

Universidade Federal do Rio de Janeiro
Centro de Ciências Matemáticas e da Natureza
Instituto de Matemática
Departamento de Métodos Estatísticos

Modelos para detecção de observações extremas no Ceará

Autor: Ingrid Christyne Luquett de Oliveira
Orientador: Thaís Cristina O. Fonseca

PROJETO FINAL DE CURSO COMO PARTE
DOS REQUISITOS NECESSÁRIOS PARA
OBTENÇÃO DO TÍTULO DE ESTATÍSTICO.

Rio de Janeiro, 19 de Maio de 2013.

Resumo

Neste projeto busca - se estudar a ocorrência de eventos extremos em dados obtidos no espaço e no tempo tais como tornados, precipitações elevadas e temperaturas fora do usual. Analisaremos a temperatura do ar, a umidade relativa e a pressão atmosférica no Ceará, obtidas a cada 3 horas para estações ao longo do estado. Apresentamos neste projeto uma aplicação de modelos com caudas mais pesadas que as da distribuição Normal, os quais são capazes de acomodar observações não usuais.

A modelagem dos dados será baseada no artigo *Non-Gaussian Spatiotemporal Modelling through Scale Mixing* desenvolvido por Thais C. O. Fonseca e aceito para publicação em 2011, tratando - se de propostas recentes e com vários aspectos ainda em desenvolvimento. O modelo é obtido através de misturas na escala de processos gaussianos, o que permite acomodar observações aberrantes tais como observações atípicas pontuais e regiões com maior variabilidade no tempo e/ou no espaço. Esse modelo tem obtido um desempenho preditivo melhor que os usuais modelos gaussianos usados para modelagem de dados georreferenciados e pode ser de grande ganho no desempenho preditivo de fenômenos como precipitações elevadas e inesperadas.

Lista de Tabelas

3.1	Estações de monitoramento	12
5.1	Interpretação do Fator de Bayes	26
6.1	DIC	30
6.2	Fator de Bayes	30
6.3	Estimativas dos parâmetros	31
6.4	Comparação de modelos - DIC	31
6.5	Comparação de modelos - Fator de Bayes	31

Lista de Figuras

3.1	Estações de monitoramento do estado do Ceará	11
3.2	Médias das estações - Temperatura do ar	13
3.3	Médias das estações - Umidade Relativa do Ar	13
3.4	Médias das estações - Pressão Atmosférica	13
3.5	Variância das estações - Temperatura do ar	14
3.6	Variância das estações - Umidade Relativa do Ar	15
3.7	Variância das estações - Pressão Atmosférica	15
3.8	Relação entre covariáveis e temperatura do ar	16
3.9	Relação entre covariáveis e Umidade Relativa do Ar	16
3.10	Relação entre covariáveis e Pressão Atmosférica	16
3.11	Análise de Resíduos - Tempo versus Resíduos	17
6.1	Variabilidade por estação - Temperatura	28
6.2	Variabilidade por estação - Umidade relativa	29
6.3	Variabilidade por estação - Pressão atmosférica	29

Conteúdo

1	Introdução	5
1.1	O projeto e seus objetivos	5
1.2	Descrição do conjunto de dados	5
2	Metodologia	7
2.1	Estatística Espacial	7
2.2	Modelagem Não Gaussiana	8
2.2.1	Proposta do projeto	10
3	Análise exploratória dos dados	11
4	Inferência	18
4.1	Inferência Bayesiana	18
4.2	Métodos MCMC	19
4.3	Procedimento de Inferência	21
5	Comparação de modelos	24
5.1	Distribuição preditiva	24
5.1.1	Estimação pela Gama Translatada	25
5.1.2	Fator de Bayes	26
5.1.3	Deviance Information Criterion (DIC)	26
6	Resultados	28
7	Conclusões	33
A	Código R	37

Capítulo 1

Introdução

É frequente a ocorrência de observações extremas em dados espaço-temporais. Por exemplo, temperaturas mais altas que as usuais podem ocorrer ao longo de certa região. Neste contexto, modelos usuais como o gaussiano não seriam capazes de se adaptar a essas observações de forma que as previsões obtidas por modelos gaussianos podem ser muito afetadas por observações fora do usual. Neste projeto, modelos não gaussianos serão utilizados para ilustrar como a modelagem dos dados pode se apresentar mais realista, permitindo que melhores previsões sejam obtidas em aplicações meteorológicas para dados brasileiros.

1.1 O projeto e seus objetivos

O projeto tem por objetivo o estudo de eventos extremos em dados observados no espaço e no tempo tais como tornados, precipitações elevadas e temperaturas fora do usual. A presença de observações aberrantes é frequente em finanças e economia e, recentemente, grande atenção tem sido dada à modelagem robusta de dados na área de meteorologia, geologia, etc. Isso é devido, principalmente, ao impacto das grandes catástrofes decorrentes de fenômenos climáticos inesperados tais como terremotos e tempestades. Dessa forma, modelos capazes de detectar, modelar e prever situações como essas são de grande interesse para pesquisadores das áreas de meteorologia, estatística, entre outras.

1.2 Descrição do conjunto de dados

Nosso estudo será realizado com base no conjunto de dados contendo informações climáticas registradas nas estações de monitoramento localizadas no estado do Ceará.

Os dados climáticos fornecidos pelas estações monitoradoras são referentes à temperatura, umidade relativa do ar e pressão atmosférica (medida em Bar), registradas nas estações de monitoramento do Ceará. Os dados foram coletados a cada 3 horas, totalizando 84 observações em cada estação de monitoramento envolvida em nosso estudo.

A temperatura atmosférica é o grau de aquecimento do ar na atmosfera. Fatores geográficos exercem grande influência sobre a temperatura, bem como os movimentos da Terra.

Como consequência da rotação da Terra, a temperatura varia ao longo de um dia, atingindo menores temperaturas com o aumento da inclinação dos raios solares. Por outro lado, o movimento de translação da Terra permite que a temperatura varie com a latitude e ao longo do ano.

Podemos, ainda, estabelecer relações entre temperatura e altitude. À medida que a altitude aumenta observamos a diminuição da temperatura. Isto decorre do fato da quantidade de componentes atmosféricos grandes absorvedores de calor (vapor de água, dióxido de carbono, poeiras) ser menor em grandes altitudes. A latitude também influencia a temperatura uma vez que em maiores latitudes são registradas menores temperaturas. Em regiões mais afastadas da linha do Equador, onde estão as maiores latitudes, a incidência dos raios solares se dá de maneira mais inclinada, provocando menores temperaturas.

A umidade relativa é a relação entre a pressão de vapor do ar (medida em Pascal) e a pressão de vapor do ar obtida em condições de equilíbrio ou saturação sobre uma superfície de água na forma líquida ou sólida. Podemos pensar a umidade relativa do ar como a relação entre a quantidade de água existente no ar (umidade absoluta) e a quantidade máxima que poderia haver na mesma temperatura (ponto de saturação). Convencionalmente é denotada em porcentagem. Existe uma relação direta entre temperatura e umidade, visto que o ar cedo aquece e arrefece mais rapidamente que o ar úmido. Deste modo, lugares mais próximos do mar apresentam - se mais amenos.

A pressão atmosférica é a força exercida pela atmosfera em cada unidade de superfície. Por influência do aumento da altitude verificamos uma queda na pressão atmosférica, pois em altitudes maiores a coluna de ar que vai até a superfície da terra é menor. Além disso, a densidade do ar diminui, tornando o ar mais leve. A pressão atmosférica, por sua vez, varia em razão inversa da temperatura visto que quando o ar dilata torna - se mais leve. Por fim, relacionamos pressão atmosférica e latitude, visto que a pressão atmosférica dispõe - se por faixas, mais ou menos paralelas segundo a latitude e alternadamente de baixas e de altas pressões.

O estado do Ceará abrange um total de 47 estações de monitoramento. Porém somente serão utilizadas em nosso estudo 37 destas estações. A exclusão de algumas estações se deve ao fato dessas estações terem em seu conjunto de dados um número de observações inferior às demais. Em especial, a estação monitoradora São Gonçalo foi retirada do conjunto de dados em estudo por possuir um número expressivo de observações de valor zero, o que acredita -se ter sido ocasionado por erro de medição ou mesmo pela falta de registro. Tanto o fato de algumas estações de monitoramento terem dados faltantes quanto o da estação São Gonçalo possuir muitas observações de valor zero introduziriam erros na modelagem dos dados, razão pela qual optou - se pela exclusão das mesmas.

Capítulo 2

Metodologia

2.1 Estatística Espacial

Entende-se por Estatística Espacial a área da Estatística que abrange o estudo de eventos ao longo do espaço, isto é, consiste em um conjunto de métodos quantitativos em que a posição geográfica dos eventos é relevante para a análise do processo em questão. Tem-se interesse em analisar e modelar tais eventos a fim de compreender o comportamento dos processos espaciais, assim como fazer previsões acerca dos mesmos. Os dados espaciais podem ser de três tipos:

- **Dados pontuais:** processos onde a região de interesse $D \in \mathfrak{R}^k$ é aleatória, com o conjunto de índices informando as localizações de ocorrência dos eventos aleatórios. O processo espacial $Y(s)$ assume valor igual a 1 para todo $s \in D$ onde há ocorrência do evento ou fornece alguma informação adicional acerca de covariáveis que sejam importantes para a caracterização do processo. Neste tipo de dado, estamos interessados em estudar se o processo espacial tende a formar regiões no espaço onde haja aglomeração de ocorrências ou se ele se comporta homogeneamente ao longo de D .
- **Dados de área:** processos onde a região de interesse $D \in \mathfrak{R}^k$ é fixa, mas particionada em um número finito de áreas com fronteiras bem definidas, onde $Y(s)$ corresponde ao número de eventos ocorridos ao longo de uma determinada partição s de D . Com este tipo de observação é possível investigar a relação entre as diversas partições da região de interesse.
- **Dados geoestatísticos:** processos onde $Y(s)$ é um vetor aleatório de $s \in \mathfrak{R}^k$, com s variando continuamente em $D \in \mathfrak{R}^k$ fixado. Dados geoestatísticos consistem em observações obtidas a partir de pontos fixados em D .

Em nosso estudo, as observações são de natureza geoestatística, onde s_i representa a estação de monitoramento i localizada no Ceará.

2.2 Modelagem Não Gaussiana

Estamos interessados em modelar uma determinada região do espaço \mathfrak{R}^k ao longo do tempo, baseados em um subconjunto de pontos fixados da região de interesse $D \subset \mathfrak{R}^k$ e no tempo $T \subset \mathfrak{R}$.

Seja $Y(s,t)$ um vetor aleatório na localidade $s \in D$ e no instante de tempo $t \in T$, onde s varia continuamente em D , com D um subconjunto de \mathfrak{R}^k e t varia em $T \subset \mathfrak{R}$. Sejam $s_1, s_2, \dots, s_n \in D$ o conjunto de localidades onde será observada a variável aleatória de interesse e t_1, t_2, \dots, t_n o conjunto de instantes de tempo nos quais $Y(s, t)$ será observada. Temos, então, uma realização parcial de (D,T) .

O processo $Y(s,t)$ é caracterizado pelo modelo

$$Y(s, t) = \mu(s, t) + \varepsilon(s, t)$$

onde $\mu(s, t)$ representa a média do processo espaço-temporal e $\varepsilon(s, t)$ diz respeito à flutuação aleatória em microescala do processo.

Usualmente, modelos para esses dados são baseados em processos Gaussianos. Porém, para muitos conjuntos de dados, o uso de processos Gaussianos para modelagem não é adequado, pois as amostras podem assumir distribuições assimétricas ou mesmo apresentar caudas mais pesadas em relação à distribuição gaussiana.

Seja, agora, o processo espaço-temporal definido por

$$Z(s, t) = f(s, t)' \beta + \sigma \varepsilon(s, t) \quad (2.1)$$

onde $\mu(s, t) = f(s, t)' \beta$ é uma função linear de $f(s, t)' = (f_1(s, t), f_2(s, t), \dots, f_k(s, t))$ vetor k -dimensional de funções conhecidas das coordenadas espaço-temporais de $(s, t) = (s_1, s_2, t)$ com $\beta \in \mathfrak{R}^k$ desconhecido. Ainda, $\varepsilon(s, t)$ é um processo estacionário de segunda ordem com média zero, variância 1 e função de correlação dependendo somente das distâncias entre duas localidades s_i e $s_{i'}$ e entre dois instantes no tempo t_j e $t_{j'}$

$$\text{corr}[\varepsilon(s_i, t_j), \varepsilon(s_{i'}, t_{j'})] = C_\theta(\|s_i - s_{i'}\| \|t_j - t_{j'}\|) = C_1(\|s_i - s_{i'}\|) C_2(\|t_j - t_{j'}\|)$$

onde $C_1(d_s)$ e $C_2(d_t)$ são funções de correlação válidas da distância d_s entre s_i e $s_{i'}$ e d_t entre t_j e $t_{j'}$, respectivamente.

Uma realização desse processo em I localidades distintas em J instantes de tempo é denotada por $z_{i,j} = Z(s_i, t_j)$, com $i=1, 2, \dots, I$ e $j=1, 2, \dots, J$. Em muitas aplicações em Geoestatística, assume-se $\varepsilon(s, t)$ um processo Gaussiano, o que implica que $z \sim N_N(X\beta, \sigma^2 C_\theta)$, onde C_θ é uma matriz $IJ \times IJ$ com $C_1(\|s_i - s_{i'}\|) C_2(\|t_j - t_{j'}\|)$ seu elemento (i, j) e $X = (f(s_1, t_1), f(s_1, t_2), \dots, f(s_I, t_J))'$.

