modelo inferencial que considera explicitamente os relacionamentos espaciais presentes no fenômeno. Um fenômeno espacial é um processo que pode ser representado por meio de uma distribuição no espaço, que pode ser contínuo ou discreto (FOOK, 2005). É considerado discreto quando se trata de entidades do mundo real como cidades, escolas, ginásios, igrejas, dentre outros. Um fenômeno é contínuo quando se refere a uma informação presente em todas as posições, como por exemplo temperatura, pressão e grau de acidez do solo.

2.1 Tipos de dados em análise espacial

A taxonomia mais utilizada para descrever o problema da análise espacial considera três tipos de dados: processos pontuais, superfícies contínuas e áreas com contagens. Processos pontuais são fenômenos expressos por meio de ocorrências identificadas como pontos localizados no espaço. A análise espacial é composta por um conjunto de procedimentos encadeados, cuja finalidade é a escolha de um modelo inferencial que considere explicitamente os relacionamentos espaciais presentes no fenômeno (OLINDA, 2008).

Áreas com contagens são fenômenos associados aos dados de levantamentos populacionais, como censos, e que, originariamente, referem-se a indivíduos localizados em pontos específicos no espaço. Normalmente, esses dados são agregados em unidades de análises. Áreas com contagens são fenômenos associados aos dados de usualmente delimitadas por polígonos fechados, tais como: setores censitários, municípios e microrregiões (OLINDA, 2008).

Superfícies contínuas(geoestatística) são fenômenos que se distribuem continuamente em uma região. Usualmente, esses tipos de dados são resultantes de levantamentos de

recursos naturais e que incluem mapas geológicos, topográficos e ecológicos. Segundo Olinda (2008), medidas da concentração de um elemento químico no solo são exemplo desse tipo de dados.

A geoestatística é uma área da estatística aplicada que se caracteriza pela modelagem da continuidade espacial de fenômenos naturais (GUIMARÃES, 2004). Leva-se em consideração a localização geográfica e a dependência espacial. Os procedimentos geoestatísticos estão fundamentados na Teoria das Variáveis Regionalizadas.

2.1.1 Teoria das variáveis regionalizadas

Segundo Genú (2004), a interpretação probabilística de que a variável aleatória regionalizada $Z(s_i)$ seja uma realização particular de uma certa função aleatória $Z(s_i)$ é consistente quando pode-se inferir toda ou pelo menos parte da lei de distribuição de probabilidade que define esta função aleatória. No entanto, em problemas práticos, em cada ponto contido em S , tem-se apenas uma realização $Z(s_i)$ e o número de pontos é sempre finito. Isto torna usualmente impossível inferir sobre a distribuição de $Z(s_i)$. Em vista disto, certas hipóteses são necessárias, as quais envolvem diferentes graus de homogeneidade espacial, sendo comumente denominadas hipóteses de estacionaridade.

2.2 Estacionaridade

De acordo com Vieira et al. (2000), uma variável regionalizada é estacionária se os momentos estatísticos da variável aleatória $Z(s_i + \mathbf{h})$ forem os mesmos para qualquer vetor de distância euclidiana \mathbf{h} . De acordo com o k -ésimo momento estatístico que são constantes, ou seja, a variável é chamada de estacionária de ordem k . Se um processo é estacionário na ordem k ele também será estacionário para as ordens inferiores a k . Por exemplo, se o processo é estacionário de ordem 4, ele também será estacionário nas ordens 1, 2 e 3. Estudos de geoestatística necessita-se, como restrição máxima, que o primeiro e o segundo momento em relação à origem sejam constantes, ou seja, exigem no máximo a estacionaridade de segunda ordem.

2.2.1 Hipótese de estacionaridade de primeira ordem

Segundo Genu (2004), a hipótese de estacionaridade de primeira ordem é definida como sendo o momento de primeira ordem da distribuição da função aleatória $Z(s_i)$, que

por sua vez, é constante em toda a região de estudos, ou seja,

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

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.

2.2.2 Hipótese de estacionariedade de segunda ordem

De acordo com Guimarães (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.

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

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

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:

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

Segundo Guedes et al. (2008), uma variável regionalizada é isotrópica quando o padrão de continuidade espacial é idêntico em todas as direções, ou seja, depende apenas do vetor \mathbf{h} . 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 o mesmo autor, a anisotropia é encontrada por meio de semivariogramas, que, construídos em diferentes direções, revelam uma privilegiada direção com um distinto padrão de dependência espacial.

2.3 Variograma e semivariograma

O variograma é a ferramenta básica, que permite descrever quantitativamente a variação no espaço de um fenômeno regionalizado (JOURNEL et al., 1978). A natureza estrutural de um conjunto de dados (assumido pela variável regionalizada) é definida por meio da comparação de valores tomados simultaneamente em dois pontos, sob uma determinada direção (GENÚ, 2004).

O semivariograma pode ser 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

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

sendo 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}) .

2.4 Modelos geoestatísticos gaussianos

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 $Y(\mathbf{s})$ de $n \times 1$ valores observados em $\mathbf{s} = s_i; i = (1, \dots, n)$, que segundo Diggle e Ribeiro Jr. (2007), sob o pressuposto de gaussianidade, pode ser modelado como

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

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 .

