RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo desta tese será disponibilizado somente a partir de 28/02/2023.

MÉTODO DE PARTIÇÃO PRODUTO APLICADO A KRIGAGEM

Maria de Fátima Ferreira Almeida

Tese apresentada à Universidade Estadual Paulista "Júlio de Mesquita Filho" para fins de obtenção do título de Doutora em Biometria.

BOTUCATU
São Paulo - Brasil
Fevereiro - 2019

MÉTODO DE PARTIÇÃO PRODUTO APLICADO A KRIGAGEM

Maria de Fátima Ferreira Almeida

Orientador: Prof. Dr. **José Sílvio Govone**

Coorientador: Prof. Dr. **Gérson Rodrigues dos Santos**

Tese apresentada à Universidade Estadual Paulista "Júlio de Mesquita Filho" para fins de Exame Geral de defesa de tese, como parte dos requisitos para obtenção do título de Doutora em Biometria.

BOTUCATU
São Paulo - Brasil
Fevereiro - 2019

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Almeida, Maria de Fátima Ferreira.
Método de partição produto aplicado a Krigagem / Maria de Fátima Ferreira Almeida. - Botucatu, 2019

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu
Orientador: José Sílvio Govone
Coorientador: Gérson Rodrigues dos Santos
Capes: 90000005

1. Krigagem. 2. Geologia - Métodos estatísticos. 3. Biometria - Estudos longitudinais. 4. Análise espacial (Estatística). 5. Teoria bayesiana de decisão estatística.

Palavras-chave: Modelo hierárquico Bayesiano; Acurácia; Krigagem; Ponto de mudança.

Dedicatória

Á Deus pela infinita bondade, pelas bênçãos recebidas, pela saúde e equilíbrio meu e de minha família, pelo dom da coragem e pela crença de que a distância física no trajeto para o doutorado era mais curta que nos mapas de voo e nas rotas dos ônibus e pela tranquilidade para a realização do meu trabalho.

Ao meu orientador José Sílvio Govone, por ter acreditado em mim e não medir esforços, orientando com paciência e sempre ponderando algumas limitações de distância, trabalho e familiares, que muitas vezes me impediram cumprir as tarefas em tempo hábil.

Aos meus pais, Xisto F. dos Santos e Percília B. dos Santos, que mesmo desconhecendo o teor do meu trabalho sempre valorizaram a escola e com simplicidade transmitiram mensagens de apoio que se tornaram a cada dia, mais importantes que os ensinamentos dos livros.

Às minhas filhas, Magaly Stefânia e Luma Fabiane e ao meu marido Geraldo A. Almeida, aos meus irmãos, sobrinhos e netos: Miguel e Maria Luiza, por ficarem do meu lado apoiando e tornando mais leves e alegres os meus dias.

Aos amigos que se mantiveram ao meu lado ao longo de todo o curso e aos que mesmo distantes compartilharam conhecimentos e apoio.

À minha querida professora primária, Dona Ailce Almeida, pela dedicação e a todos os demais professores que ao longo da minha vida estudantil, contríbuiram para que eu chegasse até aqui.

...um dia, conversando com Deus, eu disse que chegava,...e ele me trouxe até aqui, ...ai eu cheguei. Obrigada meu Deus.

Palavras não são insuficientes para demonstrar a minha gratidão!

"*Bom mesmo é ir à luta com determinação, abraçar a vida com paixão, perder com classe e vencer com ousadia, porque o mundo pertence a quem se atreve e a vida é* ***muito*** *para ser insignificante*"

(Charlie Chaplin)

Agradecimentos

Aos professores, técnicos e auxiliares do Departamento de Bioestatística/ UNESP pelos conhecimentos e dedicação.

À Universidade Federal de Viçosa (UFV) pelo apoio à mobilidade acadêmica, em especial ao professor Dr. Gerson Rodrigues dos Santos (Programa de Pós Graduação em Estatística Aplicada e Biometria - PPESTBIO) pela coorientação, ao professor Fabyano Fonseca e Silva (Programa de Pós Graduação em Zootecnia) pelo apoio e discussões relevantes para esta tese e ao Grupo de Estudos e Pesquisas em Levantamentos Hidrográficos- GEPLH na pessoa do professor Dr. Ítalo Ferreira (Departamento de Engenharia Civil) pelo apoio e por ceder os dados para a pesquisa.

Ao Instituto Federal de Educação do Norte de Minas Gerais (IFNMG) inclusive colegas de trabalho e estudantes, pelo apoio.

Aos amigos que se mantiveram ao meu lado ao longo de todo o curso, e aos que mesmo distantes se fizeram presentes, em especial: Jacqueline Domingues, Vívian Brancaglioni, Felipe Teles, Alex Santos, Márcio Rodrigues, Egídio Martins e Clênia Toletino.

À CAPES, IFNMG e CNPQ, pelo apoio financeiro.

Sumário

Lista de Figuras

Lista de Tabelas

MÉTODO DE PARTIÇÃO PRODUTO APLICADO A KRIGAGEM

Maria de Fátima Ferreira Almeida: MARIA DE FÁTIMA FERREIRA ALMEIDA
Orientador: Prof. Dr. JOSÉ SÍLVIO GOVONE