Em nosso estudo, observamos variâncias não constantes ao longo da região de interesse $D \subset \mathfrak{R}^k$, o que sugere que o processo espaço-temporal que estamos analisando não seria bem caracterizado pelo processo gaussiano $Z(s,t)$ descrito em (2.1). Uma maneira de acomodar e modelar de forma mais adequada nosso processo é adotar modelos que possuam caudas mais pesadas que as da distribuição gaussiana, o que permite acomodar heterogeneidade espacial em D , além de detectar observações muito discrepantes em relação ao conjunto de dados, ao que chamamos outlier.

Os modelos mencionados acima podem ser obtidos por meio de misturas do processo gaussiano. Uma das maneiras de transformar um processo gaussiano em um processo não gaussiano é proceder a mistura na escala do processo gaussiano. Como exemplo para um processo unidimensional, podemos tomar

$$\begin{aligned} z | \lambda &\sim N\left(\mu, \frac{\sigma^2}{\lambda}\right) \\ \lambda &\sim Ga\left(\frac{\eta}{2}, \frac{\eta}{2}\right) \end{aligned}$$

Assim, obtemos

$$z \sim t - Student(\eta),$$

que possui caudas mais pesadas que as da distribuição normal.

O fato de termos encontrado nos dados do Ceará regiões do espaço com maior variabilidade em relação às demais nos motiva ao uso de processos não gaussianos. Em nosso estudo, procederemos a modelagem por meio de processos GLG, Gaussian Log Gaussian, abordados em (5).

Este artigo propõe uma mistura na escala do processo Gaussiano $\varepsilon(s, t)$, introduzindo o processo $\lambda(s) \in \mathfrak{R}_+$, originando

$$z(s_i, t_j) = f(s_i, t_j)' \beta + \sigma \frac{\varepsilon(s_i, t_j)}{\sqrt{\lambda(s_i)}} \quad (2.2)$$

com $\lambda(s_i)$ a variável do processo $\lambda(s)$ correspondente à estação i e $\varepsilon = (\varepsilon(s_1, t_1), \dots, \varepsilon(s_I, t_J))' \sim N_n(0, C_\theta)$. Ainda, $\lambda(s_i)$ independente de $\varepsilon(s_i, t_j)$. Por simplicidade, denotaremos $z(s_i, t_j) = z_{i,j}$, $\varepsilon(s_i, t_j) = \varepsilon_{i,j}$ e $\lambda(s_i) = \lambda_i$.

Devemos ter os λ_i 's correlacionados a fim de garantir continuidade em média quadrática, ou seja, por definição $\lim_{\|x_i - x_j\| \rightarrow \infty} E[\lambda_i^{-1/2} \lambda_j^{-1/2}] = E[\lambda^{-1}]$. Utilizando os λ_i 's correlacionados garantimos que em localidades próximas os λ_i 's possuam valores próximos.

O modelo (2.2) é capaz de acomodar e identificar observações aberrantes que seriam consideradas outliers em processos Gaussianos. Como estamos interessados em saber se cada uma das observações é um outlier, introduzimos λ_i 's distintos para cada localidade.

Para o processo $\lambda(s)$, utilizaremos

$$\ln(\lambda(s)) = (\ln(\lambda_1), \ln(\lambda_2), \dots, \ln(\lambda_I))' \sim N_I \left(\frac{-v}{2} \mathbf{1}, vC_\theta \right)$$

onde $\mathbf{1}$ é um vetor $(1, 1, \dots, 1)'$ I-variado, os elementos de $\ln(\lambda(s))$ são correlacionados por C_θ , a mesma matriz de correlação de $\varepsilon(s, t)$, $v > 0$.

A escolha deste modelo $Z(s, t)$ com λ_i 's diferentes requer suavidade no processo λ , implicando que para localidades próximas os valores dos λ_i 's tendem a ser parecidos. O fato de termos λ_i pequeno implica que a variância nesta localidade é grande. Quando λ_i 's pequenos formam conglomerados, identificamos regiões de variância maior do que as outras regiões, o que caracteriza heterocedasticidade espacial.

2.2.1 Proposta do projeto

O modelo (2.2) assume que o processo $\lambda(s)$ possui mesma matriz de correlação que os dados obtidos nas estações de monitoramento. Essa suposição nos diz que $C_\theta(\|s_i - s_{i'}\|)$, função de correlação espacial, é a mesma função C_θ que correlaciona o processo $\ln(\lambda(s))$.

Esta suposição não parece plausível visto que não podemos garantir que o comportamento da correlação entre estações de monitoramento distintas se dá da mesma maneira que o processo $\lambda(s)$. Ao escolhermos uma função para caracterizar a relação espacial entre estações estamos especulando sobre como essa relação acontece na realidade. Supor que temos igualdade de funções de correlação para processos diferentes parece pouco realista. Por este motivo, avaliaremos o quão aceitável é essa suposição de igualdade.

Seja a função de correlação entre estações de monitoramento dada por

$$C_{\phi_1}(\|s_i - s_{i'}\|) = \exp\left(\frac{-d_s}{\phi_1}\right),$$

onde d_s é a distância entre estações.

Seja, ainda,

$$C_{\phi_3}(\|s_i - s_{i'}\|) = \exp\left(\frac{-d_s}{\phi_3}\right)$$

a função que correlaciona os parâmetros $\lambda_i, i = 1, \dots, I$.

Inicialmente analisaremos a diferença de qualidade no ajuste de dados espaciais feita pelo modelo gaussiano e pelo modelo não gaussiano com função de correlação iguais. Posteriormente, estudaremos empiricamente a importância do uso de funções de correlação distintas para caracterizar os dados e o processo $\lambda(s)$.

Capítulo 3

Análise exploratória dos dados

A análise exploratória dos dados consiste em um conjunto de técnicas gráficas e quantitativas cujo objetivo principal é extrair do conjunto de dados informações gerais, sintetizando os dados e ilustrando o comportamento das variáveis envolvidas no processo em estudo.

Primeiramente, observaremos como as estações de monitoramento se relacionam espacialmente, ou seja, como se dá a disposição das estações no estado do Ceará.

As figuras a seguir ilustram, respectivamente, a disposição das estações no mapa do Brasil e sua localização dentro do Ceará. Como pode ser visto na figura (3.1(a)), algumas estações de monitoramento localizam - se na região litorânea do Ceará, por exemplo, as estações Acarau e Caucaia. A região que abrange o centro do estado em direção ao litoral comporta a maior parte das estações de monitoramento e nessa região as estações se encontram mais próximas umas das outras. Em direção ao centro do país encontramos estações como Auiaba e Santa Quiteria. Na figura (3.1(b)) podemos visualizar de forma mais clara a disposição das estações ao longo do estado do Ceará. Todas as estações estão numeradas e relacionadas na tabela (3.1).

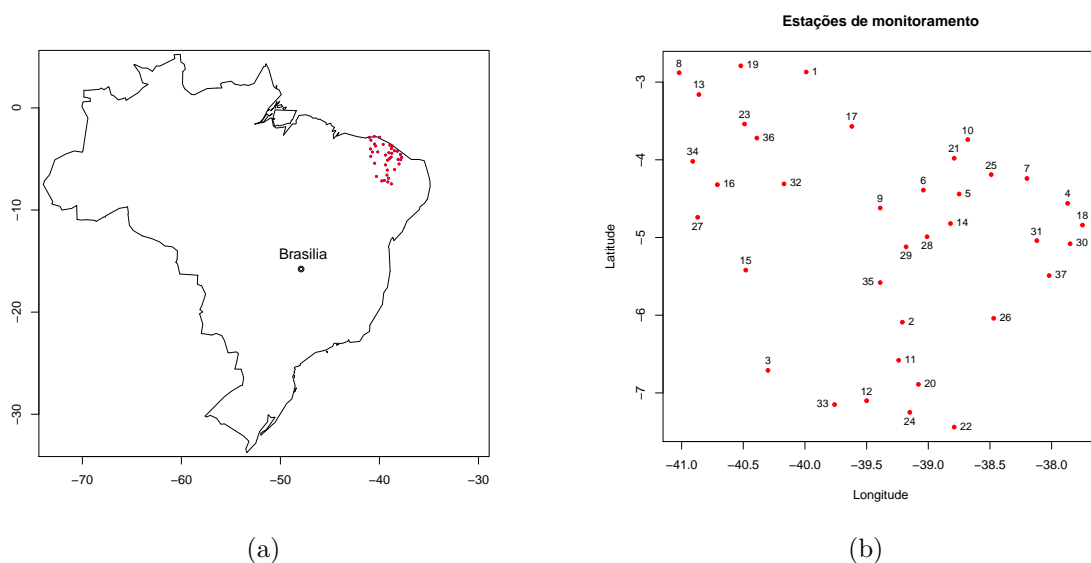


Figura 3.1: Estações de monitoramento do estado do Ceará

Numeração	Nome da estação	Numeração	Nome da estação
1	Acarau	20	Lavras
2	Acopiara	21	Maranguape
3	Aiuaba	22	Mauriti
4	Aracati	23	Meruoca
5	Aracoiaba	24	Missão
6	Aratuba	25	Pacajus
7	Beberibe	26	Pereiro
8	Camocim	27	Poranga
9	Caninde	28	Quixada
10	Caucaia	29	Quixeramobim
11	Cedro	30	Quixere
12	Crato	31	Russas
13	Granja	32	Santana
14	Ibaretama	33	Santa Quiteria
15	Independência	34	São Benedito
16	Ipu	35	Senador Pompeu
17	Itapipoca	36	Sobral
18	Jaguaruana	37	Tabuleiro
19	Jipoca		

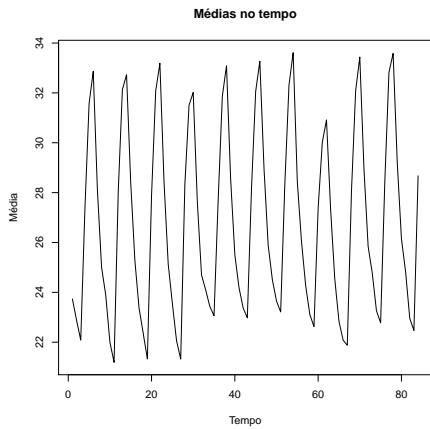
Tabela 3.1: Estações de monitoramento

Conhecida a disposição espacial das estações de monitoramento no Ceará, passamos a olhar para o comportamento das variáveis temperatura, umidade e pressão barométrica em relação à média e à variância, tanto espacialmente quanto temporalmente.

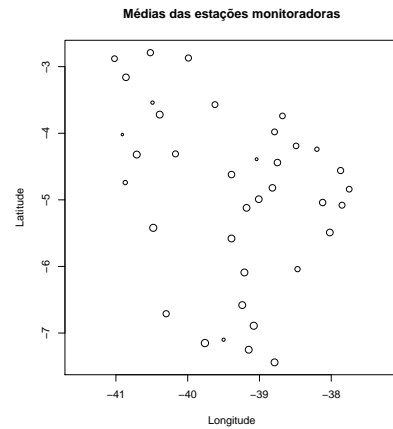
Voltando nossas atenções para a média, podemos observar ciclos temporais para os processos de temperatura e umidade, os quais se devem à forma como os dados foram coletados. Como estamos lidando com observações a cada 3 horas é intuitivo pensar em variações de temperatura ao longo do dia e em ciclos diários. Sendo a umidade diretamente relacionada à temperatura, espera-se que o comportamento da média ao longo do tempo se assemelhe ao encontrado para a temperatura. Para a pressão atmosférica observamos um padrão cíclico, ligeiramente decrescente, para a média ao longo do tempo de observação. Não há uma razão específica para o decréscimo desta média ao longo do tempo e, por esta razão, não levaremos em consideração essa característica temporal.

Espacialmente, parece haver discrepâncias entre as médias das estações para os três processos em estudo, o que nos motiva à busca por relações entre esses processos e as variáveis espaciais envolvidas em eventos climáticos, tais como latitude, longitude e altitude. O fato de lidarmos com uma porção relativamente pequena do espaço nos leva a acreditar que a latitude e a longitude não exerçam grande influência sobre as variáveis em estudo, porém é sabido que a altitude é importante na explicação de todas as variáveis (temperatura, pressão atmosférica e umidade relativa). Mais adiante, aprofundaremos o estudo dessas relações.

As figuras 3.2(a), 3.3(a) e 3.4(a) mostram o comportamento da média ao longo do tempo e as figuras 3.2(b), 3.3(b) e 3.4(b) no espaço. Os pontos são proporcionais às médias das estações.