Por simplicidade matemática, normalmente utiliza-se o logaritmo da função de verossimilhança para fazer a estimação, que associada a (1) tem a seguinte forma

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

No contexto de geoestatística, Diggle e Ribeiro Jr. (2007) adotam 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 $\boldsymbol{\beta}$ 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). Para ϕ e σ^2 não existem formas analíticas, utilizando-se $\hat{\beta}$ e $\hat{\sigma}^2$, tem-se o logaritmo da função de verossimilhança concentrada, que depende apenas de $\boldsymbol{\theta}_c = (\nu, \phi)'$ (FONSECA, 2008).

2.4.1 Funções de correlação para o semivariograma

Adicionalmente, é útil incluir no modelo, alguma flexibilidade na forma geral da função de correlação. A seguir pode-se observar algumas das principais funções de correlação do semivariograma utilizadas na geoestatística são:

Função de correlação gaussiana

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

Função de correlação exponencial

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

A função de correlação exponencial normalmente 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 (DIGGLE; RIBEIRO Jr., 2007). 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

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

A 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 (DIGGLE; RIBEIRO Jr., 2007).

Função de correlação Circular

Segundo McBratney e Webster (1986), a função de correlação circular é válida em R e em R^2 mas não em R^3 e define-se da seguinte forma

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

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*.

2.4.2 Critério de avaliação de modelos

A escolha do modelo apropriado, do ponto de vista estatístico, é um tópico extremamente importante na análise de dados (BOZDAGAN, 1987). Busca-se o modelo mais

parcimonioso, isto é, o modelo que envolva o mínimo de parâmetros possíveis a serem estimados e que explique bem o comportamento da variável resposta. Neste trabalho adota-se o critério de informação de Akaike (AIC).

Utilizando-se o AIC admite-se que dentre os "modelos avaliados" nenhum é considerado o que realmente descreve a relação entre a variável dependente e as variáveis explanatórias, ou o "modelo verdadeiro" e então, tenta-se escolher o modelo que minimize a divergência. O Critério de Informação de Akaike admite a existência de um modelo "real" que descreve os dados que é desconhecido, e tenta escolher dentre um grupo de modelos avaliados, o que minimiza a divergência de Kullback-Leibler.

O critério de informação de Akaike é expresso por

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

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 .

2.5 Interpolação de dados

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* (ZIMBACK, 2001). Existem diferentes métodos de interpolação de dados, tais como: funções globais, polinômios, equações multi-quadráticas, triangulação, inverso da potência das distâncias, e base radial (krigagem, multiquádrica-biharmônica, splines) entretanto, a maioria desses métodos não fornecem o algoritmo dos erros associados aos resultados obtidos.

A Krigagem é considerada uma boa metodologia de interpolação de dados. Ela utiliza o dado tabular e sua posição geográfica para calcular as interpolações. Utilizando-se o princípio da Primeira Lei de Geografia de Tobler, que diz que unidades de análise mais próximas entre si são mais parecidas do que unidades mais afastadas, a krigagem utiliza funções matemáticas para acrescentar pesos maiores nas posições mais próximas aos pontos amostrais e pesos menores nas posições mais distantes, e criar assim os novos pontos interpolados com base nessas combinações lineares de dados (JAKOB, 2002).

Para a aplicação da krigagem assume-se que sejam conhecidas as realizações $z(s_1)$,

$z(s_2), \dots, z(s_n)$ 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 Trøngmar 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 Material e métodos

A algarobeira é uma planta seletiva, pouco exigente em água, cuja ocorrência, em sua forma natural, se dá em zonas tropicais áridas, que não chegam a alcançar índices de 100mm. Sendo uma espécie adaptada a climas áridos e semiáridos, a germinação da algaroba, geralmente acontece durante a estação chuvosa. A espécie possui dois sistemas de raiz, uma raiz funda e um tapete de raízes laterais onde absorve águas de chuvas. Sua raiz principal chega a atingir grandes profundidades, pois, sendo uma planta de regiões áridas, busca encontrar o lençol freático, retirando água do subsolo para a superfície, pode sobreviver em áreas com baixa precipitação anual ou períodos secos muito prolongados (CORREIA, 2013).

3.1 Região de estudo

O estudo foi realizado em 1996, em uma fazenda de gado de corte pertencente à Agropecuária Jaçanã, na Cidade de Custódia, PE. A precipitação média de chuva no local é 740 milímetros e a temperatura média anual é de $26^{\circ}C$. 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).

Conforme Tiessen et al. (2003), em uma área da fazenda com aproximadamente 3.000 ha, foram plantadas no ano seguinte as árvores *Prosopis juliflora* Swartz DC. (Algarobas), a uma distância de 10m por 10m, consorciadas com *Cenchrus ciliaries* L. (Capim buffel). O pasto nunca recebeu adubação, aplicação de agrotóxicos e nem foram registrados incêndio no local.

Em outra área da fazenda, com cerca de 2.000 ha, foram preservadas algumas espécies nativas da caatinga, 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).

4 Resultados e Discussão

De acordo com Correia (2013), 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 15 cm, 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) e fósforo(P).

