

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE CIÊNCIAS EXATAS E DA TERRA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA APLICADA E ESTATÍSTICA

Daniel Matos de Carvalho

***DOWNSCALING* ESTOCÁSTICO PARA EXTREMOS
CLIMÁTICOS VIA INTERPOLAÇÃO ESPACIAL**

Natal, junho de 2010

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

DOWNSCALING ESTOCÁSTICO PARA EXTREMOS CLIMÁTICOS VIA INTERPOLAÇÃO ESPACIAL

Daniel Matos de Carvalho

Dissertação apresentada ao Corpo Docente do Programa de Pós-Graduação em Matemática Aplicada e Estatística - CCET - UFRN, como requisito parcial para obtenção do título de Mestre em Matemática Aplicada e Estatística.

Área de Concentração: Métodos Estatísticos

Orientador: Prof. Dr. Paulo Sérgio Lucio

Natal, junho de 2010

DOWNSCALING ESTOCÁSTICO PARA EXTREMOS CLIMÁTICOS VIA INTERPOLAÇÃO ESPACIAL

Este exemplar corresponde à redação final da dissertação devidamente corrigida, defendida por Daniel Matos de Carvalho e aprovada pela comissão julgadora.

Natal, junho de 2010

Banca examinadora:

- Prof. Dr. Paulo Sérgio Lucio
- Prof. Dr. Paulo Justiniano Ribeiro Junior
- Prof. Dr. Francisco Alexandre da Costa

Resumo

Em diversos segmentos da sociedade, os resultados de previsões meteorológicas têm sido utilizados para a tomada de decisão, exigindo a melhoria da qualidade das previsões de variáveis tanto na escala espacial quanto temporal. Na comunidade científica é aceito que o processo mais adequado de se obter cenários dos fenômenos meteorológicos é através dos Modelos de Circulação Global (MCG). Entretanto, os resultados destas previsões são fornecidos em redes com espaçamentos de nós da grade de algumas centenas de quilômetros. Visando obter soluções para este problema foram desenvolvidas técnicas de pós-processamento de saídas de modelos dinâmicos. Uma delas é chamada de *downscaling*, que nada mais é que uma metodologia de refinamento da escala espacial de um Modelo de Circulação Global. Este trabalho apresenta um método de *downscaling* que proporciona a modelagem de eventos extremos onde os Modelos de Circulação não fornecem informação relevante. Para isto, foram utilizadas a teoria dos valores extremos e o método geoestatístico denominado *krigagem*. Será realizado um estudo de caso utilizando séries de temperatura máxima do ar do modelo de reanálise do NCEP/NCAR - *National Centers for Environmental Prediction/National Center for Atmospheric Research*, que fornece os produtos em uma grade regular de $2,5^\circ \times 2,5^\circ$. Para a validação do método proposto, será realizada uma comparação com os resultados de uma análise de regressão quantílica, método tradicional de *downscaling*. Os ajustes aos modelos GEV e GPD apresentaram-se satisfatórios para as 72 séries de temperatura. A interpolação espacial utilizando a *krigagem* mostrou uma certa discrepância entre a intensidade dos parâmetros de forma para o modelo GEV. Para o método de regressão quantílica, curiosamente, o parâmetro de forma é igual ao estimado através da *krigagem* tanto para o quantil 95% quanto para o quantil 99%. Para os outros parâmetros a média e o limiar foram os que mais se distanciaram dos valores ajustado através dos dados observacionais, com uma diferença de aproximadamente 4° tanto para o limiar quanto para a média. Esta diferença se deve ao fato da menor variabili-

dade espacial com um menor intensidade da média e limiar. Tais discrepâncias podem ser também devido a qualidade da informação fornecida pelos dados do NCEP/NCAR. Levando em consideração que a dificuldade no conhecimento do comportamento da estrutura climática leva a incertezas nas informações dos modelos de circulação, e que as informações fornecidas pelos modelos são representações da média do quadrante da grade do modelo. Podemos concluir que os resultados apresentados são eficientes devido a toda esta problemática.

Palavras-chave: Distribuição Generalizada de Valores Extremos; Distribuição Generalizada de Pareto; Krigagem

Abstract

Present day weather forecast models usually cannot provide realistic descriptions of local and particular extreme weather conditions. However, for lead times of about a small number of days, they provide reliable forecast of the atmospheric circulation that encompasses the subscale processes leading to extremes. Hence, forecasts of extreme events can only be achieved through a combination of dynamical and statistical analysis methods, where a stable and significant statistical model based on prior physical reasoning establishes posterior statistical-dynamical model between the local extremes and the large scale circulation. Here we present the development and application of such a statistical model calibration on the basis of extreme value theory, in order to derive probabilistic forecast for extreme local temperature. The downscaling applies to NCEP/NCAR re-analysis, in order to derive estimates of daily temperature at Brazilian northeastern region weather stations.

Palavras-chave: Generalized Extreme Value, Generalized Pareto Distribution, Krigagem

Agradecimentos

A toda a minha família por todo carinho e apoio.

A minha noiva Emanuella por estar sempre ao lado nos momentos difíceis.

A todos os professores do curso de Matemática Aplicada e Estatística pelos ensinamentos transmitidos.

A todos os amigos do curso Matemática Aplicada e Estatística pelo bom convívio e amizade.

Sumário

| | | |
|----------|--|-----------|
| 1 | Introdução | 1 |
| 1.1 | Organização do Trabalho | 4 |
| 1.2 | Material e Métodos | 4 |
| 2 | <i>Downscaling</i> por Regressão Quantílica | 6 |
| 2.1 | <i>Downscaling</i> | 6 |
| 2.1.1 | Variáveis Predictoras e Preditandos | 7 |
| 2.1.2 | Estratégias de <i>Downscaling</i> | 9 |
| 2.2 | Regressão Quantílica | 9 |
| 2.2.1 | Propriedades | 10 |
| 3 | <i>Downscaling</i> por Interpolação Espacial | 12 |
| 3.1 | Teoria dos Valores Extremos - TVE | 13 |
| 3.1.1 | Distribuição Generalizada dos Valores Extremos | 13 |
| 3.1.3 | Distribuição Generalizada de Pareto | 19 |
| 3.2 | Análise Espacial | 22 |
| 3.2.1 | Geoestatística | 23 |
| 3.2.2 | Variograma | 25 |
| 3.2.3 | Krigagem | 28 |
| 4 | Resultados | 31 |
| 4.1 | Análise Exploratória dos Dados | 31 |
| 4.1.1 | Dados de Reanálise - NCEP/NCAR | 31 |
| 4.1.2 | Dados Observacionais - INMET | 32 |

| | | |
|----------|---|-----------|
| 4.2 | Teoria dos Valores Extremos | 33 |
| 4.2.1 | Análise Via Distribuição Generalizada de Valores Extremos - GEV | 34 |
| 4.2.2 | Análise via Distribuição Generalizada de Pareto - GPD | 34 |
| 4.3 | Análise Geoestatística | 36 |
| 4.3.1 | Análise Exploratória dos Dados | 36 |
| 4.3.2 | Estimação dos Parâmetros | 40 |
| 4.3.3 | Krigagem | 43 |
| 4.3.4 | Análise de Resíduos | 44 |
| 4.4 | Regressão Quantílica | 45 |
| 4.4.1 | TVE Aplicada aos Quantis 99% e 95% | 48 |
| 4.5 | Comparações dos Resultados | 54 |
| 5 | Considerações Finais | 55 |
| | Referências Bibliográficas | 57 |
| A | Programas | 63 |
| B | Gráficos | 77 |

Capítulo 1

Introdução

Em diversos segmentos da sociedade, os resultados de previsões meteorológicas têm sido utilizados para a tomada de decisão, exigindo a melhoria da qualidade das previsões de variáveis tanto na escala espacial quanto temporal. Na comunidade científica é aceito que o processo mais adequado de se obter cenários dos fenômenos meteorológicos é através dos Modelos de Circulação Global (MCG). Os MCG consistem em formulações matemáticas dos processos atmosféricos e das superfícies terrestre e oceânica, baseadas nos princípios físicos clássicos da hidrodinâmica. Tais modelos constituem o único meio disponível para consideração simultânea da grande quantidade de processos físicos interativos que caracterizam o sistema climático. Entretanto, os resultados gerados por tais modelos climáticos são apresentados por grades regulares com baixa resolução, ou seja, grades com redes esparsas, com espaçamentos típicos entre os nós da rede de algumas centenas de km. Em algumas situações é necessário obter uma informação mais detalhada no tempo e no espaço dos produtos de previsão numérica de clima. Visando obter soluções para este problema foram desenvolvidas técnicas de pós-processamento de saídas de modelos dinâmicos. Uma delas é chamada de *downscaling*, que nada mais é que uma metodologia de refinamento da escala espacial de um Modelo de Circulação Global. Ela também é conhecida como uma ferramenta de interpolação de variáveis em escala espacial para a escala pontual, por exemplo, através da obtenção de uma relação entre os dados do MCG e os dados de uma estação meteorológica, pode-se obter

o cenário futuro sobre a estação. Na literatura são encontrados diversos estudos (Benestad, 2007; Galvão, 1999; Ramos, 2000) que utilizam o *downscaling* para a previsão de variáveis meteorológicas. Entretanto, em determinadas variáveis, a aplicação do método não agrega a variabilidade da intensidade ou até mesmo a sua continuidade em toda região de estudo, por exemplo, a precipitação. Visando solucionar este problema, este trabalho tem como objetivo introduzir a teoria dos valores extremos e o método geoestatístico denominado *krigagem* como forma de *downscaling*, realizando um estudo de caso de temperatura do ar no Nordeste Brasileiro - NEB.