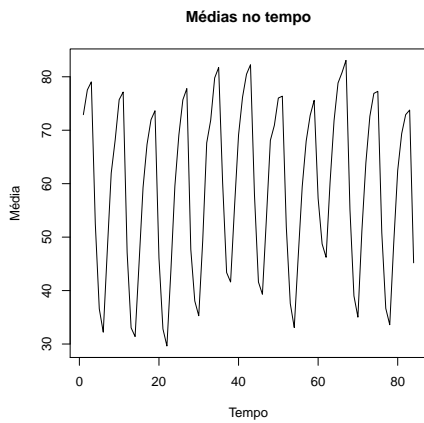


(a) Média no tempo - Temperatura

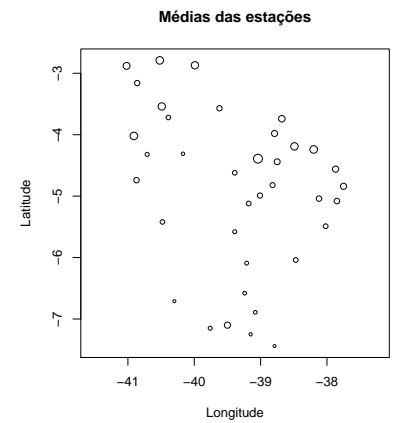


(b) Médias espaciais - Temperatura

Figura 3.2: Médias das estações - Temperatura do ar

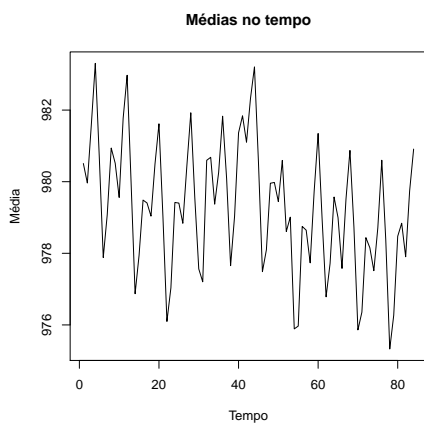


(a) Média no tempo - Umidade Relativa

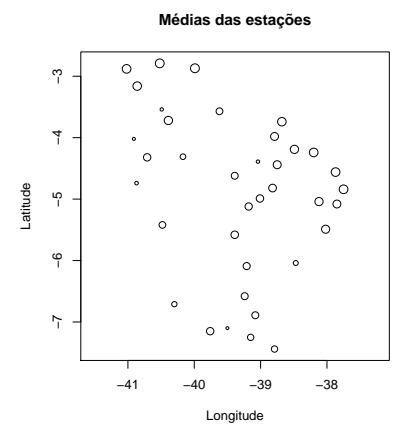


(b) Médias espaciais - Umidade Relativa

Figura 3.3: Médias das estações - Umidade Relativa do Ar



(a) Média no tempo - Pressão



(b) Médias espaciais - Pressão

Figura 3.4: Médias das estações - Pressão Atmosférica

A mesma análise feita anteriormente para a média foi realizada para a variância das observações, temporal e espacialmente. A seguir temos ilustrado o comportamento da variância das observações ao longo do tempo (figuras 3.5(a), 3.6(a) e 3.7(a)) e ao longo do espaço (figuras 3.5(b), 3.6(b) e 3.7(b)). Como ressaltado acima, a dimensão do ponto é proporcional à variância da estação de monitoramento em questão.

Observando a variância ao longo do tempo, não identificamos padrões explícitos em nenhum dos três processos. As variâncias encontram - se em geral na mesma faixa de valores com picos isolados, não nos remetendo a nenhum padrão claro.

Espacialmente, observamos diferença entre a variância da temperatura nas estações de monitoramento do litoral e as mais ao centro do estado do Ceará. As estações litorâneas apresentam variâncias menores em relação às demais estações. Como visto para a temperatura, a variância das estações em relação à umidade relativa não apresenta nenhum comportamento padrão, porém podemos notar discrepâncias entre estações. Para a pressão atmosférica encontramos um ponto de grande variância, porém não há justificativa para a ocorrência desse pico. Ainda em relação à pressão atmosférica, observamos que a estação localizada em Jaguaruana apresenta variância maior do que as demais. Ainda que não exista nenhuma razão evidente para esta variância ser tão discrepante, percebemos que um modelo gaussiano não acomodaria esta diferença de forma satisfatória.

Apesar de não termos encontrado nenhum padrão explícito para o comportamento da variância ao longo do espaço para nenhum dos três processos, podemos notar que existem discrepâncias relevantes em relação à variância das estações, justificando nossa escolha pelo uso de modelos não gaussianos.

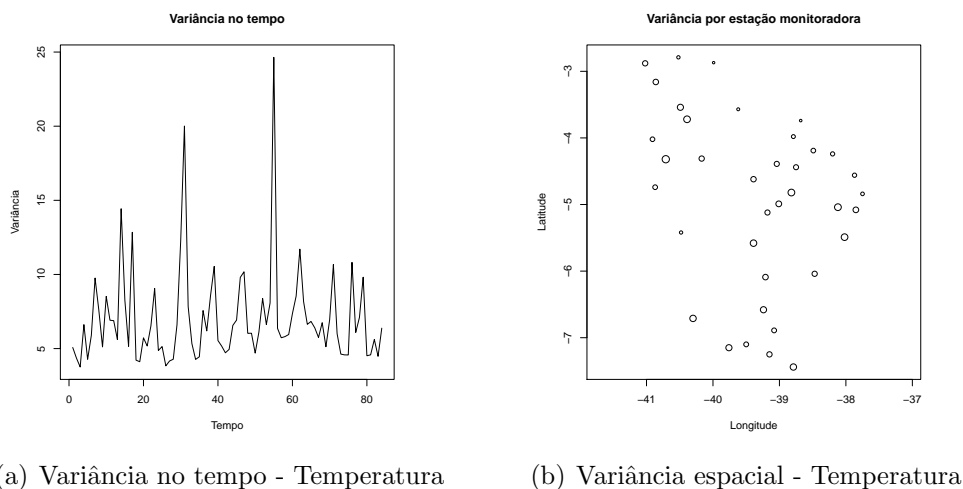
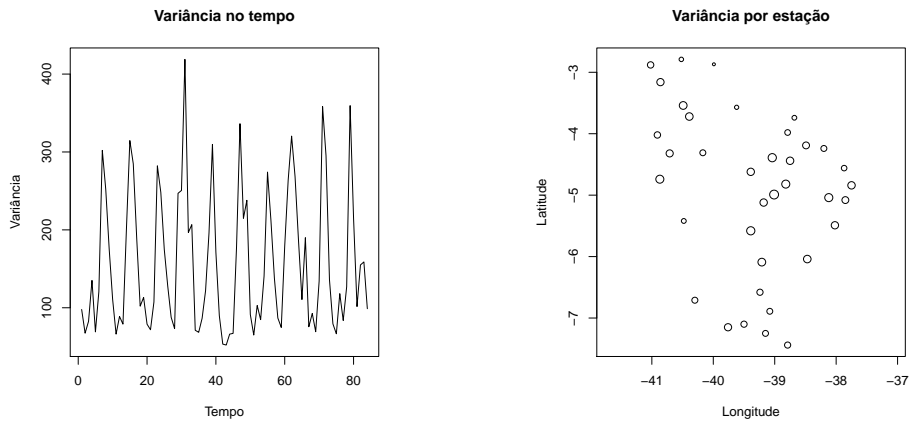
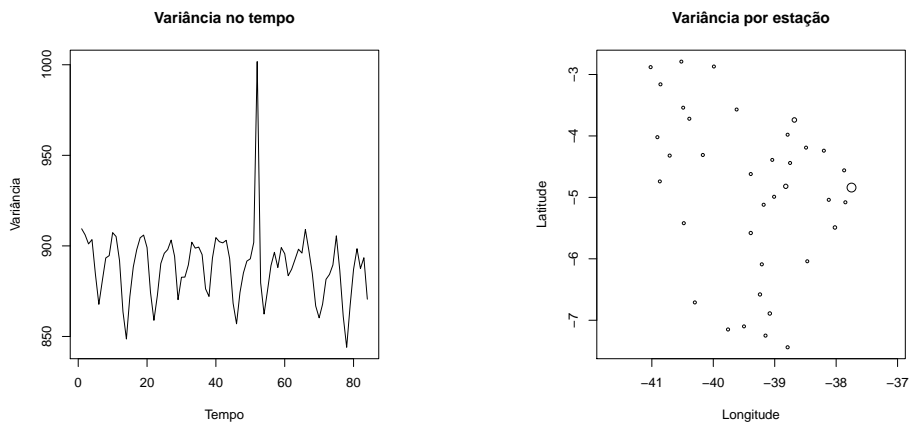


Figura 3.5: Variância das estações - Temperatura do ar



(a) Variância no tempo - Umidade Relativa (b) Variância espacial - Umidade Relativa

Figura 3.6: Variância das estações - Umidade Relativa do Ar

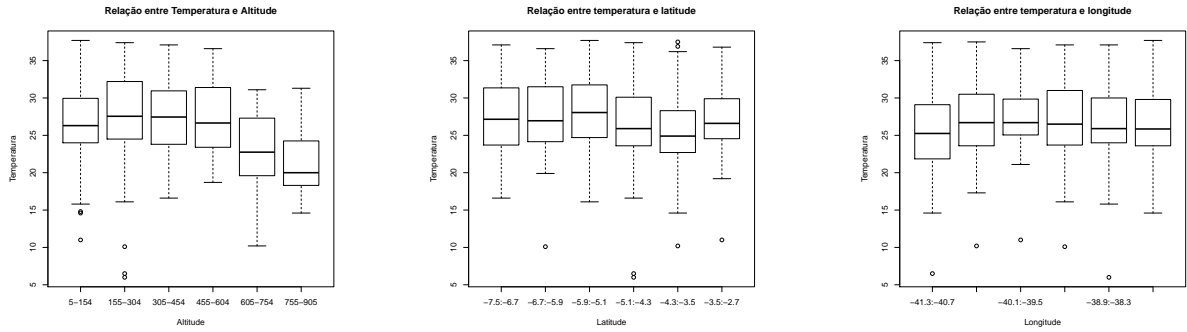


(a) Variância no tempo - Pressão (b) Variância espacial - Pressão

Figura 3.7: Variância das estações - Pressão Atmosférica

No modelo (2.2) temos que $\mu(s, t) = f(s, t)' \beta$ é a média do processo, sendo $f(s, t)'$ um vetor k - dimensional de funções conhecidas das coordenadas espaço - temporais de (s, t) , com β em R^k desconhecido. Para modelar essa média para cada um dos processos (temperatura, umidade relativa e pressão atmosférica) utilizaremos como técnica exploratória a regressão linear múltipla.

As variáveis inseridas inicialmente no modelo como covariáveis são latitude, longitude, altitude e tempo por motivos já explicitados. Motivando nossa escolha por essas covariáveis, os boxplots a seguir ilustram as relações das observações de cada processo com a latitude, longitude e altitude.

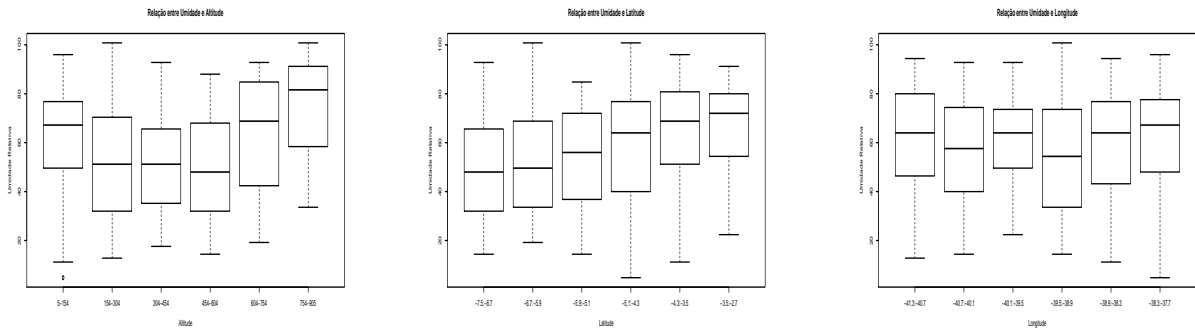


(a) Altitude versus Temperatura do ar

(b) Latitude versus Temperatura do ar

(c) Longitude versus Temperatura do ar

Figura 3.8: Relação entre covariáveis e temperatura do ar

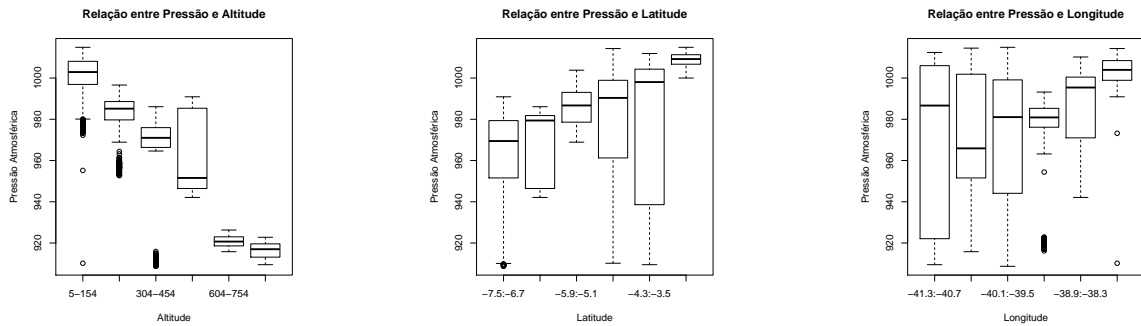


(a) Altitude versus Umidade Relativa

(b) Latitude versus Umidade Relativa

(c) Longitude versus Umidade Relativa

Figura 3.9: Relação entre covariáveis e Umidade Relativa do Ar



(a) Altitude versus Pressão

(b) Latitude versus Pressão

(c) Longitude versus Pressão

Figura 3.10: Relação entre covariáveis e Pressão Atmosférica

A figura (3.8) nos confirma a existência de uma relação linear entre temperatura e altitude. Como esperado, a temperatura diminui com o aumento da altitude. Em contrapartida, não há uma relação bem estabelecida da temperatura com a longitude e a latitude.

O efeito das covariáveis sobre a umidade relativa está ilustrada na figura (3.9). Analisando os boxplots, observamos uma relação linear crescente com a latitude e nenhuma relação com a longitude. A relação entre umidade relativa e altitude não nos remete a nenhum padrão e,

ao após o ajuste de um modelo de regressão com essa covariável, encontramos um coeficiente muito pequeno indicando que não há influência da altitude na umidade.