RESUMO

As variáveis aleatórias no espaço estão definidas por *funções aleatórias* sujeitas à teoria das variáveis regionalizadas. Para assumir continuidade espacial com um número limitado de realizações da variável aleatória são necessárias as hipóteses de estacionariedade, as quais envolvem diferentes graus de homogeneidade espacial. Formalmente, uma variável regionalizada Z é estacionária se os momentos estatísticos de $Z(s+h)$ forem os mesmos para qualquer vetor h. A hipótese de estacionaridade de primeira ordem é definida como a hipótese de que o momento de primeira ordem da distribuição da função aleatória $Z(s)$ é constante em toda a área. A hipótese intrínseca é baseada no cálculo de médias globais das semivariâncias, com a pressuposição de estacionaridade de 1^a ordem e da estacionaridade da variância dos incrementos. Embora muitas variáveis sejam sucetíveis a dupla ou múltipla estacionaridade, estas estruturas espaciais não são levadas em consideração pelo semivariograma usual, e, consequentemente, causam sérios problemas de acurácia nos mapas de Krigagem. Na perspectiva de solucionar o problema apontado, buscou-se identificar os locais dos pontos de mudança na média que definem mais de uma estrutura de semivariância, com o objetivo de melhorar a qualidade dos mapas de Krigagem Ordinária. Para isso, foi utilizado o Método de Partição Produto (MPP), com enfoque espacial, denominado Método de Partição Produto Espacial (MPPs). Para separar os grupos, foi criada uma função de busca de ponto de mudança na média utilizando o modelo hierárquico bayesiano, denominado Modelo de Partição Produto Espacial (MPPs). Utilizou-se dois bancos de dados para testar o potencial do modelo em separar grupos espacialmente dependentes, em que no primeiro havia suspeita de uma mudança na média, enquanto no segundo, "Dados2", havia suspeita de dois pontos de mudança na média. Na análise do primeiro banco de dados, o primeiro grupo, apesar de não obter um ajuste do semivariograma totalmente satisfatório, ainda assim obteve boa acurácia no mapa, enquanto que o segundo grupo, observou-se um ajuste satisfatório a um modelo diferente de semivariograma e obteve-se melhor acurácia, superando o primeiro grupo e o conjunto de amostra completa. No segundo banco de dados, "Dados 2", os três grupos se ajustaram a três semivariogramas distintos e geraram três mapas de Krigagem, nos quais, os mapas gerados para as três subáreas mostraram resultados satisfatórios, comprovando que a qualidade de ambos superaram a qualidade do mapa de krigagem feito para a amostra completa. Por meio dos resultados obtidos nos dois bancos de dados, concluiu-se que o método MPPs aplicado à Krigagem Ordinária garante mapas de melhor qualidade, por apresentar estimativas mais acuradas.

Palavras-chave: Ponto de Mudança, Método hierárquico Bayesiano espacial de partição produto - MPPs, Krigagem, Acurácia de mapas.

KRIGING INDUCED VIA METHOD OF SPATIAL PRODUCT PARTITION

Author: MARIA DE FÁTIMA FERREIRA ALMEIDA
Adviser: Prof. Dr. JOSÉ SÍLVIO GOVONE

SUMMARY

The random variables in space are defined by random functions subject to regionalized variable theory. To assume spatial continuity with a limited number of realization of the random variable, we need to assume stationarity hypotheses, which involve different degrees of spatial homogeneity. Formally, a regionalized variable Z is stationary if statistical moments of $Z(s+h)$ are the same for any vector h. The first order stationarity hypothesis is defined to be the hypothesis that first order moment of the distribution of the random function $Z(s)$ is constant throughout the area. The intrinsic hypothesis is based on the computation of global means of semivariate models, with the assumption of 1^{st} order stationarity and incremental variation stationarity. Although many variables are capable of double or multiple stationarity, these spatial structures are not taken into account by the usual semivariogram, and, consequently, cause acuracy problems in Kriging maps. In order to solve the described problem, it was identify the points of change in the average with the objective of improving the quality and accuracy of the maps of Ordinary Kriging. To separate the groups, a mean change point search function was created using the Bayesian hierarchical model, called the Space Product Partition Model (MPPs). Two databases were used to test the model's potential to separate spatially dependent groups, in which the former suspected a change in mean while in the latter. " Data2 ", there were suspicion of two points of change in the average. n the analysis of the first database, the first group, although it did not obtain a totally satisfactory semivariogram adjustment, nevertheless obtained good accuracy in the map, whereas the second group, a satisfactory adjustment was observed to a different model of semivariogram and we obtained better accuracy, surpassing the first group and the complete sample set. In the second database, Data 2, the three groups conformed to three distinct semivariograms and generated three Kriging maps. In the three subareas, referring to the second database, "Data2", the results proved that the quality of the three maps exceeded the quality of the kriging map made for the complete sample. Based on the results obtained, it was concluded that the MPPs method applied to Ordinary Kriging guarantees maps of better quality, since they present more accurate estimates. **Keywords**: Point of Change, Bayesian Hierarchical Model MPPs, Ordinary Kriging, Accuracy.

1 INTRODUÇÃO

A representação quantitativa de fenômenos físicos, assim como a predição de valores desconhecidos, sempre foi preocupação da ciência. Para tais representações, predições são feitas por meio de métodos probabilísticos (que levam em consideração a incerteza através de modelos probabilísticos) ou mesmo determinísticos (que desconsideram a incerteza que possa circundar o fenômeno em estudo). A Geoestatística utiliza de modelos probabilísticos para descrever os processos estocásticos espaciais contínuos.

De acordo com Vieira (2000) a Geoestatística teve seu limiar entre os anos de 1957 a 1962, quando o engenheiro de minas, francês, *G. Matheron*, de posse das observações de D.G. Krige, engenheiro de minas Sul africano, desenvolveu a *Teoria das variáveis regionalizadas*, representadas na prática por certa quantidade de dados brutos disponíveis, a partir dos quais foram obtidas as informações sobre as características do fenômeno. A partir do limiar da Geoestatística, muitos métodos de krigagem foram propostos e formalizados obedecendo a subdivisão em Krigagens lineares e não lineares. As Krigagens Lineares são assim denominadas porque os estimadores representam combinações lineares de seus parâmetros.

