Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Avaliação de modelos geoestatísticos multivariados

Ana Julia Righetto

Dissertação apresentada para obtenção do título de Mestre em Ciências. Área de concentração: Estatística e Experimentação Agronômica

Piracicaba
2012

Ana Julia Righetto
Estatístico

Avaliação de modelos geoestatísticos multivariados
versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **PAULO JUSTINIANO RIBEIRO Jr**

Dissertação apresentada para obtenção do título de Mestre em Ciências. Área de concentração: Estatística e Experimentação Agronômica

Piracicaba
2012

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - ESALQ/USP**

Righetto, Ana Julia

Avaliação de modelos geoestatísticos multivariados / Ana Julia Righetto. - - versão
revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2013.
92 p: il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2013.

1. Análise multivariada 2. Cokrigagem 3. Função exponencial 4. Geoestatística 5. Modelo
linear de correção regionalização I. Título

CDD 519.5
R571a

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

Dedicatória

Aos meus pais,
Edison e Luzia,
por nunca me deixarem desanimar.

AGRADECIMENTOS

Agradeço à Deus e Santo Expedito, sem os quais eu não teria fé suficiente para acreditar e alcançar meus sonhos.

Aos meus pais, Edison Righetto e Luzia Francisca Pedrazzi Righetto. Mais do que pais, vocês são meus amigos e sempre estiveram ao meu lado, me motivando e apoiando. Obrigada por nunca medirem esforços em me ajudar; ajuda que foi fundamental para mais esse passo na minha vida. Esta conquista é nossa. Amo vocês!

Ao meu grande amigo Luiz Ricardo Nakamura, pessoa de cara zangada e de um coração enorme. Obrigada pelas conversas, pelos conselhos, pelas risadas, pelos segredos e angustias compartilhadas. São muitos anos de amizade, companheirismo e eu só tenho a agradecer por você estar sempre presente.

Ao meu amigo e “irmão-acadêmico” Daniel Grados. Agradeço muito a Deus por ter te colocado nessa minha caminhada do mestrado. Você estava sempre pronto pra me ajudar; debatíamos sobre geoestatística na sala de estudo ou até mesmo pelo skype: eu no Brasil e você no Peru! Muito obrigada pela paciência, pela ajuda e mais do que isso, obrigada pela sua amizade.

Aos amigos da turma de mestrado Daniel, Ítalo, João Vitor, Kuang, Luiz Ricardo, Maria Joseane, Natalie Veronika e Patricia Ueda. Dividimos muitos momentos de angustias, em época de provas e seminários, mas também tivemos muitos momentos de lazer e alegria. Foi muito bom dividir esses momentos com vocês.

Aos amigos que Piracicaba e a ESALQ me proporcionaram, em especial: Mariana Urbano, Rodrigo Pescim, Pedro Henrique, Everton Batista, Thiago Gentil, Cássio Dessotti, Guilherme Biz, Renato Nunes, Ezequiel Lopez, Edilan Quaresma, Elizabeth Hashimoto, Simone Sartório, Fabiane Silva, Rafaelly Santos, Camila Rossetti, Lucas Santana, Tiago Santana, Débora de Godoi. Pessoas maravilhosas com as quais compartilhei momentos inesquecíveis. “Amizade é o amor que nunca morre.”

A todos os alunos do curso de Pós-graduação em Estatística e Experimentação Agrônômica da ESALQ/USP, com os quais eu tive o prazer de conviver.

À Nani e Rebecca, pelos momentos de diversão e alegria! Sempre fiéis!

Ao meu orientador Paulo Justiniano Ribeiro Júnior, muito obrigada pela orientação no decorrer deste trabalho. O senhor por muitas vezes enfatizou que um estatístico precisa programar e isso foi a grande luta no decorrer deste trabalho mas, graças a

sua ajuda e orientação, eu criei gosto em programar e isso eu devo à você. Obrigada pela confiança depositada. Foi muito satisfatório trabalhar com você!!

À todos os professores do Departamento de Ciências Exatas da ESALQ: Prof. Dra. Clarice Garcia Borges Demétrio, Prof. Dr. Edwin Moises Marcos Ortega, Prof. Dr. Gabriel Adrián Sarriés, Prof. Dra. Roseli Aparecida Leandro, Prof. Dr. Silvio Sandoval Zocchi e Prof. Dra. Sonia Maria De Stefano Piedade; por todo apoio, ensinamento, sabedoria e dedicação nesses anos. E em especial à Prof. Dra. Taciana Villela Savian, mais que uma professora, se mostrou uma grande amiga.

Aos professores do Departamento de Estatística da Faculdade de Ciências e Tecnologia de Presidente Prudente da “Universidade Estadual Paulista Júlio de Mesquita Filho”. Em especial a professora Vilma Mayumi Tachibana, minha orientadora de graduação, que sempre me motivou. Obrigada!

Aos funcionários do Departamento de Ciências Exatas da ESALQ/USP Eduardo Bonilha, Jorge Alexandre Wiendl e Rosni Honofre Aparecido Pinto pelo apoio técnico e pela amizade.

Às secretárias Luciane Brajão e Solange de Assis Paes Sabadin, pela amizade, pela confiança, pelo apoio constante ao longos desses anos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pelo apoio financeiro em forma de bolsa de estudos.

*"A mente que se abre a uma nova idéia
jamais voltará ao seu tamanho original."*

Albert Einstein

*"'Venha para a beira', disse ele.
Eles responderam: 'Nós temos medo'.
'Venha para a beira', disse ele.
Eles vieram.
Ele os empurrou ... e eles voaram."*

Apollinaire

SUMÁRIO

RESUMO	11
ABSTRACT	13
LISTA DE FIGURAS	15
LISTA DE TABELAS	17
1 INTRODUÇÃO	19
2 REVISÃO BIBLIOGRÁFICA	21
2.1 Geoestatística	21
2.1.1 Modelos Geoestatísticos Univariados	21
2.1.2 Estimação dos Parâmetros do Modelo	23
2.1.3 Variáveis Regionalizadas	25
2.1.3.1 Semivariograma	26
2.1.3.2 Krigagem	29
2.2 Geoestatística Multivariada	30
2.2.1 Cokrigagem	31
2.2.2 Modelos Lineares de Corregionalização	32
2.2.3 Semivariograma Cruzado	34
2.2.4 Modelos Geoestatísticos Bivariados Gaussianos	34
2.2.5 Modelo Gaussiano Bivariado com Componente Comum	35
2.2.6 Modelagem por Função de Correlação	37
2.2.6.1 Modelo Mátern Multivariado	38
2.2.6.1.1 Modelo Mátern Multivariado Parcimonioso	38
2.2.7 Geoestatística Bayesiana	40
3 MATERIAL E MÉTODOS	45
3.1 Estudo de Simulação	45
3.2 Descrição dos dados	46
4 RESULTADOS E DISCUSSÃO	49
4.1 Estudo de Simulação	49
4.2 Análise dos dados do Pacífico Noroeste Norte Americano	50
4.2.1 Análise exploratória	50
4.2.2 Análise univariada	52
4.2.3 Análise Bivariada	54
4.2.3.1 Análise pelo Modelo Linear de Corregionalização	54

4.2.3.2 Análise pelo Modelo Gaussiano Bivariado com Componente Comum	56
4.2.3.3 Análise pelo Modelo Matérn Multivariado Parcimonioso	57
4.2.3.4 Análise por métodos Bayesianos	58
4.2.4 Avaliação dos Modelos	62
5 CONSIDERAÇÕES FINAIS	71
REFERÊNCIAS	73
APÊNDICES	77
ANEXOS	87

RESUMO

Avaliação de Modelos Geoestatísticos Multivariados

Questões centrais em diversas áreas do conhecimento como ciências ambientais, geologia, agronomia, dentre outras, envolvem a compreensão da distribuição espacial de processos a partir de dados espacialmente referenciados. Os interesses de pesquisa podem estar na descrição espacial de duas ou mais variáveis e, desta forma, tem-se dois ou mais atributos para modelar. Modelos multivariados são propostos para o estudo se há evidências e/ou explicações contextuais de que os processos não são independentes. Diferentes modelos propostos na literatura foram avaliados e comparados ao modelo Matérn multivariado, recentemente proposto na literatura. Foram considerados o modelo linear de correionalização, o modelo bivariado gaussiano de componente comum e um modelo bayesiano de regressão espacial. Estes modelos foram ajustados e utilizados para predição espacial geoestatística (krigagem) em um conjunto de dados com duas variáveis climáticas no qual uma parte dos dados foi separada para avaliação das predições. Além disso, foi realizado um estudo de simulação para avaliar a estimação e predição sob o modelo Matérn multivariado.

Palavras-chave: Geoestatística Multivariada; Cokrigagem; Modelo Linear de Correionalização; Modelo Matérn Multivariado

ABSTRACT

Evaluation of Multivariate Geostatistic Models

Key issues in a diversity of subject areas such as environmental sciences, geology, agronomy, among other, require the understanding of the spatial distribution of natural processes from spatially referenced data. Research interests may include the spatial description of two or more variables and therefore, there are two or more attributes to be modeled. Multivariate models are adopted when there is evidence and/or contextual explanations the two processes are not independent. Different models presented in the literature are assessed and compared to the recently introduced multivariate Matérn model. The linear model of coregionalization, the bivariate Gaussian common component model and a bayesian spatial regression model were considered. The models were fitted and used for geostatistical spatial prediction (kriging) for a pair of weather related variables with part of the data used only for comparing the predictions. Additionally a simulation study assessed estimation and prediction under the multivariate Matérn model.

Keywords: Multivariate Geostatistics; Cokriging; Linear Model of Coregionalization; Multivariate Matérn Model

LISTA DE FIGURAS

Figura 1 - Pacífico Noroeste Norte Americano com as 157 observações. FONTE: Gneiting, T.; Kleiber, W.; Schlather, M., 2009	46
Figura 2 - Gráfico de círculos - (a) para a variável pressão e (b) para variável tem- peratura.	50
Figura 3 - Gráfico descritivo do padrão espacial - (a) da variável pressão e (b) da variável temperatura	52
Figura 4 - Gráfico dos envelopes do variograma - (a) da variável pressão e (b) da variável temperatura	52
Figura 5 - Localização dos 52 pontos retirados (em vermelho)	53
Figura 6 - Gráfico dos variogramas experimental e teórico - (a) da variável pressão e (b) da variável temperatura	53
Figura 7 - Gráfico das predições espaciais (krigagem) univariada - (a) da variável pressão e (b) da variável temperatura	54
Figura 8 - Gráfico dos variogramas univariados e do variograma cruzado	54
Figura 9 - Gráfico das predições espaciais (krigagem) após o ajuste do MLC - (a) da variável pressão e (b) da variável temperatura	55
Figura 10 - Gráfico das predições espaciais (krigagem) para os 52 pontos retirados da análise - (a) da variável pressão e (b) da variável temperatura	56
Figura 11 - Gráfico das predições espaciais (krigagem), em uma malha de pontos, pelo BGCCM - (a) da variável pressão e (b) da variável temperatura	57
Figura 12 - Gráfico das predições espaciais (krigagem) em uma malha de pontos pelo modelo Matérn multivariado parcimonioso - (a) da variável pressão e (b) da variável temperatura	58
Figura 13 - Gráfico do traço e da densidade do intercepto e de parte da matriz K	59
Figura 14 - Gráfico do traço e da densidade da matriz K e de parte da matriz Ψ	60
Figura 15 - Gráfico do traço e da densidade da matriz Ψ e do ϕ_1	60
Figura 16 - Gráfico do traço e da densidade do ϕ_2	61
Figura 17 - Gráfico de krigagem das variáveis em estudo e para a predição das mes- mas no modelo bayesiano - primeira variável (pressão), segunda variável (temperatura)	61

Figura 18 -Gráfico de krigagem das variáveis do grupo controle e da predição das mesmas no modelo bayesiano - primeira variável (pressão), segunda va- riável (temperatura)	62
Figura 19 -Gráfico de dispersão da variável pressão versus valores preditos pelos modelo: (a)univariado, (b)LMC, (c)BGCCM, (d)Matérn Parcimonioso, (e)Bayesiano	67
Figura 20 -Gráfico de dispersão da variável temperatura versus valores preditos pelos modelo: (a)univariado, (b)LMC, (c)BGCCM, (d)Matérn Parcimonioso, (e)Bayesiano	68

LISTA DE TABELAS

Tabela 1 - Estatísticas descritivas das estimativas dos parâmetros obtidas com 1001 simulações	49
Tabela 2 - Estatísticas descritivas do erro quadrático (EQ) das variáveis pelo estudo de simulação	50
Tabela 3 - Estatísticas descritivas do erro absoluto (EA) das variáveis pelo estudo de simulação	50
Tabela 4 - Estatística descritiva das variáveis em estudo, com variável pressão transformada	51
Tabela 5 - Valores do parâmetros estimados no modelo univariado	53
Tabela 6 - Valores do parâmetros estimados para o variograma cruzado pelo modelo linear de correionalização	55
Tabela 7 - Valores do parâmetros estimados para o modelo Gaussiano Bivariado com Componente Comum	56
Tabela 8 - Valores do parâmetros estimados para o modelo Matérn multivariado parcimonioso	57
Tabela 9 - Quantis e Média para cada parâmetro por método bayesiano	59
Tabela 10 - Valor observado e valores preditos, do grupo controle, da variável Pressão	63
Tabela 11 - Valor observado e valores preditos, do grupo controle, da variável Temperatura	65
Tabela 12 - Média e desvio padrão dos erros de krigagem do grupo controle para Pressão	66
Tabela 13 - Média e desvio padrão dos erros de krigagem do grupo controle para Temperatura	66
Tabela 14 - Medidas de Pressão e Temperatura em 157 localizações (coordenadas) no Pacífico Noroeste Norte Americano	89

1 INTRODUÇÃO

Compreender a distribuição espacial de dados provenientes de fenômenos ocorridos no espaço, constitui um grande desafio para esclarecer questões centrais em diversas áreas do conhecimento, seja na parte da saúde, do ambiente, de geologia, na agronomia, entre tantas outras áreas. Tais estudos tornam-se cada vez mais comuns, devido a avanços recentes no Sistema de Informação Geográfica (SIG) e no Sistema de Posicionamento Global (GPS), que são sistemas relacionados a informação espacial bem como avanço nos procedimentos computacionais que permitem e facilitam a análise ou representação do espaço e dos fenômenos que nele ocorrem. Assim, é crescente o interesse em analisar e ajustar modelos para dados georreferenciados.

Os métodos de estatística espacial têm o intuito de mensurar propriedades e relacionamentos, levando em conta a localização espacial do fenômeno estudado de forma explícita. Essa área da estatística é dividida em três sub-áreas de estudo: geoestatística, dados de área e processos pontuais. Como o que distingue cada uma dessas categorias é o tipo de dado utilizado, é normal que existam métodos estatísticos diferentes para analisar e descrever os dados em cada uma destas três áreas. Neste trabalho será enfatizado a geoestatística, que une o conceito de variáveis aleatórias com o conceito de variáveis regionalizadas, sendo estas últimas, variáveis que possuem um comportamento espacial. Este ramo da estatística espacial, incorpora tanto a interpretação da distribuição estatística como a correlação espacial das amostras. Segundo Matheron (1963), geoestatística é a aplicação do formalismo de funções aleatórias para o reconhecimento e estimação de fenômenos naturais.

A modelagem estatística é utilizada em muitas áreas, com o intuito de descrever o comportamento de um ou mais atributos que não podem ser descritos exclusivamente por modelos determinísticos. No geral, modelos estatísticos explicam, o máximo possível, a variabilidade dos processos estocásticos através de uma ou mais variáveis explanatórias que possuam alguma associação com a resposta de interesse. Já a modelagem geoestatística se diferencia pois as observações não são assumidas independentes e existe efeito aleatório latente na parte explanatória do modelo. Nesta modelagem, existe um conjunto de técnicas para encontrar uma função aleatória para um ou mais atributos dos quais suas localizações espaciais são conhecidas. Desta forma, essas técnicas são importantes para capturar a correlação entre as observações das variáveis aleatórias de interesse, em que exista uma forte suspeita de que pontos espaciais próximos possuam valores parecidos da

variável em estudo.

Em muitos casos, o interesse não está em apenas uma variável mas em estudar duas ou mais variáveis e desta forma, tem-se dois ou mais atributos para modelar. Ocorrendo evidências de que esses processos não são independentes e existindo explicação prática para isto, modelos multivariados devem ser propostos para o estudo. Muitas vezes, em estudos na geoestatística, se faz necessário a modelagem de várias variáveis espaciais simultaneamente. O passo crítico é então identificar a estrutura de dependência espacial, não somente em cada variável como também entre as variáveis e assim, trabalha-se com modelos geoestatísticos multivariados. Estes modelos, em geral, possuem uma estrutura complexa combinando mais de um processo latente com elevado número de parâmetros e que podem apresentar problemas para estimação paramétrica.

Na metodologia, existem vários modelos geoestatísticos multivariados. O modelo que motivou a realização deste trabalho, foi desenvolvido por Gneiting, Kleiber e Schlather (2009) e consiste no ajuste por estrutura de covariância e covariância cruzada para campo aleatório gaussiano, utilizando como função de correlação a função Matérn multivariada, que é uma função bastante flexível. Em um artigo intitulado “Matérn Cross-Covariance Functions for Multivariate Random Fields”, os autores comparam este modelo com o modelo linear de correionalização. Neste trabalho um conjunto de dados é analisado por diversos modelos geoestatísticos multivariados e avaliar qual modelo que melhor explica os dados. Além disso, o modelo proposto por Gneiting, Kleiber e Schlather (2009) foi implementado pelos autores no pacote RandomFields (Schlather, 2012) do software R (R Development Core Team, 2006) e neste trabalho, realizou-se uma reimplementação deste modelo no mesmo software.

Os dados utilizados referentes a observações de temperatura e de pressão em 157 localizações no Pacífico Noroeste Norte Americano estão disponíveis no pacote RandomFields (Schlather, 2012). Os dados foram ajustados com os diferentes modelos utilizados neste trabalho. Além disso, um estudo de simulação foi realizado com o modelo Matérn parcimonioso com o propósito de avaliar a qualidade das estimativas dos parâmetros e das previsões das variáveis.