Para a pressão atmosférica, na figura (3.10) observamos uma relação linear decrescente com a altitude, comportamente este já esperado. O aumento da latitude e da longitude provoca o aumento da pressão atmosférica, também como explicado anteriormente. Ainda que os boxplots ilustrem uma relação entre latitude e longitude com a pressão atmosférica, a regressão linear múltipla resultou em coeficientes não significativos a um nível de 5% de confiança. Deste modo, somente usaremos a altitude para explicar a média do processo de pressão atmosférica em estudo.

Introduzimos, também, o tempo como covariável para melhor compreender as relações espaço - temporais com o objetivo de modelar tanto a temperatura quanto a umidade relativa e a pressão atmosférica no domínio espacial de interesse.

Ajustamos, então, um modelo de regressão múltipla com as covariáveis espaciais citadas acima para os processos de temperatura do ar, umidade relativa e pressão atmosférica no estado do Ceará. O resultado obtido mostra serem as três covariáveis significativas no entendimento do comportamento da temperatura e umidade relativa, enquanto para pressão utilizaremos somente a altitude.

A análise de resíduos da regressão revela um comportamento cíclico nos resíduos em relação à variável tempo em todos os processos. Tal padrão cíclico encontrado nos remete ao uso da variável tempo transformada. Como foram observados ciclos, somos motivados a aplicar uma transformação harmônica. Com o uso de funções seno e cosseno traduzimos a influência do tempo de uma forma mais correta, eliminando o comportamento cíclico dos resíduos no tempo. A figura (3.11) ilustra o comportamento observado nos resíduos em relação ao tempo.

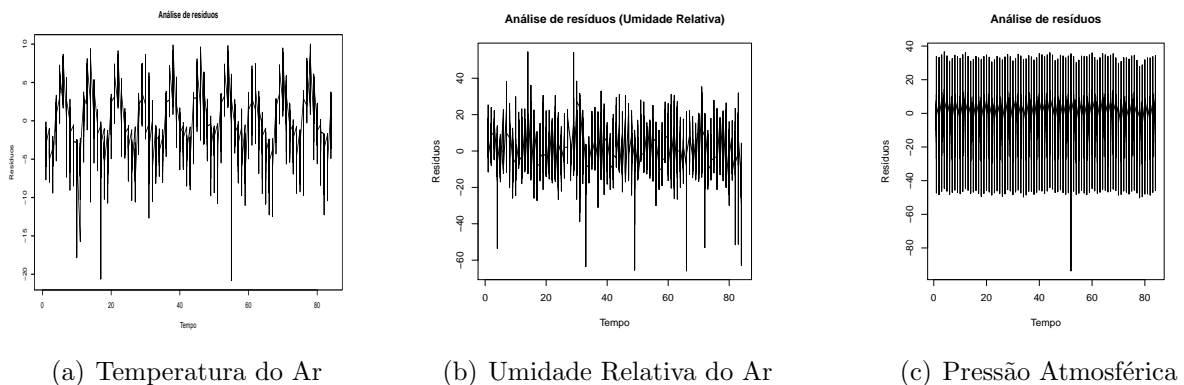


Figura 3.11: Análise de Resíduos - Tempo versus Resíduos

Concluída a etapa de conhecimento do nosso conjunto de observações e modelagem da média do processo em questão, passaremos à inferência acerca dos parâmetros envolvidos em nossa modelagem não gaussiana.

Capítulo 4

Inferência

4.1 Inferência Bayesiana

Quando tentamos explicar a ocorrência de determinados fenômenos através de modelos probabilísticos nos deparamos com a incerteza acerca dos parâmetros desses modelos. O verdadeiro valor do vetor de parâmetros $\underline{\theta}$ será sempre desconhecido, mas podemos inferir sobre ele.

Em Inferência Bayesiana atribuímos um modelo probabilístico para $\underline{\theta}$ a fim de caracterizar nossa incerteza sobre ele. A esse modelo chamamos distribuição à priori. Cada pesquisador pode atribuir uma priori à $\underline{\theta}$ com base em informações subjetivas que possua. Em casos onde pouco se sabe sobre o processo que gostaríamos de explicar, opta-se por adotar uma priori vaga para $\underline{\theta}$, ou seja, que possua distribuição de probabilidade com variância grande, deixando que a verossimilhança tenha mais influência do que a priori na obtenção da posteriori.

Podemos aumentar nosso conhecimento sobre $\underline{\theta}$ obtendo uma coleção de observações do fenômeno relacionado a $\underline{\theta}$. A distribuição amostral $p(x | \underline{\theta})$ define essa relação e através do teorema de Bayes quantificamos esse aumento de informação.

Seja X a quantidade aleatória relacionada a $\underline{\theta}$ da qual obteremos uma amostra x . O teorema de Bayes define a distribuição a posteriori de $\underline{\theta}$ como

$$p(\underline{\theta} | x) = \frac{p(\underline{\theta}, x)}{p(x)} = \frac{p(x | \underline{\theta})p(\underline{\theta})}{p(x)}.$$

Após observarmos $X=x$, a função $p(x | \underline{\theta})$ nos fornece a verossimilhança de cada um dos possíveis valores para $\underline{\theta}$. Como $p(x)$ não depende de $\underline{\theta}$, uma maneira útil de aplicar o teorema de Bayes é através da proporcionalidade

$$p(\underline{\theta} | x) \propto p(x | \underline{\theta})p(\underline{\theta}).$$

A escolha da priori não é restritiva, ou seja, podemos optar pela priori que subjetivamente acreditamos caracterizar melhor $\underline{\theta}$. Em alguns casos, escolhemos a distribuição a priori para $\underline{\theta}$ de maneira a encontrar uma forma conhecida para a posteriori.

Existe uma classe de funções de distribuição de probabilidade que conjugam com a verossimilhança e obtemos facilmente a posteriori. Como exemplo considere uma amostra aleatória X_1, X_2, \dots, X_n de variáveis independentes e identicamente distribuídas segundo uma distribuição $Poisson(\lambda)$. A verossimilhança $p(x | \lambda)$ é dada por

$$p(x | \lambda) = \prod_{i=1}^n f_{X_i}(x_i) = \prod_{i=1}^n \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}.$$

Ao escolhermos a distribuição $Gama(\alpha, \beta)$ como priori para λ , obtemos como posteriori uma distribuição $Gama(\sum_{i=1}^n x_i + \alpha, n + \beta)$. De fato,

$$p(\lambda) \propto \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$p(x | \lambda) \propto \lambda^{(\sum_{i=1}^n x_i)} e^{-n\lambda}$$

E, portanto,

$$p(\lambda | x) \propto \lambda^{(\sum_{i=1}^n x_i) + \alpha} e^{-(\beta+n)\lambda},$$

que é o núcleo de uma distribuição Gama com parâmetro de forma igual a $\sum_{i=1}^n x_i + \alpha$ e parâmetro de escala $n + \beta$.

Em alguns casos não conseguimos escolher prioris conjugadas, tornando difícil a obtenção de uma forma analítica fechada para a distribuição a posteriori. Nesse contexto necessitamos de métodos aproximados para estimar a distribuição a posteriori. Os métodos aproximados mais conhecidos em inferência bayesiana são os métodos de Monte Carlo via cadeias de Markov (MCMC), os quais serão explicados na próxima seção.

4.2 Métodos MCMC

Uma cadeia de Markov (X_n) é uma sequência de variáveis aleatórias $X_0, X_1, \dots, X_n, \dots$ tal que

$$P(X_{k+1} \in A | X_0 = x_0, X_1 = x_1, \dots, X_k = x_k) = P(X_{k+1} \in A | X_k = x_k),$$

para qualquer k . Cadeias de Markov que possuem distribuição invariante são tais que se $X_n \sim \pi$ então $X_{n+1} \sim \pi$.

Métodos de Monte Carlo via cadeias de Markov são classes de algoritmos para amostragem de distribuições de probabilidade baseados na construção de uma cadeia de Markov que tem como distribuição invariante a distribuição da qual desejamos amostrar. Em especial, desejamos amostrar da distribuição a posteriori do processo.

Os métodos MCMC mais utilizados são o Amostrador de Gibbs e o algoritmo Metropolis-Hastings.

- **Amostrador de Gibbs**

Desejamos amostrar da distribuição a posteriori $p(\underline{\theta} | z)$. Baseados na distribuição a posteriori podemos encontrar as distribuições condicionais de $\theta_i | \underline{\theta}_{-i}$, as quais chamamos condicionais completas, onde $\underline{\theta}_{-i}$ corresponde ao vetor de parâmetros $\underline{\theta}$ a menos do parâmetro θ_i .

Quando as distribuições condicionais completas são distribuições de probabilidade conhecidas usamos o amostrador de Gibbs para realizar a amostragem da distribuição à posteriori.

Algoritmo:

1. Seja $\underline{\theta}^{(0)}$ o valor inicial para $\underline{\theta} = (\theta_1, \dots, \theta_p)$;
2. Faça $i=1$ e $j=1$;
3. Simule $\theta_i^{(j)} \sim p(\cdot | \underline{\theta}_{-i}^{(j)}, z)$;
4. Se $i < p$ faça $i=i+1$ e volte ao passo 3;
5. Faça $j=j+1$;

- **Metropolis - Hastings**

Quando não conhecemos a distribuição condicional completa de $\theta_i | \underline{\theta}_{-i}$ utiliza-se o algoritmo de Metropolis - Hastings.

Considere uma densidade condicional $q(y | x)$ e uma densidade objetivo $p(x)$ (no nosso caso $p(\underline{\theta} | z)$). O algoritmo de Metropolis - Hastings produz uma cadeia de Markov (X_n) através do seguinte algoritmo:

1. Seja x_0 um valor inicial para x ;
2. Gere X^{prop} com distribuição $q(\cdot | x_k)$;
3. Tome

$$x_{k+1} = \begin{cases} x^{prop}, & \text{com probabilidade } \alpha \\ x_k, & \text{com probabilidade } 1 - \alpha, \end{cases}$$

$$\text{onde } \alpha = \min \left\{ 1, \frac{p(x^{prop})q(x_k|x^{prop})}{p(x_k)q(x^{prop}|x_k)} \right\}.$$

A densidade q é conhecida como densidade proposta e α como probabilidade de aceitação. A escolha da função de transição proposta é crucial para a implementação do algoritmo de Metropolis - Hastings.

Uma das propostas de transição mais utilizadas é o Passeio Aleatório, que consiste em usar o estado x onde a cadeia se encontra para propor o movimento para outro estado. Com isso, exploramos a vizinhança em torno do valor atual da cadeia. Podemos considerar o movimento da cadeia dependendo de $q(\cdot | x_k)$, por exemplo, como

$$X^{prop} = X_k + \epsilon,$$

onde ϵ é uma perturbação aleatória simétrica em torno de 0. Deste modo, temos que $q(x^{prop} | x_k) = q(x_k | x^{prop})$ e a probabilidade de aceitação torna - se apenas $\alpha = \min \left\{ 1, \frac{p(x^{prop})}{p(x_k)} \right\}$.

É comum, ao utilizarmos a proposta Passeio Aleatório, adotar a distribuição Normal para a perturbação aleatória ϵ . Neste caso, é preciso escolher a variância da distribuição com cautela. Ao optarmos por uma variância pequena aceitaremos o valor proposto com probabilidade alta, permitindo maior movimentação da cadeia, porém esses movimentos são lentos no domínio dos parâmetros, o que significa que demoraremos para percorrer todo o espaço dos parâmetros. Em contrapartida, se escolhermos uma variância grande, aceitaremos menos mudanças de estado, mas visitaremos mais rapidamente o domínio de interesse. Na prática, um bom percentual de aceitação α está em torno de 30%.

Métodos MCMC são, portanto, uma ferramenta de grande importância para amostragem de distribuições de probabilidade complexas, permitindo a inferência acerca dos parâmetros.

4.3 Procedimento de Inferência

Seja $Z(s,t)$ o processo definido por (2.2) e $\underline{\theta} = (\beta, \sigma^2, \phi_1, \phi_2, \lambda, \nu)$ o vetor de parâmetros do modelo. Temos que

$$Z | \underline{\theta} \sim N_N(X\beta, \sigma^2 \Sigma_{\phi_1}),$$

onde $N = IJ$ e $\Sigma = \frac{C_{\phi_1}(d_s)C_{\phi_2}(d_t)}{\sqrt{\lambda_i \lambda_{i'}}}$ com $C_{\phi_1}(d_s) = \exp\left(\frac{-d_s}{\phi_1}\right)$ e $C_{\phi_2}(d_t) = \exp\left(\frac{-d_t}{\phi_2}\right)$.

As distribuições a priori adotadas para os parâmetros foram

$$\beta \sim N_p(0, \tau_0 I_p)$$

$$\sigma^2 \sim GI\left(\frac{a}{2}, \frac{b}{2}\right)$$

$$\phi_1 \sim Gama(c, d)$$

$$\phi_2 \sim Gama(e, f)$$

$$\nu \sim Gama(g, h)$$

$$\lambda = (\lambda_1, \dots, \lambda_I) \sim \text{lognormal}\left(\frac{-\nu}{2} \mathbf{1}, \nu C_1\right)$$

A amostragem à posteriori foi feita com base em métodos MCMC, mesclando passos de Metropolis - Hastings com passos de Amostrador de Gibbs.