Landim (2011) destaca que até 1968 a Geoestatística foi utilizada para obter estimativas de reservas de hidrocarbonetos, e entre 1968 a 1970 foi fundamentada a Teoria da Krigagem Universal cujo nome foi dado por G. Matheron, em homenagem a D.G. Krige, para aplicação à cartografia submarina com tendência sistemática. De acordo com o autor, partir de 1970 muito tem se desenvolvido na Geoestatística, destacando sua ampla utilização no campo das ciências agrárias, geologia aplicada à agricultura de precisão e à preservação ambiental, dentre outros autores, cita-se (Ramirez & Colin, 1994; da Silva et al., 2012).

Huijbregts (1975) destaca que a Geoestatística explora a aparente aleatoriedade dos dados para avaliar as medidas de correlação espacial dos mesmos, considerando uma determinada vizinhança.

Dentre as técnicas de Krigagem linear as mais utilizadas são: a krigagem simples, a krigagem ordinária e a krigagem universal, enquanto que as krigagens não lineares, destacam-se a krigagem Disjuntiva, a krigagem Indicativa e a krigagem Lognormal (seção 2.3 do capítulo 2). Diversas aplicações do método de Krigagem Ordinária (KO) e Krigagem Indicativa (KI) são encontradas na agricultura, entre outros trabalhos, destacam-se (Imai et al., 2003; da Silva et al., 2012).

Os métodos geoestatísticos de Krigagem são baseados na função semivariograma a qual é calculada por meio de médias globais das semivariancias para cada distância entre os pontos. Estes métodos utilizam os valores estimados do semivariograma com a pressuposição de estacionaridade de 1^a ordem, por esse motivo a estacionaridade da média é condição essencial para se aplicar geoestatística. Atendida a pressuposição da estacionaridade da média, tem-se a garantia de variância mínima dos erros e média dos erros tendendo a zero.

Apesar da estacionaridade da média, denominada estacionaridade de primeira ordem, ser condição essencial para se aplicar geoestatística, muitos fatores tais como, custo de pesquisa, ausência de delineamento e a inexistência de método objetivo para medí-la, impedem ou dificultam comprovar a falta de estacionaridade para qualquer variável em diferentes tamanhos de malha amostral.

O método usual para medir a estacionaridade é o semivariograma, que é uma função baseada nas médias da dependência espacial em função das distâncias dentro de uma janela móvel. Por ser baseada em médias móveis, esta função não possibilita calcular os valores das dependências espaciais em todas as sub-regiões da malha amostral. A baixa dependência espacial pode ser devida a presença de dupla ou múltipla estacionaridade na estrutura de semivariancia, que exige ajuste de mais de um semivariograma, ficando um único modelo de semivariograma, inadequado para cobrir todas as sub-regiões da malha amostral.

Devido a dificuldade em detectar a estacionaridade de primeira ordem, ou sua ausência em algumas partes da malha amostral, a geoestatística tem sido usada, mesmo quando esta estacionaridade nãoé totalmente atendida, resultando em mapas de Krigagem de pouca acurácia.

Na expectativa de encontrar maneiras de melhorar a acurácia dos mapas de krigagem, muitos pesquisadores têm se dedicado ao estudo da qualidade dos mapas de krigagem e as relações existentes com o tamanho e comportamento amostral, como por exemplo, o trabalho de (Uribe-Opazo et al., 2012) mostrarando que o tamanho amostral não é fator determinante, pois, outras condições precisam ser atendidas para garantir a qualidade dos mapas.

Uma suposição sobre os fatores que interferem na qualidade dos mapas de krigagem podem estar associados ao comportamento dos dados. Tentativas de otimização da malha utilizando a geoestatística foram também propostas por (Oliver & Webster, 2015), porém, limitando-se ao estudo de caso específico para delineamentos planejados.

Muitos dos estudos feitos até então não avançaram em relação ao problema da perda de qualidade dos mapas de Krigagem em geral, por não associarem o problema dos mapas gerados com a possibilidade da presença de duas ou mais médias estacionárias, que causam perda da dependência espacial, devido a presença de pontos de mudança na estrutura espacial.

Page & Quintana (2016) apontaram problemas de acurácia em mapas de krigagem Ordinária os quais atribuíram aos problemas de agrupamento implícitos na estrutura de dependência espacial, que frequentemente não são levados em consideração nos modelos de predição, ficando tais mapas restritos a estruturas de covariância globais que mascaram tendências nos dados e produzem mapas suavizados.

Problemas como os descritos no parágrafo anterior foram abordados recentemente por Page & Quintana (2016) os quais propuseram modelos bayesianos baseadas em Modelos de Partição Produto(MPP), que modelando diretamente o agrupamento de locais em *clusters* espacialmente dependentes garantem que a configuração espacial em locais distantes tenham probabilidades pequenas, enquanto que em locais próximos tenham altas probabilidades. Para isto, propuseram modelos que geram distribuições a posteriori capazes de captar as características locais de cada cenário e estabelecer os locais de pontos de mudança estruturais na média, baseando-se nas distâncias entre as amostras.

A metodologia sugerida por Page & Quintana (2016) apresenta-se dificuldade computacional quando o tamanho da amostra a ser utilizada é grande, porque o artigo foi discutido por meio de exemplo utilizando um conjunto de poucos pontos amostrais fictícios fazendo-se uso de um método de partição aleatória. Porém uma partição aleatória requer testar todas as combinações possíveis das partes do conjunto de dados, o que torna inviável computacionalmente em bancos de dados grandes, como é o caso comumente utilizado em geoestatística.

Os Modelos de Partição Produto(MPP), foram utilizados até então para identificar pontos de mudança na média em conjuntos amostrais cujas variáveis se referiam a taxas de doença em dados de área (Barry & Hartigan, 1992) utilizando matriz de contiguidade de borda e em amostras ordenadas em função do tempo(Yao, 1984b; Loshi & Cruz, 2005), dentre outros casos envolvendo séries temporais.