2 REVISÃO BIBLIOGRÁFICA

2.1 Geoestatística

De acordo com Diggle, Ribeiro Jr e Christensen (2003), a geoestatística é um ramo da estatística espacial, em que os dados consistem de mensurações $y_1, \dots, y_n$ obtidas em localizações $\mathbf{x}_1, \dots, \mathbf{x}_n$, amostradas de uma região A espacialmente contínua, sendo que $A \subset \mathbb{R}^d$. Estas medidas estão associadas a um fenômeno espacial, tratado como a realização de um processo estocástico $S(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$, que não é diretamente observado, dado por qualquer conjunto de variáveis aleatórias $Y(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$ definidas em um espaço de probabilidade comum, em que d é o número de linhas de $\mathbf{x}$, no qual os elementos são as localizações amostrais. Além disso, segundo Soares (2006), um processo estocástico é um modo de abordagem do conjunto de dados provenientes das amostras experimentais. Então, o primeiro passo da modelagem geoestatística é a concepção de um processo aleatório que melhor caracteriza o conjunto de dados experimentais das observações.

Diggle, Tawn e Moyeed (1998) introduziram o termo “geoestatística baseada em modelos”, na qual a modelagem é realizada especificando um modelo paramétrico para a(s) variável(véis) em estudo, e utiliza-se métodos de inferência estatística baseada na função de verossimilhança aplicada a estudos geoestatísticos.

2.1.1 Modelos Geoestatísticos Univariados

Dos modelos geoestatísticos em que $S(\mathbf{x})$ é contínuo e diferenciável, o mais simples é o Gaussiano estacionário, cujas suposições são definidas por Diggle e Ribeiro Jr (2007):

- i. $\{S(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^2\}$ é um processo Gaussiano com média μ e variância $\sigma^2 = Var\{S(\mathbf{x})\}$ e função de correlação $\rho(h) = Corr\{S(\mathbf{x}), S(\mathbf{x}')\}$, em que $h = \|\mathbf{x} - \mathbf{x}'\|$ e $\|\cdot\|$ representa a distância entre as localizações;
- ii. Condicional a $\{S(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^2\}$, y_i são as realizações mutuamente independentes da variável aleatória Y_i , com distribuição normal, média condicionada dada por $E[Y_i|S(\cdot)] = S(\mathbf{x}_i)$ e variância condicional τ^2 .

Considerando que na região de estudo A exista um campo aleatório Gaussiano latente, S , é necessário realizar uma amostragem, em n localizações, para obter valores da variável de interesse. Desta forma, cria-se um vetor $\mathbf{Y}(\mathbf{x})$ de dimensão $n \times 1$, dos valores

observados nas localizações amostradas $\mathbf{x} = (x_1, \dots, x_n)$. Assim, obtem-se o modelo:

$$Y(\mathbf{x}) = \mu + S(\mathbf{x}) + \epsilon; \quad (1)$$

em que ϵ é o ruído branco, um vetor de dimensão $n \times 1$, com distribuição normal de média zero e desvio padrão τ ; $S(\mathbf{x})$ é um campo aleatório gaussiano, definido nas suposições acima, que possui vetor de médias nulo, de dimensão $n \times 1$ e matriz de covariância Σ , $n \times n$ e $\mu = \mathbf{D}\boldsymbol{\beta}$, sendo $\mathbf{D}$ uma matriz de dimensão $n \times q$, contendo $q - 1$ covariáveis e $\boldsymbol{\beta}$ é um vetor $q \times 1$ de parâmetros associados a essas covariáveis. Ainda, segundo Martins (2010), esse vetor de médias pode ser escrito por:

$$\mu = \beta_0 + \sum_{j=1}^q \beta_j D_j.$$

Utilizando o modelo (1), pode-se encontrar a distribuição de probabilidade de $Y(\mathbf{x})$, com vetor de médias $\mathbf{D}\boldsymbol{\beta}$ e matriz de covariâncias dada por $\Sigma_Y = \Sigma + \tau^2 I$, em que I é uma matriz identidade $n \times n$. Então, tem-se um vetor a ser estimado, $\boldsymbol{\theta} = (\beta, \sigma^2, \boldsymbol{\phi}^*, \tau^2)$, no qual $\boldsymbol{\phi}^*$ é um vetor de parâmetros associados a função de correlação utilizada.

Segundo Martins (2010), quando a variável observada $Y(\mathbf{x})$ não segue uma distribuição Gaussiana, o valor do erro quadrático médio mínimo do preditor é afetado, podendo assim, obter aproximações ruins. Para contornar esse problema, realiza-se uma transformação nos dados, para que estes passem a seguir uma distribuição Gaussiana. Uma transformação muito conhecida, foi proposta por Box e Cox (1964):

$$Y^*(\mathbf{x}) = \begin{cases} \frac{Y^\lambda(\mathbf{x}) - 1}{\lambda}, \lambda \neq 0 \\ \log[Y(\mathbf{x})], \lambda = 0. \end{cases}$$

no qual o parâmetro λ introduz flexibilidade ao modelo. Alguns valores interpretáveis desse parâmetro são:

$$\begin{cases} \lambda = 1 : & \text{sem transformação} \\ \lambda = 0,5 : & \text{transformação raiz quadrada} \\ \lambda = -1 : & \text{transformação recíproco} \end{cases}$$

2.1.2 Estimação dos Parâmetros do Modelo

Para estimação dos parâmetros do modelo Gaussiano utiliza-se o método dos mínimos quadrados ou o método da máxima verossimilhança. Neste último, admitindo-se uma tendência polinomial para $\mu(\mathbf{x})$, tem-se:

$$\mathbf{Y}(\mathbf{x}) \sim N_n(\mathbf{D}\boldsymbol{\beta}; \sigma^2\mathbf{R} + \tau^2\mathbf{I}), \quad (2)$$

em que $\mathbf{D}$ é a matriz $n \times q$ de covariáveis, $\boldsymbol{\beta}$ é o vetor de parâmetros e $\mathbf{R}$ é a matriz de correlação que depende de ϕ e da distância.

Desta forma, a função de verossimilhança da variável dada em (2) é expressa por:

$$L[\mathbf{Y}(\mathbf{x})] = (2\pi)^{-\frac{n}{2}} |\sigma^2\mathbf{R} + \tau^2\mathbf{I}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{Y}(\mathbf{x}) - \mathbf{D}\boldsymbol{\beta})' (\sigma^2\mathbf{R} + \tau^2\mathbf{I})^{-1} (\mathbf{Y}(\mathbf{x}) - \mathbf{D}\boldsymbol{\beta}) \right\} \quad (3)$$

e o logaritmo da função de verossimilhança é dado por:

$$\begin{aligned} l(\boldsymbol{\beta}, \tau^2, \sigma^2, \phi) &= \ln[(2\pi)^{-\frac{n}{2}}] - \ln(|\sigma^2\mathbf{R} + \tau^2\mathbf{I}|^{\frac{1}{2}}) - \\ &\quad - \frac{1}{2} (\mathbf{Y}(\mathbf{x}) - \mathbf{D}\boldsymbol{\beta})' (\sigma^2\mathbf{R} + \tau^2\mathbf{I})^{-1} (\mathbf{Y}(\mathbf{x}) - \mathbf{D}\boldsymbol{\beta}) \\ &= -\frac{1}{2} \{ n \ln(2\pi) + \ln(|\sigma^2\mathbf{R} + \tau^2\mathbf{I}|) + \\ &\quad + (\mathbf{Y}(\mathbf{x}) - \mathbf{D}\boldsymbol{\beta})' (\sigma^2\mathbf{R} + \tau^2\mathbf{I})^{-1} (\mathbf{Y}(\mathbf{x}) - \mathbf{D}\boldsymbol{\beta}) \} \end{aligned}$$

Para a maximização do log-verossimilhança, considera-se $\mathbf{Y}(\mathbf{x}) \sim N_n(\mathbf{D}\boldsymbol{\beta}; \sigma^2\mathbf{V})$ no qual tem-se a reparametrização na variância:

$$\sigma^2\mathbf{R} + \tau^2\mathbf{I} = \sigma^2 \left(\mathbf{R} + \frac{\tau^2}{\sigma^2} \mathbf{I} \right) = \sigma^2 (\mathbf{R} + v^2\mathbf{I}) = \sigma^2\mathbf{V}$$

Assim, pode-se reescrever o log-verossimilhança como:

$$l(\boldsymbol{\beta}, \sigma^2) = -\frac{1}{2} \{ n \ln(2\pi) + \ln(|\sigma^2\mathbf{V}|) + (\mathbf{Y}(\mathbf{x}) - \mathbf{D}\boldsymbol{\beta})' (\sigma^2\mathbf{V})^{-1} (\mathbf{Y}(\mathbf{x}) - \mathbf{D}\boldsymbol{\beta}) \}. \quad (4)$$

Em (4) tem-se que:

$$\begin{aligned} (\mathbf{Y}(\mathbf{x}) - \mathbf{D}\boldsymbol{\beta})' (\sigma^2\mathbf{V})^{-1} (\mathbf{Y}(\mathbf{x}) - \mathbf{D}\boldsymbol{\beta}) &= \mathbf{Y}'(\mathbf{x}) (\sigma^2\mathbf{V})^{-1} \mathbf{Y}(\mathbf{x}) - \mathbf{Y}'(\mathbf{x}) (\sigma^2\mathbf{V})^{-1} \mathbf{D}\boldsymbol{\beta} - \\ &\quad - \boldsymbol{\beta}' \mathbf{D}' (\sigma^2\mathbf{V})^{-1} \mathbf{Y}(\mathbf{x}) + \boldsymbol{\beta}' \mathbf{D}' (\sigma^2\mathbf{V})^{-1} \mathbf{D}\boldsymbol{\beta}. \end{aligned}$$

Então, pode-se reescrever (4) da seguinte forma:

$$l(\boldsymbol{\beta}, \sigma^2) = -\frac{1}{2} \left\{ n \ln(2\pi) + \ln(|\sigma^2 \mathbf{V}|) + \frac{1}{\sigma^2} \left[\mathbf{Y}'(\mathbf{x}) \mathbf{V}^{-1} \mathbf{Y}(\mathbf{x}) - 2 \mathbf{Y}'(\mathbf{x}) \mathbf{V}^{-1} \mathbf{D} \boldsymbol{\beta} + \boldsymbol{\beta}' (\mathbf{D}' \mathbf{V}^{-1} \mathbf{D}) \boldsymbol{\beta} \right] \right\} \quad (5)$$

Segundo Martins (2010), deve-se considerar os seguintes resultados da álgebra matricial:

$$\frac{\partial}{\partial \mathbf{T}} (\mathbf{C} \mathbf{T}) = \mathbf{C}' \quad (6)$$

$$\frac{\partial}{\partial \mathbf{T}} (\mathbf{T}' \mathbf{C} \mathbf{T}) = 2 \mathbf{C} \mathbf{T} \quad (7)$$

em que $\mathbf{C}$ é uma matriz quadrada de ordem $n \times n$ e $\mathbf{T}$ é um vetor de ordem $n \times 1$.

Desta forma, aplicando-se (6) e (7) em (5), tem-se:

$$\frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta}, \sigma^2) = -\frac{1}{\sigma^2} \left[-(\mathbf{Y}'(\mathbf{x}) \mathbf{V}^{-1} \mathbf{D})' + (\mathbf{D}' \mathbf{V}^{-1} \mathbf{D}) \boldsymbol{\beta} \right].$$

Considerando:

$$\frac{1}{\sigma^2} (\mathbf{D}' \mathbf{V}^{-1} \mathbf{Y}(\mathbf{x}) - \mathbf{D}' \mathbf{V}^{-1} \mathbf{D} \hat{\boldsymbol{\beta}}) = 0,$$

tem-se que o estimador de máxima verossimilhança de $\boldsymbol{\beta}$ é dado por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{D}' \mathbf{V}^{-1} \mathbf{D})^{-1} \mathbf{D}' \mathbf{V}^{-1} \mathbf{Y}(\mathbf{x}). \quad (8)$$

Observa-se na equação (8) que o estimador de $\boldsymbol{\beta}$ só depende de ϕ e v^2 . Por outro lado, para obter o estimador de σ^2 , considera-se que $|\text{Var}[\mathbf{Y}(\mathbf{x})]| = |\sigma^2 \mathbf{V}| = (\sigma^2)^n |\mathbf{V}|$. Assim, a partir desse resultado, o logaritmo da função de verossimilhança em (4) pode ser reescrito como:

$$l(\boldsymbol{\beta}, \sigma^2) = -\frac{1}{2} \left\{ n \ln(2\pi) + n \ln(\sigma^2) + \ln |\mathbf{V}| + \frac{1}{\sigma^2} (\mathbf{Y}(\mathbf{x}) - \mathbf{D} \boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{Y}(\mathbf{x}) - \mathbf{D} \boldsymbol{\beta}) \right\}. \quad (9)$$

Derivando parcialmente (9) com relação a σ^2

$$\frac{\partial}{\partial \sigma^2} l(\boldsymbol{\beta}, \sigma^2) = -\frac{1}{2} \left\{ \frac{n}{\sigma^2} + (\mathbf{Y}(\mathbf{x}) - \mathbf{D}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{Y}(\mathbf{x}) - \mathbf{D}\boldsymbol{\beta}) \left(-\frac{1}{(\sigma^2)^2} \right) \right\}$$

e igualando esta derivada parcial a zero, o estimador de σ^2 é:

$$\hat{\sigma}^2 = n^{-1} \left[\mathbf{Y}(\mathbf{x}) - \mathbf{D}\hat{\boldsymbol{\beta}} \right]' \mathbf{V}^{-1} (\mathbf{Y}(\mathbf{x}) - \mathbf{D}\hat{\boldsymbol{\beta}}). \quad (10)$$

Substituindo (8) e (10) em (4), obtém-se o log-verossimilhança concentrada nos parâmetros v^2 e ϕ , dada por:

$$l_c(\phi, v^2) = -\frac{1}{2} \{ n \ln(2\pi) + n \ln(\hat{\sigma}^2) + \ln|\mathbf{V}| + n \}. \quad (11)$$

Assim, para obter as estimativas dos parâmetros, a equação (11) deve ser otimizada numericamente com relação a ϕ e v^2 . Em seguida obtém-se a matriz $\hat{\mathbf{V}}$, que ao ser substituída em (8) retorna $\hat{\boldsymbol{\beta}}$ e, conseqüentemente, retorna $\hat{\sigma}^2$ ao ser substituída em (10). Logo, substituindo $\hat{v}^2$ e $\hat{\sigma}^2$ em $v^2 = \frac{\tau^2}{\sigma^2}$ estima-se $\hat{\tau}^2$.

2.1.3 Variáveis Regionalizadas

Segundo Landim (2006), o termo geoestatística está relacionado como um tópico da estatística aplicada em que se trabalha com problemas envolvendo variáveis regionalizadas. Estas variáveis possuem um comportamento espacial, com características intermediárias entre as variáveis totalmente determinísticas e as verdadeiramente aleatórias.

As variáveis regionalizadas possuem duplo aspectos contraditórios, pois é impossível uma previsão exata do valor da variável em um determinado ponto (aspecto aleatório), mas é esperado que o valor da variável seja parecido para pontos próximos (aspecto estrutural). Na teoria das variáveis regionalizadas, $Y(\mathbf{x})$ pode ser definida como uma variável aleatória que assume diferentes valores Y em função da posição $\mathbf{x}$ dentro de uma região A . Esta teoria tem por objetivos, segundo Yamamoto (2001), estudar e representar as propriedades estruturais dessas variáveis para a resolução de problemas de estimativa.

O conjunto dessas variáveis $Y(\mathbf{x})$ medidas na totalidade da região em estudo, pode ser considerada como uma função aleatória $Y(\mathbf{x})$ pois, conforme Isaaks e Srivastava

(1989), são variáveis aleatórias regionalizadas e assim, assume-se que a dependência entre elas seja especificada por algum mecanismo probabilístico. Segundo Journel e Huijbregts (1978), a interpretação probabilística da variável regionalizada $Y(\mathbf{x})$ ser uma particular realização de alguma função aleatória $Y(\mathbf{x})$ é válida quando pode-se inferir toda ou parte da lei de distribuição de probabilidade que define esta função aleatória. Porém, na prática, em cada localização $\mathbf{x}$ tem-se somente uma realização $Y(\mathbf{x})$ e, como o número de pontos é finito, fica usualmente impossível inferir sobre a distribuição de $Y(\mathbf{x})$. Assim, hipóteses que envolvem diferentes graus de homogeneidade espacial, denominadas de hipóteses de estacionariedade, se fazem necessárias.

As hipóteses de estacionariedade são (GRIPP, 1992):

- 1) Hipótese de estacionariedade restrita: a função aleatória $Y(\mathbf{x})$ é estacionária se a lei de probabilidades de $Y(\mathbf{x})$ for a mesma em qualquer localidade $\mathbf{x}$ considerada.
- 2) Hipótese de estacionariedade de ordem 2: a função aleatória $Y(\mathbf{x})$ é estacionária de ordem 2 se $E[Y(\mathbf{x})] = m$, isto é, a esperança matemática de $Y(\mathbf{x})$ é constante e não depende de $\mathbf{x}$.

Para o estudo do comportamento das variáveis regionalizadas, é usual utilizar ferramentas dos métodos geoestatísticos como o semivariograma e a krigagem.

2.1.3.1 Semivariograma

Considerando uma variável regionalizada $Y(\mathbf{x})$ amostrada em diversos pontos que estão regularmente distribuídos em uma região de estudo, tem-se que o valor de cada ponto está relacionado, de alguma maneira, com valores obtidos a partir de pontos situados a uma certa distância. Seguindo o princípio da Primeira Lei da Geografia de Tobler que diz: “Todas as coisas são relacionadas, porém coisas mais próximas são mais parecidas que coisas distantes”, pode-se dizer que a influência entre os pontos é tanto maior quanto menor for a distância entre eles.

Para expressar essa influência é definido o vetor de distância $\Delta\mathbf{h}$, que possui uma orientação específica e o grau de relação entre pontos em uma certa direção pode ser expresso pela covariância e, embora a covariância exista entre todas as distâncias possíveis ao longo de $\mathbf{h}$, pode ser estipulado que somente sejam considerados valores entre pontos regularmente espaçados por múltiplos inteiros de $\Delta\mathbf{h}$ (LANDIM, 2006).

A covariância calculada entre os valores encontrados, separados pela distância $\Delta \mathbf{h}$ ao longo de $\mathbf{h}$, é expressa por:

$$C(\mathbf{h}) = C(\Delta \mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum Y(\mathbf{x})Y(\mathbf{x} + \mathbf{h}) - m^2,$$

em que m é a média da variável regionalizada e $N(\mathbf{h})$ é o número de pares de valores medidos separados pela distância $\mathbf{h}$.