Para amostrar dos parâmetros β e σ^2 foi utilizado o algoritmo de Gibbs, com as seguintes condicionais completas:

$$\beta \mid \underline{\theta}_{-\beta} \sim N_p(\mu_\beta, \Sigma_\beta)$$

$$\sigma^2 \mid \underline{\theta}_{-\sigma^2} \sim GI\left(\frac{N+a}{2}, \frac{(Y-X\beta)'\Sigma^{-1}(Y-X\beta)+b}{2}\right)$$

Enquanto para os parâmetros ϕ_1, ϕ_2, ν e λ utilizamos passos de Metropolis - Hastings para amostrar das condicionais completas $p(\phi_1 \mid \underline{\theta}_{-\phi_1}, z)$, $p(\phi_2 \mid \underline{\theta}_{-\phi_2}, z)$, $p(\nu \mid \underline{\theta}_{-\nu}, z)$ e $p(\lambda \mid \underline{\theta}_{-\lambda}, z)$. O passeio aleatório foi usado como proposta para amostrar desses quatro parâmetros. Assim, as funções de distribuição de transição propostas são:

$$\phi_1^{prop} \sim \log - Normal(\log(\phi_1^k), \gamma_1)$$

$$\phi_2^{prop} \sim \log - Normal(\log(\phi_2^k), \gamma_2)$$

$$\nu^{prop} \sim \log - Normal(\log(\nu^k), \gamma_3)$$

A amostragem de $p(\lambda \mid \underline{\theta}_{-\lambda}, z)$ não é trivial, pois os elementos de $\lambda(s)$ não são condicionalmente independentes dados os outros parâmetros e os dados observados.

Para facilitar a amostragem dividimos os elementos de $\lambda(s)$ em blocos, onde as observações de cada bloco são similares. Assim, grande parte da dependência entre os λ_i^k s ficará retida no mesmo bloco.

Seja $\lambda_{(i)}$ o bloco composto por n_i elementos de $\lambda(s)$ e $\lambda_{(-i)}$ os elementos restantes. Temos então:

$$i = \begin{pmatrix} \lambda_{(-i)} \\ \lambda_{(i)} \end{pmatrix} e C_\theta = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$$

Como proposta para a etapa de Metropolis - Hastings usamos, também, o passeio aleatório, com

$$\lambda^{prop} \sim \log - Normal(\log(\lambda^k), \gamma_4).$$

Para o modelo não - gaussiano com funções de correlação distintas temos para o modelo (2.2) o vetor de parâmetros $\underline{\theta} = (\beta, \sigma^2, \phi_1, \phi_2, \phi_3, \lambda, \nu)$. Deste modo,

$$Z \mid \underline{\theta} \sim N_N(X\beta, \sigma^2 \Sigma_{\phi_1}),$$

onde $N = IJ$ e $\Sigma_{\phi_1} = \frac{C_{\phi_1}(d_s)C_{\phi_2}(d_t)}{\sqrt{\lambda_i \lambda_{i'}}$ com $C_{\phi_1}(d_s) = \exp\left(\frac{-d_s}{\phi_1}\right)$ e $C_{\phi_2}(d_t) = \exp\left(\frac{-d_t}{\phi_2}\right)$.

As distribuições a priori adotadas para o novo vetor de parâmetros são as mesmas adotadas anteriormente, a menos da priori para λ que é dada por

$$\lambda = (\lambda_1, \dots, \lambda_I) \sim \text{lognormal} \left(\frac{-\nu}{2} \mathbf{1}, \nu C_{\phi_3} \right),$$

onde $C_{\phi_3}(d_s) = \exp\left(\frac{-d_s}{\phi_3}\right)$.

Além disso, adotaremos

$$\phi_3 \sim \text{Gama}(g, h)$$

como priori para o novo parâmetro ϕ_3 . Ainda, a proposta utilizada para a amostragem via Metropolis - Hastings do parâmetro ϕ_3 será

$$\phi_3^{prop} \sim \text{log - Normal} \left(\log(\phi_3^k), \gamma_5 \right).$$

Capítulo 5

Comparação de modelos

Na modelagem de observações reais não podemos afirmar, de forma absoluta, que encontramos o modelo verdadeiro. Deste modo, devemos escolher entre os modelos ajustados aquele que melhor se adequa ao nosso conjunto de dados, tendo como referência o nosso objetivo principal que pode ser, por exemplo, a qualidade de ajuste ou a qualidade de previsão. Este tópico é de extrema importância sob a ótica estatística. Buscamos modelos que expliquem bem o comportamento da nossa variável de interesse, mas que possuam o menor número de parâmetros possível.

Existem diferentes maneiras de comparar os modelos estimados. Nas seções abaixo elucidamos dois critérios de comparação de modelos, o DIC (Deviance Information Criterion) e o Fator de Bayes. Utilizaremos ambos os critérios para escolher entre os modelos gaussiano, não gaussiano com funções de correlação similares e não gaussiano com funções de correlação distintas.

5.1 Distribuição preditiva

Uma quantidade importante na comparação de modelos via métodos Bayesianos é a chamada verossimilhança marginal ou constante de normalização. O fator de Bayes, por exemplo, consiste na razão entre essas quantidades para os dois modelos a serem comparados.

Sejam z o conjunto de observações, $p(z | \underline{\theta})$ a função de verossimilhança dos dados indexada pelo vetor de parâmetros $\underline{\theta}$ e distribuição a priori $p(\underline{\theta})$. A verossimilhança marginal é dada por

$$p(z) = \int p(z | \underline{\theta})p(\underline{\theta})d\underline{\theta}.$$

A fórmula acima decorre de

$$p(\underline{\theta} | z) = \int \frac{p(z | \underline{\theta})p(\underline{\theta})d\underline{\theta}}{p(z)}.$$

Como $p(z)$ é função do conjunto de dados, podemos defini-la como densidade preditiva.

A resolução analítica da integral de $p(z)$ pode ser bastante complicada. Pelo teorema de Bayes,

$$\frac{1}{p(z)} = \int \frac{p(\underline{\theta} | z)d\underline{\theta}}{p(z | \underline{\theta})} = E \left(\frac{1}{p(z | \underline{\theta})} | z \right).$$

Chamamos a essa esperança de média harmônica a posteriori da verossimilhança e podemos aproximá-la via integração de Monte Carlo utilizando a média harmônica amostral da verossimilhança

$$\hat{p}(z) = \left[\frac{1}{M} \sum_{k=1}^M \frac{1}{p(z | \underline{\theta}^k)} \right]^{-1}$$

baseada em uma amostra de tamanho M da posteriori de $\underline{\theta}$, ou seja, $\underline{\theta}^1, \dots, \underline{\theta}^M$.

Apesar da simplicidade do estimador de média harmônica definido acima, ele se mostra computacionalmente instável. O estimador é consistente, porém pode possuir variância infinita sobre as simulações. Uma das abordagens para solucionar este problema se encontra em (6) e baseia-se no fato de a distribuição a posteriori do log da verossimilhança seguir uma distribuição Gama translada aproximadamente.

5.1.1 Estimação pela Gama Translatada

Suponha que o resultado obtido a posteriori por meio de simulações de Monte Carlo são tais que a sequência das logverossimilhanças $l_t, t = 1, \dots, M$ é uma sequência de termos independentes.

Assintoticamente com a dimensão da matriz de observações temos que

$$l_{max} - l_t \sim Gama(\alpha, 1),$$

onde l_{max} é o máximo atingido pela logverossimilhança e $\alpha = d/2$ com d a dimensão de $\underline{\theta}$.

Poderíamos utilizar diretamente o resultado assintótico acima para estimar a média harmônica a posterior, porém nos deparamos com três problemas principais. Em geral não conhecemos l_{max} a posteriori porque normalmente esse máximo não é atingido. Ainda, existem modelos para os quais não sabemos o efetivo número de parâmetros. Além disso, usando a aproximação da logverossimilhança a posteriori pela Gama deslocada obtemos média harmônica a posteriori e distribuição preditiva infinitas.

Assumindo que a logverossimilhança a posteriori assume distribuição Gama translada, $E(l_{max} - l_t) = \alpha$ e $Var(l_t) = \alpha$. Usando a amostra a posteriori podemos obter as estimativas para $\hat{\alpha} = s_l^2$ e $\hat{l}_{max} = \bar{l} + s_l^2$, onde \bar{l} e s_l^2 são a média e a variância amostral dos l_t s, respectivamente. Estando esses momentos amostrais disponíveis, resolvemos as duas primeiras dificuldades supracitadas.

Uma das maneiras para solucionar a falta de acurácia da aproximação por uma distribuição $Gama(\alpha, 1)$ é permitir que o parâmetro de escala não seja exatamente 1. Portanto,

$$l_{max} - l_t \sim Gama(\alpha, \tau^{-1}),$$

com $\tau < 1$. Em problemas práticos escolhemos τ próximo a 1.

Usando o fato de que a função geradora de momentos de uma variável aleatória $W \sim Gama(\alpha, \tau^{-1})$ é $m_W(t) = (1 - \tau t)^{-\alpha}$, chegamos à fórmula para obtenção da preditiva de z

$$\log p(z) = l_{max} + \alpha \log(1 - \tau).$$

5.1.2 Fator de Bayes

Desejamos comparar dois modelos e verificar qual deles se ajusta melhor ao nosso conjunto de observações. Suponha, portanto, que existam duas hipóteses H_0 e H_1 de acordo com as probabilidades $p(z | H_0)$ e $p(z | H_1)$. Dadas $p(H_0)$ e $p(H_1) = 1 - p(H_0)$ as probabilidades a priori, depois de observados os dados temos que as probabilidades a posteriori são dadas por $p(H_0 | z)$ e $p(H_1 | z) = 1 - p(H_0 | z)$. Pelo teorema de Bayes temos

$$\frac{p(H_1 | z)}{p(H_0 | z)} = \frac{p(z | H_1) p(H_1)}{p(z | H_0) p(H_0)}.$$

Deste modo podemos ver que simplesmente multiplicando a razão das prioris por $\frac{p(z|H_1)}{p(z|H_0)}$ obtemos a razão das posterioris. A esse fator de transformação chamamos fator de Bayes e denotamos por B_{01} .

Na tabela 5.1 temos a interpretação do estimativa do fator de Bayes.

$2\ln(B_{01})$	Evidência contra H_0
0 a 2	Não vale a pena mencionar
2 a 6	Evidência positiva
6 a 10	Forte
> 10	Muito forte

Tabela 5.1: Interpretação do Fator de Bayes

A estimativa do fator de Bayes será obtida usando a logverossimilhança calculada por meio da aproximação da preditiva $p(z)$ pela distribuição Gama translada.

5.1.3 Deviance Information Criterion (DIC)

A comparação de modelos utilizando o DIC baseia - se na verossimilhança e na esperança a posteriori para avaliar a qualidade de ajuste de um determinado modelo.

Seja $\underline{\theta}^* = E(\underline{\theta} | z)$ e $D(\underline{\theta}) = -2\log(p(z | \underline{\theta}))$, onde z representa o conjunto de observações, $E(\underline{\theta} | z)$ é a esperança a posteriori e $p(z | \underline{\theta})$ corresponde à verossimilhança. Temos então que

$$DIC = \bar{D} + p_d,$$

onde $\bar{D} = E(D(\underline{\theta}) | z)$ e $p_d = \bar{D} - D(\underline{\theta}^*)$. A esperança \bar{D} mensura o quão bem o modelo ajusta os dados, ou seja, quanto maior o valor de \bar{D} pior é o ajuste feito pelo modelo. Por outro lado, p_d corresponde ao número efetivo de parâmetros do modelo, portanto quanto maior seu valor mais facilmente o modelo ajusta os dados. A medida p_d penaliza a complexidade do modelo, favorecendo modelos que apresentam um número pequeno de parâmetros.

Ao comparar modelos por meio do DIC é preferível optar por modelos com menores valores de DIC, visto que os dados são melhor ajustados por estes modelos.

Uma das vantagens de utilizar o DIC para comparação de modelos é o fato deste ser computacionalmente atrativo, uma vez que pode ser facilmente calculado a partir das saídas do MCMC. Sejam $\underline{\theta}^1, \underline{\theta}^2, \dots, \underline{\theta}^k$ as saídas de um algoritmo MCMC. Então:

$$\bar{D} = E(D(\underline{\theta}) | z)$$

pode ser calculado via método de Monte Carlo como

$$\bar{D} \approx \frac{1}{M} \sum_{k=1}^M D(\underline{\theta}^{(k)}),$$

onde M é o tamanho da amostra obtida pelo algoritmo MCMC.

Também

$$D(\underline{\theta}^*) = D(\bar{\underline{\theta}}),$$

com $\bar{\underline{\theta}} = \frac{1}{M} \sum_{k=1}^M \underline{\theta}^k$.

Podemos, então, avaliar a qualidade de ajuste dos modelos gaussiano e não gaussiano.

Capítulo 6

Resultados

A primeira análise a ser feita é verificar se o modelo não gaussiano de fato ajusta melhor os dados do que o modelo gaussiano. A análise exploratória nos forneceu uma intuição acerca de qual o melhor modelo na direção do modelo não gaussiano, visto que, observando as figuras 3.5(b), 3.6(b) e 3.7(b) notamos discrepâncias entre as variâncias das estações. Esse comportamento heterogêneo da variabilidade ao longo do espaço nos leva a acreditar que processos gaussianos não conseguirão explicar essa heterogeneidade espacial de forma satisfatória.

Realizada a etapa de amostragem via métodos MCMC somos capazes de analisar os resultados obtidos e decidir pelo melhor modelo a ser usado para cada processo em estudo (temperatura, umidade do ar e pressão atmosférica).

Seja $\sigma_i^2 = \frac{\sigma^2}{\lambda_i}$, onde σ_i^2 representa a variância relativa a cada estação de monitoramento. Nas figuras 6.1(a), 6.2(a) e 6.3(a) podemos observar os diferentes comportamentos da variância para estações distintas. Este fato colabora com a nossa hipótese de não normalidade dos conjuntos de dados.