Após a coleta dos dados, utilizou-se o pacote *geoR* com auxílio do software R versão 2.11.1. Utiliza-se a média dos dados como comparativo para análises. A análise exploratória espacial do teor de fertilidade do solo, inicia-se po meio da variável nitrogênio. Logo após faz-se a mesma análise para variável fósforo nas mesmas condições (CORREIA).

Variável Nitrogênio

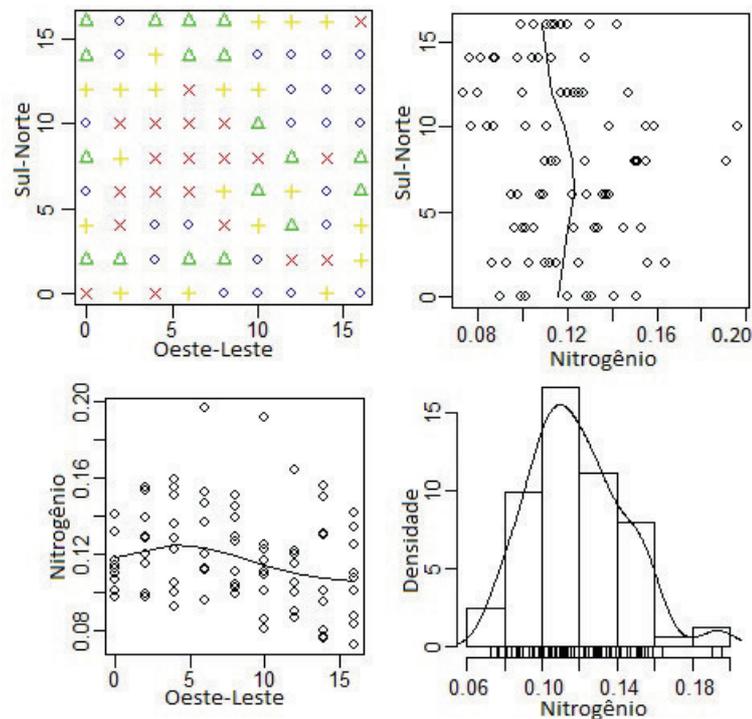


Figura 1: Gráfico descritivo do padrão espacial do nitrogênio no solo sob árvores de algaroba, com os dados originais.

O gráfico do canto superior esquerdo da Figura 1 representa os dados nos quartis da variável nitrogênio, ou seja, “+” 1º quartil, “Δ” 2º quartil, “o” 3º quartil e “x” 4º quartil,

essa imagem condiz com a ideia de que existe padrão espacial na variável. O gráfico do canto superior direito e do canto inferior esquerdo, apresentam a dispersão dos dados em torno da média geral, observa-se que os dados estão dispersos. Com base no gráfico do canto inferior direito, a análise exploratória do pressuposto de gaussianidade da resposta foi analisado, o qual mostra a densidade amostral dos dados desconsiderando-se o possível padrão espacial. Esta inspeção não garante a tal hipótese mas serve para mostrar desvios óbvios, caso existam. Como existe uma fuga da distribuição dos resíduos, necessita-se de possíveis transformações na variável.

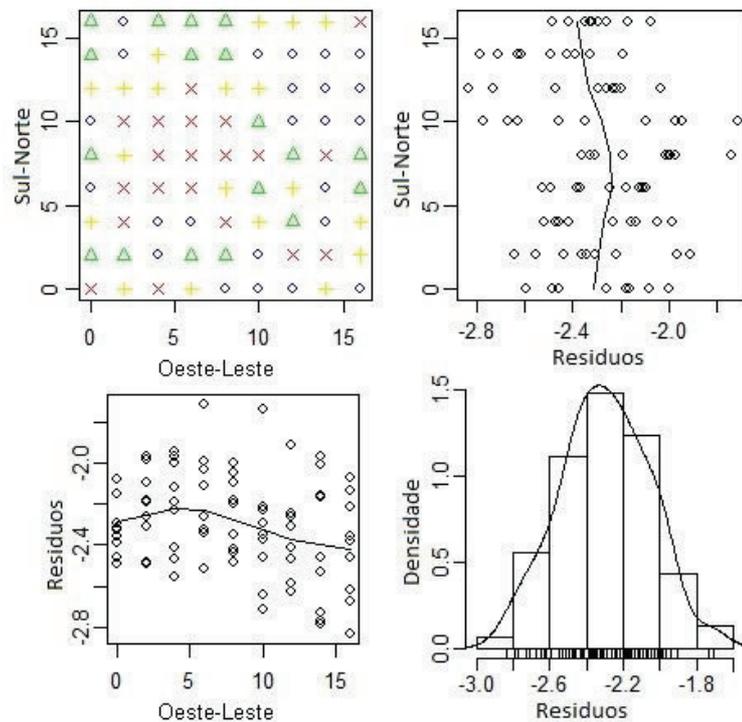


Figura 2: Gráfico descritivo do padrão espacial do nitrogênio no solo sob árvores de algaroba, com os dados transformados.