A covariância depende do tamanho de $\mathbf{h}$. Então, se $\mathbf{h} = \mathbf{0}$, $C(\mathbf{h})$ representará a variância, isto é

$$C(0) = E[Y^2] - E^2[Y] = Var[Y]$$

Desta forma, calcula-se a função denominada de semivariância, dada pela metade da variância das diferenças, representada por $\gamma(\mathbf{h})$, em que

$$\gamma(\mathbf{h}) = \frac{1}{2N} \sum_{i=1}^{N(\mathbf{h})} [Y(\mathbf{x}_i + \mathbf{h}) - Y(\mathbf{x}_i)]. \quad (12)$$

Para analisar como os valores dos dados estão se comportando conforme várias distâncias estipuladas e assim, analisar o grau de dependência espacial da variável e definir os parâmetros necessários para estimar características em locais não amostrados, utiliza-se um gráfico de $\gamma(\mathbf{h})$ por $\mathbf{h}$ denominado de semivariograma, a principal ferramenta para o diagnóstico da existência de correlação entre os pontos em estudo.

O semivariograma analisa o grau de dependência espacial entre amostras dentro de um campo experimental, além de definir parâmetros necessários para a estimativa de valores para locais não amostrados, através da técnica de krigagem (SALVIANO, 1996). Ainda, segundo Ribeiro Jr. (1995), os semivariogramas são preferidos para caracterizar a estrutura de continuidade espacial da característica avaliada, pois exigem hipóteses de estacionaridade menos restritivas (hipótese intrínseca).

O ajuste de um modelo matemático aos dados no gráfico de $\gamma(\mathbf{h})$ por $\mathbf{h}$, isto é, à uma função, é um dos possíveis métodos para estimar os parâmetros do semivariograma, os quais são: efeito pepita (C_o), denominado “nugget” no inglês, que representa o valor de γ quando $\mathbf{h} = \mathbf{0}$; quando o valor de $\mathbf{h}$ aumenta, frequentemente aumenta até uma distância a , denominada de alcance, (alcance da dependência espacial, do inglês “range”), também expresso por ϕ ; na distância a , $\gamma(\mathbf{h})$ é chamado de patamar, do inglês “sill”, dado

por $C + C_o$, sendo C a variância estrutural.

O alcance é utilizado para definir o raio de ação máximo de interpolação pela krigagem, (*range*), em que os pesos utilizados podem afetar os valores estimados. Pontos amostrados que estão separados por distância menor do que o alcance são ditos espacialmente dependentes, o que não é o caso de pontos separados por uma distância maior.

Segundo Vieira (2000), em um semivariograma ajustado o valor da semi-variância aumenta na medida em que aumenta a distância de separação entre os pontos, aumentando até atingir um patamar no qual se estabiliza. Este patamar é um parâmetro importante, pois permite a determinação da distância limite entre dependência e independência entre as amostras e é atingido quando a variância dos dados fica constante com a distância entre os pontos amostrados.

O parâmetro C_o , efeito pepita, reflete o erro analítico, isto é, indica uma variabilidade não explicada de um ponto para outro que pode ocorrer devido a erros de medidas ou microvariação não detectada por causa da distância de amostragem utilizada.

Uma outra convenção por vezes adotada é determinar o grau de aleatoriedade presente nos dados pela fórmula $E = \frac{C_o}{C}$ (GUERRA, 1988 apud LANDIM, 1996), em que:

- i. $E < 0,15$: componente aleatória pequena;
- ii. $0,15 \leq E \leq 0,30$: componente aleatória significativa;
- iii. $E > 0,30$: componente aleatória muito significativa.

O modelo matemático ajustado é dividido em duas categorias: modelo com patamar e sem patamar.

Um modelo com patamar é aquele em que a semivariância aumenta com o aumento da distância entre amostras, até atingir o patamar, no qual se estabiliza. Exemplos de modelos nessa categoria são:

- i. Esférico:
$$\begin{cases} \gamma(\mathbf{h}) = C[\frac{3}{2}(\frac{h}{a}) - \frac{1}{2}(\frac{h}{a})^3], & \text{para } \mathbf{h} < a \\ \gamma(\mathbf{h}) = C, & \text{para } \mathbf{h} \geq a \end{cases}$$
- ii. Exponencial: $\gamma(\mathbf{h}) = C[1 - e^{\frac{-3h}{a}}]$
- iii. Gaussiano: $\gamma(\mathbf{h}) = C[1 - e^{(\frac{-3h}{a})^2}]$

Um outro modelo que vem sendo bastante utilizado, devido a sua flexibilidade, é com a função Matérn que é dada por

$$\gamma(\mathbf{h}) = M(\mathbf{h}|v, a) = \frac{2^{1-v}}{\Gamma(v)} (a||\mathbf{h}||)^v K_v(a||\mathbf{h}||)$$

no qual $\Gamma(\cdot)$ é a função gama, K_v é a função Bessel de ordem v , e $||\mathbf{h}||$ é a distância euclidiana entre duas localizações. Um caso particular da função Matérn é quando o valor de v é $\frac{1}{2}$ em que a função se iguala a função exponencial.

Os modelos sem patamar, são modelos que como o próprio nome diz, não se estabilizam em um patamar. Como exemplo dessa categoria de modelo tem o modelo linear e o modelo potencial.

Após o ajuste de um variograma experimental, é importante realizar a validação cruzada, em que depois de obter o modelo variográfico, cada valor original da amostra é removido do domínio espacial e, utilizando os demais valores, um novo valor é estimado para o ponto retirado, sendo possível assim realizar uma comparação entre o real valor e o estimado, verificando se o erro de estimação é muito grande. Esta técnica da validação cruzada, não prova que o modelo escolhido é o mais correto, mas sim que o mesmo não é de todo incorreto.

Depois de estipulado o semivariograma da variável em estudo e ocorrendo a dependência espacial entre os pontos amostrados, pode-se interpolar valores em qualquer posição na área em estudo, sem tendência e com variância mínima. O método de interpolação utilizado é a krigagem e quando se trata de um modelo com mais de uma variável envolvida utiliza-se a cokrigagem.

2.1.3.2 Krigagem

Krigagem é um processo de estimativa de valores de variáveis distribuídas no espaço e/ou no tempo, a partir de valores adjacentes enquanto considerados como interdependentes pelo semivariograma (LANDIM, 2006).

Utilizado como algoritmo interpolador, esta técnica pode prever o valor pontual de uma variável regionalizada em um determinado local dentro da área de estudo, além de realizar o cálculo médio de uma variável regionalizada para um volume maior que o suporte geométrico. Em todos estes casos, este método fornece o erro associado a estimativa realizada, tornando-o um algoritmo melhor que os demais existentes.

Os tipos de krigagem mais utilizados são krigagem simples, na qual assume-se que a função média do processo estocástico é conhecida e o processo é estacionário e krigagem ordinária, na qual a função média do processo é estimada. Quando tem-se a estimação de uma variável em um estudo conjunto com mais variáveis faz-se o uso da cokrigagem, que será discutida mais adiante.

Pela krigagem, o valor estimado da variável é expresso por:

$$Y_{\mathbf{x}_0^*} = \lambda_0 + \sum_{i=1}^n \lambda_i Y_{\mathbf{x}_i}$$

em que n é o número de vizinhos medidos, $Y_{\mathbf{x}_i}$ valores utilizados na estimativa e λ_i são os ponderadores aplicados a cada $Y_{\mathbf{x}_i}$ que são selecionados de forma que a estimativa obtida seja não tendenciosa.

2.2 Geoestatística Multivariada

Segundo Chilès e Delfiner (1999), as informações disponíveis de um fenômeno natural raramente são limitadas por valores de uma única variável medida em um conjunto de pontos. Assim, nas aplicações de ciência da terra existem pelo menos mais quatro fontes adicionais de informação, sendo elas:

- i) Valores numéricos de outras possíveis variáveis em outras localizações;
- ii) Valores numéricos da mesma variável estudada em outra época;
- iii) Relações físicas entre variáveis;
- iv) Aplicações de conhecimento.

A maioria dos estudos envolvem mais de uma variável e então, trabalha-se com modelos multivariados. De acordo com Bognola et al. (2008), os modelos multivariados são particularmente atrativos quando existe uma variável primária a ser estudada de difícil acesso, seja por causa de custo elevado ou algum outro motivo; mas existe uma outra variável, secundária, correlacionada com a variável de interesse, sendo que essa variável secundária é de fácil acesso. Desta forma, frequentemente combina-se algumas mensurações da variável de interesse com um número maior da secundária e a partir disso, realiza-se um estudo da variável de interesse sem muitas medições da mesma.

No caso de observações bi ou multivariadas, os modelos e procedimentos são estendidos de diferentes formas, visando modelar as covariâncias simples e cruzadas entre as variáveis envolvidas (WACKERNAGEL, 2003 apud BOGNOLA et al., 2008). Existem várias técnicas de modelagem multivariada como: modelagem através do variograma e variograma cruzado, em que um modelo muito utilizado é o modelo linear de correionalização; modelagem por modelos que implicam estrutura de correlação, como o “Bivariate Gaussian Common Component Model” (BGCCM); modelagem através de função de correlação bi (ou multivariada); utilizando métodos bayesianos, entre outros tipos de modelagem.

2.2.1 Cokrigagem

O método de predição na modelagem geoestatística multivariada é a cokrigagem. Segundo Bognola et al.(2008), a partir do ajuste de um modelo é possível obter predições espaciais das variáveis de interesse através da cokrigagem, no qual o valor predito em um determinado ponto é dado por uma média ponderada dos dados das variáveis envolvidas. Isto é, cokrigagem é uma extensão multivariada do método da krigagem, quando para cada local amostrado obtém-se um vetor de valores ao invés de um único valor. Esta técnica é bastante útil e evidente, quando duas ou mais variáveis são amostradas nos mesmos pontos amostrais, dentro de um mesmo domínio espacial e apresentam significativo grau de correlação entre elas.

A cokrigagem, segundo Isaaks e Srivastava (1989), considera uma ou mais variáveis secundárias na estimativa de uma variável primária, podendo ser escrita em termos das covariâncias simples e cruzadas, sendo que os pesos W , usados na interpolação, são dados por $W = J^{-1}G$, em que:

$$J = \begin{pmatrix} Cov(Y_1Y_1) & Cov(Y_1Y_2) & 1 \\ Cov(Y_2Y_1) & Cov(Y_2Y_2) & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}; W = \begin{pmatrix} \mathbf{e} \\ \mathbf{f} \\ -\lambda_1 \\ -\lambda_2 \end{pmatrix} \text{ e } G = \begin{pmatrix} Cov(Y_1Y_0) \\ Cov(Y_2Y_0) \\ 1 \\ 0 \end{pmatrix}$$

tem-se que Y_1 e Y_2 são as variáveis primária e secundária, respectivamente, Y_0 é referente aos valores que serão preditos; $\mathbf{e}$ e $\mathbf{f}$ são os vetores de pesos atribuídos às variáveis primária e secundária, respectivamente, sendo que $\sum e_i = 1$ e $\sum f_i = 1$ são restrições para garantir

a não-tendenciosidade, introduzidas pelo uso dos operadores de Lagrange λ_1 e λ_2 .

2.2.2 Modelos Lineares de Corregionalização

A modelagem de corregionalização das variáveis usa o denominado modelo linear de corregionalização (WACKERNAGEL, 2003). Este modelo, segundo Bellier et al. (2007), permite a modelagem conjunta do variograma simples e do variograma cruzado, sendo que o número total desses variogramas é dado por $\frac{p(p+1)}{2}$ em que p é o número de variáveis no conjunto em estudo. Para que este modelo seja compreendido, precisa-se que o modelo básico, utilizado no variograma cruzado, seja o mesmo utilizado nos auto-variogramas, mantendo-se assim o valor para o alcance.

O modelo linear de corregionalização fornece um método para modelagem do auto-variograma e do variograma cruzado entre duas ou mais variáveis desde que a variância de qualquer combinação linear possível dessas variáveis seja sempre positiva. Cada variável é caracterizada por seu próprio auto-variograma amostral e cada par de variáveis por seu variograma cruzado amostral (ISAACS; SRIVASTAVA, 1989). Segundo Wackernagel (2003), um conjunto real de funções aleatórias estacionárias de segunda-ordem dado por $Y_i(\mathbf{x}); i = 1, \dots, N$ pode ser particionado em sub-conjuntos $Y_u^i(\mathbf{x}); u = 1, \dots, O$ de componentes espaciais não correlacionados

$$Y_i(\mathbf{x}) = \sum_{u=0}^O Y_u^i(\mathbf{x}) + m_i,$$

em que para todos os índices i, j, u e v , tem-se

$$E[Y_i(\mathbf{x})] = m_i, \quad E[Y_u^i(\mathbf{x})] = 0,$$

$$\text{cov}(Y_u^i(\mathbf{x}), Y_u^j(\mathbf{x} + \mathbf{h})) = E[Y_u^i(\mathbf{x})Y_u^j(\mathbf{x} + \mathbf{h})] = C_{ij}^u(\mathbf{h})$$

$$\text{cov}(Y_u^i(\mathbf{x}), Y_u^j(\mathbf{x} + \mathbf{h})) = 0 \quad \text{para} \quad u \neq v$$

em que $C_{ij}^u(\mathbf{h})$ são funções de covariância cruzada associadas com os componente espaciais e são compostas por coeficientes b_{ij}^u . Essas funções são proporcionais as funções de correlação

$\rho_u(\mathbf{h})$, ou seja,

$$C_{ij}(\mathbf{h}) = \sum_{u=0}^O C_{ij}^u(\mathbf{h}) = \sum_{u=0}^O b_{ij}^u \rho_u(\mathbf{h})$$

As matrizes de correionalização, denotadas por $\mathbf{B}_u$ de dimensão $N \times N$, podem ser estabelecidas e assim, segundo Wackernagel (2003), tem-se o modelo de função de covariâncias aninhados multivariado, em que a matriz de correionalização é positiva semi-definida, dado por:

$$C(\mathbf{h}) = \sum_{u=0}^O \mathbf{B}_u \rho_u(\mathbf{h}).$$

Cada componente espacial $Y_u^i(\mathbf{x})$ pode ser representado como um conjunto de fatores não correlacionados W_u^p com coeficientes de transformação w_{up}^i ,

$$Y_u^i(\mathbf{x}) = \sum_{p=1}^N w_{up}^i W_u^p,$$

sendo que para todos os valores do índices i, j, u, v, p e q tem-se

$$\begin{aligned} E[Y_u^p(\mathbf{x})] &= 0, \\ cov(Y_u^p(\mathbf{x}), Y_u^p(\mathbf{x} + \mathbf{h})) &= \rho_u(\mathbf{h}) \quad \text{e} \\ cov(Y_u^p(\mathbf{x}), Y_v^q(\mathbf{x} + \mathbf{h})) &= 0 \quad \text{para} \quad u \neq v \quad \text{ou} \quad p \neq q \end{aligned}$$

Desta forma, combinando a decomposição multivariada com a espacial, obtém-se o modelo linear de correionalização:

$$Y_i(\mathbf{x}) = \sum_{u=0}^O \sum_{p=1}^N w_{up}^i W_u^p \quad (13)$$

O modelo do variograma multivariado, associado ao modelo linear de funções aleatórias estritamente estacionárias pode ser representado por:

$$\Gamma(\mathbf{h}) = \sum_{u=0}^O \mathbf{B}_u g_u(\mathbf{h}),$$

em que $g_u(\mathbf{h})$ representa os variogramas normalizados e $\mathbf{B}_u$ é matriz positiva semi-definida.

Assim, ao examinar o comportamento do variograma empírico, do simples e do variograma cruzado, o pesquisador escolhe o número e tipo de base de modelos. Cada função do variograma utilizado para definir o modelo linear de correção regionalização tem diferentes famílias e intervalos, em que cada um caracteriza uma escala específica da distribuição espacial da variável resposta em estudo.

2.2.3 Semivariograma Cruzado

Sejam p variáveis $\{Y_j(\mathbf{x}); j = 1, \dots, p\}$ que foram medidas dentro de uma mesma região de estudo e nos mesmos pontos amostrais. Dessas variáveis medidas, considerando apenas duas, temos $\{[Y_1(\mathbf{x})]$ e $[Y_2(\mathbf{x})]\}$ de tal forma que a covariância cruzada entre essas variáveis seja dada por:

$$Cov_{12}(h) = E\{[Y_1(\mathbf{x}+\mathbf{h})Y_2(\mathbf{x})]\} - m_1m_2$$

e

$$Cov_{21}(h) = E\{[Y_2(\mathbf{x}+\mathbf{h})Y_1(\mathbf{x})]\} - m_2m_1$$

Desta forma, o semivariograma cruzado será:

$$\gamma_{12}^*(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} \{[Y_1(\mathbf{x}+\mathbf{h}) - Y_1(\mathbf{x})][Y_2(\mathbf{x}+\mathbf{h}) - Y_2(\mathbf{x})]\}$$

Segundo Silva et al.(2003 apud Vieira, 1996) a forma de obter os gráficos do semivariograma cruzado experimental tem significado diferente do obtido pelo semivariograma simples, pois no variograma cruzado o alcance (a) representa a distância máxima de dependência espacial entre as duas variáveis; o patamar, se existir, deve aproximar-se do valor da covariância entre as duas variáveis utilizadas. Quando estas duas variáveis apresentarem uma correlação inversa, a covariância será negativa e, desta forma, o semivariograma cruzado produzirá resultados negativos.

2.2.4 Modelos Geoestatísticos Bivariados Gaussianos

Quando o interesse está em mais de um campo aleatório, suponha-se dois, acredita-se que o passo inicial é o de ajustar um modelo geoestatístico para cada um dos dois vetores das variáveis observadas porém, nem sempre é o melhor caminho pois há casos em que existe correlação entre as variáveis aleatórias estudadas e nestes casos, é

interessante utilizar modelos geoestatísticos bivariados.

Existindo dois campos aleatórios gaussianos, utiliza-se o modelo dado por:

$$Y_i = \mu_i + S_i, \quad i = 1, 2 \quad (14)$$

em que Y_i é um vetor $n_i \times 1$ dos valores observados no campo aleatório latente S_i , o qual possui um vetor de médias nulo de dimensão $n_i \times 1$ e matriz de covariância Σ_i com dimensão $n_i \times n_i$, μ_i é um vetor dos parâmetros de média associado a Y_i .