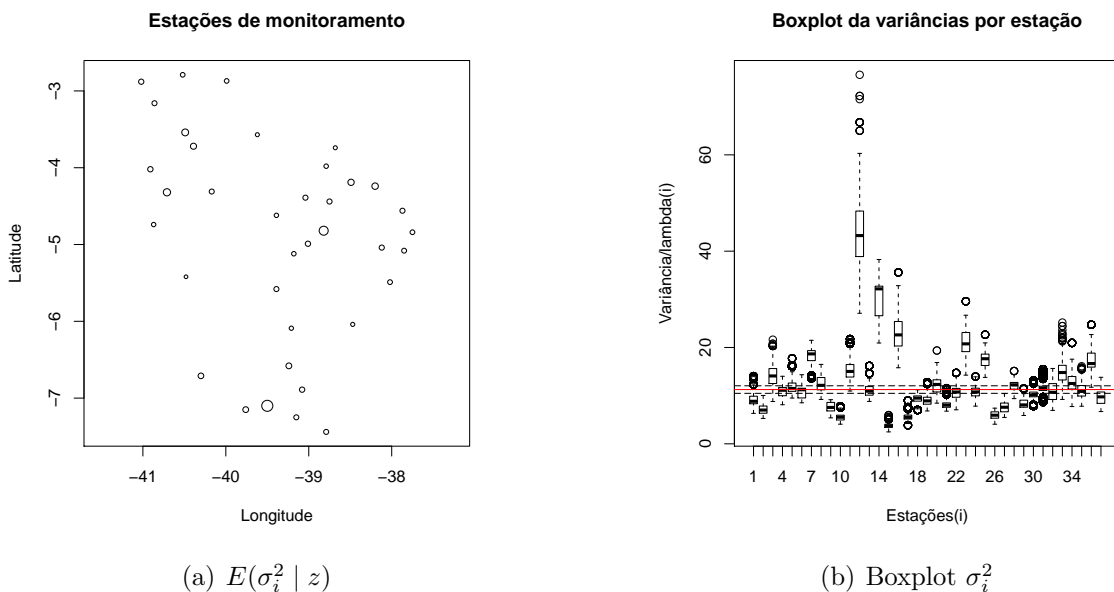


Figura 6.1: Variabilidade por estação - Temperatura

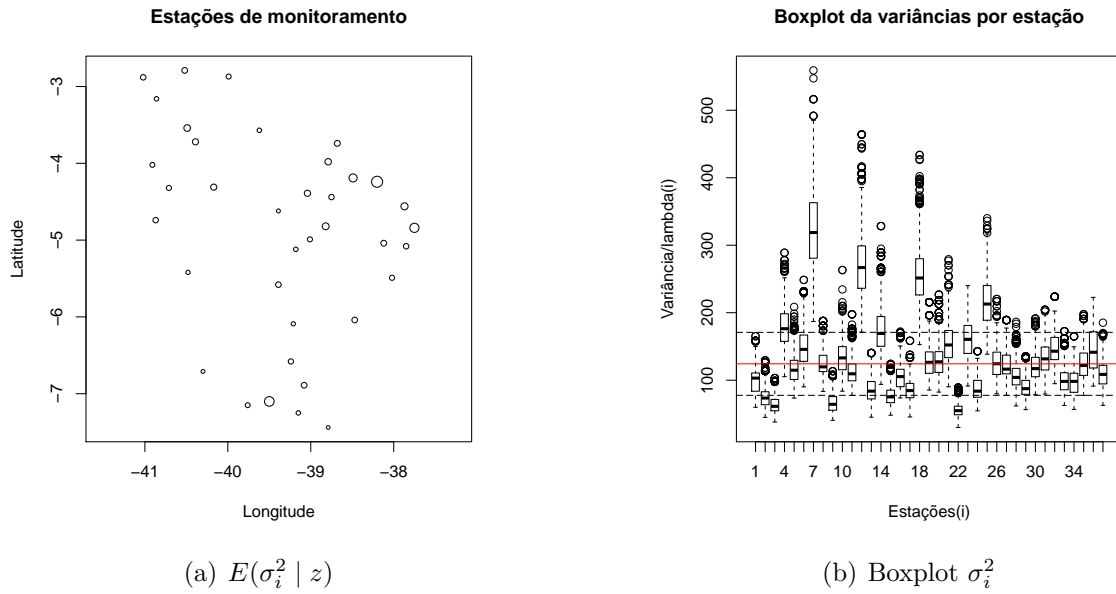


Figura 6.2: Variabilidade por estação - Umidade relativa

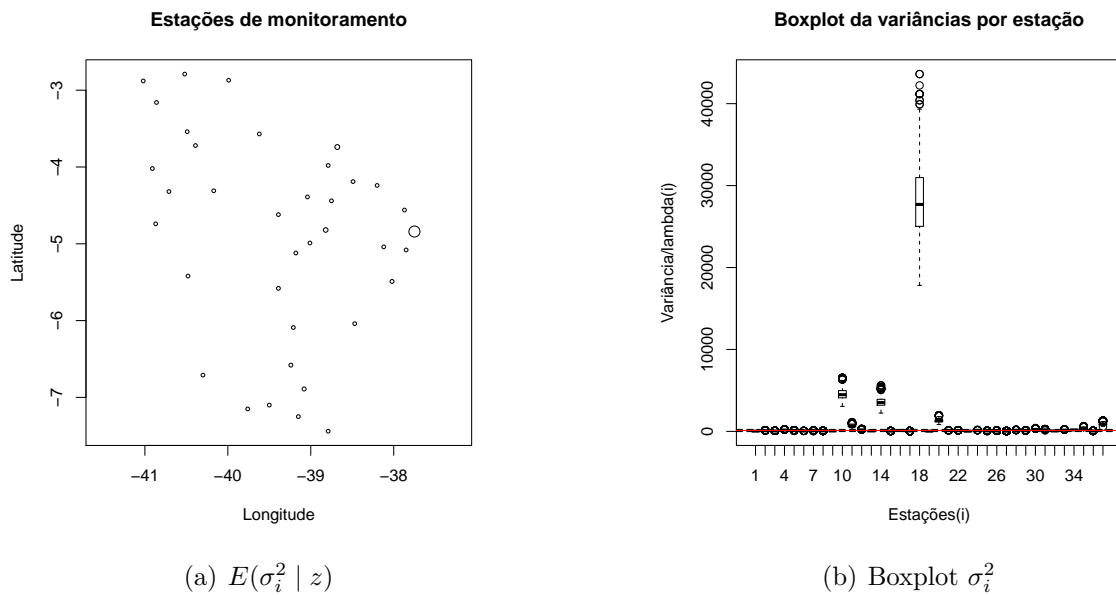


Figura 6.3: Variabilidade por estação - Pressão atmosférica

Para também justificar a adoção de modelos não gaussianos, as figuras 6.1(b), 6.2(b) e 6.3(b) mostram de maneira clara a existência de estações de monitoramento com variâncias muito diferentes em relação às outras estações. Está ilustrada, ainda, a mediana de σ^2 bem como seu respectivo intervalo de credibilidade para o modelo gaussiano. Podemos constatar, então, que o uso de modelos gaussianos comprometeria a adaptação de regiões de grande variabilidade.

Colaborando com as conclusões obtidas qualitativamente, utilizamos o DIC e o fator de Bayes como critérios de comparação entre a qualidade do ajuste aos dados pelo modelo gaussiano e pelo modelo não gaussiano para os todos os processos estudados. Na tabela (6.1) temos os valores de DIC obtidos para ambos os modelos e a tabela (6.2) traz as estimativas do fator

de Bayes obtido assumindo como H_0 a hipótese onde os dados seguem um processo Gaussiano e H_1 aquela onde os dados seguem um processo não Gaussiano.

	Modelo Gaussiano	Modelo Não Gaussiano
Temperatura	7429.83	3500.06
Umidade relativa	16610.23	13332.21
Pressão atmosférica	11341.38	-29649.36

Tabela 6.1: DIC

	Fator de Bayes
Temperatura	290.201
Umidade relativa	125.62
Pressão atmosférica	2782.62

Tabela 6.2: Fator de Bayes

Analisando a tabela (6.1) notamos que para todos os processos (temperatura, umidade relativa e pressão atmosférica) o modelo não gaussiano obteve um menor valor de DIC, indicando ser o melhor modelo. Ainda, comparando as estimativas do fator de Bayes com a tabela (5.1) concluímos que há evidências muito fortes para escolher o modelo não gaussiano referente à hipótese H_1 .

Comprovamos, quantitativamente, que o modelo não gaussiano ajusta melhor os dados, acomodando de forma mais adequada a variabilidade espacial.

Visto que o modelo não gaussiano se mostrou mais adequado para explicar nossos conjuntos de dados, passaremos agora a analisar a suposição de que as funções de correlação das observações e do processo espacial $\lambda(s)$ são iguais.

Como argumentado no capítulo 2, não parece plausível assumir que ambas as funções de correlação possuam mesma estrutura com mesmo parâmetro. Deste momento em diante nos dedicaremos a testar empiricamente essa suposição.

Adotamos a função exponencial para modelar a estrutura de correlação entre estações. Suponha que estejamos de fato trabalhando com funções de correlação diferentes para o conjunto de observações e o processo $\lambda(s)$. Temos então

$$C_{\phi_1}(\|s_i - s_{i'}\|) = \exp\left(\frac{-d_s}{\phi_1}\right),$$

onde d_s é a distância entre estações, e

$$C_{\phi_3}(\|s_i - s_{i'}\|) = \exp\left(\frac{-d_s}{\phi_3}\right)$$

a função que correlaciona os parâmetros $\lambda_i, i = 1, \dots, I$.

Com base em ambos os processos, foram feitas simulações via MCMC para o processo com mesma função de correlação e o processo que assume C_{ϕ_1} e C_{ϕ_3} distintas, separadamente.

A tabela (6.3) explicita os valores obtidos a partir de uma amostra a posteriori. As estimativas dos parâmetros ϕ_1 e ϕ_3 correspondem à mediana da amostragem a posteriori via MCMC. O parâmetro ϕ_1 refere - se à função de correlação espacial C_{ϕ_1} para o processo que julga semelhantes as funções de correlação de $\lambda(s)$ e das observações. Enquanto o parâmetro ϕ_3 corresponde à função de correlação C_{ϕ_3} para o processo que admite funções de correlação diferentes.

Processo	Parâmetro	Mediana	Limite inferior	Limite superior
Temperatura	ϕ_1	1.81	1.76	1.85
	ϕ_3	0.08	0.03	0.13
Umidade relativa	ϕ_1	0.797	0.783	0.790
	ϕ_3	0.22	0.13	0.32
Pressão atmosférica	ϕ_1	1.34	1.32	1.35
	ϕ_3	0.098	0.083	0.112

Tabela 6.3: Estimativas dos parâmetros

Observamos que os parâmetros ϕ_1 e ϕ_3 são consideravelmente diferentes. Além disso, vemos que os intervalos de 95% de credibilidade para ϕ_1 e ϕ_3 possuem interseção vazia. Portanto, percebemos que não é razoável assumir que o processo de covariância de $\lambda(s)$ seja o mesmo que para o conjunto de observações.

Novamente faremos a comparação entre modelos através de dois critérios: DIC e fator de Bayes. As tabelas (6.4) e (6.5) apresentam os resultados obtidos. Para o cálculo do fator de Bayes assumimos H_0 a hipótese onde o processo com função de correlação igual ajusta melhor os dados e H_1 o processo com funções de correlação distintas.

	Modelo	DIC
Temperatura	Não Gaussiano (C_{ϕ_1})	3500.061
	Não Gaussiano (C_{ϕ_1}, C_{ϕ_3})	-10733.15
Umidade relativa	Não Gaussiano (C_{ϕ_1})	13332.21
	Não Gaussiano (C_{ϕ_1}, C_{ϕ_3})	13671.89
Pressão atmosférica	Não Gaussiano (C_{ϕ_1})	-29649.36
	Não Gaussiano (C_{ϕ_1}, C_{ϕ_3})	-27226.1

Tabela 6.4: Comparação de modelos - DIC

Processo	Fator de Bayes
Temperatura	203.03
Umidade relativa	-0.04
Pressão atmosférica	-28.76

Tabela 6.5: Comparação de modelos - Fator de Bayes

Para as observações de temperatura do ar, obtivemos o resultado esperado comparando os modelos tanto pelo DIC quanto pelo fator de Bayes. Como podemos observar, o DIC do

modelo com funções de correlação distintas é menor do que o outro modelo, significando melhor qualidade de ajuste por este. O fator de Bayes ratifica a decisão de escolha pelo modelo com correlações diferentes visto que seu valor é superior a 10. Apesar de tanto o DIC quanto o fator de Bayes acusarem que para o processo de umidade relativa o modelo não gaussiano com funções de correlação iguais seja o melhor modelo, parece que não existem grande diferenças entre os dois modelos, visto que o fator de Bayes possui um valor muito pequeno. Para a pressão atmosférica existe uma forte evidência de que o modelo com funções de correlação iguais seja o que melhor ajusta as observações de pressão. Ainda assim, como notoriamente os parâmetros da tabela (6.3) são diferentes, acreditamos que o modelo não gaussiano com funções de correlação distintas explique melhor os dados e seja capaz de se adequar à variabilidade espacial. Em um contexto preditivo, possivelmente o uso de modelos com funções de covariância distintas apresentará um desempenho melhor. Como ϕ_1 e ϕ_3 são discrepantes o uso de modelos com a mesma função de covariância representaria uma subestimação do processo $\lambda(s)$ no sentido que estaríamos inferindo erroneamente sobre a covariância de $\lambda(s)$.

Capítulo 7

Conclusões

A análise exploratória dos nossos conjuntos de observações de temperatura, umidade relativa e pressão atmosférica mostrou que estações de monitoramento distintas apresentam variâncias discrepantes entre si para todos os três processos estudados. Tal volatilidade espacial sugerida por essa análise inicial é maior do que a esperada por um modelo gaussiano. Este fato nos motiva ao uso de modelos não gaussianos a fim de acomodar de forma adequada essa variabilidade.