A transformação ótima de Box-Cox (lado esquerdo) da Figura 2, por meio do logaritmo da função da verossimilhança, mostra qual o valor do parâmetro que mais aproxima os resíduos à distribuição normal. Foi necessário utilizar uma transformação na variável em estudo, pois os intervalos de confiança $[-0,99; 0,84]$ para λ não contém o valor 1, neste caso, para a variável, foi encontrado o valor de $\lambda = -0,06$.

Após a transformação dos dados de nitrogênio, percebe-se que os dados estão menos dispersos em torno da média comparando-se com o gráfico da Figura 1, estes gráficos estão localizados no canto superior direito e no canto inferior esquerdo da Figura 2. A análise exploratória do pressuposto de gaussianidade da resposta foi analisado, no gráfico do canto inferior direito, percebe-se uma mudança em relação ao anterior.

Para verificar o grau de dependência espacial entre os teores de nitrogênio no solo, utilizou-se o efeito pepita τ^2 sobre o patamar $\sigma^2 + \tau^2$. Obteve-se então, um grau de dependência espacial moderada de 43%, conforme a classificação de Cambardella et al. (1994) quando a relação está entre $25\% \geq IDE \leq 75\%$, há uma moderada dependência espacial. Nota-se que a variável nitrogênio apresenta dependência espacial, que pode ser descrita pelo modelo circular com alcance de aproximadamente 7,884m, ou seja, amostras de nitrogênio de solo selecionadas a distâncias inferiores são espacialmente dependentes.

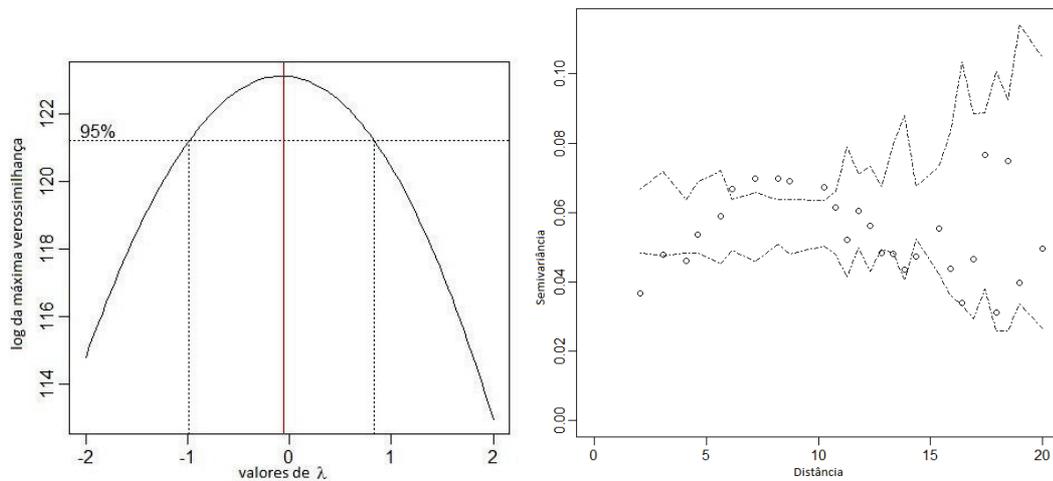


Figura 3: Gráfico de possíveis transformações de variáveis por meio da transformação ótima de Box-Cox (Esquerdo) e a verificação de dependência espacial por meio do envelope simulado (direito) da variável nitrogênio da espécie algaroba.

A dependência espacial pode ser verificada por meio do gráfico utilizando-se envelopes simulados e variogramas empíricos, como vê-se na Figura 3. Nessa pesquisa foram utilizadas 1.000 simulações. Verifica-se que alguns pontos estão fora do envelope simulado (lado direito) indicando assim uma dependência espacial, pois, para que haja dependência, deve haver pelo menos um ponto do variograma fora do envelope simulado, quando isto ocorre rejeita-se H_0 , que supõe a não dependência espacial (DIGGLE; RIBEIRO Jr., 2007), deste modo as análises clássicas devem ser substituídas por análises espaciais, que levam em consideração as relações entre observações vizinhas, ou seja, dependência espacial.

Na Tabela 1, observa-se as estimativas da máxima verossimilhança para os parâmetros dos modelos ajustados, considerando-se a média constante sob a região de estudo. O estimador do parâmetro β apresenta valores próximos para todas as funções de correlações.

O modelo que se ajustou melhor ao semivariograma experimental para a variável nitrogênio onde está inserida a algaroba foi a função de correlação circular indicando o

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

Modelos	$\hat{\beta}$	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$	AIC	IDE
Matérn kappa = 0,5	0,1173	0,0001	0,0005	2,207	-376,4	17 %
Matérn kappa = 1,0	0,1173	0,0003	0,0004	1,857	-377,6	43 %
Esférico	0,1178	0,0003	0,0004	8,591	-378,6	43 %
Gaussiano	0,1175	0,0004	0,0003	4,452	-377,9	57 %
Circular	0,1790	0,0003	0,0004	7,884	-378,9	43 %

menor valor de AIC para todos os modelos espaciais como mostra na Figura 4 , encontra-se o semivariograma experimental ajustado conforme o método da máxima verossimilhança.