Nas últimas décadas estão sendo registrados diversos eventos extremos, com consequências catastróficas para a sociedade tais como: enchentes, furacões, tsunamis, incêndios florestais entre outros (Hann, 2003; Coles, 2003; Hosking, 1985). Estudos que estimem as probabilidades de ocorrência destes eventos são de primordial importância para que planos de proteção sejam desenvolvidos. Estas estimativas são obtidas, geralmente, através da Teoria dos Valores Extremos (Fisher e Tippett, 1928). Em certas pesquisas científicas é necessário realizar estudos acerca das caudas das distribuições de variáveis aleatórias (Joyce, 2008). Portanto, é de grande importância a definição de um critério que caracterize as observações que são extremas (Beijo, 2005; Chowdhury, 1991). A modelagem estatística sobre acontecimentos extremos deve ser baseada no critério definido sobre tais observações. Na literatura são encontradas diferentes formas para definir observações extremas e sua respectiva forma de modelagem. Por exemplo, a modelagem associada a distribuição conjunta de r estatísticas de ordem (Smith, 1986); a análise de valores por meio de modelos baseados em processos pontuais proposta por Smith (1989); por fim os mais difundidos, a metodologia clássica de Gumbel, que extrai o máximo por blocos; e a metodologia baseada na modelagem de valores acima de um certo limiar, denominada Distribuição Generalizada de Pareto. Apesar da larga utilização atual da teoria dos valores extremos esta é uma teoria que, segundo Bautista (2002), iniciou-se por volta de 1928 com o trabalho de Fisher e Tippett, determinando os três tipos possíveis de distribuições assintóticas dos valores extremos. Assim como a grande maioria das metodologias da teoria estatística apresentam suposições

iniciais para a construção de um modelo, a teoria dos valor extremos não é diferente. Com a existência de diferentes metodologias é de grande importância que o pesquisador tenha as especificações e pressuposições requeridas pelo modelo, pois o não atendimento destas necessidades leva à inconsistência dos resultados inviabilizando assim o objetivo fixado. A grande utilidade da aplicação da teoria do valor extremo em variáveis meteorológicas é a possibilidade de conhecer o comportamento das ocorrências de eventos extremos. Diante do exposto, a contribuição dessa dissertação consiste na tentativa de aprimoramento da metodologia *downscaling* no estudo de eventos extremos, através da utilização da teoria dos valores extremos e métodos geoestatísticos. O interesse principal da abordagem proposta consiste na obtenção da distribuição de eventos extremos onde os MCG não fornecem informação relevantes. Para o estudo de caso serão utilizados dados de temperatura do ar do modelo de reanálise do *National Centers for Environmental Prediction/National Center for Atmospheric Research* - NCEP/NCAR e a região considerada será o Nordeste Brasileiro - NEB. Em resumo a abordagem consiste nos seguintes passos: 1 - Selecionado o retângulo envolvente como os pontos da grade do NCEP/NCAR que envolvem o NEB, será ajustada uma distribuição de eventos extremos para cada série histórica de temperatura do ar correspondente a cada ponto da grade. O ajuste produzirá estimativas dos parâmetros da Distribuição Generalizada de Valores Extremos - GEV (Generalized Extreme Value) e a Distribuição Generalizada de Pareto - GPD (Generalized Pareto Distribution) para cada série; 2 - Considerando os parâmetros das distribuição Generalizada de Valores Extremos e de Pareto variáveis contínuas no espaço, podemos considerar as estimativas produzidas com ajuste amostras destes parâmetros sobre o espaço. O georeferenciamento das estimativas permite a aplicação das técnicas de geoestatística, mais especificamente a interpolação espacial; 3 - Aplicar a *krigagem* para cada amostra das estimativas dos parâmetros e assim, obter a distribuição de eventos extremos em localizações onde o NCAR/NCEP não fornece informação. A validação desta abordagem será realizada através do seguinte processo: 1 - Ajustar a Distribuição Generalizada de Valores Extremos e a Distribuição Generalizada de Pareto aos dados de temperatura máxima

do ar observados da estação de meteorológica do Recife; 2 - Comparar os com a distribuição ajustada através da interpolação espacial. 3 - Finalmente, obter um modelo de regressão quantílica como método tradicional de *dowscaling*.

1.1 Organização do Trabalho

No Capítulo 2, apresentamos os princípios tradicionais de *dowscaling* e revisão bibliográfica da regressão quantílica. No capítulo 3 apresentamos a revisão bibliográfica dos métodos utilizados para o *dowscaling* via interpolação espacial: teoria dos valores extremos e *krigagem*. No Capítulo 4 os principais resultados obtidos são apresentados para as duas formas de *dowscaling*. Finalmente, o Capítulo 5 descreve as considerações finais e as comparações dos resultados obtidos nos capítulos anteriores.

1.2 Material e Métodos

Devido a grande dificuldade de obter séries de dados meteorológicos para o planeta como um todo, vários centros de previsão de tempo e clima vêm oferecendo produtos de reanálise. Estes produtos são definidos como séries de dados meteorológicos obtidos através da assimilação e reanálise de dados observados em todo o planeta (Pinto, 2006). O NCEP/NCAR (*National Center for Environmental Prediction / The National Center for Atmospheric Research*) fornece séries de variáveis meteorológicas de reanálise de 1948 até os dias atuais, podendo ser encontrados no endereço eletrônico <http://www.esrl.noaa.gov/psd/data/gridded/reanalysis/>. Os dados são fornecidos em uma grade regular sobre o globo com resolução $2,5^\circ$ de latitude por $2,5^\circ$ de longitude (Figura 1.1). Para cada ponto da grade existe uma série histórica da variável climática de interesse. Na literatura, os dados de reanálise são considerados os dados mais próximos dos reais. Devido a estas considerações os dados de reanálise do NCEP/NCAR de temperatura do ar serão utilizados para o estudo de caso. A Figura 1.1 apresenta os 72 pontos da grade do NCEP que serão utilizados neste estudo, envolvendo todo o Nordeste Brasileiro. Os dados observados foram obtidos no Instituto Nacional de

Meteorologia - INMET órgão federal de administração direta do Ministério da Agricultura, Pecuária e Abastecimento responsável em prover informações meteorológicas. O período dos dados fornecidos pelo INMET corresponde a 1 de janeiro de 1951 a 31 de dezembro de 2008.

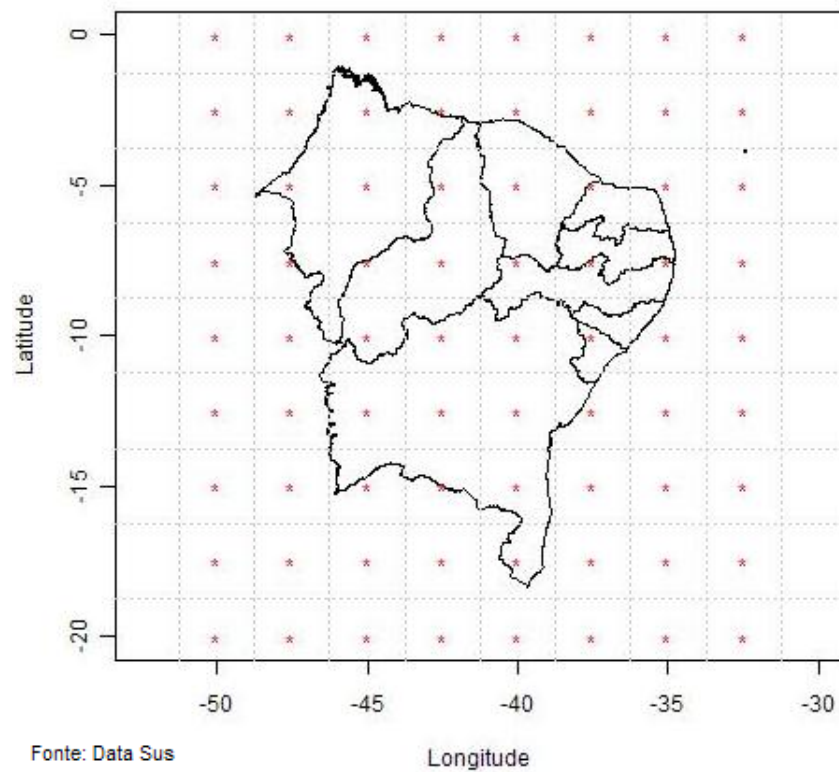


Figura 1.1: Grade do NCEP sobre o Nordeste Brasileiro.

Para o desenvolvimento desta dissertação foi utilizado o *software* R obtido no endereço eletrônico www.r-project.org (R DEVELOPMENT CORE TEAM, 2009).

Capítulo 2

Downscaling por Regressão Quantílica

Neste capítulo serão apresentados alguns aspectos das técnicas de pós-processamentos de modelos de circulação, conhecidos como *downscaling* e da regressão quantílica. Todas estas abordagens estatísticas estão muito bem desenvolvidas na literatura. Sendo assim, neste capítulo serão apresentados apenas os principais conceitos.

2.1 *Downscaling*

Cenários de mudanças climáticas são representações plausíveis do futuro sendo consistentes com as tendências de emissão de gases de efeito estufa e outros poluentes, e com o entendimento científico atual sobre o efeito desses gases sobre o clima global (IPCC-TGCI, 1999). Os MCG são ferramentas poderosas, com grande potencial para o estudo dos cenários de mudanças climáticas, descrevendo o clima utilizando *grid* tridimensional sobre o planeta, tendo resolução global variando entre 250 e 600 km. Assim, os MCG oferecem informações sobre mudanças de clima de grande utilidade para a escala continental.

Devido à baixa resolução horizontal da malha desses MCG, variáveis relevantes às escalas local que influenciam o clima de uma região, são mal representadas levando a resultados irrealistas se aplicados diretamente em pequena escala. Para solucionar este

problema surgiram técnicas de transferência de informação meteorológicas para as escalas menores chamadas técnicas de *downscaling*, também chamadas de desagregação espacial ou refinamento estatístico. As técnicas de *downscaling* se classificam em dois grupos (Hewiston e Crane, 1996):

1. Métodos dinâmicos, que são modelos numéricos semelhantes aos MCG, com resolução mais fina, usando como dados iniciais advindos de um MCG (Benestad, 2007; Giorgi, 1990).
2. Métodos estocásticos, que transferem a previsão meteorológica fornecida pelo MCG estatisticamente (Benestad, 2007).

Os métodos dinâmicos consistem da utilização de aninhamentos de modelos numéricos. São modelos climáticos de mesoescala ou regionais que, inicializados com os resultados de um MCG, descrevem melhor as condições climáticas de uma região. Esses modelos podem apresentar ainda a peculiaridade de possuir várias grades de espaçamentos diferentes, que podem ser aninhadas, proporcionando resultados de alta resolução (Pielke, 1992). O refinamento de variáveis meteorológicas através de métodos estatísticos parte da suposição de que relações empíricas estáveis podem ser estabelecidas entre processos atmosféricos que ocorrem em escalas espaciais e/ou temporais diferentes (Wilby, 1998). Trata-se de uma abordagem que oferece soluções em curto espaço de tempo, e que não necessitam de grande poder computacional, constituindo uma solução viável para os pesquisadores da área. A principal limitação desta abordagem é que o modelo de *downscaling* fica restrito a uma região e às séries de dados utilizados. Com isto, os métodos estocásticos tem sido largamente utilizados por ser um método viável de baixo custo computacional.

2.1.1 Variáveis Predictoras e Preditandos

Downscaling estocástico de cenários de clima, consiste em identificar relações entre as variáveis de larga escala (predictoras) e as variáveis de pequena escala (preditando), e então aplicar tal relação nas saídas dos dados de modelos de clima (Hanssen e Bauer,

2005). Uma hipótese fundamental para estabelecer estas aproximações é que relações podem ser estabelecidas entre processos atmosféricos, sendo que estas correm em diferentes escalas de espaço e tempo (Wilby e Wigley, 1997). Matematicamente, a relação entre as variáveis de grande e pequena escala é representada da seguinte forma: $Y = f(X)$ onde, X representa as variáveis predictoras definidas como os dados de entrada do modelo, Y corresponde ao preditando definidos como os dados de saída do modelo e f representa uma função de transferência estatística, matemática ou física. As variáveis predictoras caracterizam-se por serem de larga escala e descreverem o regime de circulação sobre a região de interesse, enquanto que o preditando caracteriza-se por ser de pequena escala, por exemplo, a temperatura ou precipitação de uma estação climática. A Figura 2.1 apresenta a estruturação básica de aplicação dos métodos tradicionais de *downscaling* estocástico.