O objetivo principal da utilização de modelos GLG é a capacidade que estes modelos possuem de se adaptar a observações destoantes do esperado. Assim, modelos GLG são interessantes na modelagem de dados onde observações aberrantes devem ser levadas em consideração para explicar o processo que originou os dados. Além disso, por possuírem caudas mais pesadas em relação ao modelo gaussiano os modelos geoestatísticos não gaussianos são eficazes na detecção e previsão de observações extremas. Modelos GLG possuem grande destaque na modelagem de dados climáticos, visto que é de grande importância a possibilidade de previsão de observações extremas a fim de minimizar os danos causados por grandes alterações climáticas.

A média de cada processo $\mu_i(s, t) = f_i(s, t)' \beta_i, i = 1, 2, 3$ foi modelada via regressão linear múltipla, tendo como covariáveis a altitude, longitude, latitude e transformações harmônicas do tempo. A escolha das covariáveis, explicada no capítulo sobre a análise exploratória, é natural visto que a posição geográfica de cada estação influencia a medição de variáveis climáticas. A modelagem da média $\mu_i(s, t)$ explica a variabilidade em macroescala, deixando para $\epsilon(s, t)$ a variabilidade em microescala não explicada por $\mu_i(s, t)$. Esta flutuação aleatória em microescala é modelada via modelos não gaussianos.

Para transformar um modelo gaussiano usual em não gaussiano introduzimos o processo $\lambda(s)$ na escala do modelo gaussiano (2.1). A inclusão de $\lambda(s)$ aumenta a complexidade do modelo uma vez que a cada estação está associada uma variável $\lambda(s_i)$. Tal complexidade torna - se explícita na estimação do vetor de parâmetros $\underline{\theta}_1 = (\beta, \sigma^2, \phi_1, \phi_2, \phi_3, \nu, \lambda)$ do modelo (2.2) com função de correlação entre estações diferente da função de correlação do processo $\lambda(s)$. Nosso modelo não possui distribuição de probabilidade a posteriori com forma analítica fechada, necessitando do uso de métodos MCMC para amostragem de $p(\underline{\theta}_1 | z)$ e consequente estimação do vetor $\underline{\theta}_1$.

O modelo não gaussiano proposto em *Non-Gaussian Spatiotemporal Modelling through Scale Mixing* assume que, dentro de cada um dos processos de temperatura, umidade relativa e pressão atmosférica, a estrutura de correlação das variáveis observadas segue o mesmo padrão da estrutura de covariância do processo espacial $\lambda(s)$ inserido na escala do modelo gaussiano. Essa suposição, porém, não parece razoável uma vez que não existe nenhuma justificativa para que, de fato, as funções de correlação sejam iguais. Intuitivamente espera-se que variáveis diferentes apresentem matrizes de correlação espacial diferentes, visto as especificidades de cada variável. Podemos testar empiricamente a validade desta suposição usando funções de correlação distintas para os dados e para $\lambda(s)$.

Após amostrarmos do modelo não gaussiano proposto utilizando métodos MCMC somos capazes de fazer inferência acerca do vetor $\underline{\theta}_2 = (\beta, \sigma^2, \phi_1, \phi_2, \nu, \lambda)$ referente ao modelo não gaussiano que assume igualdade entre as funções de covariância. Tendo estimado os parâmetros do nosso modelo, podemos comparar a qualidade de ajuste de ambos os modelos gaussiano e não gaussiano. Duas maneiras de comparar modelos são o uso do DIC e do Fator de Bayes como critérios quantitativos de escolha. Comparando os modelos (2.1) e (2.2), verificamos a suposição inicial de que o modelo não gaussiano (2.2) modela mais adequadamente o nosso conjunto de observações.

O principal objetivo deste trabalho foi verificar, empiricamente, se a suposição de que a função exponencial $C_\theta(d_s)$ possui mesmo parâmetro tanto para os dados quanto para $\lambda(s)$ é válida. Em outras palavras, queremos testar se ϕ_1 e ϕ_3 , parâmetros das funções de correlação $C_{\phi_1}(d_s)$ e $C_{\phi_3}(d_s)$, respectivamente, podem ser considerados iguais.

Através de métodos MCMC podemos obter uma amostra a posteriori do vetor $\underline{\theta}_1 = (\beta, \sigma^2, \phi_1, \phi_2, \phi_3, \nu, \lambda)$ e, conseqüentemente, encontrar uma estimativa $\hat{\underline{\theta}}_1$ e seus intervalos de credibilidade. Na tabela (6.3) estão as estimativas a posteriori obtidas para ϕ_1 e ϕ_3 em cada um dos conjuntos de observações. Os valores dos intervalos de credibilidade para $\hat{\phi}_1$ e $\hat{\phi}_3$ são razoavelmente diferentes para todos os três processos observados, levando à conclusão que a suposição de igualdade entre funções de covariância não é válida.

Ainda que alguns critérios de comparação de modelos não nos forneçam evidências de que o modelo não gaussiano com $C_{\phi_1}(d_s)$ e $C_{\phi_3}(d_s)$ distintas é melhor (como era esperado), escolhemos este modelo como o que melhor ajusta os dados baseados nas estimativas dos parâmetros ϕ_1 e ϕ_3 .

Concluimos, portanto, que o modelo não gaussiano que assume diferentes as funções de correlação das observações e do processo espacial $\lambda(s)$ se mostrou com melhor qualidade de ajuste aos dados, modelando bem a variabilidade destes processos em estudo e permitindo a detecção de observações destoantes do esperado.

Uma extensão ao presente trabalho poderia ser a análise da qualidade de previsão do modelo, pois a previsão de eventos extremos é de suma importância na área meteorológica. Algumas estações de monitoramento foram excluídas da modelagem devido à falta de observações para todo o intervalo de tempo observado. Estas estações poderiam ser usadas para mensurar a discrepância entre o previsto e o observado para determinada estação em determinado instante

de tempo, permitindo o cálculo de medidas para comparação entre modelos.

O estudo sugerido como extensão deste trabalho complementa o conhecimento sobre a eficácia do modelo não gaussiano na modelagem de dados com observações discrepantes. Com a análise da qualidade de previsão aumentamos nosso conhecimento sobre as diferenças entre os modelos gaussiano e não gaussiano e estamos melhor capacitados para escolher o melhor modelo a ser usado.

Bibliografia

- [1] Eberly Shelly, Swall Jenise, Holland David, Cox Bill, Baldrige Ellen Developing Spatially Interpolated Surfaces and Estimating Uncertainty .
- [2] Carlin Bradley P., Gelfand Alan E. Hierarchical Modeling and Analysis for Spatial Data.
- [3] Cressie Noel, Wikle Christopher K. Statistics for Spatio-Temporal data.
- [4] Fonseca, Thais C.O., Steel, M F J (2011) Non-Gaussian Spatiotemporal Modelling through Scale Mixing. *Biometrika*, **4**, 761-774.
- [5] Palacios, M B, Steel, M F J (2006) Non-Gaussian Bayesian Geostatistical Modeling. *Journal of the American Statistical Association*, **474**, 604–618.
- [6] Raftery, A. E., Newton, M. A., Satagopan, J. M., Krivitsky, P. N. (2007). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. In Bayesian Statistics 8, J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, M. West, eds. Oxford: Oxford University Press.

Apêndice A

Código R

```
# Calculo da matriz de covariancia Sigma

DistS = function(dists,I){
  valor = matrix(0,I,I);
  for (i in 1:I){ for (j in 1:I){
    norma = sqrt((dists[i,1]-dists[j,1])^2+(dists[i,2]-dists[j,2])^2);
    valor[i,j] = norma;
    valor[j,i] = valor[i,j];}}
  valor;}

DistT = function(distt,J){
  valor = matrix(0,J,J);
  for (i in 1:J){ for (j in 1:J){
    norma = abs(distt[i]-distt[j]);
    valor[i,j] = norma;
    valor[j,i] = valor[i,j];}}
  valor;}

Cov12 = function(distance,phi1){
  value1 = exp(-distance/phi1);
  value1}

# Calculo das distancias

I = 37; J = 84;
distanceS = DistS(locat,I)
distanceT = DistT(times,J)
```

```

# MODELO NORMAL

# Calculo da verossimilhanca

LogVero=function(vecz, distanceS, distanceT, sig2, phi1, phi2, mu, I, J){
  C1=Cov12(distanceS,phi1)
  C2=Cov12(distanceT,phi2)
  Inv.C1=solve(C1); Inv.C2=solve(C2); Inv.Sig2=kronecker(Inv.C2, Inv.C1);
  ldet.C1=determinant(C1,log=T)$modulus[[1]];
  ldet.C2=determinant(C2,log=T)$modulus[[1]];
  lvero=-J/2*ldet.C1-I/2*ldet.C2-((I*J)/2)*log(sig2)
  -1/(2*sig2)*(t(vecz-mu)%*%Inv.Sig2%*(vecz-mu));
  lvero}

# MCMC

# rphi1, rphi2 e rs2 variancias de phi1, phi2 e sigma2 respectivamente
# M o numero de iteracoes
# sig2, phi1k, phi2k, betak chutes iniciais
cont1 = cont2 = cont3 = 0
theta = matrix(0, M, p) # p a dimensao do vetor theta = (beta, sigma2, phi1, phi2)
muk=X%*%betak # X matriz com as covariaveis
lvero=matrix(NA,M,1)

for (k in 1:M){

##### gerando beta

C1=Cov12(distanceS,phi1k)
C2=Cov12(distanceT,phi2k)
Inv.C1=solve(C1); Inv.C2=solve(C2); Inv.Sig2=kronecker(Inv.C2, Inv.C1)
mu.1=t(X)%*%Inv.Sig2%*X
mu.beta=solve(mu.1)%*%t(X)%*%Inv.Sig2%*vecz
sigma.beta=sig2k*solve(mu.1)
betak=rmvnorm(1,mean=mu.beta,sigma=sigma.beta)
betak=matrix(betak,20,1)
muk=X%*%betak
lverok=LogVero(vecz,distanceS,distanceT,sig2k,phi1k,phi2k, mu=muk, I, J);

```

gerando sig2

```
sig2prop = exp(log(sig2k)+rs2*rnorm(1));
mu=X%*%betak
lveroprop = LogVero(vecz,distanceS,distanceT,sig2prop,phi1k,phi2k, mu, I, J);
auxprop = lveroprop+log(sig2prop)+dinvgamma(sig2prop,2.1,1.1);
auxk = lverok+log(sig2k)+dinvgamma(sig2k,2.1,1.1);
ratio = auxprop-auxk;
test = runif(1);
if (ratio>log(test)) {
  sig2k = sig2prop;
  lverok = lveroprop;
  cont1 = cont1 + 1;}

```

gerando phi1 e phi2

```
phi1prop = exp(log(phi1k)+rphi1*rnorm(1));
phi2prop = exp(log(phi2k)+rphi2*rnorm(1));
mu=X%*%betak
lveroprop = LogVero(vecz,distanceS,distanceT,sig2k,phi1prop,phi2prop,mu, I, J);
auxprop = lveroprop+log(phi1prop)+log(dgamma(phi1prop,0.5,0.5));
auxk = lverok+log(phi1k)+log(dgamma(phi1k,.5,.5));
ratio = auxprop-auxk;
test = runif(1);
if (ratio>log(test)) {
  phi1k = phi1prop;
  lverok = lveroprop;
  cont2 = cont2 + 1;}

```

```
lveroprop = LogVero(vecz,distanceS,distanceT,sig2k,phi1prop,phi2prop,mu, I, J);
auxprop = lveroprop+log(phi2prop)+log(dgamma(phi2prop,2,2));
auxk = lverok+log(phi2k)+log(dgamma(phi2k,2,2));
ratio = auxprop-auxk;
test = runif(1);
if (ratio>log(test)) {
  phi2k = phi2prop;
  lverok = lveroprop;
  cont3 = cont3 + 1;}

```

```
theta[k,1] = sig2k;
```

```

theta[k,2] = phi1k;
theta[k,3] = phi2k;
theta[k,4] = betak[1,]
...
theta[k,p] = betak[q,] #q numero de coeficientes da regressao
lvero[k]=lverok}

# MODELO NAO NORMAL (FUNCOES DE COVARIANCIA IGUAIS)

# criandos regioes (blocos) para lambda

s1=locat[,1]; s2=locat[,2];
ind = 1:I
region3 = ind[which((s1<4.570359)&s2<94.34557)]
region2 = ind[which((s1>=4.570359)&s2>94.34557)]
region1 = ind[which((s1<4.570359)&s2>94.34557)]
region4 = ind[which((s1>=4.570359)&s2<94.34557)]
indr = c(region1,region2,region3,region4)
n=NA

n[1] = length(region1)
n[2] = length(region2)
n[3] = length(region3)
n[4] = length(region4)
nr = c(0,cumsum(n))

# Calculo da covariancia

C1=function(distanceS, phi1,lambda1,I){
  c1=matrix(0,I,I)
  for(k in 1:I){
    for(j in 1:I){
      c1[k,j]=(exp(-distanceS[k,j]/phi1))/sqrt(lambda1[k]*lambda1[j]) }}
  c1;}

Cov12 = function(distance,phi1){
  value1 = exp(-distance/phi1);
  value1}