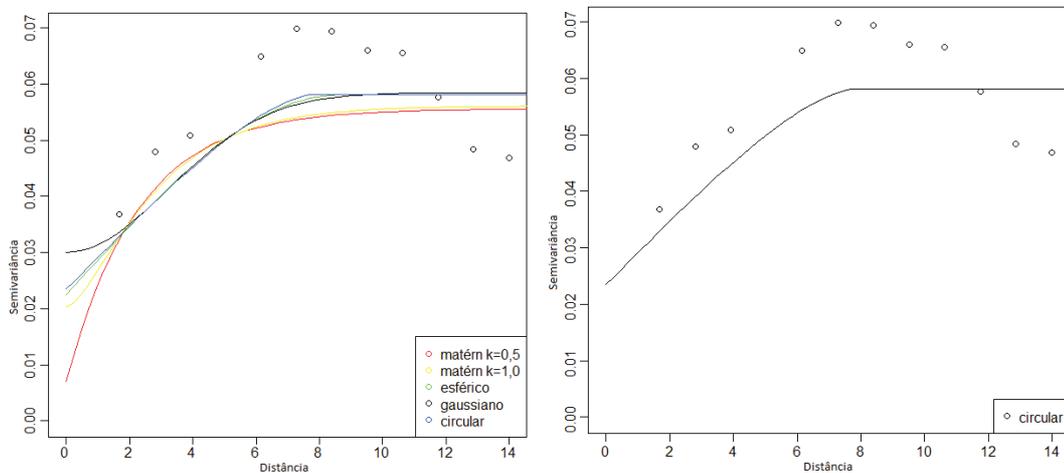


Figura 4: Modelo ajustado ao semivariograma conforme o método da máxima verossimilhança para a variável nitrogênio da espécie algaroba.

Na Figura 4 (lado direito) encontra-se o semivariograma experimental ajustado conforme o método da máxima verossimilhança, a análise variográfica apresentada mostrou que todas as características estudadas nos procedimentos até aqui descritos, apresentam estruturas espaciais bem definidas, ou seja, existe uma função estrutural com semi-variância de comportamento modelável.

Com os modelos finais estabelecidos, foram calculadas as predições espaciais para o índice de fertilidade de solo em uma malha de 81 localizações espaciais, sendo que as estimativas paramétricas foram substituída para a fórmula de krigagem.

Pode-se observar por meio da Figura 5 o mapa preditivo do campo aleatório para o efeito de crescimento de capim. Observa-se pelo gráfico do lado esquerdo da figura 5 uma variabilidade dentro da área onde está inserida a algaroba. Percebe-se uma concentração muito forte de nitrogênio na cor mais clara (de branca à rosa) em uma pequena região de

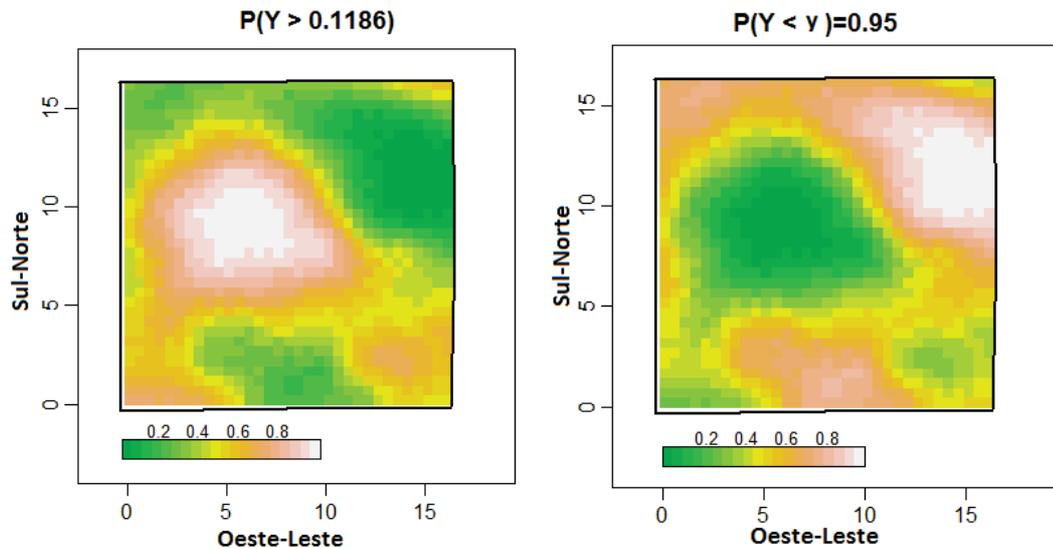


Figura 5: Comparação dos valores observados com o valor de referência para solos com baixo teor de nitrogênio, na espécie algaroba

estudos, abaixo da copa da árvore. Pode-se dizer com uma probabilidade entre 80 e 100% que o índice de nitrogênio nesta área é maior que a média. Na área em amarelo pode-se dizer com uma probabilidade entre 50 e 70% que o índice de nitrogênio nesta área é maior que a média.

Variável Fósforo

Os resultados obtidos demonstraram que há indícios de dependência espacial do fósforo no solo sob árvores de algaroba.