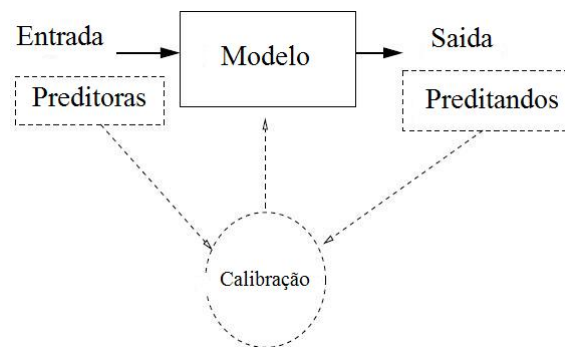


Figura 2.1: Estrutura Básica de *Downscaling* Estocástico

A aplicação do *downscaling* Estocástico pode ser resumida em alguns passos (Hewiston e Crane, 1996):

1. Seleção do domínio atmosférico e dos dados locais;
2. Redução dos dados atmosféricos;
3. Validação do MCG;
4. Suavização temporal e espacial dos dados de grande e pequena escalas;

5. Ponderação histórica dos dados;
6. Obtenção da função de transferência, $f(X)$, entre os preditores e preeditandos;
7. Teste da acurácia do *downscaling*;
8. Aplicação do modelo sobre cenários de mudanças climáticas.

2.1.2 Estratégias de *Downscaling*

Os métodos estatísticos mais largamente utilizados na literatura são os métodos lineares de regressão multivariada, análise de correlação canônica, decomposição de vetores singulares; e os métodos não-lineares modelo analógicos, métodos de classificação e redes neurais. Entre os métodos estatísticos de regressão multivariada se destacam: *perfect prog method* - PPM (Klein, 1958), *model output statistics* - MOS (Glahn e Lowry, 1972) e *model output calibration* - MOC (Mao, 1998). Recentemente o método de regressão quantílica e a teoria de valores extremos foram utilizados como ferramenta de *downscaling* (Friederichs, 2010; Friederichs, 2007).

Todas as técnicas de *downscaling* descritas são lineares, sendo que a relação entre as características de escala sinótica ou climáticas e campos observados não apresentam geralmente tal comportamento (Ramirez, 2003). Assim, a seleção do método deve levar em consideração as relações entre as variáveis de interesse.

2.2 Regressão Quantílica

A regressão quantílica consiste em uma extensão natural do modelo de regressão linear. Enquanto que o modelo de regressão linear trata a média condicional da variável dependente associada com as mudanças em suas co-variáveis, o modelo de regressão quantílica trata as mudanças em uma função quantílica condicional. Koenker e Basset (1978) listam algumas vantagens na utilização nos modelos de Regressão Quantílica:

1. A técnica de regressão quantílica permite caracterizar toda distribuição condicional de uma variável resposta a partir de um conjunto de regressores;

2. A regressão quantílica pode ser usada quando a distribuição não é gaussiana;
3. A regressão quantílica é robusta a *outliers*.

O trabalho apresentado por Bassett e Koenker em 1978, sobre regressão quantílica, pode ser visto como uma tentativa de estender para o modelo linear a idéia de quantil amostral. Um caso particular importante de regressão quantílica é a regressão L_1 , isto é, a minimização dos desvios absolutos (Pinheiro, 1992). No modelo de posição o estimador de norma L_1 é a mediana, que sabemos ser um estimador melhor do que a média, quando os dados provêm de distribuições com caudas alongadas ou quando temos *outliers* no conjunto dos dados. A seguir iremos apresentar alguns dos seus principais resultados. Basicamente o modelo de regressão quantílica será ajustado através da obtenção de uma relação (calibração, Figura 2.1) entre a série de temperatura máxima do ar do modelo de reanálise do NCEP/NCAR (variável preditora) mais próxima da estação do Recife do NCEP/NCAR, e a série de temperatura máxima do ar observada no Recife (Preditandos). Através desta calibração o modelo de regressão quantílica é estimado e o *downscaling* dos quantis dos dados de reanálise do NCEP/NCAR podem ser obtidos.

2.2.1 Propriedades

Definição 2.2.2. *Seja $(y_i : i = 1, \dots, n)$ uma amostra da variável aleatória Y , com distribuição F . O θ -ésimo quantil, $0 < \theta < 1$ amostral pode ser definido como solução do problema de minimização*

$$\underbrace{\min}_{\tilde{\beta} \in \mathfrak{R}} \left[\sum_{i \in (i: y_i \geq \tilde{\beta})} \theta |y_i - \tilde{\beta}| + \sum_{i \in (i: y_i < \tilde{\beta})} (1 - \theta) |y_i - \tilde{\beta}| \right] \quad (2.1)$$

onde $0 < \theta < 1$. Para o caso que $\theta = \frac{1}{2}$ o quantil amostral de ordem $\frac{1}{2}$ é a mediana amostral. Visto que a regressão quantílica é uma generalização do modelo linear definiremos o modelo de regressão como:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}$$

sendo $\mathbf{Y} = (y_1, \dots, y_n)$, β é um vetor de parâmetros desconhecidos, \mathbf{X} é uma matriz $n \times p$ e \mathbf{Z} é um vetor de variáveis aleatórias independentes e identicamente distribuídas com função de distribuição F_z .

Definição 2.2.3. *A θ -ésima regressão quantílica, $0 < \theta < 1$, é definida como qualquer solução para o problema de minimização:*

$$\underbrace{\min}_{\tilde{\beta} \in \mathbb{R}^k} \left[\sum_{i \in (i: y_i \geq x_i \tilde{\beta}_\theta)} \theta |y_i - x_i \tilde{\beta}_\theta| + \sum_{i \in (i: y_i < x_i \tilde{\beta})} (1 - \theta) |y_i - x_i \tilde{\beta}| \right] = \underbrace{\min}_{\tilde{\beta}_\theta} \sum_{i=1}^n \rho_\theta(y_i - \tilde{\beta}_\theta x_i) \quad (2.2)$$

sendo β é um vetor coluna de dimensão $p \times 1$ e ρ é uma função de ponderação dada por:

$$\rho_\theta(u) = \begin{cases} u\theta & , u \geq 0 \\ (\theta - 1)u & , u < 0 \end{cases}$$

O parâmetro θ pondera a regressão localizando-a no conjunto de dados de forma que ao menos 100% das observações encontra-se acima da regressão quantílica e ao menos $100(1 - \theta)$ das observações encontra-se abaixo da regressão. O resultado da Equação 2.2 é denotado por $\hat{\beta}$, do qual é retirado o θ^{th} quantil condicional $Q_{Y|X}(\theta) = X\beta_\theta$. As formas matemáticas decorrentes da regressão quantílica são muito diferentes dos que surgem no método dos mínimos quadrados. O método dos mínimos quadrados conduz a uma análise dos problemas de um espaço de produto interior, envolvendo projeção para subespaços e, assim, o problema de minimizar os erros ao quadrado pode ser reduzido a um problema numérico de álgebra linear. A regressão quantílica não tem esta estrutura, e em vez disso leva a problemas de programação linear que podem ser resolvidos, por exemplo, pelo método simplex.

Capítulo 3

Downscaling por Interpolação Espacial

A contribuição central desta dissertação consiste na obtenção do ajuste da distribuição de eventos extremos de dados de Modelos de Circulação Global, onde o mesmo não fornece informação. Para alcançarmos este objetivo pretende-se aplicar os seguintes procedimentos:

1. Após a obtenção dos dados de Temperatura do ar do NCEP/NACR, iremos ajustá-los a Distribuição Generalizada de Valores Extremos e a Distribuição de Generalizada de Pareto para cada série de Temperatura máxima do ar. Com isto, teremos as estimativas dos parâmetros destas distribuições para cada série de temperatura do ar.
2. Utilizando a técnica de interpolação espacial denominada *krigagem* iremos obter a superfície interpolada de cada parâmetro estimado.
3. A validação dos resultados será realizada comparando as estimativas produzidas através do ajuste dos extremos de temperatura do ar observados da estação meteorológica localizada no Recife.
4. A validação do método será realizada utilizando a regressão quantílica.

3.1 Teoria dos Valores Extremos - TVE

O objetivo desta seção é a introdução da teoria dos valores extremos - TVE, que possui grande importância na modelagem de eventos extremos. Para isto serão apresentados definições e conceitos para o melhor entendimento do teorema central da TVE: o Teorema de Fisher e Tippet (1928).

3.1.1 Distribuição Generalizada dos Valores Extremos

Seja X_1, X_2, \dots uma sequência de variáveis aleatórias (v.a) independentes e identicamente distribuídas (i.i.d.) com função de distribuição F_X , e seja M_n o máximo de n variáveis aleatórias, ou seja, $M_n = \max(X_1, \dots, X_n)$. A variável aleatória X é representada por uma série de valores no tempo, por exemplo, o registro diário do nível do oceano ou a temperatura do ar em uma determinada região. Com isto, M_n representa o máximo de n observações em um dado período de tempo. Sob a suposição de independência e igualdade de distribuição, a função distribuição do máximo, $F_{M_N}(x)$, é dada por:

$$\begin{aligned}
 P(M_n \leq x) &= P(X_1 \leq x, \dots, X_n \leq x) \\
 &= P(X_1 \leq x) \times \dots \times P(X_n \leq x) \\
 &= [P(X \leq x)]^n \\
 &= [F(x)]^n
 \end{aligned} \tag{3.1}$$

para $x \in \mathbb{R}$ $n \in \mathbb{N}$

Na prática a obtenção da distribuição do máximo não é simples devido a alguns problemas:

1. A distribuição F_X não é conhecida.
2. A distribuição do máximo diverge - (3.2).

Um solução viável seria a estimação de F_X através de dados observados. Entretanto, pequenas discrepâncias nas estimativa de F_X levam a substâncias discrepâncias

em $[F(x)]^n$ (Coles, 2001). Uma alternativa para este problema seria aceitar o desconhecimento da distribuição de F_X e obter famílias aproximadas de modelos para o máximo. Esta alternativa tem como analogia o Teorema Central do Limite.

O comportamento assintótico de M_n está relacionado com a cauda de F_X próximo do limite superior do suporte da distribuição de X , pois os valores extremos se localizam próximos deste limites. A abordagem clássica da teoria dos valores extremos consiste em caracterizar cauda superior (ou inferior) da distribuição de F_X a partir da distribuição do máximo (ou mínimo). Seja x_{F_X} o limite do suporte da distribuição de F_X , ou seja, $x_{F_X} = \sup \{x \in \mathfrak{R} : F(x) < 1\}$. Desta maneira quando $n \rightarrow \infty$ temos:

$$\lim_{n \rightarrow \infty} F^n(x) = \begin{cases} 1 & , x \geq x_{F_X} \\ 0 & , x < x_{F_X} \end{cases} \quad (3.2)$$

no caso que $x_{F_X} < \infty$. Este comportamento não fornece informação relevante. Com isto, estudos assintóticos para o máximo seriam de grande utilidade.