Segundo Fonseca (2008), a distribuição de $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$ é o interesse final e esta distribuição é gaussiana n -variada, sendo $n = n_1 + n_2$, com vetor de médias $\mu = (\mu_1, \mu_2)$ e matriz de covariância Σ_Y , sendo que esta matriz é positiva definida e possui comportamento empírico de correlações utilizado na geoestatística, podendo ser particionada por:

$$\Sigma_Y = \begin{bmatrix} \Sigma_1 & \Sigma_{1,2} \\ \Sigma_{1,2}^t & \Sigma_2 \end{bmatrix}$$

sendo Σ_i uma matriz de covariância de dimensão $n_i \times n_i$ da variável Y_i com $i = 1, 2$ e $\Sigma_{1,2}$ a matriz de covariâncias cruzadas entre as variáveis Y_1 e Y_2 , de dimensão $n_1 \times n_2$.

2.2.5 Modelo Gaussiano Bivariado com Componente Comum

Para problemas bivariados, Diggle e Ribeiro Jr (2007) descrevem o modelo denominado “Bivariate Gaussian common component model” (BGCCM), no qual o termo S_i , campo aleatório latente, em (14), é decomposto em:

$$S_i = \sigma_{0i}T_0 + \sigma_iT_i, \quad i = 1, 2 \quad (15)$$

sendo $\sigma^* = (\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2)$ um vetor de parâmetros de dispersão associados ao modelo (14); T_0 , T_1 e T_2 são campos aleatórios gaussianos mutuamente independentes, com vetores de médias nulos, variância unitárias e correlações obtidas por funções de correlação válidas. As correlações destes campos aleatórios, quando combinadas desta forma, vão gerar correlações cruzadas válidas entre as variáveis estudadas, Y_1 e Y_2 , já que o campo aleatório T_0 é comum às duas variáveis, isto é, o BGCCM induz uma estrutura de correlação. Desta

forma, deve-se adotar três funções de correlação para T_0 , T_1 e T_2 e assim gerar uma matriz Σ_Y válida.

Então, com a decomposição de S_1 e S_2 , o modelo (1) pode ser reescrito por:

$$\begin{cases} Y_1 = \mu_1 + \sigma_{01}T_0 + \sigma_1T_1 \\ Y_2 = \mu_2 + \sigma_{02}T_0 + \sigma_2T_2 \end{cases}$$

Seja $Y_i(\mathbf{x}_l)$ e $Y_i(\mathbf{x}_k)$ observações da variável Y_i realizadas nas localizações $\mathbf{x}_l$ e $\mathbf{x}_k$, separadas por uma distância euclidiana $h = h_{lk}$, com $l, k = 1, \dots, n_i$ e $i = 1, 2$. Assim, o elemento $\Sigma_{i,(l,k)}$ é representado por $Cov(h) = \sigma_{0i}^2\rho_0(h) + \sigma_i^2\rho_i(h)$, no qual ρ_0 e ρ_i são as funções de correlação para T_i , $i = 0, 1, 2$. Pelas propriedades básicas de covariância, encontra-se que $\Sigma_{1,2}$ é igual a $\sigma_{01}\sigma_{02}R_0$, em que R_0 é uma matriz, de dimensão $n_1 \times n_2$, das correlações cruzadas entre as variáveis e depende da função de correlação atribuída para T_0 .

Considerando apenas duas localizações amostradas, $\mathbf{x}_k$ e $\mathbf{x}_l$, e sendo $h_{kl} = h$ a distância entre essas localizações, tem-se que a matriz de covariâncias para o vetor $(Y_i(\mathbf{x}_l), Y_i(\mathbf{x}_k), Y_j(\mathbf{x}_l), Y_j(\mathbf{x}_k))$ é dada por:

$$\Sigma_Y = \begin{bmatrix} \sigma_{01}^2 + \sigma_1^2 & \sigma_{01}^2\rho_0(h) + \sigma_1^2\rho_1(h) & \sigma_{01}\sigma_{02} & \sigma_{01}\sigma_{02}\rho_0(h) \\ \sigma_{01}^2\rho_0(h) + \sigma_1^2\rho_1(h) & \sigma_{01}^2 + \sigma_1^2 & \sigma_{01}\sigma_{02}\rho_0(h) & \sigma_{01}\sigma_{02} \\ \sigma_{01}\sigma_{02} & \sigma_{01}\sigma_{02}\rho_0(h) & \sigma_{02}^2 + \sigma_2^2 & \sigma_{02}^2\rho_0(h) + \sigma_2^2\rho_2(h) \\ \sigma_{01}\sigma_{02}\rho_0(h) & \sigma_{01}\sigma_{02} & \sigma_{02}^2\rho_0(h) + \sigma_2^2\rho_2(h) & \sigma_{02}^2 + \sigma_2^2 \end{bmatrix}$$

Desta forma, a distribuição de probabilidade do vetor $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$ está estabelecida e depende do vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\beta}^*, \boldsymbol{\sigma}^*, \phi_0^*, \phi_1^*, \phi_2^*)$ no qual $\boldsymbol{\beta}^*$ é o vetor de parâmetros associados a μ e $\boldsymbol{\phi}_j^*$ é o vetor de parâmetros relacionados a escolha da função ρ_j , com $j = 0, 1, 2$.

Para a estimação dos parâmetros desse modelo e a predição espacial, utiliza-se as técnicas dos modelos geoestatísticos univariados porém, é necessário uma reparametrização adicional em $\boldsymbol{\theta}$ para obter a função de verossimilhança concentrada.

2.2.6 Modelagem por Função de Correlação

Como foi descrito acima, o BGCCM induz uma estrutura de correlação entre as variáveis através de um campo aleatório comum. O MLC (modelo linear de correção) possui uma falta de flexibilidade, por esta falta de flexibilidade do MLC, os modelos espaciais multivariados, estão sendo estudados por muitos autores como revisado em Cressie(1993) e Goovaerts(1992). Jun, Knutti e Nychka (2008), estudaram estrutura espacial dos desvios individuais de modelos climáticos, usando covariâncias do tipo Matérn, ao invés de estrutura conjunta.

Estas mudanças e desafios motivaram Gneiting, Kleiber e Schlather (2009) a propor um ajuste por estrutura de covariância e covariância cruzada para um campo aleatório gaussiano, preservando a função Matérn que permite estrutura de covariância cruzada flexíveis e não triviais. É uma proposta potencialmente vantajosa, devido ao fato da facilidade em mudar os parâmetros do modelos e ajustá-los da melhor forma aos dados. A família Matérn tem sido muito utilizada na atualidade, por ser uma função bastante flexível.

Esta família especifica uma função de covariância dada por $\sigma^2 M(\mathbf{h}|k, a)$ em que $\sigma^2 > 0$ é a variância marginal e

$$M(\mathbf{h}|k, a) = \frac{2^{1-k}}{\Gamma(k)} (a||\mathbf{h}||)^k K_k(a||\mathbf{h}||)$$

é a correlação espacial na distância $||\mathbf{h}||$, e como já mencionado, K_k é a função Bessel.

Em um processo multivariado, seja o vetor de variáveis $\mathbf{Y}(\mathbf{x}) = (\mathbf{Y}_1(\mathbf{x}), \dots, \mathbf{Y}_p(\mathbf{x}))'$, em que cada localização $\mathbf{x} \in \mathbb{R}^d$ existem p componentes. Assumindo que o processo é gaussiano multivariado e estacionário de segunda ordem, com vetor de média nulo e matriz de função de covariância dada por

$$C(\mathbf{h}) = \begin{pmatrix} C_{11}(\mathbf{h}) & \cdots & C_{1p}(\mathbf{h}) \\ \vdots & \ddots & \vdots \\ C_{p1}(\mathbf{h}) & \cdots & C_{pp}(\mathbf{h}) \end{pmatrix} \quad (16)$$

em que cada $C_{ii}(\mathbf{h}) = E(Y_i(\mathbf{x} + \mathbf{h})Y_i(\mathbf{h}))$ é a função de covariância univariada, enquanto que $C_{ij}(\mathbf{h}) = E(Y_i(\mathbf{x} + \mathbf{h})Y_j(\mathbf{h}))$ é a função de covariância cruzada entre os processos $1 \leq i \neq j \leq p$.

No modelo proposto por Gneiting, Kleiber e Schlather (2009), a função de

covariância marginal é dada por:

$$C_{ii}(\mathbf{h}) = \sigma_i^2 M(\mathbf{h}|k_i, a_i), \quad (17)$$

em que σ_i^2 é o parâmetro de variância, k_i é relacionado a suavidade e a_i é parâmetro de escala. A covariância cruzada é dada por

$$C_{ij}(\mathbf{h}) = C_{ji}(\mathbf{h}) = \rho_{ij} \sigma_i \sigma_j M(\mathbf{h}|k_{ij}, a_{ij}), \quad (18)$$

sendo ρ_{ij} o coeficiente de correlação, k_{ij} é o parâmetro de suavidade e a_{ij} é parâmetro de escala espacial. Segundo Gneiting, Kleiber e Schlather (2009), precisa-se ter cuidado com os valores destes parâmetros, pois estes valores precisam fornecer uma estrutura de segunda ordem válida, isto é, positiva definida, o que é garantido impondo restrições aos parâmetros.

2.2.6.1 Modelo Mátern Multivariado

Como já mencionado, se o parâmetro k da função Mátern for igual $\frac{1}{2}$, então esta função se reduz ao produto da função exponencial e de um polinômio, isto é

$$M(\mathbf{h}|n + \frac{1}{2}) = \exp(-a||\mathbf{h}||) \sum_{m=0}^n \frac{(n+m)!}{(2n)!} \binom{n}{m} (2a||\mathbf{h}||)^{n-m}$$

com $n = 0, 1, \dots$

Segundo Gneiting, Kleiber e Schlather (2009), a flexibilidade e popularidade da classe Mátern se deve ao fato do parâmetro k , sendo que $k > 0$, que controla a suavidade e é crítico em problemas de interpolação espacial.

2.2.6.1.1 Modelo Mátern Multivariado Parcimonioso

Gneiting, Kleiber e Schlather (2009), definem um modelo Matérn parcimonioso multivariado, que possui função marginal e covariância cruzada dada por (17) e (18). Para este modelo os autores definiram alguns teoremas que fazem com que este modelo seja válido.

Teorema 1: Seja $d \geq 1$ e $p \geq 2$, supondo que $k_{ij} = \frac{1}{2}(k_i + k_j)$, para $1 \leq i \neq j \leq p$, e considerando que exista um parâmetro de escala comum, a , isto é, $a_1 = \dots = a_p = a$ e $a_{ij} = a$. Desta forma, o modelo Matérn multivariado fornece uma

estrutura de segunda ordem válida, em $\mathbb{R}^d$, se

$$\rho_{ij} = \beta_{ij} \frac{\Gamma(k_i + \frac{d}{2})^{\frac{1}{2}}}{\Gamma(k_i)^{\frac{1}{2}}} \frac{\Gamma(k_j + \frac{d}{2})^{\frac{1}{2}}}{\Gamma(k_j)^{\frac{1}{2}}} \frac{\Gamma(\frac{1}{2}(k_i + k_j))}{\Gamma(\frac{1}{2}(k_i + k_j) + \frac{d}{2})}$$

em que $(\beta_{ij})_{i,j=1}^p$, com diagonal composta por $\beta_{ii} = 1$, $i = 1, \dots, p$, elementos fora da diagonal dados por β_{ij} , para $1 \leq i \neq j \leq p$. Essa matriz é simétrica e positiva definida.

A condição de ter apenas um parâmetro de escala para todos os componentes pode parecer restritiva, porém, Zhang (2004) provou que os parâmetros σ^2 e a da função de covariância Matérn, não podem ser consistentemente estimados sob algumas suposições, mas a composição $\sigma^2 a^{2k}$ pode ser estimada, isso quando parâmetro k da função é fixo.

Teorema 2: Supondo que $c_1, \dots, c_p$ funções de valores reais em $\mathbb{R}^d$ e essas funções são integráveis e o quadrado delas também é integrável. Se

$$C_{ij}(\mathbf{h}) = (c_i * c_j)(\mathbf{h}) \quad \text{para } i, j = 1, \dots, p$$

em que o $*$ denota o operador de convolução e a matriz dada em (16) é a função de covariância multivariada em $\mathbb{R}^d$.

Este método de convolução para a construção da função de covariância multivariada tem sido estudada por Ver Hoef e Barry (1998), Gaspari e Cohn (1999), Oliver (2003) e, mais recentemente, por Ver Hoef, Cressie e Barry (2004).

O Teorema 1 acima, fornece uma flexível estrutura de covariância cruzada para campos aleatórios multivariados para qualquer número de componetes p do processo. Quando $p = 2$, isto é, apenas dois componentes, os parâmetros $\beta_{12} = \beta_{21}$ admite valores entre -1 e 1 . Assim, o modelo Matérn bivariado parcimonioso é dado por:

$$C_{11}(\mathbf{h}) = \sigma_1^2 M(\mathbf{h}|k_1, a) \quad C_{22}(\mathbf{h}) = \sigma_2^2 M(\mathbf{h}|k_2, a) \quad (19)$$

e

$$C_{12}(\mathbf{h}) = C_{21}(\mathbf{h}) = \rho_{12} \sigma_1 \sigma_2 M(\mathbf{h}|\frac{1}{2}(k_1 + k_2), a) \quad (20)$$

em que

$$|\rho_{12}| \leq \frac{\Gamma(k_1 + \frac{d}{2})^{\frac{1}{2}}}{\Gamma(k_1)^{\frac{1}{2}}} \frac{\Gamma(k_2 + \frac{d}{2})^{\frac{1}{2}}}{\Gamma(k_2)^{\frac{1}{2}}} \frac{\Gamma(\frac{1}{2}(k_1 + k_2))}{\Gamma(\frac{1}{2}(k_1 + k_2) + \frac{d}{2})} \quad (21)$$

Considerando $p = 2$ e $d = 2$, de um campo aleatório espacial bivariado, a inequação dada em (21) se resume a

$$|\rho_{12}| \leq \frac{(k_1 k_2)^{\frac{1}{2}}}{\frac{1}{2}(k_1 + k_2)} \quad (22)$$

Gneiting, Kleiber e Schlather (2009) ainda definem um modelo Matérn bivariado completo. Neste modelo, cada função de covariância e covariância cruzada possuem um parâmetro de escala, isto é, tem-se a_1, a_2 e a_{12} , além de possuir algumas restrições.

O modelo Matérn parcimonioso multivariado é um modelo cujo suporte não é compacto. A estimação, computacionalmente, dos parâmetros é bastante delicada. Modelos cuja função de correlação são de suporte compacto foram estudados por Gneiting (2002) e a vantagem deste modelo, com o suporte compacto, é na facilidade e rapidez na simulação e a predição computacional é bastante eficiente.

2.2.7 Geoestatística Bayesiana

Em geoestatística, os parâmetros dos modelos são estimados e utilizados na obtenção das predições espaciais das variáveis em estudo. Desta forma, pode-se dizer que a incerteza sobre estes parâmetros pode ser incorporada na incerteza associada as predições realizadas. No enfoque bayesiano a incerteza sobre os parâmetros é representada por distribuições de probabilidades. Ainda, as associações espaciais são melhores capturadas quando se faz uso de modelos hierárquicos que constróem dependência em diferentes estágios e, estes modelos seguem o paradigma bayesiano de inferência estatística.

Quando se utiliza a metodologia da geoestatística clássica na estimação de parâmetros, através do método da máxima verossimilhança, utiliza-se apenas as informações provenientes dos dados, do modelo adotado e das teorias assintóticas. Porém, quando se utiliza os métodos bayesianos, além de todos esses conhecimentos, também se utiliza informações a priori sobre os parâmetros, isto é, pode existir um conhecimento inicial que é expresso por meio das distribuições a priori e desta forma, na prática, não existe diferença entre parâmetros e observações, sendo que ambos são tratados como quantidades aleatórias.

Através do Teorema de Bayes obtém-se a distribuição a posteriori dos parâ-

metros, dada por:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)} \quad (23)$$

em que θ é o parâmetro, ou vetor de parâmetros, de interesse; $p(y) = \int_{\theta} p(y|\theta)p(\theta)d\theta$ e este termo é constante sob a distribuição de probabilidade de $\theta|y$; $p(y|\theta)$ é a função de verossimilhança de $y|\theta$ e $p(\theta)$ é a distribuição a priori de θ , que representa o conhecimento prévio da distribuição de probabilidade dos parâmetros.

Segundo Kolman (2004), retirando o termo constante de (23) tem-se:

$$p(\theta|y) \propto p(y|\theta)p(\theta) \quad (24)$$

sendo que $p(\theta|y)$ dado em (24) não é uma distribuição de probabilidades, porém $\omega p(\theta|y)$ é, se houver uma escolha adequada para a constante ω , que é obtida por métodos analíticos ou numéricos.

Como já mencionado, após observar $Y = y$, existe o interesse em realizar previsões para uma quantidade Y_o não observada que, relacionada a θ é dado por:

$$p(y_o|y) = \int p(y_o, \theta|y)d\theta = \int p(y_o|\theta, y)p(\theta|y)d\theta,$$

sendo que na maioria das vezes, essa integral não possui solução analítica e faz-se necessário métodos de aproximação para resolvê-la.

Depois de definir a distribuição a posteriori dos parâmetros é possível resumir informações sobre os mesmos através de técnicas inferenciais, utilizando normalmente probabilidades e esperanças. Nesta etapa, muitas vezes é necessário utilizar métodos computacionais intensivos como o Monte Carlo via Cadeias de Markov, denominada de MCMC. Na geoestatística, o amostrador de Gibbs e o algoritmo Metropolis-Hastings são os métodos mais utilizados.

Neste trabalho, estudou-se o modelo de regressão espacial bayesiana multivariada. Inicialmente, considerando um modelo de regressão espacial univariado tem-se que:

$$Y(\mathbf{x}) = X^t(\mathbf{x})\boldsymbol{\beta} + \mathbf{W}(\mathbf{x}) + \epsilon(\mathbf{x})$$

em que $\mathbf{W}(\mathbf{x})$ é um campo aleatório espacial, sendo considerado um campo aleatório espacial válido somente se para qualquer conjunto de localidades $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ de tamanho

arbitrário, o vetor $\mathbf{W} = [W(\mathbf{x}_i)]_{i=1}^n$ segue uma distribuição de probabilidade conjunta bem definida.