```

```

# Calculo da verossimilhanca

LogVeroNG=function(vecz, distanceS, distanceT, sig2, phi1,
phi2, mu, I, J, loglambda1){
  C.1=C1(distanceS, phi1, exp(loglambda1), I)
  C.2=Cov12(distanceT,phi2)
  Inv.C1=solve(C.1); Inv.C2=solve(C.2); Inv.Sig2=kronecker(Inv.C2, Inv.C1);
  ldet.C1=determinant(C.1,log=T)$modulus[[1]];
  ldet.C2=determinant(C.2,log=T)$modulus[[1]];
  lvero=-J/2*ldet.C1-I/2*ldet.C2-((I*J)/2)*log(sig2)
  -1/(2*sig2)*(t(vecz-mu)%*%Inv.Sig2%*(vecz-mu));
  lvero}

LogVeroLogLamb1 = function(nu1,loglamb1,distanceS,phi1){
  mu.loglamb1 = rep(-nu1/2,I)
  M1.loglamb1 = nu1*Cov12(distanceS,phi1);
  dmvnorm(loglamb1,mean=mu.loglamb1,sigma=M1.loglamb1,log=T)}

# MCMC

# rphi1, rphi2, rs2, rnu1 e rlambq variancias de phi1, phi2,
# sigma2, nu1 e lambda1 respectivamente
# M o numero de iteracoes
# sig2, phi1k, phi2k, betak, nu1k e loglamb1k chutes iniciais
cont1 = cont2 = cont3 = cont4 = cont5 = 0
theta = matrix(0, M, p+1) # p+1 numero de parametros do modelo
muk=X%*%betak # X matriz com as covariaveis
lvero=matrix(NA,M,1)
lambda1 = matrix(NA,M,I)

for (k in 1:M){

##### gerando beta
c.1=C1(distanceS, phi1k, exp(loglamb1k), I)
c.2=Cov12(distanceT,phi2k)
Inv.c1=solve(c.1); Inv.c2=solve(c.2); Inv.sig2=kronecker(Inv.c2, Inv.c1)
mu.1=t(X)%*%Inv.sig2%*%X
mu.beta=solve(mu.1)%*%t(X)%*%Inv.sig2%*%vecz
sigma.beta=sig2k*solve(mu.1)
betak=rmvnorm(1,mean=mu.beta,sigma=sigma.beta)

```

```

betak=matrix(betak,20,1)
muk=X%*%betak
lverok = LogVeroNG(vecz, distanceS, distanceT, sig2k, phi1k, phi2k,
muk, I, J, loglamb1k);

##### gerando sig2
sig2prop = exp(log(sig2k)+rs2*rnorm(1));
lveroprop = LogVeroNG(vecz, distanceS, distanceT, sig2prop, phi1k, phi2k,
muk, I, J, loglamb1k)
auxprop = lveroprop+log(sig2prop)+dinvgamma(sig2prop,1.1,2.1);
auxk = lverok+log(sig2k)+dinvgamma(sig2k,1.1,2.1);
ratio = auxprop-auxk;
test = runif(1);
if (ratio>log(test)) {
  sig2k = sig2prop;
  lverok = lveroprop;
  cont1 = cont1 + 1;}

##### gerando phi1 e phi2

phi1prop = exp(log(phi1k)+rphi1*rnorm(1));
phi2prop = exp(log(phi2k)+rphi2*rnorm(1));
mu=X%*%betak
lveroprop = LogVeroNG(vecz,distanceS,distanceT,sig2k,phi1prop,phi2k,
mu, I, J, loglamb1k);
auxprop = lveroprop+log(phi1prop)+log(dgamma(phi1prop,0.5,0.5));
auxk = lverok+log(phi1k)+log(dgamma(phi1k,.5,.5));
ratio = auxprop-auxk;
test = runif(1);
if (ratio>log(test)) {
  phi1k = phi1prop;
  lverok = lveroprop;
  cont2 = cont2 + 1;}

lveroprop = LogVeroNG(vecz,distanceS,distanceT,sig2k,phi1k,phi2prop,
mu, I, J, loglamb1k);
auxprop = lveroprop+log(phi2prop)+log(dgamma(phi2prop,2,2));
auxk = lverok+log(phi2k)+log(dgamma(phi2k,2,2));
ratio = auxprop-auxk;
test = runif(1);

```

```

if (ratio>log(test)) {
  phi2k = phi2prop;
  lverok = lveroprop;
  cont3 = cont3 + 1;}

##### gerando nu1

nulprop = exp(log(nu1k)+rnu1*rnorm(1));
lveroprop2 = LogVeroLogLamb1(nulprop,loglamb1k,distanceS,phi1k)
auxprop = lveroprop2+log(nulprop)+log(dgamma(nulprop,1,5))
auxk = lverok2+log(nu1k)+log(dgamma(nu1k,1,5));
ratio = auxprop-auxk;
test = runif(1);
if (ratio>log(test)) {
  nu1k = nulprop;
  lverok2 = lveroprop2;
  cont4 = cont4 + 1;}

##### gerando lambda1

for (j in 1:4){
  loglamb1prop = loglamb1k;
  for (jj in 1:4){
    if (j==jj){
      region = indr[(nr[j]+1):nr[j+1]]
      loglamb1prop[region] = loglamb1k[region] + rlamb1[j]*rnorm(length(region));}}
  lveroprop2 = LogVeroLogLamb1(nu1k,loglamb1prop,distanceS,phi1k)
  lveroprop = LogVeroNG(vecz, distanceS, distanceT, sig2k, phi1k, phi2k,
  muk, I, J, loglamb1prop)
  auxprop = lveroprop+lveroprop2;
  auxk = lverok+lverok2;
  ratio = auxprop-auxk;
  test = runif(1);
  if (ratio>log(test)) {
    loglamb1k = loglamb1prop;
    lverok2 = lveroprop2;
    lverok = lveroprop;
    cont5 = cont5 + 1;}}

theta[k,1] = sig2k;

```

```

theta[k,2] = phi1k;
theta[k,3] = phi2k;
theta[k,4] = nu1k;
theta[k,5] = betak[1,]
...
theta[k,p+1] = betak[q,] # q o numero de coeficientes da regressao
lambda1[k,] = exp(loglamb1k);
lvero[k]=lverok
}

# MODELO NAO NORMAL (FUNCOES DE COVARIANCIA IGUAIS)

# Criandos regioes (blocos) para lambda

s1=locat[,1]; s2=locat[,2];
ind = 1:I
region3 = ind[which((s1<4.570359)&s2<94.34557)]
region2 = ind[which((s1>=4.570359)&s2>94.34557)]
region1 = ind[which((s1<4.570359)&s2>94.34557)]
region4 = ind[which((s1>=4.570359)&s2<94.34557)]
indr = c(region1,region2,region3,region4)
n=NA

n[1] = length(region1)
n[2] = length(region2)
n[3] = length(region3)
n[4] = length(region4)
nr = c(0,cumsum(n))

# Calculo da covariancia

C1=function(distanceS, phi1,lambda1,I){
  c1=matrix(0,I,I)
  for(k in 1:I){
    for(j in 1:I){
      c1[k,j]=(exp(-distanceS[k,j]/phi1))/sqrt(lambda1[k]*lambda1[j]) }}
  c1;}

```

```

Cov12 = function(distance,phi1){
  value1 = exp(-distance/phi1);
  value1}

LogVeroNG=function(vecz, distanceS, distanceT, sig2, phi1, phi2, mu,
I, J, loglambda1){
  C.1=C1(distanceS, phi1, exp(loglambda1), I)
  C.2=Cov12(distanceT,phi2)
  Inv.C1=solve(C.1); Inv.C2=solve(C.2); Inv.Sig2=kronecker(Inv.C2, Inv.C1);
  ldet.C1=determinant(C.1,log=T)$modulus[[1]];
  ldet.C2=determinant(C.2,log=T)$modulus[[1]];
  lvero=-J/2*ldet.C1-I/2*ldet.C2-((I*J)/2)*log(sig2)
  -1/(2*sig2)*(t(vecz-mu)%*%Inv.Sig2%*(vecz-mu));
  lvero}

LogVeroLogLamb1 = function(nu1,loglamb1,distanceS,phi3){
  mu.loglamb1 = rep(-nu1/2,I)
  M1.loglamb1 = nu1*Cov12(distanceS,phi3);
  dmvnorm(loglamb1,mean=mu.loglamb1,sigma=M1.loglamb1,log=T)}

# MCMC

# rphi1, rphi2, rphi3, rs2, rnu1 e rlambq variancias de phi1, phi2, phi3,
# sigma2, nu1 e lambda1 respectivamente
# M o numero de iteracoes
# sig2, phi1k, phi2k, phi3k, betak, nul1 e loglamb1k chutes iniciais
cont1 = cont2 = cont3 = cont4 = cont5 = cont6 = 0
theta = matrix(0, M, p+2) # p+2 numero de parametros do modelo
muk=X%*%betak # X matriz com as covariaveis
lvero=matrix(NA,M,1)
lambda1 = matrix(NA,M,I)

for (k in 1:M){

##### gerando beta
c.1=C1(distanceS, phi1k, exp(loglamb1k), I)
c.2=Cov12(distanceT,phi2k)
Inv.c1=solve(c.1); Inv.c2=solve(c.2); Inv.sig2=kronecker(Inv.c2, Inv.c1)
mu.1=t(X)%*%Inv.sig2%*%X
mu.beta=solve(mu.1)%*%t(X)%*%Inv.sig2%*%vecz

```

```

sigma.beta=sig2k*solve(mu.1)
betak=rmvnorm(1,mean=mu.beta,sigma=sigma.beta)
betak=matrix(betak,20,1)
muk=X%*%betak
lverok = LogVeroNG(vecz, distanceS, distanceT, sig2k, phi1k, phi2k,
muk, I, J, loglamb1k);

##### gerando sig2
sig2prop = exp(log(sig2k)+rs2*rnorm(1));
lveroprop = LogVeroNG(vecz, distanceS, distanceT, sig2prop, phi1k, phi2k,
muk, I, J, loglamb1k)
auxprop = lveroprop+log(sig2prop)+dinvgamma(sig2prop,1.1,2.1);
auxk = lverok+log(sig2k)+dinvgamma(sig2k,1.1,2.1);
ratio = auxprop-auxk;
test = runif(1);
if (ratio>log(test)) {
  sig2k = sig2prop;
  lverok = lveroprop;
  cont1 = cont1 + 1;}

##### gerando phi1 e phi2

phi1prop = exp(log(phi1k)+rphi1*rnorm(1));
phi2prop = exp(log(phi2k)+rphi2*rnorm(1));
mu=X%*%betak
lveroprop = LogVeroNG(vecz,distanceS,distanceT,sig2k,phi1prop,phi2k,
mu, I, J, loglamb1k);
auxprop = lveroprop+log(phi1prop)+log(dgamma(phi1prop,0.5,0.5));
auxk = lverok+log(phi1k)+log(dgamma(phi1k,.5,.5));
ratio = auxprop-auxk;
test = runif(1);
if (ratio>log(test)) {
  phi1k = phi1prop;
  lverok = lveroprop;
  cont2 = cont2 + 1;}

lveroprop = LogVeroNG(vecz,distanceS,distanceT,sig2k,phi1k,phi2prop,mu, I, J, loglamb1k)
auxprop = lveroprop+log(phi2prop)+log(dgamma(phi2prop,2,2));
auxk = lverok+log(phi2k)+log(dgamma(phi2k,2,2));
ratio = auxprop-auxk;

```

```

test = runif(1);
if (ratio>log(test)) {
  phi2k = phi2prop;
  lverok = lveroprop;
  cont3 = cont3 + 1;}

##### gerando phi3

phi3prop = exp(log(phi3k)+rphi3*rnorm(1));
lveroprop2 = LogVeroLogLamb1(nu1k,loglamb1k,distanceS,phi3prop)
auxprop = lveroprop2+log(phi3prop)+log(dgamma(phi3prop,1,5))
auxk = lverok2+log(phi3k)+log(dgamma(phi3k,1,5));
ratio = auxprop-auxk;
test = runif(1);
if (ratio>log(test)) {
  phi3k = phi3prop;
  lverok2 = lveroprop2;
  cont4 = cont4 + 1;}

##### gerando nu1

nu1prop = exp(log(nu1k)+rnu1*rnorm(1));
lveroprop2 = LogVeroLogLamb1(nu1prop,loglamb1k,distanceS,phi3k)
auxprop = lveroprop2+log(nu1prop)+log(dgamma(nu1prop,1,5))
auxk = lverok2+log(nu1k)+log(dgamma(nu1k,1,5));
ratio = auxprop-auxk;
test = runif(1);
if (ratio>log(test)) {
  nu1k = nu1prop;
  lverok2 = lveroprop2;
  cont5 = cont5 + 1;}

##### gerando lambda1

for (j in 1:4){
  loglamb1prop = loglamb1k;
  for (jj in 1:4){
    if (j==jj){
      region = indr[(nr[j]+1):nr[j+1]]
      loglamb1prop[region] = loglamb1k[region] + rlamb1[j]*rnorm(length(region));}}

```

```

lveroprop2 = LogVeroLogLamb1(nu1k,loglamb1prop,distanceS,phi3k)
lveroprop = LogVeroNG(vecz, distanceS, distanceT, sig2k, phi1k, phi2k,
muk, I, J, loglamb1prop)
auxprop = lveroprop+lveroprop2;
auxk = lverok+lverok2;
ratio = auxprop-auxk;
test = runif(1);
if (ratio>log(test)) {
  loglamb1k = loglamb1prop;
  lverok2 = lveroprop2;
  lverok = lveroprop;
  cont6 = cont6 + 1;}}

theta[k,1] = sig2k;
theta[k,2] = phi1k;
theta[k,3] = phi2k;
theta[k,4] = phi3k;
theta[k,5] = nu1k;
theta[k,6] = betak[1,]
...
theta[k,p+2] = betak[q,] # q o numero de coeficientes da regressao
lambda1[k,] = exp(loglamb1k);
lvero[k]=lverok
}

```