Definições

Nesta seção será introduzido o Teorema de Fisher-Tippett e as definições de variáveis aleatórias max-estáveis que são leis limites do máximo para a identificação da distribuição não-degenerada G . Os teoremas não serão demonstrados aqui, entretanto suas demonstrações podem ser encontradas em (Gnedenko, 1943; Viola, 2006).

A teoria dos valores extremos assegura, sob uma transformação linear, que a distribuição do máximo converge para uma distribuição assintótica não degenerada, isto é, para constantes apropriadas $a_n > 0$ e $b_n \in \mathfrak{R}$ tem-se que:

$$\frac{M_n - b_n}{a_n} \rightarrow G \quad (3.3)$$

ou ainda,

$$\lim_{n \rightarrow \infty} P(M_n \leq a_n x + b_n) = \lim_{n \rightarrow \infty} F_X^n(a_n x + b_n) = G(x) \quad (3.4)$$

Se para constantes apropriadas $a_n > 0$ e $b_n \in \mathfrak{R}$ (3.4) é verificado, F_X pertencerá

ao domínio de atração do máximo associado à distribuição de valores extremos G , denotado por $F_X \in D(G)$, ou seja, a coleção das F_X tais que os máximos associados possuem uma distribuição limite não-degenerada é denominada domínio de atração da G .

Definição 3.1.1.1. *Seja uma sequência de variáveis aleatórias X_1, \dots, X_n i.i.d. e seja X uma variável aleatória não-degenerada i.i.d. com a sequência $\{X_i\}_{i=1}^n$. A distribuição da variável aleatória X é chamada max-estável se existirem constantes apropriadas $a_n > 0$, $b_n \in \mathfrak{R}$ para cada $n \geq 2$ que satisfaça $\max(X_1, \dots, X_n) \stackrel{d}{=} a_n X + b_n$.*

Definição 3.1.1.2. *Duas funções de distribuição G_1 e G_2 são do mesmo tipo se $G_2(x) = G_1(ax + b)$ para algumas constantes $a > 0$ e $b \in \mathfrak{R}$.*

A partir das duas definições acima podemos extrapolar a definição de funções de distribuição max-estáveis.

Definição 3.1.1.3. *Uma função de distribuição G não-degenerada é max-estável se, para cada $n \geq 2$ existirem constantes $a_n > 0$ e $b_n \in \mathfrak{R}$ tais que $G^n(a_n x + b_n) = G(x)$, ou seja, G^n é do mesmo tipo da G .*

Definição 3.1.1.4. *Seja F_X a função de distribuição de uma v.a. . O limite superior x_{F_X} do suporte da distribuição F_X é definido como $x_{F_X} = \sup \{x \in \mathfrak{R}; F_X(x) < 1\}$.*

As definições, acima citas, em conjunto com outros resultados dão base a demonstração ao Teorema de Fisher-Tippett que segue.

Teorema (FISHER-TIPPETT, 1928) 3.1.2. *Seja X_n uma sequência de variáveis aleatórias independentes e identicamente distribuídas. Se existirem sequências de constantes normalizadoras $a_n > 0$ e b_n e uma distribuição não-degenerada G tal que*

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} G$$

então G é do tipo de uma das três funções de distribuição abaixo:

- Tipo I Gumbel:

$$G_I(x) = \exp \left\{ - \exp \left[- \frac{(x - \mu)}{\sigma} \right] \right\}, x \in \mathfrak{R} \quad (3.5)$$

- *Tipo II Fréchet:*

$$G_{III} = \left\{ \begin{array}{ll} 0 & , x \leq \mu \\ \exp \left\{ - \left[\frac{(x-\mu)}{\sigma} \right]^{-\xi} \right\} & , x > \mu \end{array} \right\} \quad (3.6)$$

- *Tipo III Weibull*

$$G_{III}(x) = \left\{ \begin{array}{ll} \exp \left\{ - \left[- \left(\frac{x-\mu}{\sigma} \right) \right]^\xi \right\} & , x < \mu \\ 1 & , x \geq \mu \end{array} \right\} \quad (3.7)$$

onde $\mu \in \mathfrak{R}$, $\sigma > 0$ e $\xi > 0$

O Teorema de Fisher-Tippett afirma que a estabilização do máximo, através de sequências apropriadas, convergem em distribuição para uma das três distribuições citadas, e que, pode-se estimar a distribuição assintótica de $\frac{M_n - b_n}{a_n}$ diretamente da família G sem fazer qualquer referência à distribuição de X , pois a distribuição G é de um dos três tipos. Conforme mencionado anteriormente a prova não será apresentada aqui, entretanto, provas detalhadas podem ser encontradas em Gnedenko (1943). As três distribuições possuem diferentes comportamentos caudais. Esta afirmação pode ser verificada a partir do limite do suporte das distribuições. Para a distribuição *Weibull* x_{G_X} é finito, enquanto para as distribuições Fréchet e Gumbel o limite superior do suporte é infinito. Entretanto, a densidade da distribuição Gumbel decai exponencialmente, enquanto que para a distribuição Fréchet o decaimento é polinomial. Apesar destas diferenças do comportamento caudal, as três distribuições de valores extremos estão fortemente relacionadas (Cerezer, 2008). Seja $X > 0$ então:

$$X \sim G_{II}(x) \Leftrightarrow \ln(X^\xi) \sim G_I(x) \Leftrightarrow -X^{-1} \sim G_{III}(x)$$

O Teorema Central do Limite - TCL é uma das ferramentas mais importantes na teoria da probabilidade e estatística, afirmando que a distribuição normal é a única distribuição da soma de variáveis aleatórias sob certas condições (Silva, 2008). Analogamente, a teoria dos valores extremos afirma que existem três tipos de distribuição para o limite da distribuição do máximo (Hann, 2003). A comparação entre as condições

dos teoremas pode ser visualizada na Tabela 3.1.

Tabela 3.1: Analogia entre TCL e Teorema de Fisher e Tippet

| Analogia | TLC | TFT |
|---------------------|--------------------|-------------------------|
| Suposições | iid com var finita | iid |
| Objeto de Estudo | S_N (Soma) | M_n (Máximo) |
| Distribuição Limite | Normal | $G(x)$ (Não Degenerada) |

Em aplicações da teoria dos valores extremos seleciona-se uma das três distribuições de valores extremos e realiza-se a estimação dos parâmetros. Entretanto, este procedimento apresenta problemas práticos: 1 - uma técnica é requerida para selecionar qual das três famílias de distribuições é mais adequada para os dados em questão; 2 - A partir da decisão tomada inferências subseqüentes presumem que esta escolha está correta não levando em consideração incertezas acerca da seleção da família.

Visando solucionar este problema Jenkinson (1955) apresentou uma importante contribuição no desenvolvimento inicial da teoria dos valores extremos. Neste trabalho o autor mostrou que as funções de distribuição acumulada de valores extremos, apresentadas por Fisher e Tippet (1928) são casos particulares da Distribuição Generalizada dos Valores Extremos (Generalized Extreme Value - GEV). As três classes de distribuições podem ser reunidas em uma única família de distribuições, tendo função de distribuição de probabilidade acumulada da forma:

$$G(x) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \quad (3.8)$$

definida em $\{x : 1 + \xi \left(\frac{x - \mu}{\sigma} \right) > 0\}$ $\mu, \xi \in \Re$ e $\sigma > 0$, sendo μ, σ e ξ os parâmetros de locação, escala e de forma, respectivamente. Esta é a família de distribuição generalizada de valores extremos (GEV). Para $\xi < 0$ temos a família da distribuição Weibull, para $\xi > 0$ temos a família da distribuição Fréchet. Como a função não está definida para $\xi = 0$ temos para $\xi \rightarrow 0$ a família da distribuição de Gumbel com função de distribuição de probabilidade acumulada dada pela Eq. 3.9.

$$G(x) = \exp \left[- \exp \left\{ - \left(\frac{x - \mu}{\sigma} \right) \right\} \right] x \in \mathfrak{R} \quad (3.9)$$

Com esta unificação a inferência pode ser realizada diretamente sobre o parâmetro ξ , ou seja, não é necessário escolher uma família inicial evitando assim o problema de sua melhor escolha. Esta abordagem dos valores extremos permite modelar a distribuição para o máximo coletado em blocos de tamanho n . Como exemplo, consideremos a variável aleatória temperatura do ar em uma determinada estação meteorológica de Natal-RN, fornecendo dados diários de 1948 a 2008. O ajuste a distribuição GEV será realizada considerando cada ano como sendo os blocos, ou seja, serão considerados 61 blocos de 365 dias (ou 366 dias para os anos bissexto). Considerando os máximo em blocos como sendo uma nova amostra, estimativas dos parâmetros podem ser obtidas a partir dos métodos mais adequados (método dos momentos, métodos bayesianos e método da máxima verossimilhança). A dificuldade dos métodos de máxima verossimilhança aplicados à distribuição GEV está relacionada com as condições de regularidade que são requeridas para as propriedades assintóticas associadas ao estimador de máxima verossimilhança. Assim, inicialmente, com o aparecimento das três famílias GEV, devido aos limites da distribuição dependerem de seus parâmetros, as condições de regularidade para a estimação pelo método da máxima verossimilhança não são necessariamente satisfeitas, tal como acontece nas distribuições Lognormal, Weibull e Gama de três parâmetros. Assim através de um estudo cuidadoso, obtiveram-se os seguintes resultados (Coles, 2001):

1. $\xi > -0,5$ os estimadores de máxima verossimilhança são completamente regulares;
2. $-1 < \xi < -0,5$ os estimadores de máxima verossimilhança, existem, mas não são regulares;
3. $\xi < -1$ os estimadores de máxima verossimilhança não existem.

Apesar dos problemas que podem ocorrer quando $\xi < -1$ esta situação é extrema-

mente rara para dados ambientais, e correspondem a distribuições com cauda superior muito curta e finita. De uma forma geral, quando se trabalha com dados reais os valores do parâmetro de forma geralmente encontram-se no intervalo $-1 < \xi < -0,5$.

3.1.3 Distribuição Generalizada de Pareto

Na seção anterior foi mostrada a modelagem de extremos utilizando a Distribuição Generalizada de Valores Extremos. Entretanto, devido a sua construção, um bloco pode conter mais eventos extremos que outro (Figura 3.1a). Assim, surge como alternativa para este problema os *Thresholds Models*, ou seja, modelos onde os extremos são definidos levando em consideração limiares 3.1b.