Estendendo o modelo do caso univariado para um modelo de regressão espacial multivariado, cada localização $\mathbf{x}$ possui um vetor resposta $Y(\mathbf{x}) = [Y_i(\mathbf{x})]_{i=1}^m$ além da matriz de regressores $X^t(\mathbf{x})$. Ainda, no modelo multivariado a matriz $\mathbf{W}$ é um processo Gaussiano multivariado centrado em zero, denotado por $\mathbf{W}(\mathbf{x}) \sim MVGP(0, \mathbf{K}(\cdot, \cdot; \theta))$, além de capturar a variação espacial; $\epsilon(\mathbf{x})$ modela o efeito do erro de medida da variável resposta, segue uma distribuição normal multivariada com média nula e matriz de dispersão Ψ .

De acordo com Finley, Banerjee, Carlin (2007) o processo Gaussiano multivariado é especificado pela função de matriz de covariância cruzada $\mathbf{K}(\mathbf{x}, \mathbf{x}'; \theta) = [Cov(W_i(\mathbf{x}), W_j(\mathbf{x}'))]$, em que o (i, j) -elemento é a covariância entre $W_i(\mathbf{x})$ e $W_j(\mathbf{x}')$, sendo θ composto por parâmetros que controlam a suavidade e a correlação do processo.

Considerando

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \alpha\hat{\mathbf{W}} + \epsilon,$$

em que $\mathbf{Y}$ é vetor $np \times 1$ de respostas, $\mathbf{X}$ é a matriz $pn \times q$ de regressores e $\boldsymbol{\beta}$ é o vetor correspondente dos coeficientes de regressão. Segundo Finley, Banerjee, Carlin (2007), a especificação dessas matrizes α e $\hat{\mathbf{W}}$ dão origem a diferentes modelos de regressão espacial multivariados.

É comum assumir que $\mathbf{W}(\mathbf{x}) = \mathbf{J}(\mathbf{x})\hat{\mathbf{W}}(\mathbf{x})$, em que $\mathbf{J}$ é uma matriz transformada de variação do espaço que é não-singular para todo $\mathbf{x}$ e desta forma, a matriz de funções de covariâncias cruzadas é dada por $\mathbf{K}(\mathbf{x}, \mathbf{x}'; \theta) = \mathbf{J}(\mathbf{x})\hat{K}(\mathbf{x})\mathbf{J}'(\mathbf{x})$.

Na literatura existem vários modelos. Em Finley, Banerjee, Carlin (2007), os autores apresentam sete modelos:

- i) de regressão linear simples sem efeito aleatório;
- ii) que assume variância espacial comum;
- iii) que permite que o efeito pepita seja o mesmo para as variáveis respostas;
- iv) uma extensão do modelo (iii), em que permite termos específicos para variância espacial e efeito pepita;
- v) específico para os elementos fora da diagonal principal da matriz de covariância cruzada $\mathbf{K}$;

- vi) modela específicos termos do alcance espacial;
- vii) permite que não exista dependencia da variância espacial entre as superfícies de resposta.

O modelo mais completo é o (vii) sendo que todos os outros são particularidades deste modelo. Para este trabalho, utilizou-se o modelo (vi) que é representado por:

$$\text{Modelo: } \mathbf{J}, \mathbf{\Psi} = \text{diag}[\tau_i^2]_{i=1}^m, \mathbf{\Phi} = [\phi_i]_{i=1}^m$$

em que $\mathbf{\Psi}$ é a matriz não espacial de covariância cruzada, $\mathbf{\Phi}$ matriz da dependência espacial e $\mathbf{K}$ que depende de $\mathbf{J}$ é a matriz espacial de covariância cruzada.

3 MATERIAL E MÉTODOS

Vários trabalhos que estão sendo publicados, utilizam modelos geoestatísticos bivariados para estruturar a dependência espacial entre e dentre campos aleatórios de interesse. Sendo assim, a idéia deste trabalho é de comparar os modelos citados na Revisão de Literatura -Modelo Linear de Corregionalização, Modelo Gaussiano Bivariado com Componente Comum, Modelo Matérn Multivariado Parcimonioso e Modelo Bayesiano- e assim, averiguar qual modelo se ajusta melhor aos dados utilizados neste trabalho.

Todas as análises e resultados foram obtidos através do ambiente R de programação (R Development Core Team, 2006), utilizando os pacotes gstat (Pebesma, 2004), spBayes (Finley, Banerjee e Carlin, 2012) e geoR (Ribeiro Jr. e Diggle, 2001). Realizou-se também a implementação do modelo Matérn Multivariado pois este modelo só está implementado no pacote RandomFields (Schlather, 2012).

Realizou-se um estudo de simulação para o modelo Matérn parcimonioso e fez-se um estudo no conjunto de dados denominado “weather” que foi obtido através do pacote RandomFields (Schlather, 2012). Os dados estão disponíveis no Anexo A.

3.1 Estudo de Simulação

Esta fase do trabalho é muito importante para detectar possíveis dificuldades no ajuste do modelo Matérn parcimonioso. Foram feitos estudos de simulação com 1001 conjuntos de dados gerados, em um quadrado de lado igual a 25 e com uma malha irregular de localizações. Para avaliar as precisões das estimativas espaciais, cada vetor $Y = (Y_1, Y_2)$ de dados simulados foi dividido em um grupo de modelagem e um grupo de controle. O primeiro grupo foi utilizado para ajustar o modelo bivariado e o segundo foi utilizado para avaliar o erro de predição.

As observações do modelo Matérn parcimonioso foram geradas considerando como valores dos parâmetros: $\kappa_1 = 0.9$, $\kappa_2 = 0.75$, $\rho = 0.7$, $\sigma_1 = \sqrt{0.7}$, $\sigma_2 = \sqrt{0.5}$, $\tau_1 = \sqrt{0.2}$, $\tau_2 = \sqrt{0.1}$, $\mu_1 = 2$ e $\mu_2 = 10$. Essas observações são co-locadas, isto é, retiradas das mesmas localizações.

Os parâmetros foram estimados por máxima verossimilhança e cada vetor de parâmetros estimados foi utilizado na predição e krigagem nas localizações do grupo de controle. Em cada amostra, cada variável era composta por 121 observações, das quais 81 foram utilizadas na análise e 40 foram retiradas para validação do modelo. Depois de

ajustar o modelo e estimar os parâmetros, fez-se a predição das 40 observações retiradas de cada variável, em cada amostra e para avaliar a precisão dessas predições, calculou-se intervalos de confiança marginais, erros quadráticos (EQ) e erros absolutos (EA).

Ressalta-se que as simulações foram realizadas em um computador Dell, Intel (R) Core (TM) i5 M 460 @ 2.53GHz 64 bits e as 1001 simulações demoraram em torno de 24h para serem realizadas.

3.2 Descrição dos dados

São dados meteorológicos, referentes a observações de temperatura e de pressão em 157 localizações no Pacífico Noroeste Norte Americano, como mostrados na Figura 1. Temperatura e pressão são negativamente correlacionados, isto é, em locais onde a pressão é alta tem-se baixa temperatura, baixa pressão indica alta temperatura. Nos dados utilizados, a correlação dessas duas variáveis é -0,47. Como as duas variáveis possuem escalas diferentes, a variável pressão foi utilizada com seus valores divididos por 100 para maior estabilidade dos métodos numéricos.

Figura 1 - Pacífico Noroeste Norte Americano com as 157 observações. FONTE: Gneiting, T.; Kleiber, W.; Schlather, M., 2009

Uma análise exploratória dos dados foi conduzida utilizando gráficos e medidas descritivas, que detectaram a existencia de padrão espacial nas duas variáveis em estudo.

Na análise bivariada, para validação dos modelos, retirou-se parte dos dados. Dos 157 pontos amostrados de cada variável, retirou-se 52 pontos (grupo controle) para

predição e os 105 pontos restantes de cada variável foram utilizados na análise. Nesta análise, ajustou-se os modelos bivariados mencionados na Revisão de Literatura, para o vetor $Y = (Y_1, Y_2)$, sendo Y_1 a variável pressão e Y_2 a variável temperatura.

No modelo linear de correionalização, ajustou-se um variograma teórico ao variograma experimental de cada uma das variáveis, depois realizou-se o variograma cruzado da análise bivariada e então, ajustou-se o modelo linear de correionalização e com os parâmetros estimados do modelo fez-se uma krigagem para a área em estudo e realizou a predição dos dados no grupo controle.

O modelo bivariado gaussiano de componente comum (BGCCM) mostrou-se muito instável para variáveis negativamente correlacionadas. Uma tentativa de correção para este problema foi de fazer uma mudança, em uma das variáveis. A mudança foi de trabalhar criando uma nova variável Y_{nova} que é dada por: $Y_{nova} = \max(Y_1) - Y_1$. Conseguiu-se até que bons valores nas predições no grupo controle porém, por problemas computacionais, não conseguiu-se realizar uma krigagem na área de estudo.

Para o modelo matern multivariado parcimonioso, ajustou-se o modelo aos dados e estimou os parâmetros. Uma krigagem na área de estudo foi realizada e além disso, realizou-se a predição dos 52 pontos retirados da análise.

Na abordagem bayesiana, ajustou-se um modelo de regressão espacial bayesiana multivariada. Calculou-se os quantis e as médias dos parâmetros e realizou a krigagem para as variáveis antes e após o ajuste do modelo. Além disso, realizou a krigagem para os valores preditos do grupo controle.

Após o ajuste de todos modelos, calculou-se o erro quadrático e o erro absoluto para cada uma das variáveis em cada um dos modelos e comparou os resultados obtidos.

4 RESULTADOS E DISCUSSÃO

4.1 Estudo de Simulação

Este estudo teve o objetivo de avaliar a qualidade das estimativas dos parâmetros e das predições espaciais obtidas com o modelo Matérn parcimonioso multivariado para o qual fez-se uma programação no software R. A estimação mostrou-se numericamente instável com dificuldades na convergência das estimações nas análises. A programação deste modelo está no Apêndice A.

Simulou-se 1001 amostras do vetor $Y = (Y_1, Y_2)$, sendo gerados 121 dados de cada variável, em cada amostra e cada vetor de dados foi dividido em grupo de modelagem e grupo de controle, sendo este último composto com 40 observações de cada variável. Estas variáveis são co-locadas, isto é, são tomadas nas mesmas localizações.

Verifica-se na Tabela 1 as estatísticas descritivas das estimativas paramétricas por máxima verossimilhança. Observa-se que os parâmetros σ_2^2 e κ_1 foram superestimados.

Tabela 1 - Estatísticas descritivas das estimativas dos parâmetros obtidas com 1001 simulações

Parâmetros	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
μ_1	1,0320	1,7600	2,0090	2,0030	2,2240	3,2230
μ_2	0,8165	9,7330	10,0000	10,0100	10,2600	13,0600
σ_1^2	0,0522	0,2408	0,2836	0,2923	0,3300	1,2060
σ_2^2	0,0000	2,3650	2,8760	5,9870	4,0870	100,7000
τ_1^2	0,0000	0,1372	0,2060	0,2258	0,2823	1,1150
τ_2^2	0,0000	0,1235	0,1725	0,2379	0,2171	3,7630
κ_1	0,0197	0,3756	0,8850	2,3300	1,7810	83,5900
κ_2	0,2220	0,9127	2,6220	3,9540	4,6090	9,5600
ϕ	0,0000	0,0660	0,1689	0,2108	0,3119	2,6020
ρ	-0,9996	0,4222	0,5472	0,4657	0,6738	0,9890

Nas Tabelas 2 e 3, estão as estatísticas descritivas do erro quadrático médio e do erro absoluto, respectivamente, das estimativas do grupo de controle. Percebe-se que o erro quadrático da segunda variável possui uma variabilidade muito alta. Intervalos de predição foram criados e, em torno de 90% de cada intervalo contemplam os valores simulados das duas variáveis dos grupos de controle.

Tabela 2 - Estatísticas descritivas do erro quadrático (EQ) das variáveis pelo estudo de simulação

EQ	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
Y_1	0,0127	0,0313	0,0488	0,1050	0,0998	5,2430
Y_2	0,0100	0,0316	0,6195	5,2470	3,1670	119,0000

Tabela 3 - Estatísticas descritivas do erro absoluto (EA) das variáveis pelo estudo de simulação

EA	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
Y_1	-0,0491	-0,0088	-0,0003	-0,0002	0,0076	0,2136
Y_2	0,5881	-0,0187	0,0003	0,0008	0,019	1,3160

4.2 Análise dos dados do Pacífico Noroeste Norte Americano

Essa seção mostra os resultados obtidos da aplicação dos modelos detalhados na Revisão de Literatura no conjunto de dados “weather”, lembrando que os valores iniciais da variável pressão foram divididos por 100.

4.2.1 Análise exploratória

A Figura 2 mostra os gráficos de círculos com diâmetro proporcional aos valores dos atributos mensurados nas 157 localidades amostradas das variáveis em estudo.

Figura 2 - Gráfico de círculos - (a) para a variável pressão e (b) para variável temperatura.

Na Tabela 4 tem-se os valores da mediana, média e do desvio padrão de cada uma das variáveis, já utilizando a variável pressão com os valores transformados.

Na Figura 3 tem-se os gráficos descritivos das variáveis em estudo. Os gráficos do canto superior esquerdo das Figuras 3(a) e 3(b), categorizam os dados nos quartis

Tabela 4 - Estatística descritiva das variáveis em estudo, com variável pressão transformada

Variável	Mediana	Média	Desvio Padrão
Pressão	1,2691	0,9461	1,9487
Temperatura	-0,6305	0,1408	2,7182

amostrais das observações, em que os símbolos “+”, “ Δ ”, “o”, “x”, nesta ordem, indicam os quartis amostrais. Esses gráficos sugerem a existência de padrão espacial nos dados, uma vez que existem conglomerados das categorias. Essa dependência espacial também é detectada na Figura 4, que é o envelope do variograma. A existência de pontos fora do envelope indicam a presença de padrão espacial. O gráfico do canto inferior esquerdo da Figuras 3(a) e 3(b), mostram a distribuição de cada uma das variáveis com relação a coordenada no sentido oeste-leste, e os gráficos do canto superior esquerdo é com relação a coordenada no sentido norte-sul. Nota-se também que a variável temperatura possui tendência. Já os gráficos do canto inferior direito, referem-se ao histograma de cada variável. Nos gráficos do canto inferior direito das Figuras 3(a) e 3(b), tem-se a densidade amostral das variáveis. Não detectou-se desvio de normalidade nos dados.

Figura 3 - Gráfico descritivo do padrão espacial - (a) da variável pressão e (b) da variável temperatura

Figura 4 - Gráfico dos envelopes do variograma - (a) da variável pressão e (b) da variável temperatura

4.2.2 Análise univariada

Para validação de cada modelo estudado, retirou-se 52 pontos amostrados (grupo controle). Os pontos em vermelho na Figura 5 representam os pontos retirados. Assim, com alguns pontos retirados da análise, realizou-se o variograma experimental para cada uma das variáveis, representados na Figura 6 e ajustou-se um modelo para esses variogramas. Nota-se que para as duas variáveis o variograma teórico explica muito bem o variograma experimental em distâncias menores. Além disso, uma estimação univariada foi realizada para as duas variáveis e também fez-se krigagem para a área de estudo. Realizou-se a estimação dos parâmetros e estes valores estão na Tabela 5 e a krigagem

realizada na área de estudo, pode ser observada na Figura 7. Além disso, predições foram feitas para o grupo controle.

Figura 5 - Localização dos 52 pontos retirados (em vermelho)

Figura 6 - Gráfico dos variogramas experimental e teórico - (a) da variável pressão e (b) da variável temperatura

Tabela 5 - Valores do parâmetros estimados no modelo univariado

Variável	beta	τ^2	σ^2	alcance
Pressão	1,898	0,003	5,597	2,138
Temperatura	-0,599	0,000	7,1317	1,2137

Figura 7 - Gráfico das previsões espaciais (krigagem) univariada - (a) da variável pressão e (b) da variável temperatura

4.2.3 Análise Bivariada

4.2.3.1 Análise pelo Modelo Linear de Correogionalização

Através do pacote gstat (Pebesma, 2004), realizou-se o estudo bivariado pelo modelo linear de correogionalização. Realizou-se o variograma cruzado das variáveis, como pode ser observado na Figura 8 onde no canto superior esquerdo e no canto inferior direito tem-se os variogramas das variáveis pressão e temperatura, respectivamente e o gráfico do canto inferior esquerdo é o variograma cruzado entre as variáveis. Pode-se notar que o mesmo é decrescente, isso ocorre devido ao fato das variáveis serem negativamente correlacionadas.

Figura 8 - Gráfico dos variogramas univariados e do variograma cruzado

Ajustou-se um modelo para cada um dos variograma da Figura 8 e os parâmetros estimados do variograma cruzado estão na Tabela 6. Percebe-se que as variâncias foram super estimadas.

Tabela 6 - Valores dos parâmetros estimados para o variograma cruzado pelo modelo linear de correionalização

Variograma	Modelo	τ^2	σ^2	alcance	kappa
Pressão	Matérn	0,160	11,209	11	0,5
Temperatura	Matérn	1,283	18,437	11	0,5
Pressão $\times$ Temperatura	Matérn	0,135	6,951	11	0,5

Após ajustar o modelo linear de correionalização, realizou-se uma nova krigagem para cada uma das variáveis, a partir do modelo ajustado. Esta nova krigagem foi realizada em uma malha de pontos para todas as localidades entre as coordenadas dos dados amostrados, dado na Figura 9.

Figura 9 - Gráfico das predições espaciais (krigagem) após o ajuste do MLC - (a) da variável pressão e (b) da variável temperatura

Na Figura 10 tem-se um gráfico das predições realizada no grupo controle, composto pelos 52 pontos retirados da análise. Pode-se observar que em localizações com valores altos da variável pressão, estão baixos valores de temperatura, corroborando o fato de serem variáveis negativamente correlacionadas.

Figura 10 - Gráfico das predições espaciais (krigagem) para os 52 pontos retirados da análise - (a) da variável pressão e (b) da variável temperatura

4.2.3.2 Análise pelo Modelo Gaussiano Bivariado com Componente Comum

Para esta análise, utilizou-se as funções disponíveis no pacote geoR (Ribeiro Jr. e Diggle, 2001) e o modelo utilizado foi o (14), com a devida decomposição do campo latente, como revisado na seção 2.2.5.