Em modelos Bayesianos, θ é o parâmetro de interesse e o seu verdadeiro valor é desconhecido. Com o intuito de tentar reduzir o grau de incerteza que se tem sobre o verdadeiro parâmetro, modelos probabilísticos são propostos para descrever o comportamento de θ.

O método proposto nesta tese tem o objetivo de encontrar o local k do ponto de mudança na média μ utilizando MPPs, para isso, é apresentado o modelo

$$\begin{aligned} \theta &= \alpha + \beta_j X + \epsilon \\ X &= (x_i - x_k) \end{aligned}$$

em que α é a esperança de Y ser o ponto de mudança, β é o parâmetro que indexa a covariável X em que $X = (x_i - x_k)$ e ϵ é o erro aleatório.

No modelo, X, a variável regressora, refere-se a diferença entre as distâncias médias de cada ponto em relação aos seus vizinhos, colocada em ordem crescente, critério adotado de acordo com (Barry & Hartigan, 1992).

Para um valor fixo de Y, a função $l(\theta; y) = p(y|\theta)$ fornece a plausibilidade ou verossimilhança de cada um dos possíveis valores de θ, enquanto $p(\theta)$ representa as informações que se supõe conhecer a priori, chamada distribuição a priori de θ. Com estas duas fontes de informação, a informação a priori e a informação da verossimilhança são expressas como produto e a distribuição a posteriori de θ, $p(\theta|y)$ (Ehlers, 2011) é traduzida em

$$p(\theta|y) \propto p(Y = y|\theta)\, p(\theta)$$

em que $p(Y = y|\theta)$ representa a verossimilhança dos dados.

A distribuição a posteriori representa o produto da verossimilhança pela distribuição a priori, em que no modelo é descrita por:

No modelo com um ponto de mudança na média (corte),

$$p_i(\alpha, \beta_1, \beta_2, k|y) \propto \text{Verossimilhança} \times \text{prioris}$$

$$p_i(\alpha, \beta_1, \beta_2, k|y) \propto N(\mu_i, \sigma^2)\Gamma(0,001; 0,001)N(0,00; 0,001)N(0,1)cat(U\{1,2\})$$

No modelo com dois pontos de mudança na média (cortes),acrescenta-se β_3 e k_1, em que k e k_1 representam as duas posições dos pontos de mudança na média no espaço e β_3 refere-se ao coeficiente que indexa o terceiro grupo.

Na construção do modelo considera-se que $Y_i \sim Normal(\mu, \sigma^2)$, $i = \{1, \ldots, n\}$, a função de distribuição dos dados escrita por meio de um modelo hierárquico normal, μ representa uma regressão nos parâmetros da covariável X e J_i representa uma variável categórica ($cat\{1,2\}$) com $J_i = 1$, para o grupo 1 se $i \leq k$ e $J_i = 2$ para o grupo 2, se $i > k$, para o caso de suspeita de um ponto de mudança na média no espaço amostral, ou J_i representa uma variável categórica ($cat\{1,2,3\}$) que assume $J_i = 1$, para o grupo 1 se $i \leq k$, $J_i = 2$ para o grupo 2, se $i < k < k_1$ e $J_i = 3$ para o grupo 3, se $i \geq k_1$, no caso de suspeita de dois pontos de mudança na média amostral.

Para o modelo espacial hierárquico bayesiano de ponto de mudança (MPPs), θ é o parâmetro que representa a esperança do ponto de mudança da média amostral, regido pela covariável espacial X(distância média entre os vizinhos), com 95% de credibilidade, em que $X = x_1 - x_k$.

Fazendo a esperança $E[Y] = \alpha$ no ponto de mudança, os coeficientes β_1 e β_2, indexam os grupos antes do ponto de mudança e após o ponto de mudança, respectivamente, e obtém-se o modelo preditivo posteriori cuja fórmula é expressa por um modelo hierárquico que não tem forma analítica conhecida, apresentado em (1):

$$\begin{aligned} Y_i &\sim N(\mu, \sigma^2) \\ \mu &= \alpha + \beta_{J_i}(x_i - x_k) \\ J_i &= 1, \text{se} \qquad i \leq k \\ J_i &= 2, \text{se} \qquad i < k_1 \\ \sigma &= \tau^{-1} \end{aligned} \tag{1}$$

em que $y_i = \gamma_i$, vetor de resposta, é composto pelas semivariancias locais dos dados, a qual é assumida seguir distribuição normal. O modelo para dois pontos de mudança na média é similar ao modelo (1) feito para um ponto de mudança na média, acrescentando-se apenas os parâmetros que descrevem um terceiro grupo, como mostrado na expressão (64) do capítulo 4.

No modelo (1) são assumidas cinco priores, sendo duas delas não informativas que representam τ e α, uma priori informativa β que assume normal padrão, uma distribuição uniforme que descreve os valores de k e k_1 e uma uniforme categórica que descreve os grupos j e uma variável categorica, $cat(U\{1,2\})$, em que foram assumidas duas categorias, $\{1,2\}$. São utilizados também valores iniciais para k e k_1, valores iniciais para os parâmetros betas e um valor inicial para τ para gerar as distribuições. O local do ponto de mudança na média, k, é a variável aleatória de interesse estimada pelo modelo, uniformemente distribuída de $1, \ldots, n$.

Os modelos foram programados em linguagem R CORE Team (2015) e WINBUGS Lunn et al. (2000), em que se verifica no modelo (1), após ser obtido k, assume-se que as observações seguem duas distribuições normais independentes $y_i|\phi \sim N(\mu_1, \tau), i = 1, \ldots, k$ e $y_i|\gamma \sim N(\mu_2, \tau), i = k+1, \ldots, n$, em que ϕ e γ representam as distribuições, uma para cada parte da partição dos dados, onde o parâmetro de média da distribuição normal de cada uma das partes é descrito por um modelo de regressão $\mu_i = \alpha + \beta_{J_i}(x_i - x_k)$ e $\tau \sim Gama(\alpha, \beta_j)$.