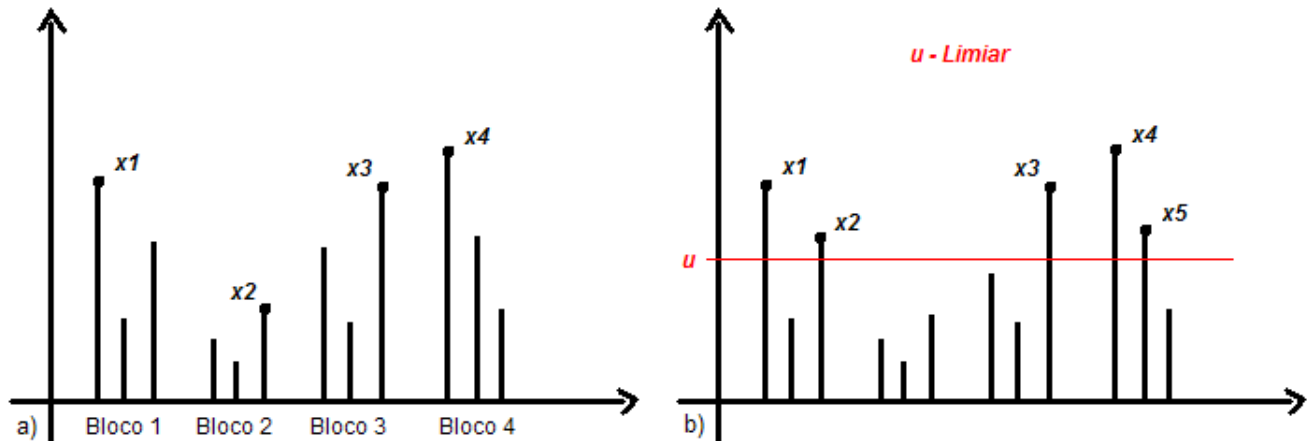


Figura 3.1: Representação dos Eventos Extremos para a GEV - *Generalized Extreme Value* a) e GPD - *Generalized Pareto Distribution* b)

Caracterização do Modelo Assintótico

Seja X_1, X_2, \dots uma sequência de variáveis aleatórias independentes e identicamente distribuídas, possuindo função de distribuição dada por F . Caracterizando um valor extremo como sendo os valores de X_i que extrapolarem um certo limiar u , ou seja, $Y = X - u$, temos interesse em determinar a função de distribuição dos excessos. Tal distribuição é determinada da seguinte forma:

$$P(X > u + y | X > u) = \frac{1 - F_X(u + y)}{1 - F_X(u)}, y > 0 \quad (3.10)$$

O conhecimento da distribuição de F implica no conhecimento da distribuição dos excessos. Entretanto, na prática, esta distribuição não é conhecida. Utilizando o teorema de Fisher-Tippett aproximações da distribuições dos excessos podem ser determinadas (Coles, 2001).

Teorema Generalized Pareto Distribution - GPD 3.1.4. *Seja X_1, X_2, \dots um sequência de variáveis aleatórias independentes e identicamente distribuídas com função de distribuição F , e seja,*

$$M_n = \max(X_1, \dots, X_n)$$

Considerando um termo qualquer da sequência X_i por X , e supondo que $F_X \in D(G)$, então para um n grande temos,

$$P(M_n \leq z) \approx G(x)$$

onde

$$G(x) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}$$

para $\mu, \sigma > 0$ e $\xi \in \mathfrak{R}$. Então, para u (limiar) elevado a função de distribuição de $Y = X - u$ condicionada a $X > u$ é aproximadamente:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma} \right)^{-\frac{1}{\xi}} \quad (3.11)$$

A família de distribuições definidas pela Eq.3.11 é denominada Família Generalizada de Pareto (Coles, 2001). O teorema acima afirma que se os máximos em blocos possuem um distribuição não degenerada G , então a distribuição dos excessos seguem a família generalizada de Pareto. Assim como a GEV associa a três distribuições limites do máximo, a GPD associa três distribuições limites para o excessos (Hann, 2003). A concordância entre GEV e GPD se deve ao fato do parâmetro de forma ξ é determinante no comportamento das distribuições generalizadas. Para a GPD se $\xi < 0$ temos a

distribuição Beta se $\xi > 0$ temos distribuição Pareto II para $\xi \rightarrow 0$ temos a distribuição Exponencial.

Na modelagem via GEV o pesquisador deve definir os tamanhos dos blocos para a estruturação dos máximos e ajuste da distribuição limite. No caso da GPD existe a necessidade de se definir o limiar que melhor representará a distribuição dos excessos. A escolha do limiar é semelhante a escolha do tamanho do bloco na modelagem GEV, onde deve-se ponderar entre vício e variabilidade. Escolhendo um limiar muito baixo corre-se o risco de violar o pressuposto assintótico; enquanto que limiares elevados geram poucos eventos extremos dificultando a estimação do modelo tendenciando a uma alta variância. Assim, existe a necessidade de um método que garanta um balanceamento entre vício e variabilidade. Na literatura o método mais utilizado para a seleção do limiar é denominado *Mean Residual Life Plot*, Gráfico da Média dos Excessos.

Gráfico da Média dos Excessos

Se Y segue a distribuição generalizada de Pareto com parâmetros σ e ξ , então

$$E(Y) = \frac{\sigma}{1 - \xi}, \quad \xi < 1 \quad (3.12)$$

Quando $\xi \geq 1$ a média é infinita. Seja X_1, \dots, X_n um sequência de variáveis aleatórias i.i.d. supondo que a distribuição generalizada de Pareto é um modelo válido para o seu excessos, dado um limiar u_0 . Pela Eq. 3.12 pode-se representar a média dos excessos da seguinte forma:

$$E(X - u_0 | X > u_0) = \frac{\sigma_0}{1 - \xi}$$

Se este modelo considerado é válido para u_0 então é válido também para todo $u > u_0$.

Assim, para $u > u_0$ temos:

$$\begin{aligned} E(X - u | X > u) &= \frac{\sigma_u}{1 - \xi} \\ &= \frac{\sigma_{u_0} + \xi u}{1 - \xi} \end{aligned} \quad (3.13)$$

De acordo com Eq.3.13 o valor esperado é linear em u . Assim, estimando empiricamente o valor esperado dos excessos para cada u e produzindo um gráfico de dispersão, espera-se que o valor do limiar mais adequado para o modelo apresente um comportamento aproximadamente linear. A partir da seleção do limiar todos os métodos clássicos de inferência estatística podem ser aplicados sobre os excessos.

3.2 Análise Espacial

Em diversas áreas do conhecimento, seja em saúde, em meio ambiente, em geologia, em agronomia entre outras, existe a necessidade de compreender a distribuição espacial de fenômenos ocorridos no espaço. O geoprocessamento utiliza técnicas estatísticas, matemáticas e computacionais para o tratamento das informações georeferenciadas (Druck, 2004). As ferramentas computacionais para Geoprocessamento, chamadas de Sistemas de Informação Geográfica (SIG), permitem realizar análises complexas, ao integrar dados de diversas fontes e ao criar bancos de dados geo-referenciados. Além da compreensão da distribuição espacial do problema, deve-se considerar também os padrões existentes com considerações objetivas e mensuráveis, tais como:

- Epidemiologistas coletam dados sobre ocorrência de doenças. A distribuição dos casos forma um padrão no espaço? Existe associação com alguma fonte de poluição? Evidência de contágio? Variou no tempo?
- Deseja-se investigar se existe alguma concentração espacial na distribuição de homicídios por arma de fogo em uma determinada região (Carvalho, 2007).

Todos estes problemas podem ser respondidos utilizando a análise espacial de dados geográficos. O objetivo da análise espacial é mensurar propriedades e relacionamentos, levando em conta a localização espacial do fenômeno em estudo de forma explícita, englobando basicamente três tipos de dados:

- Eventos Pontuais - Fenômenos expressos por ocorrências identificadas como pontos localizados no espaço, denominados processos pontuais. Ex - Localizações de

ocorrências de doenças, crimes e espécies vegetais.

- Dados de Superfícies Contínuas - Dados de amostra de campo, que podem estar regularmente ou irregularmente distribuídas. Geralmente, este tipo de dado é resultante de levantamentos de recursos naturais.
- Dados de Áreas com Contagens e Taxas Agregadas - Dados que são agregados em unidades de análise, usualmente delimitadas por polígonos fechados (municípios, setores censitários, zonas de endereçamento postal, municípios).

3.2.1 Geoestatística

Neste trabalho será utilizado exclusivamente a análise espacial voltada para dados de superfícies contínuas denominada geoestatística. O objetivo da análise de superfícies é reconstruir a superfícies da qual se realizou a amostragem levando em consideração a informação espacial, ou seja, a variabilidade espacial.

Os modelos estocásticos gaussianos são frequentemente utilizados devido ao completo conhecimento de sua estrutura de dependência, de sua simplicidade inferencial e devido também ao desenvolvimento dos métodos intensivos auxiliados pelo evolução da computação. Assim, toda a abordagem geoestatística tratada nesta dissertação será abordada utilizando os modelos Gaussianos. Nesta seção será discutido o modelo utilizado para a predição da superfície.

Modelos Gaussianos para Dados Geoestatísticos

Uma função aleatória, ou campo aleatório é um processo estocástico que existe em algum espaço real d-dimensional, dada por:

$$\{Z(s_i) : s_i \in G \subset \mathbb{R}^d\} \quad (3.14)$$

sendo que $Z(s_i)$ denota a variável aleatória Z na localização s_i do espaço sob estudo

G . Assim, um *processo espacial gaussiano*

$$\{Z(s) : s \in G \subset \mathfrak{R}^2\} \quad (3.15)$$

é um processo estocástico com as seguintes propriedades:

- Para qualquer coleção de localizações s_1, s_2, \dots, s_n onde $s_i \in \mathfrak{R}^2$ a distribuição conjunta $Z = \{Z(s_1), \dots, Z(s_n)\}$ é Gaussiana n -variada sendo completamente especificada pelo vetor de média $n \times 1$, denotado por $\mu(s) = E(Z(s))$ e pela matriz de variância e covariância $n \times n$, denotada por Σ com elementos $\Sigma_{ij} = \gamma(s_i, s_j) = Cov\{Z(s_i), Z(s_j)\}$.
- Sendo assim, em todas as localizações do espaço G , Z segue uma distribuição normal.

A construção do modelo requer a definição de uma função matemática que estruture a matriz de covariância de forma que seja positiva definida. Diggle e Ribeiro Jr. (2007) apresentam algumas possíveis funções para a construção de processos gaussianos. Para este modelo algumas suposições adicionais devem ser consideradas. A restrição mais utilizada é a de que o processo seja estacionário, ou seja, a distribuição de probabilidade associada ao processo espacial não depende da grandeza de escala das coordenadas, logo, a distribuição conjunta de $(Z(s_1), Z(s_2), \dots, Z(s_n))$ é igual a distribuição conjunta $(Z(s_1 + h), Z(s_2 + h), \dots, Z(s_n + h))$, para qualquer incremento h - estacionariedade forte (Fonseca, 2008). Diggle e Ribeiro Jr. (2007) mostram outra definição menos restritiva, onde afirmam que um *processo espacial gaussiano* é estacionário se $\mu(s) = \mu \quad \forall s \in G$ e $\gamma(s_i, s_j) = \gamma(u)$, onde $u = s_i - s_j$, ou seja, a média é constante para toda a região de estudo e a covariância entre $Z(s_i)$ e $Z(s_j)$ depende apenas da diferença entre as localizações. Na literatura esta estacionariedade é chamada de fraca ou de segunda ordem. Além disso, o processo estacionário é dito isotrópico se $\gamma(u) = \gamma(\|u\|)$, onde $\|\cdot\|$ denota a distância Euclidiana, ou seja, a covariância entre $Z(s_i)$ e $Z(s_j)$ depende apenas da distância entre os pares. Entretanto, estas suposições são de difícil verificação, logo,

outra suposição menos restritiva é assumir que os incrementos $[Z(s) - Z(s + h)]$ são estacionários, ou seja, $\forall s_i \in G, E[Z(s_i)] = \mu$ e

$$\frac{Var([Z(s) - Z(s + h)])}{2} = V(h) \quad (3.16)$$

onde $V(h)$ é chamado semivariograma. Tal suposição é denominada estacionariedade intrínseca (Schanbenberger, 2005). Como mencionado é necessário estabelecer uma função matemática que dependa apenas das distâncias entre as localizações e que defina a matriz de covariâncias. Encontrar funções definidas positivas não é procedimento trivial. Abaixo segue uma família de funções, caracterizadas por possuírem dois parâmetros, um deles controlando a suavidade do processo e outro a dependência espacial.