As variáveis em estudo possuem correlação negativa o que não é acomodado na implementação do BGCCM. Para contornar este problema redefiniu-se a variável pressão por: $Y_{1p} = \max(Y_1) - Y_1$. Além disso, foram feitas adaptações na programação do modelo BGCCM na geoR (Ribeiro Jr. e Diggle, 2001).

Na estimação dos parâmetros, considerou-se o $\kappa = 0,5$, fixo. Na Tabela 7 tem-se as estimativas dos parâmetros para este modelo, ressaltando que considerou-se média constante. Os σ_i , $i = 1, 2$ foram super estimados, principalmente o da segunda variável.

Tabela 7 - Valores dos parâmetros estimados para o modelo Gaussiano Bivariado com Componente Comum

Variável	média	σ_{01}^2	σ_{02}^2	σ_1^2	σ_2^2	alcance
Pressão	1,5287	11,2793	—	23,8579	—	1,2137
Temperatura	1,5287	—	34,6511	—	0,0000	1,2137

Após a estimação dos parâmetros, realizou-se krigagem para cada uma das variáveis, que pode ser observada nas Figuras 11 (a) e (b), em que as áreas em vermelho possuem os menores valores das variáveis.

Figura 11 - Gráfico das previsões espaciais (krigagem), em uma malha de pontos, pelo BGCCM - (a) da variável pressão e (b) da variável temperatura

4.2.3.3 Análise pelo Modelo Matérn Multivariado Parcimonioso

O parâmetro de correlação estimado foi $-0,472$ e na Tabela 8, estão os valores estimados dos outros parâmetros deste modelo. Para a variável pressão, o modelo conseguiu estimar bem os parâmetros; para a temperatura foi estimado uma média negativa sendo que a média da variável é positiva porém, o modelo conseguiu estimar bem a variância desta variável.

Tabela 8 - Valores do parâmetros estimados para o modelo Matérn multivariado parcimonioso

Variável	média	σ^2	τ^2	kappa	alcance
Pressão	1,485	5,092	0,310	2,769	1,855
Temperatura	-0,590	6,831	0,536	1,420	1,855

Na Figura 12 observa-se os mapas de krigagem, para as duas variáveis, realizado em uma malha de pontos que abrange a área estudada.

Figura 12 - Gráfico das previsões espaciais (krigagem) em uma malha de pontos pelo modelo Matérn multivariado parcimonioso - (a) da variável pressão e (b) da variável temperatura

4.2.3.4 Análise por métodos Bayesianos

Para a abordagem bayesiana, utilizou-se o pacote `spbayes` (Finley, Banerjee e Carlin, 2012) para realizar as análises.

Na análise bayesiana, considerou-se 60 número de nós, que representam as variáveis (de entrada e de saída) que se interrelacionam e representam a estrutura de raciocínio de um especialista num domínio de aplicação; o número de iterações para Monte Carlo via cadeias de Markov escolhido foi de 5000 iterações, com burning das 2500 primeiras iterações. Com isso, iniciou-se a análise bayesiana.

Para as distribuições a priori, decidiu-se por optar na distribuição Uniforme para os valores iniciais de ϕ ; optou-se pela Wishart-Inversa como distribuição a priori tanto para matriz não espacial de covariância cruzada, denominada Ψ , como para matriz espacial de covariância cruzada, denominada K .

Na Tabela 9 observa-se os quantis e as médias para cada um dos parâmetros estimados. Nas Figuras 13, 14, 15 e 16, observa-se os traços (lado esquerdo do gráfico) e a densidade (lado direito do gráfico) dos parâmetros estimados: os interceptos, as partições das matrizes de covariância cruzadas não espacial (Ψ) e espacial (K) e os ϕ .

Tabela 9 - Quantis e Média para cada parâmetro por método bayesiano

	2,5%	25%	50%	75%	97,5%	Média
(Intercepto).mod1	0,52216	0,78149	0,92246	1,07643	1,4211	0,93104
(Intercepto).mod2	-0,50882	-0,10837	0,06343	0,26122	0,5996	0,06280
K[1,1]	2,50612	3,00143	3,35305	3,68684	4,4874	3,36666
K[2,1]	-2,83684	-2,22275	-1,93240	-1,54184	-1,0857	-1,90875
K[2,2]	4,07998	4,94664	5,47120	5,96653	7,1985	5,51756
$\Psi[1,1]$	0,07714	0,13747	0,19820	0,30028	0,5911	0,23875
$\Psi[2,1]$	-0,28691	-0,06686	-0,02074	0,02164	0,1702	-0,02424
$\Psi[2,2]$	0,09272	0,17360	0,23824	0,36461	0,8082	0,29911
ϕ_1	0,57644	0,62445	0,64401	0,65116	0,6574	0,63527
ϕ_2	0,49635	0,55552	0,58157	0,60114	0,6313	0,57595

Figura 13 - Gráfico do traço e da densidade do intercepto e de parte da matriz K

Figura 14 - Gráfico do traço e da densidade da matriz K e de parte da matriz Ψ

Figura 15 - Gráfico do traço e da densidade da matriz Ψ e do ϕ_1

Figura 16 - Gráfico do traço e da densidade do ϕ_2

Na Figura 17, tem-se o gráfico de krigagem para as duas variáveis (lado esquerdo) e para a predição das mesmas (lado direito). Pode-se observar que o modelo conseguiu prever muito bem tanto a variável pressão, quanto a variável temperatura para os valores mais altos, representados pelas cores mais fortes. Depois de ajustar o modelo e realizar a krigagem das variáveis utilizadas, realizou-se a krigagem para os 52 pontos retirados para validação do modelo, dada na Figura 18. Observa-se que o modelo fez uma boa predição para a segunda variável (temperatura), mas não fez uma boa predição para a primeira, temperatura, predizendo valores maiores dos que os reais. Em ambas as krigagens, o modelos não conseguiu prever bem para valores menores da variável pressão.

Figura 17 - Gráfico de krigagem das variáveis em estudo e para a predição das mesmas no modelo bayesiano - primeira variável (pressão), segunda variável (temperatura)

Figura 18 - Gráfico de krigagem das variáveis do grupo controle e da predição das mesmas no modelo bayesiano - primeira variável (pressão), segunda variável (temperatura)

4.2.4 Avaliação dos Modelos

Como dito anteriormente, para validação dos modelos retirou-se parte dos dados - 52 pontos de cada variável. Assim, estes pontos foram estimados por todos os modelos estudados e por abordagem univariada e as predições são dadas Tabelas 10 e 11 para pressão e temperatura, respectivamente.

Tabela 10 - Valor observado e valores preditos, do grupo controle, da variável Pressão

(continua)					
Observado	Univariado	LMC	BGCCM	Matérn	Bayesiano
2,7873	3,2636	3,4252	8,4769	2,3796	2,4334
1,2219	-1,6648	-1,3562	1,1917	0,8115	-0,3392
0,7314	0,3009	0,1531	2,7173	1,0296	-0,6871
2,4039	1,7682	1,6987	6,6551	1,000	0,4602
3,4360	3,8245	3,7643	8,5900	2,5961	0,4411
0,4345	0,5077	0,5477	3,5329	0,7222	-0,0657
-0,3408	0,2399	0,0400	7,9468	0,6277	1,7582
2,1398	1,4329	1,3316	3,29901	1,5666	-0,5520
1,7962	2,2353	2,1418	6,1808	1,0659	3,5108
2,2783	3,3161	3,0560	6,7434	1,1321	0,3148
1,3567	1,8279	1,7864	2,55088	1,6005	0,7603
-5,8891	0,2513	0,0557	7,76014	1,0034	0,5245
1,1555	1,4942	1,5526	4,2368	1,0280	0,2173
1,6448	2,6524	2,4472	6,1577	1,1660	6,3204
1,7137	2,0218	1,9379	6,2007	1,0549	0,8529
0,3053	0,5074	0,5001	4,3898	1,2146	1,3774
-1,2259	-1,1335	-1,0591	3,6145	0,4749	2,5696
0,7682	0,0843	-0,0207	2,2733	1,0048	-0,4809
2,3539	3,2367	3,3812	8,0911	2,3476	3,1935
3,4928	3,2443	3,3302	6,8445	2,3861	-0,0422
-1,6754	-1,8656	-1,8076	2,0651	0,4632	1,5123
0,5027	0,2037	0,1102	2,4486	1,0166	1,6922
0,1932	0,1125	0,0954	2,4789	1,0279	-0,3931
1,8975	3,0191	3,0826	8,7541	2,3543	1,9766
0,7764	0,6225	0,4485	7,7067	0,6187	1,5085
-3,1967	1,2496	1,1854	-0,8878	1,1146	-0,4775
1,6708	3,5026	3,0578	7,4416	1,1208	2,4056
1,6993	0,8949	0,8608	4,4139	1,0212	-0,0230
1,2010	1,1451	1,0936	8,5572	1,3713	2,6325
2,6492	1,7489	1,7089	6,5687	1,0186	-0,6622
0,9163	0,8878	0,8267	4,8579	1,2156	0,5092
2,0780	3,1136	2,8365	6,5704	1,1503	1,7651
1,7616	2,4928	2,3121	6,7984	1,1167	-0,7060

Tabela 10 - Valor observado e valores preditos, do grupo controle, da variável Pressão

Observado	Univariado	LMC	BGCCM	Matérn	(conclusão)	
					Parcimonioso	Bayesiano
2,8166	2,2888	2,4465	5,4287		1,2327	3,0006
-1,0403	-1,0209	-1,1350	1,4900		0,6708	0,1736
2,6347	2,6083	2,6645	5,7433		1,1412	-0,3755
1,1201	3,0116	3,1820	9,7566		2,3743	2,2122
4,6475	5,2297	5,8485	12,3334		2,5344	-0,4422
-0,9207	-0,1226	-0,5206	4,9518		0,8094	3,6673
1,4724	-1,7949	-1,4887	2,0788		0,8422	-0,7064
-0,0492	-1,0119	-1,1050	6,5318		0,6055	-0,9455
2,0048	2,0508	2,7022	13,3396		1,5899	0,8643
1,4588	2,0080	1,9866	6,4267		1,6886	0,3098
0,0106	-0,0487	-0,2238	7,7627		0,6394	-0,1801
1,5178	-1,6879	-1,4031	2,3609		0,8471	0,1999
-1,1014	-1,6048	-1,6097	3,8279		0,7013	0,2545
2,4627	-2,6183	-2,1372	0,0467		0,8135	2,5324
-0,6531	-1,1322	-1,1862	6,4491		0,6386	1,0754
1,2677	1,3170	1,3737	3,7929		0,9679	0,4915
-0,5286	-0,9909	-1,0844	6,4416		0,5969	4,7904
1,2691	2,7720	2,6968	5,5251		1,1848	0,1353
-0,6199	-1,0080	-1,0919	6,8479		0,6243	0,2399

Tabela 11 - Valor observado e valores preditos, do grupo controle, da variável Temperatura

Observado	Univariado	LMC	BGCCM	Matérn	(continua)	
					Parcimonioso	Bayesiano
0,2420	-1,2437	-1,2626	-2,5760		-0,5570	1,8429
-0,6513	-0,4399	0,2528	-2,0575		-0,6525	-1,0365
-0,6390	-1,1700	-0,6901	-3,1759		-0,1594	0,0314
-3,3724	-1,0781	-1,2629	-2,5480		-0,8101	-0,0436
-1,1076	-0,3794	-0,7867	-0,3562		-0,6244	-0,2802
-0,7606	-0,7212	-0,5898	-2,0950		-0,5016	2,6416
5,6492	5,4078	5,0535	11,7393		1,0474	-0,0900
0,1679	-0,3046	-0,0112	-0,8901		-0,1726	-0,5152
-4,7001	-1,9489	-2,0001	-4,4029		-0,8063	-0,4302
-1,2109	-1,6111	-1,6212	-3,2896		-0,5279	0,1219
0,2502	0,7757	0,7164	1,6869		-0,0107	1,5991
4,4346	-0,2392	-1,2089	-1,0357		-0,4928	-1,5220
-1,2707	-1,4149	-1,3520	-3,3785		-0,5808	0,8879
-2,2498	-1,2719	-1,3030	-2,7217		-0,4765	-0,3632
-2,7466	-1,7909	-1,9358	-4,1059		-0,8019	1,2945
-0,4714	0,3019	0,2213	0,2314		-0,1183	1,0999
2,4354	1,9673	2,4537	3,5543		0,7224	-2,2762
-1,2178	-1,3265	-0,6875	-3,5932		-0,1291	-0,6411
-1,4580	-1,3999	-1,3386	-2,9233		-0,5358	2,0070
0,8776	-1,7902	-1,4513	-3,7414		-0,4805	-1,2817
0,3536	0,5068	1,3991	0,0761		0,4821	1,0575
-0,7264	-1,3049	-0,7293	-3,5105		-0,1019	0,3801
0,1633	-1,1095	-0,6141	-3,0936		-0,0771	3,0200
-0,4448	0,0220	-0,1724	0,2749		-0,3741	-1,6639
2,6569	3,8784	3,7963	8,3934		1,0596	-1,2382
3,0335	-1,5517	-2,6819	-3,7734		-0,6468	1,2118
-1,9787	-1,0675	-1,5510	-2,0002		-0,5060	-2,2266
-0,6073	-0,2132	-0,5919	-0,8231		-0,3771	-1,0992
3,0584	0,3096	1,1467	0,4330		0,2725	-1,9120
-0,6791	-1,2347	-1,2320	-2,8827		-0,7277	5,8980
-0,6365	0,0972	-0,1058	-0,1232		-0,2171	1,7336
-0,7015	-1,5093	-1,5447	-3,1221		-0,5131	-0,3291
-0,2640	-0,0820	-1,1209	-0,0623		-0,7227	0,2587
-1,2612	-1,8169	-1,4045	-4,0712		-0,6977	0,9678
-1,6437	-0,4874	-0,9840	2,0025		-0,2489	-1,2440
-1,7396	-1,7190	-1,6941	-3,7430		-0,7308	5,6349
-2,5967	0,1059	-0,1043	0,4231		-0,4049	-2,5301
-3,9677	-1,4639	-2,2486	-2,6276		-0,8791	-0,8157
-2,1334	0,1623	-0,5648	-0,2490		-0,4263	-2,7849
-0,1159	-0,5197	0,3278	-2,2404		-0,6563	3,2333
4,9246	6,0272	5,4061	12,7747		0,9439	-2,8285
0,6053	-0,6389	-1,2560	-1,5993		-0,6458	-0,5999
0,8991	0,2713	0,3436	0,5984		-0,1120	-2,2679
5,7472	5,8399	5,3859	12,6321		1,0233	6,0622

Tabela 11 - Valor observado e valores preditos, do grupo controle, da variável Temperatura

Observado	Univariado	LMC	BGCCM	Matérn	(conclusão)	
					Parcimonioso	Bayesiano
-0,0518	-0,5360	0,2889	-2,2424	-0,6572		0,0058
3,1238	3,3625	3,1585	6,5712	0,5960		-1,1602
-2,1738	0,3989	0,6319	-2,2398	-0,6455		-2,3186
6,6589	6,1157	5,3987	12,9423	0,9232		5,7721
-2,4662	-1,8263	-1,8413	-4,3657	-0,5188		1,2889
5,1073	5,8617	5,3052	12,4058	0,9438		-4,5394
-3,839	4,5163	-4,6979	-10,0727	-0,7767		-2,0032
6,5978	6,4556	5,6610	13,7464	0,9536		6,2389

Além das tabelas, pode-se observar a relação entre os valores reais das variáveis e os preditos por cada modelo através das Figuras 19 e 20. Analisando as figuras, pode-se notar que nenhum modelo ajustou muito bem as duas variáveis, sendo que dos modelos bivariados o LCM se destaca nas melhores previsões. Nota-se também que o modelo de regressão bayesiano superestimou os valores da variável temperatura.

Nas Tabelas 12 e 13 estão os resultados descritivos das diferenças entre os valores observados e os valores preditos do grupo controle para a variável pressão e temperatura, respectivamente. Para a variável pressão, os modelos bivariados conseguiram previsão tão boa quanto o modelo univariado; só o BGCCM que teve um desvio padrão bem maior que os demais. Para a variável temperatura os modelos bivariados também conseguiram explicá-la tão bem quanto o univariado.

Tabela 12 - Média e desvio padrão dos erros de krigagem do grupo controle para Pressão

Estatística	Univariado	LMC	BGCCM	Matérn	Parcimonioso	Bayesiano
Média	0,0764	0,0698	-0,2782		0,2022	0,0153
Desvio Padrão	1,6280	1,5482	3,5581		1,5064	2,3366

Tabela 13 - Média e desvio padrão dos erros de krigagem do grupo controle para Temperatura

Estatística	Univariado	LMC	BGCCM	Matérn	Parcimonioso	Bayesiano
Média	0,0045	-0,0537	-0,1222		0,3507	0,3860
Desvio Padrão	1,4826	1,5645	3,4150		2,3002	3,3362

(a)

(b)

(c)

(d)

(e)

Figura 19 - Gráfico de dispersão da variável pressão versus valores preditos pelos modelo: (a)univariado, (b)LMC, (c)BGCCM, (d)Matérn Parsimonioso, (e)Bayesiano

(a)

(b)

(c)

(d)

(e)

Figura 20 - Gráfico de dispersão da variável temperatura versus valores preditos pelos modelos: (a)univariado, (b)LMC, (c)BGCCM, (d)Matérn Parsimonioso, (e)Bayesiano

Com relação aos modelos bivariados estudados, através das Tabelas 10, 11, 12 e 13 e analisando os gráficos nas Figuras 19 e 20, pode-se afirmar que o modelo linear de correionalização se destaca dentre os demais pois com esse modelo se obteve as melhores predições para as duas variáveis estudadas.

5 CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo avaliar o comportamento de modelos geoestatísticos multivariados para um particular conjunto de dados. Além disso, realizou-se uma programação, no software R, para um estudo de simulação no modelo Matérn parcimonioso multivariado.

Neste modelo, existe uma restrição ao parâmetro de correlação (ρ) que depende dos parâmetros κ_i , $i = 1, 2$ e isso trás maiores dificuldades na estimação dos parâmetros. Na implementação, os procedimentos numéricos utilizados na simulação mostraram-se instáveis, indicando potenciais dificuldades na estimação dos parâmetros do modelo. Desta forma, apesar do estudo de simulação, foi bastante complicado programar o modelo de uma forma a deixá-lo mais estável e apesar de todo o trabalho demandado, o programa não ficou tão estável como se gostaria, visto que no estudo de simulação a variância da segunda variável foi superestimada. A inclusão de informação de gradiente pode melhorar o desempenho do algoritmo mas não foi possível implementá-la a tempo da conclusão deste trabalho. Fica como um possível trabalho futuro. Uma possível extensão seria implementar e avaliar o modelo Matérn completo multivariado, também proposto por Gneiting, Kleiber e Schlater (2009).