Na modelagem de dois pontos de mudança na média, a quantidade de parâmetros a ser estimados pelo modelo aumenta, como é o caso de k e β que passa a ter dois parâmetros (k e k_1) e três parâmetros, (β_1, β_2 e β_3) que definem três grupos. Consequentemente, pode se verificar que aumentando o número de pontos de mudanças na média, aumenta não somente o número de parâmetros do modelo, mas também o tempo de processamento.

Após ser definir os grupos (partes da partição) feito pelo MPPs, é aplicada a geoestatística com ajuste dos semivariogramas às partes e um mapa de krigagem para cada uma, quando possível a ambas. Se apenas algumas das partes apresentam dependência espacial suficientes, ás partes com dependências espaciais insuficientes são aplicadas outras metodologias ou a mesma metodologia com condições limitadas e assumindo que estes mapas, sob tais condições, são pouco confiáveis.

Para o modelo com dois pontos de mudança na média a quantidade de parâmetros a ser estimados pelo modelo aumenta, como é o caso de k e β que passa a ter dois e três parâmetros, respectivamente. De acordo com as características do modelo (1), pode se observar que aumentando o número de pontos de mudanças na média, o número de parâmetros do modelo e o tempo de processamento também aumentam.

O local do ponto de mudança na média, k, é a variável aleatória de interesse estimada pelo modelo, uniformemente distribuída de $1, \ldots, n$.

Para o modelo (1), dado k, assume-se que as observações seguem duas distribuições normais independentes $y_i|\phi \sim N(\mu_1, \tau), i = 1, \ldots, k$ e $y_i|\gamma \sim N(\mu_2, \tau), i = k+1, \ldots, n$, em que ϕ e γ representam as distribuições, uma para cada parte da partição dos dados, onde o parâmetro de média da distribuição normal de cada uma das partes é descrito por um modelo de regressão $\mu_i = \alpha + \beta_{J_i}(x_i - x_k)$ e $\tau \sim Gama(\alpha, \beta_j)$.

Os fenômenos abrangidos pela Geoestatística são do tipo estruturados e se caracterizam por sua variável aleatória ser contínua no espaço e possuir dependência espacial, por isso a estrutura de semivariancia é considerada função do ajuste de parâmetros globais, assumindo assim a pressuposição de estacionaridade da média e em alguns casos, a finitude da variância e da covariância.

Os casos em que as pressuposições de estacionaridade de 1^a e 2^a ordens (apresentada com maiores detalhes no capítulo 2, na seção 2.2.1), quando válidas para todas as direções, são ditos fenômenos isotrópicos, raramente encontrados. Para muitos casos, estas pressuposições são assumidas, mas não são totalmente atendidas, sendo consideradas válidas por falta de um método objetivo de mensuração.

Os métodos de krigagem adotam o semivariograma ominidirecional que é uma média da estrutura global de semivariancia em todas as direções para cada distância entre pontos. No entanto, a semivariancia calculada nem sempre é a mesma em todas as direções e locais da área e quando a pressuposição de estacionaridade de primeira ordem não é totalmente atendida, os mapas gerados explicam muito pouco do fenômeno em estudo, embora feitos para cobrir toda a área porque quando a variável espacial tem aparentemente, mais de uma estrutura de covariância, ao se ajustar um modelo único de semivariograma estas estruturas não são levadas em consideração e mascaram a dependência espacial pela contaminação dos dados. Em malhas amostrais com estas características, ao se ajustar o semivariograma global ocorre uma suavização exagerada nos pontos estimados, provocada pela contaminação dos dados. Além disso, ao assumir um modelo único, impõe-se uma estrutura de covariância enganosa na estimação, levando a baixa acurácia nos mapas gerados por Krigagem.

A constatação apresentada no parágrafo anterior leva a suposição que o MPP possa ser adaptado a um método denominado Modelo de Partição Produto Espacial (MPPs), para identificar mudanças na estrutura espacial da média em dados com dependência espacial, ou seja com presença de duas ou mais médias estacionárias, que mascaram a estrutura de semivariancia espacial.

Na perspectiva de solucionar o problema apontado nos parágrafos anteriores, pretende-se identificar pontos de mudança na média que provocam a mudança da estrutura de semivariancia e estabelecer cortes na malha amostral, com o objetivo de melhorar a qualidade e acurácia dos mapas gerados por Krigagem Ordinária. Para isso, nesta tese é desenvolvido o MPP dando enfoque espacial, que denomina-se aqui, Modelo de Partição Produto espacial(MPPs), que é um método bayesiano, cuja teoria está fundamentada em (Smith, 1975) e primeiro uso prático foi proposto em sua forma geral por Hartigan (1990) para identificar pontos de mudanças na média.

O objetivo desta tese consiste em criar um novo Método de Partição Produto Espacial (MPPs) direcionado à aplicação de Krigagem Ordinária como uma metodologia para resolver o problema da dupla ou múltipla estacionaridade da média de uma variável espacialmente distribuída para garantir que os mapas de Krigagem tenham maior acurácia.

O MPP convencional garante que as médias de cada grupo sejam mais homogêneas, porém não garante que os grupos tenham dependência espacial, como proposta de garanti-la, sugere-se na tese, a transformação dos dados utilizando como variável resposta as semivariancias locais, cada uma correspondente a um ponto, e como covariável a distância média entre vizinhos, para que o ponto de mudança, sendo regido pela distancia média entre vizinhos, possa identificar o local da mudança da média em função da distancia entre as amostras. Os pontos de mudança na média serão os locais dos cortes na malha amostral que formam as partes da partição.