Função de Matèrn

Proposta por Bertil Matèrn (1986) possui a seguinte forma:

$$\rho(h) = \{2^{\kappa-1}\Gamma\kappa\}^{-1} \left(\frac{h}{\phi}\right)^{\kappa} K_{\kappa}\left(\frac{h}{\phi}\right)$$

onde $\Gamma(\cdot)$ é a função gama, $K_{\kappa}(\cdot)$ é a função Bessel de ordem κ , $\phi, \kappa > 0$ e h é a distância euclidiana entre duas localizações do campo aleatório. Como mencionado o parâmetro ϕ controla o alcance das correlações, quanto mais elevado maior a dependência espacial de maior alcance. Enquanto que κ controla a suavidade do processo, quanto mais elevado maior a suavidade do processo.

3.2.2 Variograma

O variograma é uma ferramenta básica de suporte à técnica de interpolação espacial denominada *krigagem*, que permitem representar quantitativamente a variação de um fenômeno regionalizado no espaço (Huijbergts, 1975). Com ele podemos caracterizar a dependência espacial do *processo espacial gaussiano*. O variograma para um processo intrinsecamente estacionário dado pela Eq.3.16 pode ser reescrito da seguinte forma:

$$\begin{aligned} V(s_i, s_j) &= \frac{1}{2} [Var \{Z(s_i)\} + Var \{Z(s_j)\} - 2Cov \{Z(s_i), Z(s_j)\}] \quad (3.17) \\ &= \sigma^2 \{1 - \rho(u)\} \end{aligned}$$

onde $\sigma^2 = Var \{Z(s)\}$ e $\rho(u) = Corr \{Z(s_i), Z(s_j)\}$. Como tratado nas secções anteriores, a suposição de estacionariedade e que seu padrão de correlação não depender das direções (isotrópico) são necessários para o *processo espacial gaussiano*. O processo estocástico com estas suposições modelam inúmeros fenômenos naturais (geográficos). A especificação deste modelo necessita de um função matemática definida positiva e que dependa apenas das distâncias entre as localizações. Na literatura são encontradas diversas funções de correlação tais como: Matèrn, exponencial, gaussiana, esférico, circular, cúbica, *Wave*, Potência entre outras. Schlather (1999) apresenta uma descrição sobre funções de correlação.

Considere, agora, uma amostra de n localizações espaciais de um região G com os valores dos atributos nestas localizações dados por $(s_i, y_i) : i = 1, \dots, n$. Diggle e Ribeiro Jr. (2007) afirmam que o vetor de observações $Y(s)$ $n \times 1$ sob as suposições de um processo Gaussiano, pode ser modelado como:

$$Y_s = \mu + Z(s) + \epsilon \quad (3.18)$$

sendo $\mu = X\beta$ e X uma matriz $n \times q$ contendo $q - 1$ possíveis covariáveis, β um vetor $qx1$ de parâmetros associados a X , $Z(s)$ um *Processo Espacial Gaussiano* que possui vetor de médias nulo $n \times 1$ e matriz de covariância $\sum n \times n$ e por fim ϵ são mutuamente independentes e identicamente distribuídos com média zero e variância τ^2 . Para este modelo define-se o variograma como, $V(u_{ij}) = \frac{1}{2}E[(Y_i - Y_j)^2] = Var \{Y_i - Y_j\}$, onde $u_{ij} = \|x_i - x_j\|$. Sob estacionariedade o variograma é dado por:

$$V(u) = \tau^2 + \sigma^2 \{1 - \rho\} \quad (3.19)$$

Os parâmetros do variograma definem a estrutura da dependência espacial do

processo. O típico semivariograma é uma função monótona crescente com as seguintes características (Figura3.2):

- $\rho\{u\}$ é a função de correlação com $\rho\{0\} = 1$ e $\rho\{0\} \rightarrow 0$ com $u \rightarrow \infty$
- O τ^2 chamado de *nugget variance* - Teoricamente espera-se que $V(0) = 0$. Entretanto, na prática, à medida que $|h|$ tende para 0, $V(h)$ se aproxima do valor positivo denominado Efeito Peita, que revela a descontinuidade do variograma para distâncias menores do que a menor distância entre as amostras.
- E a soma $\tau^2 + \sigma$ corresponde a variância do processo observado Y chamado de patamar (*sill*).
- Por fim, o *Practical Range* - É definido como a distância u_0 o qual $\rho(u_0) = 0.05$

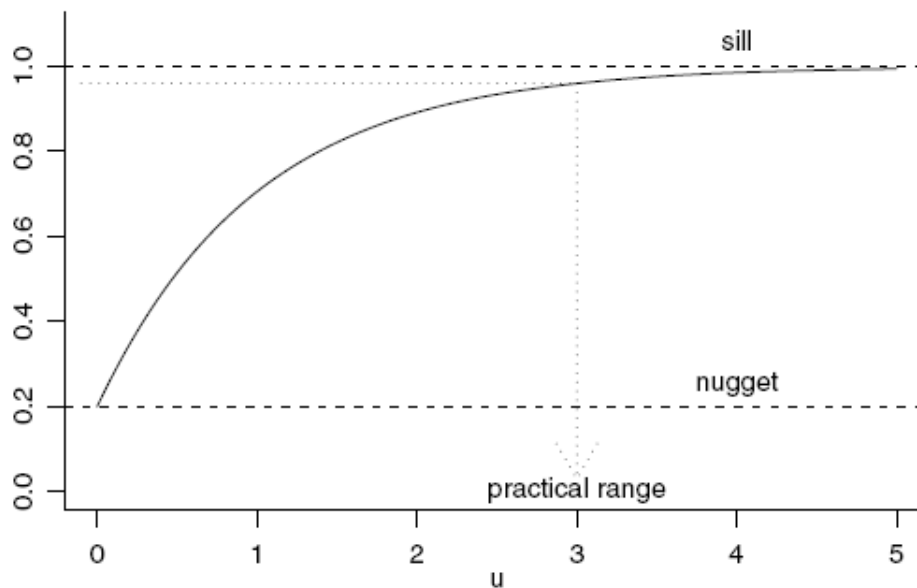


Figura 3.2: Comportamento típico de um variograma teórico.

Assim, o modelo da Eq.3.18 está totalmente especificado como $Y(s) \sim N(X\beta, \Sigma + \tau^2 I)$ sendo I uma matriz identidade $n \times n$. Portanto, para se construir a superfície de um dado fenômeno é necessário estimar o vetor de parâmetros $\theta = (\beta, \sigma^2, o^*, \tau^2)$ onde o^* é um vetor de parâmetros associados à função de correlação utilizada.

Estimação dos Parâmetros

Muitos trabalhos utilizam um estimativa empírica do semivariograma dada por:

$$2V(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [z(s_i) - z(s_i + h)]^2 \quad (3.20)$$

sendo $N(h)$ o número de pontos contidos no raio da distância h e $2V(h)$ é o semi-variograma estimado. Em alguns estudos a estimação é realizada utilizando a melhor função que se ajuste ao variograma empírico. Essa abordagem de estimação pode não ser confiável devido a sua subjetividade. Diggle e Ribeiro, Jr (2007) recomendam a utilização desta abordagem como método inicial de análise. Neste trabalho será utilizado o método da máxima verossimilhança como os métodos iterativos.

3.2.3 Krigagem

Seja $Z(s)$ a realização de um processos estocásticos em alguns pontos s_1, s_2, \dots, s_n , $s_i \in D \subseteq \mathfrak{R}^2$, $s_i \neq s_j$, gerando as realizações $Z(x_1), \dots, Z(x_n)$. Deseja-se prever o valor $Z(x_0)$ em pontos $s_0 \in D \subseteq \mathfrak{R}^2$ não amostrados, utilizando as informações $Z(s_1), \dots, Z(s_n)$. Esta predição é realizada utilizando o conceito de erro quadrático médio definida por:

$$EQM(\hat{Z}(s_0)) = E \left[\left\{ Z(\hat{s}_0) - Z(s_0) \right\}^2 \right]$$

Deve-se obter uma função dos valores observados tal que o erro quadrático médio seja mínimo. A função geral que minimiza o EQM é dado pelo seguinte resultado: o erro quadrático médio é mínimo quando $\hat{Z}(s_0) = \psi(Z(s_1), \dots, Z(s_n)) = E[Z(s_0)|z(s_1), \dots, z(x_n)]$. Para um *processo espacial gaussiano* estacionário a função de predição ψ que minimiza o EQM é uma combinação linear dada por:

$$\hat{Z}(s_0) = \mu_0 + \mathbf{K}_0 \Sigma^{-1} [\mathbf{Z}(s) - \mu]$$

onde $\mathbf{Z}(s) = Z(s_1), \dots, Z(s_n)$, Σ é a matriz de covariâncias de $\mathbf{Z}(s)$, \mathbf{K}_0 é o vetor de

covariâncias entre $Z(s_0)$ e $Z(s_i)$, μ é o vetor de esperanças e por fim $\mu_0 = E[Z(s_0)]$.

A *krigagem* engloba diversos métodos de predição diferindo nas suposições utilizadas no processo estocástico espacial. Os métodos que envolvem a função de predição linear são:

- Krigagem Simples - utilizada para predição no caso em que a esperança e a covariância do processo estocástico são conhecidas. Reescrevendo a equação preditiva sobre estas suposições temos: $\hat{Z}(s_0) = m + \sum_{i=1}^n \lambda_i [\mathbf{Z}(s_i) - m]$. Journel (1988) mostra que minimizando a variância do erro os pesos λ_i são obtidos a partir do seguinte sistemas de equações:

$$\sum_{i=1}^n \lambda_i COV_{ij} = COV_{i0}$$

para $i = 1, \dots, n$ sendo COV_{ij} referente à função covariância correspondente a um vetor, $|\mathbf{h}|$, com origem em s_i , e extremidade em s_j e COV_{i0} refere-se à função covariância correspondente a um vetor, $|\mathbf{h}|$, com origem em s_i , e extremidade no ponto a ser estimado s_0 .