Na análise dos dados weather percebeu-se que os modelos bivariados conseguiram prever bem as variáveis, ocorrendo um destaque para o modelo linear de correção regionalização. Desta forma, na falta de recursos para medir e estudar as duas variáveis, o estudo de uma pode ser baseado nas medições e análises realizados na outra.

REFERÊNCIAS

- BELLIER, E.; MONESTIEZ, P.; DURBEC, J.P.; CANDAU, J.N. Identifying spatial relationships at multiple scales: principal coordinates of neighbor matrices (PCNM) and geostatistical approaches. **Ecography**, Malden, v.30, p.385-399, 2007.
- BOGNOLA, I.A.; RIBEIRO JÚNIOR, P.J.; SILVA, E.A.A; LINGNAU, C.; HIGA, A.R. Modelagem Uni e Bivariada Da Variabilidade Espacial de Rendimento De Pinus taeda l. **Floresta**, Curitiba, v.38, n. 2, 13p, abr/jun.2008.
- BOX, G.; COX, D. An analysis of transformation. **Journal of the Royal Statistical Society, Series B**, London, v.26, n. 2, p. 211-252, 1964.
- CHILÈS, J. -P.; DELFINER, P. **Geostatistics: modeling spatial uncertainty**. New York: John Wiley & Sons, 1999. 722p.
- CRESSIE, N.A.C. (Ed.). **Statistics spatial data**. New York: wiley, 1993.
- DIGGLE, P. J.; RIBEIRO JR., P. J. **Model-based geostatistics**. USA: Springer Series in Statistics, 2007. 241p.
- DIGGLE, P.J.; RIBEIRO JR., P.J.; CHRISTENSEN, O.F. An introduction to model-based geostatistics. In: MØLLER, J. (Ed.). **Spatial Statistics and Computational Methods**. New York: Springer, 2003. p. 43-86.
- DRUCK, S.; CARVALHO, M.S.; CÂMARA, G.; MONTEIRO, A.M.V. **Análise espacial de dados geográficos**. Brasília, EMBRAPA, 2004. 209p.
- FINLEY, A. O.; BANERJEE, S.; CARLIN, B. P. spBayes: An R Package for Univariate and Multivariate Hierarchical Point-referenced Spatial Models. **Journal of Statistical Software**, Minnesota, v.19, n.4, 2007.
- FINLEY, A. O.; BANERJEE, S.; CARLIN, B. P. **spBayes**: Univariate and Multivariate Spatial Modeling. R package version 0.2-2. 2012. Disponível em: <<http://CRAN.R-project.org/package=spBayes>>. Acesso em: 01 ago. 2011.
- FONSECA, B.H.F. **Um estudo sobre estimação e predição em modelos geoestatísticos bivariados**. 2008. 74p. Dissertação (Mestrado em Ciências Exatas) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2008
- GASPARI, G.; COHN, S.E. Construction of correlation functions in two and three dimensions. **Quarterly Journal of the Royal Meteorological Society**, Malden, n.125, p. 723-757, 1999.
- GNEITING, T. Compactly Supported Correlation Functions. **Journal of Multivariate Analysis**, Amsterdã, n. 83, p. 493-508, 2002.

GNEITING, T.; KLEIBER, W.; SCHLATHER, M. Matérn Cross-Covariance Functions for Multivariate Random Fields. **Technical Report n.549. Department of Statistics**, University of Washington. 2009

GOOVAERTS, P. **Geostatistics for Natural Resources Evaluation**. New York: Oxford University Press, 1997. 496p.

GRIPP, A.H. Utilização de Geoestatística para o tratamento de dados de prospecção geoquímica. **Revista Brasileira de Geociências**, Visconde, v. 22, n. 2, p. 248-251, jun. 1992

ISAAKS, E.H.; SRIVASTAVA, R.M. **An Introduction to Applied Geostatistics**. New York: Oxford University Press, 1989. 561 p.

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

JUN, M.; KNUTTI, R.; NYCHKA, D.W. Spatial Analysis to Quantify Numerical Model Bias and Dependence. **Journal of the American Statistical Association**, London, v. 103, n. 483, p. 934-947, 2008.

KOLMAN, B. **Bayesian statistics: an introduction**. New York: Hodder Arnold, 2004. 462p.

LANDIM, P.M.B. **Sobre Geoestatística e mapas**. Terrae Didatica, v. 2, n. 1, p.19-33, 2006. Disponível em: <<http://www.ige.unicamp.br/terradidatica/>>. Acesso em: 05 jan. 2011.

MARTINS, A.B.T. **Análise geoestatística de dados composicionais**. 2010. 200 p. Tese (Doutorado) - Universidade Federal do Paraná, Curitiba, 2010.

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

OLIVER, D.S. Gaussian cosimulation: Modelling of the cross-covariance. **Mathematical Geology**, Montreal, v. 35, p. 681-698, 2003.

PEBESMA, E. J. **Multivariable geostatistics in S: the gstat package**. Computers & Geosciences, v. 30, p. 683-691. Disponível em: <<http://cran.r-project.org/web/packages/gstat/>>. Acesso em: 01 ago. 2011.

R: A Language and Environment for Statistical Computing, R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2006, ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>. Acesso em: 01 jan. 2011.

RIBEIRO JÚNIOR, P. J. **Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo**. 1995. 99 p. Dissertação (Mestrado em Ciências Exatas) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1995.

RIBEIRO Jr, P.J.; DIGGLE, P.J. geoR: A package for geostatistical analysis. **R-NEWS**, Vienna, v.1, n.2, p.14-18, 2011.

SALVIANO, A.A.C. **Variabilidade de atributos de solo e de *Crotalaria juncea* em solo degradado do município de Piracicaba-SP**. 1996. 91 p. Tese (Doutorado em Ciências Exatas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1996.

SCHLATHER, M. **RandomFields**: Simulation and Analysis of Random Fields. R package version 2.0.54. Disponível em:
<<http://CRAN.R-project.org/package=RandomFields>>. Acesso em: 01 jan. 2011.

SILVA, E.A.A.; URIBE-OPAZO, M.A.; SOUZA, E.G.; ROCHA, J.V. Um estimador robusto e o semivariograma cruzado na análise de variabilidade espacial de atributos de solo e planta. **Acta Scientiarum**. Agronomy. Maringá, v. 25, n. 2, p. 365-371, 2003.

SOARES, A. **Geoestatística para ciências da terra e do ambiente**. 2.ed. Portugal: IST Press, 2006. 214 p.

VER HOEF, J.M.; BARRY, R. P. Constructing and fitting models for cokriging and multivariable spatial prediction. **Journal of Statistical Planning and Inference**, Amsterdã, v. 69, p. 275-294, 1998.

VER HOEF, J.M.; CRESSIE, N.; BARRY, R.P. Flexible spatial models for kriging and cokriging using moving averages and the Fast Fourier Transform (FFT). **Journal of Computational and Graphical Statistics**, Washington, v. 13, p. 265-282, 2004.

VIEIRA, S.R. Geoestatística em estudos de variabilidade espacial do solo. **Tópicos em ciência do solo**. Viçosa, v. 1, p. 1-54, 2000.

WACKERNAGEL, H. **Multivariate Geostatistics**- An Introduction with applications. Berlin: Springer-Velag, 2003. 403p.

YAMAMOTO, J.K. **Avaliação e Classificação de Reservas Minerais**. São Paulo: EDUSP, 2001. 226p.

ZHANG, H. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. **Journal of the American Statistical Association**, London, V. 99, p. 250-261, 2004.

APÊNDICES

APÊNDICE A - Programação no pacote geoR do Modelo Matérn Multivariado Par-cimonioso

```

require(geoR)
require(RandomFields)

LP <- list(nu1 = 0.9 , nu2 = 0.75,
          a = 1/4 ,
          rho = 0.7 ,
          sig1 = sqrt(0.7), sig2 = sqrt(0.5) ,
          tau1 = sqrt(0.2), tau2 = sqrt(0.1)
          )

nx      = 121
ny      = 121
nsample = 40

sim<-150
sim1<-NULL
sim2<-NULL

var1<- matrix(NA,ncol=1,nrow=(nx-nsample))
var2<-matrix(NA,ncol=1,nrow=(ny-nsample))

pred1<-matrix(NA,ncol=1,nrow=(nsample))
pred2<-matrix(NA,ncol=1,nrow=(nsample))

pred.var1 <-matrix(NA,ncol=1,nrow=(nsample))
pred.var2 <-matrix(NA,ncol=1,nrow=(nsample))

krige.var1 <-matrix(NA,ncol=1,nrow=(nsample))
krige.var2 <-matrix(NA,ncol=1,nrow=(nsample))

#0 que o programa retorna:
gerado <- matrix(NA,ncol=sim,nrow= (nx+ny) )
variavel <-matrix(NA,ncol=sim,nrow=length(c(var1,var2)))
predi <- matrix(NA,ncol=sim,nrow=length(c(pred1,pred2)))
ajustesim <- matrix(NaN,ncol=10,nrow=sim) #parametros
predicao <-matrix(NA,ncol=sim,nrow=length(c(pred.var1,pred.var2)))
krige.var <- matrix(NA,ncol=sim,nrow=length(c(krige.var1,krige.var2)))
llik <- matrix(NA,ncol=sim,nrow=1)

#Inicio da simulação:
for (nsim in 1:sim) {

## grid irregular
gs <- cbind(runif(121, 0, 25), runif(121, 0, 25))

#distancia
dM <- dist(as.matrix(gs), upper=T, diag=T)

##FUNCOES:

```



```

lower2matrix <- function(x){
  n <- (1+sqrt(1+8*length(x)))/2
  M <- diag(n)
  M[lower.tri(M)] <- x
  M <- t(M)
  M[lower.tri(M)] <- x
  M
}

#Montando Sigma
montaSigma <- function(parList, distPares){
  evalq({
    n <- (1+sqrt(1+8*length(distPares)))/2
    ## condicao/restricao
    if(abs(rho) > sqrt(nu1 * nu2)/((nu1+nu2)/2))
      stop("restricao nos parametros nao satisfeita")
    c11 <- cov.spatial(distPares,cov.model="matern",
      cov.pars=c(1, 1/a),kappa=nu1)
    c11 <- sig1^2 * lower2matrix(c11)
    diag(c11) <- diag(c11) + tau1^2
    c22 <- cov.spatial(distPares,cov.model="matern",
      cov.pars=c(1, 1/a),kappa=nu2)
    c22 <- sig2^2 * lower2matrix(c22)
    diag(c22) <- diag(c22) + tau2^2
    c12 <- cov.spatial(distPares,cov.model="matern",
      cov.pars=c(1, 1/a),kappa=(nu1+nu2)/2)
    c12 <- rho * sig1 * sig2 * lower2matrix(c12)
    S <- rbind(cbind(c11, c12), cbind(c12, c22))
  }, env=parList)
}

dV <- as.vector(dist(gs))
system.time(SIGMA <- montaSigma(LP, distPares=dV))
dM <- as.matrix(dist(gs, diag=T,up=T)); dimnames(dM) <- list(NULL, NULL)

cholSIGMA <- chol(SIGMA)
sim <- drop(crossprod(cholSIGMA, rnorm(dim(cholSIGMA)[1])))
rm(SIGMA)

sim <- sim + c(rep(2, 121), rep(10, 121))
sim1<-sim[1:(length(sim)/2)]
sim2<-sim[((length(sim)/2)+1):length(sim)]
conj<-cbind(gs,sim1,sim2)
amostra<-sample(seq(1:nrow(conj)),40,rep=FALSE)

###PREDIÇÃO
conjunto.pred<-conj[amostra,]
pred1<-conjunto.pred[,3]
pred2<-conjunto.pred[,4]

#UTILIZADO
conjunto<-conj[-amostra,]

```

```

var1<-conjunto[,3]
var2<-conjunto[,4]

logpars2pars <- function(pars){
  pars <- as.list(pars)
  if(!is.null(pars$sig1)) pars$sig1 <- exp(pars$sig1)
  if(!is.null(pars$sig2)) pars$sig2 <- exp(pars$sig2)
  if(!is.null(pars$tau1)) pars$tau1 <- exp(pars$tau1)
  if(!is.null(pars$tau2)) pars$tau2 <- exp(pars$tau2)
  if(!is.null(pars$a)) pars$a <- exp(pars$a)
  if(!is.null(pars$rho)) pars$rho <- (exp(pars$rho)-1)/(1+exp(pars$rho))
  return(unlist(pars))
}

geral <- function(pars1, pars2, logpars=FALSE,
                  printpars = FALSE, conc = FALSE, ...){
  pars<-as.list(c(pars1,pars2))
  rodaPARS <- as.vector(pars)
  TODOS <- c("mu1","mu2", "nu1","nu2", "sig1","sig2","tau1","tau2","a","rho")
  if(any(is.na(match(TODOS, names(pars)))))
    stop("tá faltando parametro!!!")
  if(!all(c("mu1","mu2","sig1") %in% names(pars1))) conc <- FALSE
  #1 ou
  #if(logpars){
    # pars$sig1 <- exp(pars$sig1)
    # pars$sig2 <- exp(pars$sig2)
    # pars$tau1 <- exp(pars$tau1)
    # pars$tau2 <- exp(pars$tau2)
    # pars$rho <- (exp(pars$rho)-1)/(1+exp(pars$rho))
  # }
  #2
  if(logpars){
    pars <- as.list(logpars2pars(pars))
  }
  if(pars$nu1 <=0 || pars$nu2 <=0) return(-.Machine$double.xmax^0.5)
  if(pars$tau1 <=0 || pars$tau2 <=0) return(-.Machine$double.xmax^0.5)
  if(pars$sig1 <=0 || pars$sig2 <=0) return(-.Machine$double.xmax^0.5)
  if(evalq(abs(rho)>1||abs(rho) > sqrt(nu1 * nu2)/((nu1+nu2)/2), env=pars))
    return(-.Machine$double.xmax^0.5)
  if(pars$a <=0 ) return(-.Machine$double.xmax^0.5)
  ll<-function(pars, dados) {

    # parsList recebe desvios (e não variâncias!)
    #padrão com opção para log(desvio padrao)
    # dados : list(d, X, n1, n2, n)1e-10

    if(conc){
      pars$sig1 <- 1
    }
    tchSIGMA <- t(chol(montaSigma(pars, dados$d)))

    if(conc){

```

```

## matrix (supondo n1 = n2) !!!!!
ONE <- kronecker(diag(1, 2), rep(1, dados$n1))
## se n1 != n2:
## ONE <- cbind(c(rep(1, dados$n1), rep(0, dados$n2)),
# c(rep(0, dados$n1), rep(1, dados$n2)))
ONEchSIGMA <- forwardsolve(tchSIGMA, ONE)
mus <- solve(crossprod(ONEchSIGMA),
             crossprod(ONEchSIGMA,
             forwardsolve(tchSIGMA, dados$X)))
}
else mus <- c(pars$mu1,pars$mu2) #pars[1:2]
Mu <- c(rep(mus[1],dados$n1), rep(mus[2],dados$n2))
res <- dados$X - Mu
z <- forwardsolve(tchSIGMA,res)
HQ <- sum(z*z)
detSIGMA <- 2*sum(log(diag(tchSIGMA))) ## det. calculado a partir da choleski

if(conc)
  llik <- -0.5 * (dados$n * (log(2*pi) + log(HQ/dados$n))
                + detSIGMA + dados$n)
else
  llik<- -0.5 * (dados$n * log(2*pi) + detSIGMA + HQ)

return(llik)
}
  llik <- ll(pars, ...)
if(printpars){
  if(logpars){
    cat("original pars: \n")
    print(unlist(pars)[1:length(pars1)])
  }
  cat("logpars: \n")
  print(unlist(pars1))
}
return(llik)
}

lista<-list(n=length(c(var1,var2)),
           n1=length(var1),
           n2=length(var2),
           X=c(var1,var2), d=as.vector(dist(conjunto[,1:2])))

ini <-c(tau1=log(sqrt(0.4)), tau2=log(sqrt(0.3)),
      rho=log((0.75+1)/(1-(0.75))),
      nu1=1.3, nu2=0.5,
      sig2=log(sqrt(0.6)),
      a=log(5))
fixos <- c(mu1=2,mu2=8,sig1=log(sqrt(0.65)))

ajuste <- optim(ini, geral, pars2=fixos, dados=lista,#method="BFGS",
control=list(fnscale=-1, maxit=3000), conc=T,

```

```

printpars =T,logp=T)

ajuste$par<-logpars2pars(ajuste$par)

pars.est<-list(a<-ajuste[[1]][[7]],
rho<- (ajuste[[1]][[3]]),
nu1<-ajuste[[1]][[4]],
nu2<-ajuste[[1]][[5]],
tau1<-(ajuste[[1]][[1]]),
tau2<-(ajuste[[1]][[2]]),
sig2<-(ajuste[[1]][[6]]),
sig1<-1
)

names(pars.est) <- c('a','rho','nu1','nu2','tau1','tau2','sig2','sig1')

dados<-list(n1=length(var1), n2=length(var2),
            X=c(var1,var2), d=as.vector(dist(conjunto[,1:2])))

param<-function(pars.est,dados){
  SIGMA <- montaSigma(pars.est, dados$d)
  tchSIGMA <- t(chol(SIGMA))
  ONE <- kronecker(diag(1, 2), rep(1, dados$n1))
  ONEchSIGMA <- solve(tchSIGMA, ONE)
  mus <- solve(crossprod(ONEchSIGMA),
               crossprod(ONEchSIGMA, solve(tchSIGMA, dados$X)))

  dif<-dados$X-ONE%*%mus
  sig1.<-(1/(dados$n1+dados$n2))*(t(dif)%*%solve(tchSIGMA,dif))

  sig1<-sig1.^2
  sig2<-(pars.est$sig2)^2
  tau1<-(pars.est$tau1)^2
  tau2<-(pars.est$tau2)^2
  nu1<-pars.est$nu1
  nu2<-pars.est$nu2
  a<-pars.est$a
  rho<-pars.est$rho
  mu1<-mus[1,]
  mu2<-mus[2,]
  parametros <-c(mu1,mu2,sig1,sig2,tau1,tau2,nu1,nu2,a,rho)
  names(parametros) = c('mu1','mu2','sig1','sig2','tau1',
                        'tau2','nu1','nu2','a','rho')

  return(parametros)
}

parametros <- param(pars.est,dados)
gerado[,nsim]<-c(sim1,sim2)
variavel[,nsim]<-c(var1,var2)
predi[,nsim]<-c(pred1,pred2)