Esta tese foi pensada com o potencial de atingir várias áreas que utilizam-se da geoestatística e por isso, pretende-se propor uma metodologia para usuários de geoestatística e estatística espacial aplicada a diversos campos das ciências, sendo assim, a ordenação dos capítulos foi pensada com o propósito de apresentar a metodologia de forma simples, clara e prática.

A escolha de não adotar métodos clássicos de agrupamento, facilmente disponíveis em alguns trabalhos, deve-se ao fato que nestes métodos não se tem a garantia que os grupos terão dependência espacial e nem tão pouco que serão mais prováveis de obter melhores estimativas, por não se basearem em distribuições de probabilidades, enquanto, que o MPPs é um método bayesiano, por isso, baseia-se em distribuições de probabilidades por assumirem funções de probabilidade que geram os dados (de onde obtém-se a verossimilhança) e funções de probabilidade a priori para os parâmetros; podendo incorporar informações externas aos dados, chamadas informações a priori para se obter a distribuição a posteriori para as estimativas.

A aplicação da métodologia MPPs tem como primeiro passo a transformação da variável resposta em semivariancias locais e o uso da covariável espacial " Distancia Média entre Vizinhos". Essa transformação e a restrição do número de cortes à supostas quantidades de pontos de mudança na média, estabelecidas para permitir gerar agrupamentos com mais pontos amostrais localizados na mesma subregião do seu grupo e com dependência espacial, condição necessária para aplicação em geoestatística.

O conjunto completo da tese está dividido em seis capítulos: Introdução, Geoestatística, Método de Partição Produto, Material e métodos, Resultados e discussão e Conclusão, ficando no primeiro capítulo reservado à Introdução, o segundo capítulo incumbido de apresentar um estudo sobre Geoestatística iniciando-se com o tema "Processos Estocásticos Espaciais", seguindo por "Variáveis Regionalizadas", "Semivariograma" e finalizando com o estudo dos principais "Métodos de Krigagem".

No terceiro capítulo apresenta-se uma revisão sobre o Método de Partição Produto (MPP) em sua proposta original e espacial, com um tópico específico sobre pontos de mudança na média sob o enfoque bayesiano.

No quarto capítulo apresenta-se a Metodologia MPPs (Método de Partição Produto Spacial) aplicada ao método de Krigagem, que qualifica o método utlizado nesta tese.

O quinto capítulo apresenta os Resultados e Discussão do método aplicado a dados altimétricos e batimétricos, e a seguir, no sexto capítulo apresenta a Conclusão, em que são levantadas, também, algumas sugestões de trabalhos futuros.

Finalmente, apresenta-se todas as referências consultadas para construção da Tese.

6 CONCLUSÃO

Conclui-se que o método MPPs aplicado à krigagem é capaz de identificar os pontos de mudança da média e garantir grupos com médias mais estacionárias, assim, pode se afirmar que o método MPPs é mais viável de ser utilizado para se aplicar Krigagem Ordinária em comparação ao método de krigagem Ordinária clássico, quando a distribuição amostral dos dados apresentar características de se ter dependência espacial mas a estacionaridade de primeira ordem não for totalmente satisfeita e mostrar haver possibilidade de mais de uma estacionaridade da média, sendo o método indicado por permitir encontrar os pontos de mudança na média e identificar os locais mais prováveis de haver estas mudanças, formando grupos de médias mais estacionárias no espaço, ideais para produzir mapas de krigagem mais precisos e acurados.

6.1 Sugestões de trabalhos futuros

Diante da nova concepção que se forma ao inserir um método Bayesiano de partição para buscar grupos mais estacionários em termos de média e variância, abre-se o leque de possibilidades e uma das formas de ampliação é o seu uso em dados de saúde, em estudos epidemiológicos, relacionando covariáveis espaciais juntamente com covariáveis de saúde para se criar mapas regionalizados de melhor qualidade e favorecer políticas de prevenção e tratamento de epidemias e pandemias de forma otimizada (um trabalho já iniciado).

Outra possibilidade é testar o método para outros tipos de krigagem, inclusive a krigagem indicativa, utilizando outras funções a priori e de verossimilhança.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, M. F. F. Uso da Krigagem Indicativa na seleção de áreas propícias ao cultivo de café em consorciação ou rotação com outras culturas. Viçosa-MG, 2013. 137p. Dissertação (Mestrado) - Universidade Federal de Viçosa.

ANDRIOTTI, J. L. S. **Fundamentos de Estatística e Geoestatística**. São Leopoldo: UNISINOS:Coleção Manual Universitário, 2003. 165p.

BARRY, D.; HARTIGAN, J. A. Product partition models for change point problems. **The Annals of Statistics**, v.20, n.1, p.260–279, 1992.

BARRY, D.; HARTIGAN, J. A. A Bayesian analysis for change point problem. **Journal of the American Statistical Association**, v.88, n.421, p.309–319, 1993.

BASSETO, V. F.; GONZATTO, O. A. J.; ROSSONI, D. F.; JARDEL, M. H. Estimadores de Semivariância: Uma Revisão. **Ciência e Natura**, v.38, n.3, 2016.

BEAUMONT, L. J.; HUGHES, L.; PITMAN, A. Why is the choice of future climate scenarios for species distribution modelling important? **Ecology letters**, v.11, n.11, p.1135–1146, 2008.

BESAG, J.; YORK, J.; MOLLIÉ, A. Bayesianimage restoration, with two applications in spatial statistics. **Annals of the Institute of Statistical Mathematics**, v.43, n.1–20, p.21–59, 1991.

BHATTACHARYA, R. N.; WAYMIRE, E. C. **Stochastic processes with applications**. Philadélphia: Society for Industrial and Applied Mathematics, 1990. 691p.

BURROUGH, P. A. Development of intelligent geographical information systems. **International journal of geographical information systems**, v.6, n.1, p.1–11, 1992.