O gráfico do canto superior esquerdo da Figura 6 representa os dados nos quartis da variável fósforo, ou seja, "+" 1º quartil, "Δ" 2º quartil, "o" 3º quartil e "×" 4º quartil, essa imagem condiz com a ideia de que existe padrão espacial na variável. O gráfico do canto superior direito e do canto inferior esquerdo, apresentam a dispersão dos dados em torno da média geral.

Para verificar o grau de dependência espacial entre as características de fertilidade do solo, utilizou-se o efeito pepita τ^2 sobre o patamar $\sigma^2 + \tau^2$. Obteve-se então, um grau de dependência espacial moderada de 12%, conforme a classificação de Cambardella et al. (1994) quando a relação está entre $IDE \leq 25\%$, há uma forte dependência espacial. Nota-se que a variável fósforo apresenta dependência espacial, que pode ser descrita pelo modelo esférico com alcance de aproximadamente 8,074m, ou seja, amostras de solo selecionadas a distâncias inferiores são espacialmente dependentes.

A dependência espacial pode ser verificada por meio do gráfico utilizando-se envelopes

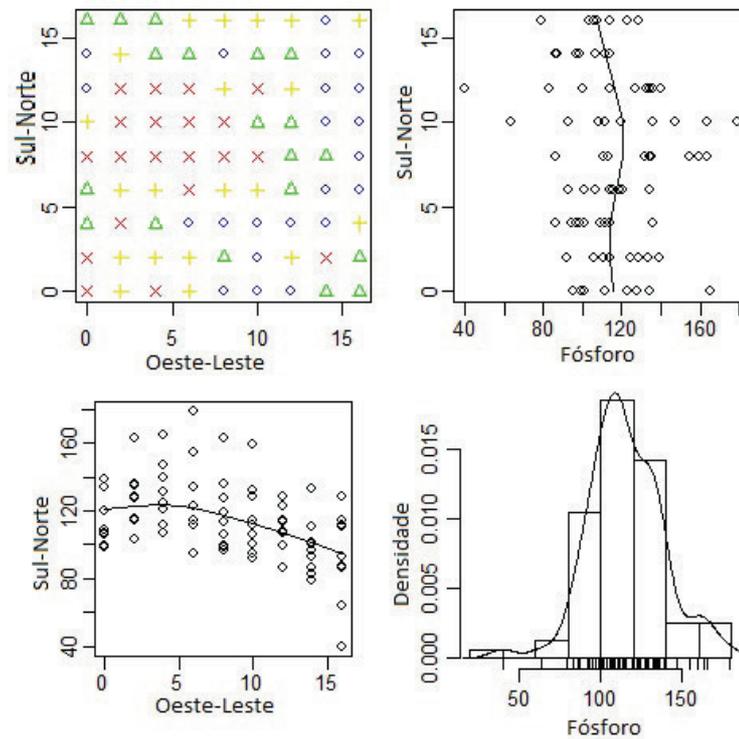


Figura 6: Gráfico descritivo do padrão espacial da algaroba, variável fósforo, com os dados originais.

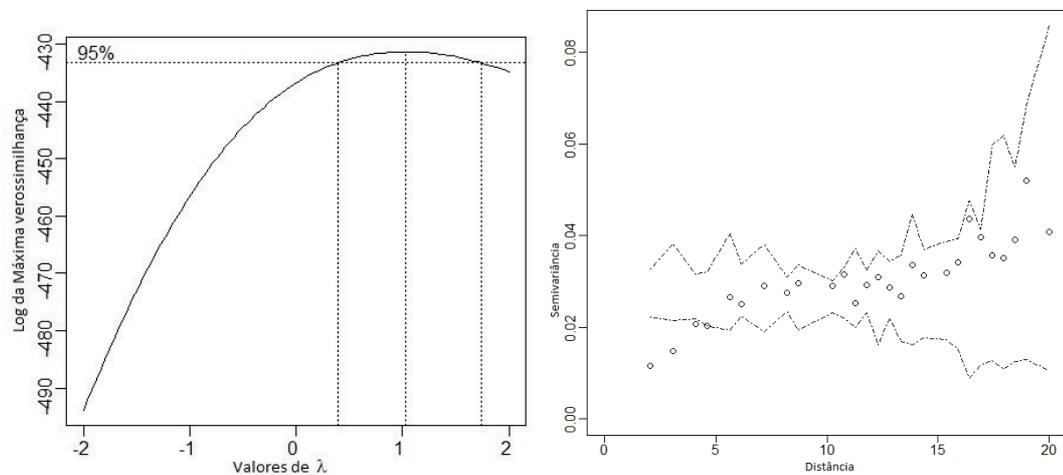


Figura 7: Gráfico descritivo do padrão espacial da algaroba, variável fósforo, com os dados originais.

simulados e variogramas empíricos, como vê-se na Figura 7. Nessa pesquisa foram utilizadas 1.000 simulações. Verifica-se que alguns pontos estão fora do envelope simulado (lado direito) indicando assim uma dependência espacial, pois, para que haja dependência, deve haver pelo menos um ponto do variograma fora do envelope simulado, quando isto ocorre rejeita-se H_0 , que supõe a não dependência espacial (DIGGLE; RIBEIRO Jr., 2007), deste modo as análises clássicas devem ser substituídas por análises espaciais, que