- Krigagem Ordinária - utilizada quando $E[Z(s)] = m$ é apenas estacionário e desconhecido para todo s . Reescrevendo a função preditiva temos:

$$\begin{aligned} \hat{Z}(s_0) &= \mu + \sum_{i=1}^n \lambda_i [\mathbf{Z}(s) - \mu] \\ &= \mu \left(1 - \sum_{i=1}^n \lambda_i \right) + \sum_{i=1}^n \lambda_i \mathbf{Z}(s) \end{aligned} \quad (3.21)$$

A Krigagem Ordinária não requer o prévio conhecimento da média. Minimizando a variância do erro, os pesos são obtidos a partir do seguinte sistemas de equações:

$$\sum_{i=1}^n \lambda_i COV_{ij} - \alpha = COV_{i0}$$

para $i = 1, \dots, n$ e

$$\sum_{i=1}^n \lambda_i = 1$$

onde α é o multiplicador de Lagrange necessário para a minimização da variância do erro.

- Krigagem Universal - utilizado quando $E[Z(s)] = m(s) \forall s \in D$, ou seja, não estacionário. Considerando $m(s) = \sum_{j=1}^s \alpha_j f_j(s)$ para $\forall s \in D$, onde α_j , para $j = 1, \dots, n$ são parâmetros desconhecidos e f_j são funções conhecidas de s .

Capítulo 4

Resultados

A primeira etapa consistiu na obtenção dos dados do NCEP/NCAR de Temperatura do Ar no retângulo envolvente de latitude -50 a -32,5 e -20 a 0 de latitude contendo todo o NEB, no período de 1 de janeiro de 1948 a 31 de dezembro de 2008. O NCEP/NCAR fornece os dados em formato *Network Common Data Form* - NETCDF, que são formatos de arquivos para armazenamento de grande quantidade de informações. Estes arquivos apresentam 4 observações diárias. Entretanto, é necessário uma série que forneça um máximo diário. Para isto, foi produzido um algoritmo no *software R* - Apêndice A. Vale ressaltar que todos os resultados mostrados aqui foram produzidos utilizando o *Software R* e que, todos os algoritmos produzidos podem ser encontrados no Apêndice desta dissertação.

4.1 Análise Exploratória dos Dados

4.1.1 Dados de Reanálise - NCEP/NCAR

Como tratamento inicial dos dados, foi gerado o *Box Plot* para as 72 séries de Temperatura do Ar máxima diária. A Figura 4.1 é a representação dos *Box Plot* para cada coluna da grade do NCEP/NCAR. Visto que a grade do NCEP/NCAR é formada por 72 pontos divididas em 8 colunas e 9 linhas. Nota-se que as séries situadas ao Sul do NEB apresentam maior variabilidade com valores medianos mais elevados. A medida

que se aproxima do litoral os valores medianos decrescem enquanto a variabilidade apresenta-se intacta. Este comportamento é esperado, pois à medida que afasta-se do Equador as temperaturas tendem a apresentar uma maior variabilidade. Entretanto, as temperaturas ao sul do Nordeste Brasileiro apresentam, visualmente, maior intensidade. Com esses resultados espera-se que no Norte do NEB o modelo GEV ajuste-se melhor, devido a menor variabilidade encontrada. Para a séries localizadas na região Sul do NEB, espera-se, que os modelos GPD apresentem melhores ajuste, devido a maior variabilidade.

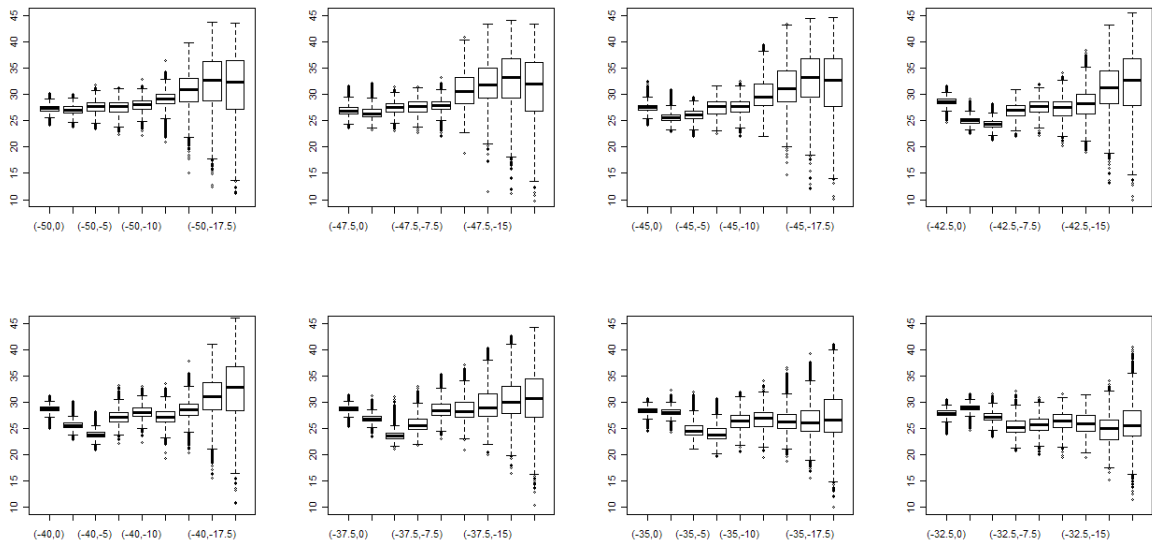


Figura 4.1: *Box Plot* das 72 séries de temperatura máxima do ar

4.1.2 Dados Observacionais - INMET

Será utilizado os dados observacionais fornecida pelo INMET para a validação do método. Os dados observacionais serão utilizados para comparação direta do ajuste da GEV e GPD e com o resultado da Regressão Quantílica. O interesse inicial dos estudo seria a obtenção de um relação dos dados dos NCEP/NCAR e os dados observados para os quantis 90% e 95% de todas as capitais do Nordeste. Contudo, a qualidade dos dados observacionais, fornecidas pelo INMET, não é satisfatória devido a grande

quantidade de informações faltantes na série de todas as capitais do Nordeste. Para contornar este problema será utilizada apenas a série da capital de Pernambuco devido a maior quantidade de informações. Para um simples verificação do comportamento da Temperatura do Ar na estação de Recife foi gerado o *Box Plot* (Figura 4.2). Nota-se um certa assimetria na distribuição com uma menor variabilidade e maior intensidade em comparação como os dados do NCEP/NCAR nas proximidades do Recife.

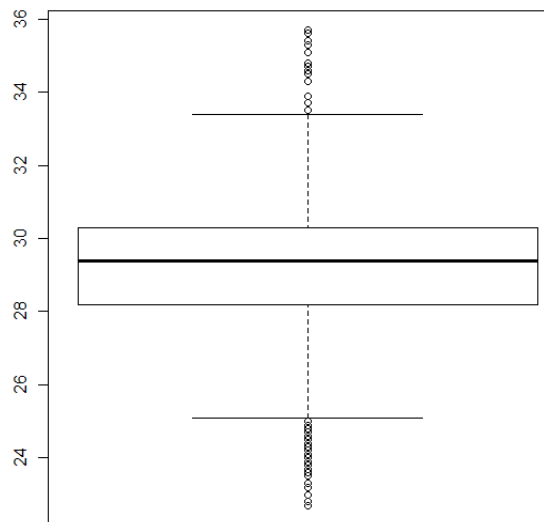


Figura 4.2: *Box Plot* dos Máximos Diários Observados sobre a estação do Recife

4.2 Teoria dos Valores Extremos

Na literatura não existe um consenso sobre quais modelos são mais adequados para extremos de temperatura máxima e mínima. Trabalhos já realizados poderiam indicar a melhor distribuição para o ajuste das Temperaturas do Ar máxima no NEB. Entretanto, poucos trabalhos foram encontrados. Sansigolo e Nery (2000) constataram que as distribuições de Gumbel foram as mais adequadas para temperaturas mínimas de 27 localidades do Estado do Paraná. Assim, não teremos nenhuma base de ajuste para comparação com os trabalhos já realizados.

4.2.1 Análise Via Distribuição Generalizada de Valores Extremos - GEV

O ajuste ao modelo GEV foi baseado no máximo para cada ano, obtendo-se assim 61 valores máximos para cada ponto da grade. Com a identificação dos eventos extremos a próxima etapa consiste em estimar os parâmetros através do método da máxima verossimilhança (Tabela 4.1). Produzindo intervalos de confiança para o parâmetro de forma (ξ), podemos constatar qual a provável distribuição do máximo de Temperatura do Ar. Para isto, foram construídos gráficos do perfil de máxima verossimilhança para ξ , resumidos na Tabela 4.1. Para os pontos em vermelho pode-se afirmar que há evidências estatísticas para rejeitar a hipótese de nulidade de $\xi = 0$, com nível de significância de 5%, ou seja, podemos rejeitar a hipótese de uma distribuição *Gumbel* para os eventos extremos de Temperatura do Ar nesta localizações. Com a finalidade de validar o ajuste da Distribuição Generalizada de Valores Extremos é necessário realizar o diagnóstico do modelo, realizado através dos gráficos de probabilidade, quantis, nível de retorno e densidade. Para todos os modelos o ajuste mostrou-se satisfatório. Devido a grande quantidade de gráficos e informações produzidos serão apresentados os principais no apendice B.

4.2.2 Análise via Distribuição Generalizada de Pareto - GPD

O primeiro passo para o ajuste da GPD consiste na seleção do limiar realizado através da análise do gráfico Média dos Excessos. O limiar “ótimo” é o menor valor de u tal que o gráfico da média dos excessos (Davison e Smith, 1990) seja aproximadamente linear. Na Tabela 4.2 encontram-se os limiares selecionados através do método proposto por Davison e Smith (1990) além das estimativas dos parâmetros ξ e σ . A hipótese de nulidade $\xi = 0$ foi testada ao nível de significância de 5%. Assim, para 10 pontos em vermelho, pode-se afirmar que existem evidências estatísticas para não rejeitarmos a hipótese de nulidade de $\xi = 0$, com nível de significância de 5%, ou seja, a distribuição *Exponencial* é indicada para os eventos extremos de Temperatura do Ar

Tabela 4.1: Estimativas de Máxima Verossimilhança dos Parâmetros $\mu^{\circ}C$, σ e ξ da GEV.