CÂMARA, G.; MONTEIRO, A. M.; FUCKS, S. D.; CARVALHO, M. S. Análise espacial e geoprocessamento. **Análise espacial de dados geográficos**, v.2, 2002.

CAMARGO, E. C. G. Geoestatística:fundamentos e aplicações. **Geoprocessamento para projetos ambientais. São José dos Campos:INPE**, 1998.

CHUNG, C. K.; CHONG, S.; VARSA, E. C. Sampling strategies for fertility on a stoy silt loam soil. **Communications in Soil Science and Plant Analysis**, v.26, n.5-6, p.741–763, 1995.

CLARKE, A. B.; DISNEY, R. L. **Probabilidade e Processos Estocásticos**. LTC, 1979.

CRESSIE, N.; HAWKINS, D. M. Robust estimation of the variogram: I. **Journal of the International Association for Mathematical Geology**, v.12, n.2, p.115–125, 1980.

CRESSIE, N. A. C. Statistics for spatial data: Wiley series in probability and mathematical statistics. **online**, 1993.

DEUTSCH, C. V.; JOURNEL, A. G. **GSLIB: Geostatistical Software Library and User's Guide. Hauptbd**. Oxford university press, 1992.

DEUTSCH, C. V.; JOURNEL, A. G.; ET AL. Geostatistical software library and users guide. **Oxford University Press, New York**, 1998.

EHLERS, R. S. Inferência bayesiana. **Departamento de Matemática Aplicada e Estatística, ICMC-USP**, p.64, 2011.

FELGUEIRAS, C. A. Modelagem ambiental com tratamento de incertezas em sistemas de informação geográfica: o paradigma geoestatístico por indicação. São José dos Campos, 1999. Tese (Doutorado) - INPE-Instituto Nacional de Pesquisas Espaciais.

FERNANDEZ, P. J. **Introdução a Teoria das Probabilidades**. Rio de Janeiro-RJ: Livros Técnicos e Científicos, 1973. 174p.

FERREIRA, I. O.; RODRIGUES, D. D.; DE P SANTOS, A. Levantamento batimétrico automatizado aplicado à gestão de recursos hídricos. Estudo de caso: Represamento do ribeirão São Bartolomeu, Viçosa -Mg. **Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação**, v.4, p.1–8, 2012.

FERREIRA, J. A.; LOSCHI, R. H.; COSTA, M. A. Detecting changes in time series: A product partition model with across-cluster correlation. **Signal Processing**, v.96, p.212–227, 2014.

GAMERMAN, D.; LOPES, H. F. **Markov Chain Monte Carlo: stochastic simulation for Bayesian inference**. Chapman and Hall/CRC, 2006.

GOOVAERTS, P. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. **Journal of hydrology**, v.228, n.1-2, p.113–129, 2000.

GUERRA, P. A. **Geoestatística Operacional**. Brasília: Ministério das Minas e Energia, 1988. 145p.

HARLAN, W. S. A quick derivation of geostatistical Kriging, 2013.

HARTIGAN, J. A. Partition models. **Communications in statistics-Theory and methods**, v.19, n.8, p.2745–2756, 1990.

HASLETT, J. On the sample variogram and the sample autocovariance for non-stationary time series. **Journal of the Royal Statistical Society: Series D (The Statistician)**, v.46, n.4, p.475–484, 1997.

HOLMES, C. C.; DENISON, D. G. T.; RAY, S.; MALLICK, B. K. Bayesian Prediction via Partitioning. **J. Comput. Graph. Statist**, v.14, n.1, p.811–830, 2005.

HUIJBREGTS, C. J. **Regionalized variables and quantitative analysis of spatial data. In: Davis, J.C. and McCullagh, M.J. (ed) Display and analysis of spatial data**. New York: John Wiley, 1975. 38-53p.

IBGE, A. M. S. Malha municipal digital do Brasil: situação em 2005. **Rio de Janeiro: IBGE**, 2006.

IMAI, N. N.; VICENTE, J.; LIMA, D. L. T.; VILMA, M.; SILVA, E. A.; VOLL, E.; OLIVEIRA Análise comparativa da interpolação por krigagem ordinária e krigagem por indicação no caso de ervas daninhas em cultura de soja. In: PROJETO MUDANÇA DO REFERENCIAL GEODÉSICO, 2003. ; resumos. Belo Horizonte: XXI Congresso Brasileiro de Cartografia. Publicação em CD-Rom sem paginação, 2003.

ISAAKS, E. H.; SRIVASTAVA, R. M. An introduction to applied geostatistics. Rel. téc., Oxford university press, 1989.

JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining geostatistics**. Academic press, 1978.

JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining Geostatistics**. Academic press, 1978.

JOURNEL, A. G.; JOURNEL, A. G. **Fundamentals of Geostatistics in Five Lessons**. American Geophysical Union Washington, DC, 1989. 8v.

KREH, M. Bessel functions. **Lecture Notes, Penn State-Göttingen Summer School on Number Theory**, v.82, 2012.

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

LANDIM, P. M. B. **Análise estatística de dados geológicos multivariados**. Oficina de Textos, 2011.

LOSHI, R. H.; CRUZ, F. R. B. Extension to the Product Partition Model Computing the Probability of a change. **Computational Statistics and data analisis**, v.48, n.2, p.255–268, 2005.

LUNN, D. J.; THOMAS, A.; BEST, N.; SPIEGELHALTER, D. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. **Statistics and computing**, v.10, n.4, p.325–337, 2000.

MANZIONE, R. L. Variabilidade espacial de atributos químicos do solo em Araguari-MG. Botucatu, 2002. 141p. Dissertação (Mestrado) - Universidade Estadual Paulista Júlio de Mesquita Filho- UNESP.