```

```

parametros<-list(mu1<-parametros[[1]],
mu2<-parametros[[2]],
sig1<-parametros[[3]],
sig2<-parametros[[4]],
tau1<-parametros[[5]],
tau2<-parametros[[6]],
nu1<-parametros[[7]],
nu2<-parametros[[8]],
a<-parametros[[9]],
rho<-parametros[[10]]
)
names(parametros) = c('mu1','mu2','sig1','sig2','tau1',
                      'tau2','nu1','nu2','a','rho')

parametros
ajustesim[nsim,1:10] <- as.matrix(as.data.frame(parametros))

#####
###Para Predição:###
#####

k = NULL
k = as.data.frame(list(mu1,mu2,sig1,sig2,tau1,tau2,nu1,nu2,a,rho))
names(k) = c('mu1','mu2','sig1','sig2','tau1',
             'tau2','nu1','nu2','a','rho')
d = as.vector(dist(conjunto[,1:2]))

S<-montaSigma(k, d)
Sinv<-solve(S)

object<-as.list(c(n1=length(conjunto[,3]),
                  n2=length(conjunto[,4]),
                  mu1=parametros$mu1,
                  mu2=parametros$mu2,
                  sigma1=parametros$sig1,
                  sigma2=parametros$sig2,
                  phi=parametros$a,
                  kappa1=parametros$nu1,
                  kappa2=parametros$nu2
                  ))

geodata1<-as.geodata(conjunto,coords.col=1:2,data.col=3)
geodata2<-as.geodata(conjunto,coords.col=1:2,data.col=4)

#####
##PREDICAO:##
#####
#Predição dos valores retirados:

locations<-conjunto.pred[,1:2]

pred<-function (object, locations, variable.to.predict = 1)

```

```

{
  call.fc <- match.call()
  if (!is.vector(variable.to.predict) || !is.numeric(variable.to.predict) ||
      length(variable.to.predict) != 1)
    stop("vari?vel deve ser um escalar igual a 1 ou 2")
  locations <- locations
  geodata1 <- geodata1
  #geodata1 <- get(attr(object, "geodata1"))
  geodata2 <- geodata2
  #geodata2 <- get(attr(object, "geodata2"))
  #dimnames(locations) <- list(NULL, NULL)
  n0 <- nrow(locations)
  n1 <- object$n1
  n2 <- object$n2
  n <- n1 + n2
  S<-montaSigma(k, d)
  Sinv<-solve(S)
  res12 <- c((geodata1$data - object$mu1), (geodata2$data - object$mu2))
  S012 <- matrix(0, nrow = length(res12), ncol = nrow(locations))
  if (any(variable.to.predict == 1)) {
    mu0 <- rep(object$mu1, n0)
    var0 <- rep(object$sigma1, n0)
    S012[1:n1, ] <- cov.spatial(loccoords(geodata1$coords,
      locations),
      cov.pars = c(object$sigma1, object$phi),
      cov.model = "matern",
      kappa = object$kappa1)
    S012[(n1 + 1):(n1 + n2), ] <- cov.spatial(loccoords(geodata2$coords,
      locations), cov.pars = c(rho * sqrt(sig1) * sqrt(sig2), object$phi),
      cov.model = "matern",
      kappa = (object$kappa1+object$kappa2)/2)
  }
  if (any(variable.to.predict == 2)) {
    mu0 <- rep(object$mu2, n0)
    var0 <- rep(object$sigma2, n0)
    S012[1:n1, ] <- cov.spatial(loccoords(geodata1$coords,
      locations), cov.pars = c(rho * sqrt(sig1) * sqrt(sig2), object$phi),
      cov.model = "matern",
      kappa = (object$kappa1+object$kappa2)/2)
    S012[(n1 + 1):(n1 + n2), ] <- cov.spatial(loccoords(geodata2$coords,
      locations), cov.pars = c(object$sigma2, object$phi),
      cov.model = "matern", kappa = object$kappa2)
  }
  res <- list()
  res$predict <- mu0 + drop(crossprod(S012, Sinv) %*%
    res12)
  res$krige.var <- var0 - geoR::diagquadraticformXAX(S012,
    lowerA = Sinv[lower.tri(Sinv)], diagA = diag(Sinv))
  res$call <- call.fc
  attr(res, "prediction.locations") <- call.fc$locations
  return(res)
}

```

```

predicao1<-pred(object,locations,variable.to.predict=1)
#plot(conjunto.pred[,3],predi??o1$pred)
pred.var1<-predicao1$pred
krige.var1<-predicao1$krige.var

predicao2<-pred(object,locations,variable.to.predict=2)
#plot(conjunto.pred[,4],predicao2$pred)
pred.var2<-predicao2$pred
krige.var2<-predicao2$krige.var

  predicao[,nsim]<-c(pred.var1,pred.var2)
  krige.var[,nsim]<-c(krige.var1,krige.var2)

  ##Loglik do modelo

loglik<-geral(pars1=c(tau1=sqrt(tau1), tau2=sqrt(tau2),
                    rho=rho, nu1=nu1, nu2=nu2, sig2=sqrt(sig2), a=a),
pars2=c(mu1=mu1, mu2=mu2, sig1=sqrt(sig1)), dados=lista,
conc=T,logpars=T, printpars=F)

  llik[,nsim]<-loglik

}

```

ANEXOS

Anexo A - Conjunto de dados utilizados

Tabela 14 - Medidas de Pressão e Temperatura em 157 localizações (coordenadas) no Pacífico Noroeste Norte Americano

(continua)				
Longitude	Latitude	Pressão	Temperatura	Pressão/100
-131.0000	46.00000	200.484375	0.60537720	2.00484375
-124.4000	41.90000	384.851562	-0.02233887	3.84851563
-124.5000	46.10000	156.898438	-0.26644897	1.56898438
-124.7000	47.30000	248.429688	-1.30670166	2.48429688
-124.5000	44.60000	253.226562	0.14398193	2.53226562
-124.9850	49.90700	159.203125	-0.27954102	1.59203125
-125.0000	49.90000	159.523438	-0.36056519	1.59523437
-127.9220	49.73200	147.242188	-0.11590576	1.47242187
-127.9000	49.70000	151.781250	-0.05181885	1.51781250
-123.7000	49.30000	176.164062	-0.26400757	1.76164063
-128.8000	51.40000	54.515625	-0.92639160	0.54515625
-126.0000	48.83300	261.898438	-1.73202515	2.61898438
-126.0000	48.80000	263.476562	-1.73962402	2.63476562
-129.9100	50.86000	46.726562	-0.73724365	0.46726563
-129.9000	50.90000	43.453125	-0.76062012	0.43453125
-124.3800	43.34000	278.734375	0.24209595	2.78734375
-123.5000	48.30000	207.804688	-0.70153809	2.07804688
-122.9500	50.13333	179.625000	-4.70013428	1.79625000
-123.3000	49.48333	299.476562	-2.84262085	2.99476562
-121.5000	51.15000	-2.101562	1.98922729	-0.02101562
-126.5500	49.38333	264.929688	-0.67916870	2.64929688
-123.8000	49.21667	241.171875	0.70336914	2.41171875
-123.0500	48.78333	218.125000	-1.67288208	2.18125000
-128.9000	50.81667	166.054688	0.63452148	1.66054688
-124.1667	49.35000	240.445312	-2.01202393	2.40445312
-122.7333	50.30000	171.375000	-2.74661255	1.71375000
-124.4333	49.48333	232.984375	-0.28515625	2.32984375
-114.9000	49.75000	-123.468750	-2.26663208	-1.23468750
-116.5000	49.08333	-285.500000	2.99017334	-2.85500000
-119.3000	50.23000	-43.148438	0.59146118	-0.43148437
-123.5833	48.58333	409.312500	-0.65832520	4.09312500
-121.5833	50.23333	156.265625	-2.77227783	1.56265625
-122.6833	49.20000	156.343750	-1.53796387	1.56343750
-117.8167	50.26667	-163.750000	1.59265137	-1.63750000
-123.4300	48.43000	227.835938	-1.21099854	2.27835937
-120.5167	50.60000	-173.789062	0.63757324	-1.73789062
-124.8333	49.25000	199.789062	-2.29895020	1.99789062
-127.9333	50.11667	122.195312	-0.65136719	1.22195312
-119.2833	50.70000	-92.070312	-2.13345337	-0.92070312
-123.9167	48.38333	163.898438	-3.01309204	1.63898437
-114.8833	49.75000	-121.617188	-2.33050537	-1.21617187
-119.6500	49.56667	-184.898438	-1.97451782	-1.84898437
-123.4667	48.53333	295.445312	-2.22024536	2.95445313
-122.7833	49.01667	115.554688	-1.27078247	1.15554687
-115.0667	51.08333	462.437500	-8.25271606	4.62437500
-123.3000	48.45000	363.054688	-0.87478638	3.63054687
-116.3333	51.45000	-319.679688	3.03353882	-3.19679687
-119.4333	49.03333	-281.875000	-2.99792480	-2.81875000
-121.7667	49.25000	252.234375	-3.39920044	2.52234375
-115.5500	51.20000	126.914062	-3.83938599	1.26914062
-125.9900	50.45000	240.398438	-3.37243652	2.40398437

Tabela 14 - Medidas de Pressão e Temperatura em 157 localizações (coordenadas) no Pacífico Noroeste Norte Americano

(continuação)				
Longitude	Latitude	Pressão	Temperatura	Pressão/100
-124.6700	51.67000	103.054688	-3.03326416	1.03054687
-125.7667	49.08333	359.054688	-2.10574341	3.59054687
-125.2667	49.95000	184.375000	-0.64474487	1.84375000
-123.8667	49.05000	188.203125	-1.71463013	1.88203125
-122.7833	49.01667	115.554688	-1.27078247	1.15554687
-115.0667	51.08333	462.437500	-8.25271606	4.62437500
-123.3000	48.45000	363.054688	-0.87478638	3.63054687
-116.3333	51.45000	-319.679688	3.03353882	-3.19679687
-119.4333	49.03333	-281.875000	-2.99792480	-2.81875000
-121.7667	49.25000	252.234375	-3.39920044	2.52234375
-115.5500	51.20000	126.914062	-3.83938599	1.26914062
-125.9900	50.45000	240.398438	-3.37243652	2.40398437
-124.6700	51.67000	103.054688	-3.03326416	1.03054687
-125.7667	49.08333	359.054688	-2.10574341	3.59054687
-125.2667	49.95000	184.375000	-0.64474487	1.84375000
-123.8667	49.05000	188.203125	-1.71463013	1.88203125
-117.6333	49.30000	-207.945312	-0.74282837	-2.07945313
-120.5167	49.46667	-10.742188	0.35998535	-0.10742188
-116.9833	51.30000	-588.914062	4.43466187	-5.88914062
-121.4833	49.36667	235.507812	-5.95294189	2.35507812
-120.4500	50.70000	35.906250	-2.15194702	0.35906250
-119.3833	49.96667	-104.031250	-1.64370728	-1.04031250
-124.9000	49.71667	189.414062	0.28295898	1.89414062
-123.1667	49.18333	158.531250	-0.90811157	1.58531250
-123.3333	48.41667	167.085938	-1.97875977	1.67085938
-115.7833	49.61667	-485.539062	9.21578979	-4.85539063
-122.3667	49.03333	126.773438	-2.46624756	1.26773438
-119.6000	49.46667	-172.015625	-3.59271240	-1.72015625
-123.4333	48.65000	189.335938	-1.24542236	1.89335937
-127.3667	50.68333	150.031250	-0.74151611	1.50031250
-124.4900	47.68000	243.304688	-0.46954346	2.43304688
-119.5510	46.68900	-52.867188	5.10739136	-0.52867188
-119.5780	46.65700	-4.929688	4.92465210	-0.04929687
-119.5210	46.55600	-83.718750	6.30679321	-0.83718750
-119.2860	46.36400	-6.648438	5.94964600	-0.06648437
-119.3600	46.43000	-72.320312	6.49157715	-0.72320312
-119.5990	46.56300	-61.992188	6.59786987	-0.61992187
-119.6330	46.54500	-107.093750	6.58248901	-1.07093750
-119.3010	46.30100	1.062500	5.74725342	0.01062500
-120.2180	45.74400	1.078125	1.98730469	0.01078125
-119.7260	46.57800	-65.312500	6.65890503	-0.65312500
-120.5644	41.49139	232.164062	1.15109253	2.32164063
-124.1058	40.97917	407.039062	-1.60119629	4.07039062
-118.2842	46.09500	77.648438	2.65695190	0.77648438
-123.8825	46.15694	148.687500	0.13101196	1.48687500
-122.3136	47.54583	50.273438	-0.72647095	0.50273437
-117.8086	44.84278	-53.687500	-0.56497192	-0.53687500
-122.5392	48.79944	83.398438	-1.34039307	0.83398438
-118.9564	43.59500	120.101562	3.05841064	1.20101562
-116.2406	43.56667	112.695312	-0.75271606	1.12695312
-124.2333	41.78333	343.601562	-1.10769653	3.43601563
-123.5053	48.12222	164.484375	-2.24981689	1.64484375
-117.8833	48.55000	-167.546875	0.35366821	-1.67546875
-121.8833	45.66667	18.218750	1.16357422	0.18218750
-117.4283	47.97417	-122.593750	2.43548584	-1.22593750
-121.1661	45.61944	-2.585938	1.20391846	-0.02585938

Tabela 14 - Medidas de Pressão e Temperatura em 157 localizações (coordenadas) no Pacífico Noroeste Norte Americano

(continuação)				
Longitude	Latitude	Pressão	Temperatura	Pressão/100
-120.2014	47.39778	-203.046875	1.78897095	-2.03046875
-115.7875	40.82639	338.648438	-3.83233643	3.38648438
-120.5292	47.03361	-110.140625	3.12380981	-1.10140625
-119.5147	47.30806	-91.507812	2.15734863	-0.91507812
-123.2144	44.13333	286.609375	-2.45333862	2.86609375
-123.0253	48.52028	173.585938	-0.05715942	1.73585937
-117.5278	47.62139	-52.773438	3.12014771	-0.52773437
-128.3000	50.00000	-357.742188	-0.27230835	-3.57742188
-122.9544	45.54806	138.328125	0.79730225	1.38328125
-123.9303	46.97278	83.125000	0.39553833	0.83125000
-119.2611	45.82583	63.945312	4.20974731	0.63945312
-128.1000	49.90000	246.273438	-2.17388916	2.46273438
-121.7242	42.14694	309.007812	1.68194580	3.09007812
-117.0144	46.37472	-12.359375	5.36535645	-0.12359375
-118.4247	45.51139	138.632812	3.22091675	1.38632812
-122.8722	42.38111	282.507812	-1.04565430	2.82507813
-122.3161	41.31583	112.015625	-2.59674072	1.12015625
-115.6697	47.45417	711.593750	2.63964844	7.11593750
-123.1322	45.19611	145.882812	0.89916992	1.45882812
-119.3133	47.19306	-97.257812	3.57839966	-0.97257812
-116.0994	44.88333	10.015625	-0.49929810	0.10015625
-122.6506	48.34917	169.937500	-0.60733032	1.69937500
-122.9033	46.97333	30.539062	-0.47146606	0.30539063
-119.5192	48.46139	-261.195312	-0.07791138	-2.61195312
-117.0097	44.01944	-207.664062	4.14086914	-2.07664063
-124.2500	43.41667	235.390625	-1.45800781	2.35390625
-115.5964	46.14417	-178.359375	-0.47375488	-1.78359375
-122.2825	47.92306	-210.007812	0.44400024	-2.10007813
-118.8342	45.69833	150.562500	4.86987305	1.50562500
-122.6003	45.59083	126.359375	1.59207153	1.26359375
-119.1167	46.26972	-34.085938	5.64929199	-0.34085938
-117.1136	46.74389	21.804688	2.71405029	0.21804688
-123.3561	43.23389	344.429688	-3.48632812	3.44429688
-121.1500	44.25417	239.195312	-0.89498901	2.39195313
-117.8644	42.59056	213.984375	0.16799927	2.13984375
-122.2128	47.49444	-15.046875	-1.04244995	-0.15046875
-122.3136	47.44472	19.328125	0.16333008	0.19328125
-117.3203	47.68556	-88.906250	2.18682861	-0.88906250
-123.1408	47.23806	91.632812	-0.63653564	0.91632813
-122.4667	41.78333	189.757812	-0.44482422	1.89757812
-117.6500	47.61667	-122.335938	2.87023926	-1.22335938
-122.9950	44.90778	243.851562	-0.14089966	2.43851562
-121.3372	47.27667	-34.070312	1.12136841	-0.34070312
-122.8622	45.76917	105.734375	1.53469849	1.05734375
-123.3642	42.60028	349.281250	0.87765503	3.49281250
-122.4833	47.15000	-1.343750	-1.29293823	-0.01343750
-122.8000	46.48333	48.046875	1.48397827	0.48046875
-122.5761	47.26750	24.359375	0.03085327	0.24359375
-122.4089	45.55111	146.406250	2.24365234	1.46406250
-122.7656	45.24889	135.679688	0.25024414	1.35679687
-124.5550	47.93750	232.578125	-0.91394043	2.32578125
-122.6564	45.62333	131.781250	2.15310669	1.31781250
-117.8072	40.90194	359.390625	-2.84185791	3.59390625

Tabela 14 - Medidas de Pressão e Temperatura em 157 localizações (coordenadas) no Pacífico Noroeste Norte Americano

Longitude	Latitude	Pressão	Temperatura	(conclusão)
				Pressão/100
-120.5336	46.56417	-163.960938	5.56695557	-1.63960938
-122.3550	47.63800	45.617188	-1.59701538	0.45617188
-124.0700	44.61000	247.835938	0.59069824	2.47835938
-122.8400	48.32000	184.500000	-0.02194214	1.84500000
-124.7400	48.39000	281.664062	-1.26123047	2.81664062
-122.3099	47.65350	76.828125	-1.21789551	0.76828125
-128.9000	42.20000	642.265625	-2.05545044	6.42265625
-129.5000	42.40000	464.757812	-3.96774292	4.64757812
-122.4350	47.65800	73.140625	-0.63900757	0.73140625