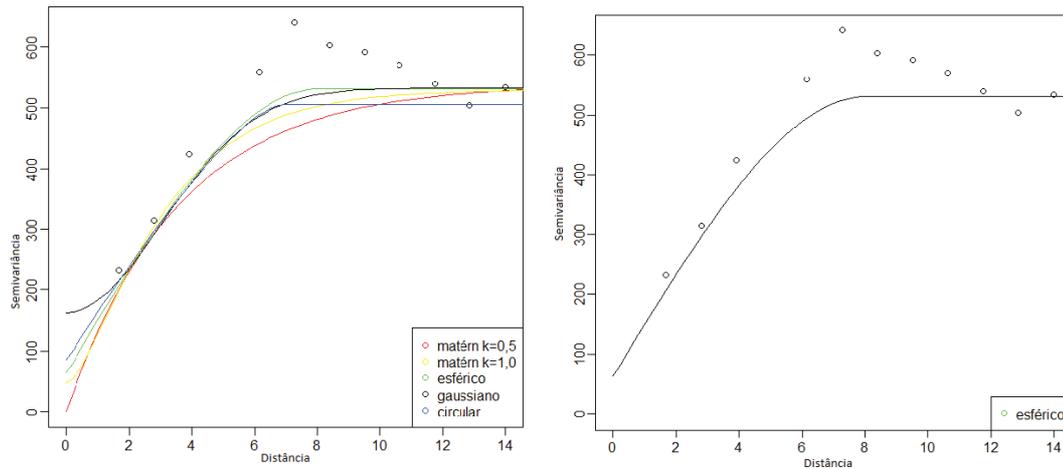


Figura 8: Modelo ajustado ao semivariograma conforme o método da máxima verossimilhança para a variável fósforo da espécie algaroba.

levam em consideração as relações entre observações vizinhas, ou seja, dependência espacial.

Tabela 2: Estimativa dos parâmetros associado aos modelos por meio da máxima verossimilhança, assumindo a média da variável fósforo constante sobre a região de estudo.

Modelos	$\hat{\beta}$	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$	AIC	IDE
Matérn kappa = 0,5	112,492	0,000	538,522	3,584	699,7	0 %
Matérn kappa = 1,0	113,110	47,467	481,468	2,071	699,2	9 %
Esférico	113,523	63,830	369,540	8,074	698,4	12 %
Gaussiano	113,330	162,320	369,540	4,260	699,1	31 %
Circular	113,458	85,862	419,507	6,897	698,7	17 %

Na Tabela 2, observa-se as estimativas da máxima verossimilhança para os parâmetros dos modelos ajustados, considerando-se a média constante sob a região de estudo. O estimador do parâmetro β apresenta valores próximos para todas as funções de correlações. Para todas as características avaliadas. O modelo que se ajustou melhor ao semivariograma experimental para a variável fósforo onde está inserida a algaroba foi a função de correlação esférico indicando o menor valor de AIC para todos os modelos espaciais conforme apresentada na Figura 8;

Na Figura 8 (lado direito) encontra-se o semivariograma experimental ajustado conforme o método da máxima verossimilhança, a análise variográfica apresentada 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.

Pode-se observar por meio da Figura 9 o mapa preditivo do campo aleatório para o

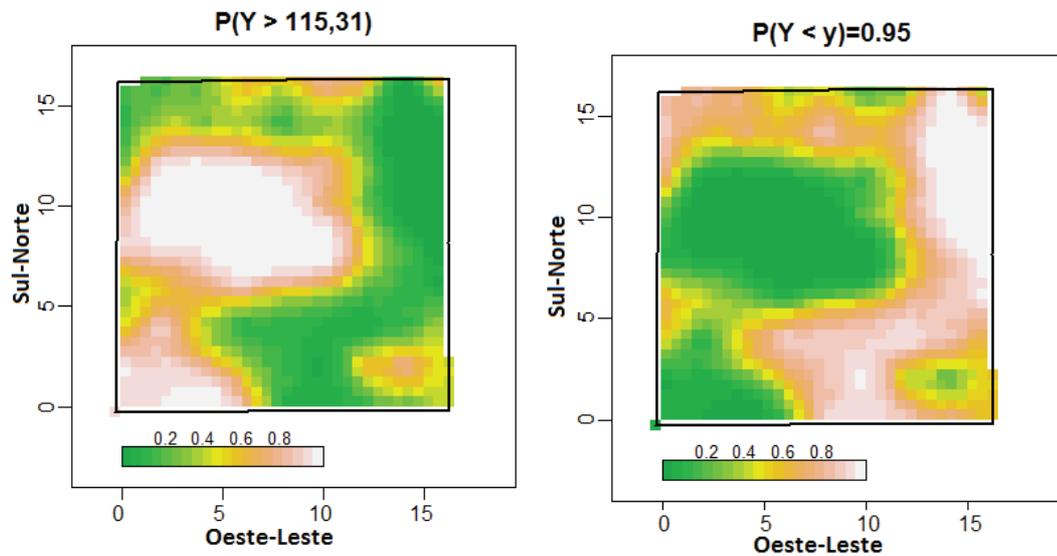


Figura 9: Comparação dos valores observados com o valor de referência para solos com baixo teor de fósforo, na espécie algaroba