MCBRATNEY, A. B.; WEBSTER, R.; BURGESS, T. M. The design of optimal sampling schemes for local estimation and mapping of of regionalized variables—I: Theory and method. **Computers & Geosciences**, v.7, n.4, p.331–334, 1981.

MOOD, A.; GRAYBILL, F. A.; DBOES, C. **Introduction to the Theory of Statistics**. United State of America: McGraw-Hill, 1974. 480p.

MÜLLER, P.; QUINTANA, F.; ROSNER, G. Bayesian clustering with regression. Rel. téc., Working paper, 2008.

MÜLLER, P.; QUINTANA, F.; ROSNER, G. L. A product partition model with regression on covariates. **Journal of Computational and Graphical Statistics**, v.20, n.1, p.260–278, 2011.

OLEA, R. A. **Optimum mapping techniques using regionalized variable theory**. Kansas Geological Survey, 1975.

OLEA, R. A. A six-step practical approach to semivariogram modeling. **Stochastic Environmental Research and Risk Assessment**, v.20, n.5, p.307–318, 2006.

OLIVER, M. A.; WEBSTER, R. **Basic Steps in Geoestatistics: The variogram and Kriging**. University Reading, United Kigdom of Great Britain and North Ireland: Springer, 2015. 100p.

PAGE, G. L.; QUINTANA, F. A. Spatial Product Partition Models. **Bayesian Analysis**, v.11, n.1, p.265–298, 2016.

PARK, J.; DUNSON, D. B. Bayesian Generalized Product Partition Model. **National Cancer Institute and Duke University**, v.20, n.1, p.1203–1226, 12010.

QUINTANA, F. A. A predictive view of Bayesian clustering. **Journal of Statistical Planning and Inference**, v.136, n.8, p.2407–2429, 2006.

QUINTANA, F. A.; IGLESIAS, P. L. Bayesian Clustering and product partition models. **Journal of the Royal Statistical Society**, v.65, n.2, p.557–574, 2003.

RAMIREZ, J. J.; COLIN, R. T. Especie nueva del género Jatropha(Euphorbiaceae) de la sección Mozinna. In: ANALES DEL INSTITUTO DE BIOLOGÍA. SERIE BOTÁNICA, 65, Instituto de Biología, Universidad Nacional Autónoma de México, 1994. ; resumos. México: Anales del Instituto de Biología, Universidad Nacional Autónoma de México:Serie Botánica, 1994. 1.

REICH, B. J.; FUENTES, M.; ET AL. A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields. **The Annals of Applied Statistics**, v.1, n.1, p.249–264, 2007.

SALVIANO, A. A. C. **Variabilidade de atributos de solo e de Crotalaria juncea em solo degradado do municipio de Piracicaba-SP.**. ESALQ, 1996.

SILVA, A. F.; QUARTEZANI, W. Z.; ZIMBACK, C. R. L.; LANDIM, P. M. B. Aplicação da geoestatística em ciências agrárias. **Botucatu, SP: FEPAF**, 2011.

DA SILVA, A. F.; ZIMBACK, C. R. L.; LANDIM, P. M. B. Classificação de imagens em áreas cultivadas com citros por técnicas de sensoriamento remoto e geoestatística. **Energia na Agricultura**, v.27, n.3, p.01–15, 2012.

DA SILVA, A. P.; LIBARDI, P. L.; VIEIRA, S. R. Variabilidade espacial da resistência à penetração de um Latossolo Vermelho-Escuro ao longo de uma transeção. **Revista Brasileira de Ciência do solo**, v.13, n.1, p.1–5, 1989.

SMITH, A. A Bayesian approach to inference about a change-point in a sequence of random variables. **Biometrika**, v.62, n.2, p.407–416, 1975.

TEAM, R. C. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing, 2015.

TRANGMAR, B. B.; YOST, R. S.; WADE, M. K.; UEHARA, G.; SUDJADI, M. Spatial Variation of Soil Properties and Rice Yield on Recently Cleared Land 1. **Soil Science Society of America Journal**, v.51, n.3, p.668–674, 1987.

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

VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. **Tópicos em ciência do solo. In: Sociedade Brasileira de Ciência do Solo**, v.1, n.1, p.1–54, 2000.

WACKERNAGEL, H.; OLIVEIRA, V.; KEDEM, B. Multivariate geostatistics. **SIAM Review**, v.39, n.2, p.340–340, 1997.

WANG, X. J.; QI, F. The effects of sampling design on spatial structure analysis of contaminated soil. **Science of the total environment**, v.224, n.1-3, p.29–41, 1998.

WEBSTER, R.; OLIVER, M. A. Sample adequately to estimate variograms of soil properties. **Journal of soil science**, v.43, n.1, p.177–192, 1992.

YAMAMOTO, J. K. **Avaliação e Classificação de Reservas Minerais**. Edusp, 2001. 38v.

YAMAMOTO, J. K.; FURUIE, R. A. Um estudo sobre estimativa de dados lognormais. **Geociências (São Paulo)**, v.29, n.1, p.5–19, 2010.

YAMAMOTO, J. K.; LANDIM, P. M. **Geoestatstica: Conceitos e Aplicações**. Oficina de Textos, 2013.

YAMAMOTO, J. K.; MAO, X. M.; KOIKE, K.; CROSTA, A. P.; LANDIM, P. M. B.; HU, H. Z.; WANG, C. Y.; YAO, L. Q. Mapping an uncertainty zone between interpolated types of a categorical variable. **Computers & Geosciences**, v.40, p.146–152, 2012.

YAO, Y. Estimation of a noise discrete-time step function: Bayes and empirical. **Bayes approaches**, v.12, n.4, p.1434–1447, 1984a.

YAO, Y. Estimation of a noise discrete-time step function:Bayes and empirical Bayes approaches. **The annals of statistics**, v.12, n.4, p.1434–1447, 1984b.