efeito de crescimento de capim. Percebe-se uma concentração acima da média da variável fósforo na cor mais clara (de branca à rosa) em uma pequena região de estudos. Pode-se dizer com uma probabilidade entre 70 e 100% que o índice de fósforo nesta área é maior que a média. Na área em amarelo pode-se dizer com uma probabilidade entre 50 e 60% que o índice de fósforo nesta área é maior que a média e na região verde pode-se dizer com uma probabilidade entre 10 e 40% que o índice de fósforo nesta área é maior que a média.

5 Conclusão

Ao término deste trabalho conclui-se que foi possível verificar e modelar a dependência espacial através de funções de correlação que expressam a estrutura de dependência espacial das características avaliadas.

Considerada uma boa metodologia de interpolação de dados para as variáveis estudadas, a krigagem tornou-se um fator importante, possibilitando realizar previsões a partir dos dados amostrais.

Os resultados mostram que há indícios de dependência espacial do nitrogênio e do fósforo no solo sob árvores de algaroba. A variável nitrogênio apresenta dependência espacial, com alcance de aproximadamente 7,884m e a variável fósforo apresenta dependência espacial com alcance de aproximadamente 8,074m.

Pode-se concluir então que a uma distância em média de 8m existe especial dependência dos nutrientes estudados no solo sob árvore de algaroba.

Referências

- BOZDANGAN. H. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. **Psychometrika**, v.52, n.3, p.345-370, 1987.
- CORREIA, M. D. **Modelagem geoestatística da distribuição de carbono do solo e biomassa de herbáceas em sistema silvopastoril**. Dissertação (Mestrado em biometria e estatística aplicada) - Universidade Federal Rural de Pernambuco, Pernambuco, 2013. 101p.
- DIGGLE, P. J.; RIBEIRO JÚNIOR, P. J. **Model Based Geostatistics**. In: SINAPE, 14., 2000, Caxambu. Anais. São Paulo: Associação Brasileira de Estatística, 2000.
- 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, p.60-83, 2004.
- FOOK. K. D. **Integração da Estatística Espacial em Ambientes Gis**, 2005.
- FONSECA, B. H. F. **Um estudo sobre estimação e predição em modelos geoestatísticos bivariados**. Piracicaba. Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, 2008. 74p.
- GUEDES, L. P. C., URIBE. A. M., JOHANN. J. A. e SOUZA. E. G.; **Anisotropia no Estudo da Variabilidade Espacial de Algumas Variáveis Químicas do Solo**. 2008.
- GENÚ. A. M.; **Geoestatística Multivariada**. 2004.
- GUIMARÃES, E. CARVALHO.; **Geoestatística Básica e Aplicada**, Universidade Federal de Uberlândia Faculdade de Matemática, Núcleo de Estudos Estatísticos e Biométricos, MG. 2004. 205p.
- GOODCHILD, M. F.; HAINING, R. P. GIS and spatial data analysis: covering perspectives. **Papers in Regional Science**, v.83, p.363-385, 2004.
- JAKOB, A. A. E., **A Krigagem como Método de Análise de Dados Demográficos**; UNICAMP/NEPO; Trabalho apresentado no XIII Encontro da Associação Brasileira de Estudos Populacionais, realizado em Ouro Preto, Minas Gerais, Brasil de 4 a 8 de novembro de 2002. Apoio do CNPq.
- JOURNEL, A.G.; HUIJBREGTS, Ch.J. **Mining Geostatistics**. London: Academic Press, 1978. 600p.
- KRUGER, M. C. **Análise geoestatística de dados meteorológicos do estado do paran utilizando um software livre**, 2005. 96p.

- MCBRATNEY, A. G.; WEBSTER, A. G. Choosing functions for semi-variograms and fitting them to sampling estimates. **Journal of Soil Science**, v.37, p.617-639, 1986.
- 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**. R. Bras. Ci. Solo, p.817-826. 1999.
- OLINDA, R. A. **Métodos para análise de Independência entre marcas e pontos em Processos Pontuais Marcados**, v.1, p39-56, 2008
- SANTOS, I. P. **Análise da variabilidade espacial dos casos de dengue na capital Rio Branco por meio de técnicas geoestatísticas**. (Trabalho de Conclusão de Curso - CCT - Universidade Estadual da Paraíba.) 2011. 45p.
- TRAGMAR, B. B., YOST, R. S., UEHARA, G. Application of geostatistics to spatial studies of soil properties. **Advances in Agronomy**, v.38, p.45-94, 1985.
- 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.3, p.195-205, 2003.
- VIEIRA, S. R.; NOVAIS, R. F.; ALVARES, V. H.; SCHAEFFER, C. E. G. R. Geoestatística em estudos de variabilidade espacial do solo. Tópicos em ciência do solo. Viçosa: **Sociedade Brasileira de Ciência do Solo**, p.1-54, 2000.
- 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.