| Ponto | μ | σ | ξ | Ponto | μ | σ | ξ |
|---------------|-------|----------|-------|---------------|-------|----------|-------|
| (-50,0) | 28,78 | 0,49 | -0,19 | (-40,0) | 30,16 | 0,42 | -0,27 |
| (-50,-2.5) | 28,7 | 0,61 | -0,33 | (-40,-2.5) | 28,23 | 0,61 | -0,13 |
| (-50,-5) | 29,63 | 0,55 | -0,11 | (-40,-5) | 26,02 | 0,64 °C | -0,16 |
| (-50,-7.5) | 29,69 | 0,43 | -0,05 | (-40,-7.5) | 30,47 | 0,61 | 0 |
| (-50,-10) | 30,26 | 0,44 | 0 | (-40,-10) | 31,08 | 0,66 | -0,24 |
| (-50,-12.5) | 33 | 0,65 | -0,08 | (-40,-12.5) | 30,23 | 0,63 | -0,04 |
| (-50,-15) | 37,34 | 0,71 | -0,14 | (-40,-15) | 33,14 | 0,87 | -0,04 |
| (-50,-17.5) | 41,84 | 0,82 | -0,3 | (-40,-17.5) | 38,71 | 0,83 | -0,18 |
| (-50,-20) | 42,16 | 0,72 | -0,41 | (-40,-20) | 42,84 | 0,99 | -0,22 |
| (-47.5,0) | 30,34 | 0,7 | -0,47 | (-40,0) | 30,1 | 0,37 | -0,12 |
| (-47.5,-2.5) | 29,84 | 1,05 | -0,36 | (-37.5,-2.5) | 28,92 | 0,59 | -0,13 |
| (-47.5,-5) | 29,29 | 0,53 | -0,07 | (-37.5,-5) | 26,18 | 0,8 | 0,12 |
| (-47.5,-7.5) | 29,93 | 0,51 | -0,2 | (-37.5,-7.5) | 30,1 | 0,63 | 0,08 |
| (-47.5,-10) | 30,76 | 0,59 | -0,14 | (-37.5,-10) | 33,74 | 0,71 | -0,38 |
| (-47.5,-12.5) | 38,3 | 0,77 | -0,14 | (-37.5,-12.5) | 34,41 | 0,78 | -0,18 |
| (-47.5,-15) | 41,04 | 0,71 | -0,05 | (-37.5,-15) | 37,48 | 1,06 | -0,21 |
| (-47.5,-17.5) | 41,94 | 0,94 | -0,37 | (-37.5,-17.5) | 40,19 | 0,91 | -0,21 |
| (-47.5,-20) | 41,55 | 0,77 | -0,3 | (-37.5,-20) | 41,4 | 1,01 | -0,27 |
| (-45,0) | 30,18 | 0,74 | -0,24 | (-37.5,0) | 29,7 | 0,4 | -0,28 |
| (-45,-2.5) | 28,52 | 0,88 | -0,16 | (-35,-2.5) | 29,81 | 0,45 | -0,04 |
| (-45,-5) | 28,15 | 0,5 | -0,28 | (-35,-5) | 28,34 | 0,95 | -0,08 |
| (-45,-7.5) | 29,98 | 0,45 | -0,12 | (-35,-7.5) | 28,25 | 0,88 | -0,22 |
| (-45,-10) | 30,87 | 0,61 | -0,23 | (-35,-10) | 30,21 | 0,59 | -0,25 |
| (-45,-12.5) | 37,31 | 0,74 | -0,17 | (-35,-12.5) | 30,84 | 0,86 | -0,15 |
| (-45,-15) | 41,13 | 1 | -0,31 | (-35,-15) | 33,05 | 1,44 | -0,29 |
| (-45,-17.5) | 42,53 | 0,9 | -0,33 | (-35,-17.5) | 35,28 | 1,23 | -0,25 |
| (-45,-20) | 42,35 | 0,72 | -0,16 | (-35,-20) | 38,75 | 1,35 | -0,54 |
| (-42.5,0) | 30,31 | 0,47 | -0,24 | (-35,0) | 29,24 | 0,41 | -0,2 |
| (-42.5,-2.5) | 27,44 | 0,67 | -0,17 | (-32.5,-2.5) | 30,25 | 0,35 | -0,04 |
| (-42.5,-5) | 26,93 | 0,56 | -0,22 | (-32.5,-5) | 29,78 | 0,55 | -0,21 |
| (-42.5,-7.5) | 29,71 | 0,57 | -0,42 | (-32.5,-7.5) | 29,39 | 0,74 | -0,17 |
| (-42.5,-10) | 30,03 | 0,44 | 0,03 | (-32.5,-10) | 29,17 | 0,52 | -0,19 |
| (-42.5,-12.5) | 31,1 | 0,71 | -0,08 | (-32.5,-12.5) | 29,7 | 0,47 | -0,15 |
| (-42.5,-15) | 34,8 | 1,02 | -0,18 | (-32.5,-15) | 29,76 | 0,58 | -0,22 |
| (-42.5,-17.5) | 39,98 | 0,82 | -0,07 | (-32.5,-17.5) | 30,2 | 0,87 | 0,15 |
| (-42.5,-20) | 42,88 | 0,93 | -0,24 | (-32.5,-20) | 35,91 | 1,49 | -0,24 |

nestas localizações. Para as 62 restantes temos a indicação de uma distribuição *Beta* para os eventos extremos de temperatura do ar. Para todos os modelos os ajustes mostraram-se satisfatórios. Vale ressaltar, que o objetivo do ajuste da GEV e GPD sobre os dados do NCEP/NCAR é a obtenção das estimativas dos parâmetros para a interpolação espacial. Na próxima seção iremos tratar a distribuição espacial das estimativas geradas.

4.3 Análise Geoestatística

O objetivo da análise geoestatística aqui, é a produção de superfícies para os três parâmetros da GEV (ξ , σ e μ) e para os três parâmetros da GPD (ξ , σ e limiar).

4.3.1 Análise Exploratória dos Dados

A análise exploratória dos dados é uma etapa fundamental dentro de estudos estatísticos. Burroughs et. al. (1996) afirma que esta é uma etapa precursora introduzida para averiguar as propriedades estatísticas dos dados. Como mencionado anteriormente, a amostra é composta por 72 estimativas de cada parâmetro da GEV e GPD. Visando verificar o comportamento da distribuição espacial das estimativas, mapas coroplético com as estimativas foram gerados (Figuras 4.3 e 4.4). O processo de construção é dado da seguinte forma: observando a Figura 1.1 nota-se que está dividida em quadrantes, assim cada observação será preenchida com a cor referente ao seu valor na escala das estimativas.

A distribuição espacial das observações das estimativas da média da GEV (Figura 4.3b) mostram que a ocorrência de extremos na região Sul do NEB são mais intensos. Para a variância σ (Figura 4.3a) a variabilidade espacial apresenta um comportamento mais contínuo. Entretanto, a região Sul do NEB apresenta maior variabilidade. Já para o parâmetro de forma ξ (Figura 4.3c) o comportamento é mais “aleatório”. A distribuição espacial dos parâmetros da GPD apresentam um comportamento semelhante aos parâmetros da GEV, apesar de possuir apenas dois parâmetros. As estimativas da

Tabela 4.2: Limiar e Estimativas de Máxima Verossimilhança dos Parâmetros $\sigma^{\circ}\text{C}$ e ξ da GPD

| Ponto | Limiar | σ | ξ | Ponto | Limiar $^{\circ}\text{C}$ | σ | ξ |
|---------------|-------------------------|----------|-------|---------------|---------------------------|----------|-------|
| (-50,0) | 27,8 | 0,45 | -0,12 | (-40,0) | 29 | 0,41 | -0,11 |
| (-50,-2.5) | 27,8 | 0,45 | -0,11 | (-40,-2.5) | 27,8 | 0,52 | -0,05 |
| (-50,-5) | 28,9 | 0,41 | -0,05 | (-40,-5) | 26 | 0,46 | 0,00 |
| (-50,-7.5) | 28,8 | 0,40 | -0,06 | (-40,-7.5) | 29 | 0,65 | -0,03 |
| (-50,-10) | 29 | 0,46 | -0,07 | (-40,-10) | 29,4 | 0,64 | -0,10 |
| (-50,-12.5) | 30,5 | 0,95 | -0,14 | (-40,-12.5) | 30 | 0,57 | -0,01 |
| (-50,-15) | 36,5 | 0,70 | -0,11 | (-40,-15) | 31 | 0,86 | -0,04 |
| (-50,-17.5) | 40 | 1,10 | -0,23 | (-40,-17.5) | 35 | 1,60 | -0,22 |
| (-50,-20) | 39 | 1,47 | -0,28 | (-40,-20) | 41,3 | 0,95 | -0,11 |
| (-47.5,0) | 28 | 1,19 | -0,29 | (-40,-20) | 29 | 0,40 | -0,10 |
| (-47.5,-2.5) | 28 | 0,89 | -0,12 | (-37.5,0) | 27,3 | 0,50 | -0,04 |
| (-47.5,-5) | 28,8 $^{\circ}\text{C}$ | 0,31 | 0,03 | (-37.5,-2.5) | 26,9 | 0,99 | -0,08 |
| (-47.5,-7.5) | 28,8 | 0,47 | -0,10 | (-37.5,-5) | 28,1 | 0,86 | -0,08 |
| (-47.5,-10) | 28,5 | 0,81 | -0,16 | (-37.5,-7.5) | 30 | 1,73 | -0,33 |
| (-47.5,-12.5) | 37,5 | 0,71 | -0,12 | (-37.5,-10) | 32 | 1,18 | -0,21 |
| (-47.5,-15) | 38 | 1,62 | -0,28 | (-37.5,-12.5) | 34 | 1,53 | -0,19 |
| (-47.5,-17.5) | 39,5 | 1,23 | -0,23 | (-37.5,-15) | 38 | 1,26 | -0,18 |
| (-47.5,-20) | 39,5 | 1,00 | -0,17 | (-37.5,-17.5) | 37 | 2,20 | -0,29 |
| (-45,0) | 28,6 | 0,67 | -0,07 | (-37.5,-20) | 28,5 | 0,48 | -0,17 |
| (-45,-2.5) | 28,4 | 0,76 | -0,18 | (-35,0) | 28,5 | 0,49 | -0,08 |
| (-45,-5) | 27 | 0,44 | -0,10 | (-35,-2.5) | 28 | 0,67 | -0,07 |
| (-45,-7.5) | 29 | 0,44 | -0,09 | (-35,-5) | 24 | 1,53 | -0,20 |
| (-45,-10) | 28,8 | 0,73 | -0,13 | (-35,-7.5) | 28 | 0,84 | -0,15 |
| (-45,-12.5) | 36 | 0,83 | -0,14 | (-35,-10) | 30 | 0,68 | -0,04 |
| (-45,-15) | 37 | 2,02 | -0,29 | (-35,-12.5) | 30 | 1,38 | -0,09 |
| (-45,-17.5) | 40 | 1,39 | -0,27 | (-35,-15) | 29,8 | 2,22 | -0,21 |
| (-45,-20) | 40 | 1,12 | -0,18 | (-35,-17.5) | 34 | 2,43 | -0,32 |
| (-42.5,0) | 29 | 0,44 | -0,10 | (-35,-20) | 28,3 | 0,40 | -0,12 |
| (-42.5,-2.5) | 25 | 0,62 | -0,05 | (-32.5,0) | 29,1 | 0,45 | -0,11 |
| (-42.5,-5) | 26 | 0,43 | 0,00 | (-32.5,-2.5) | 28,5 $^{\circ}\text{C}$ | 0,50 | -0,07 |
| (-42.5,-7.5) | 28 | 0,70 | -0,20 | (-32.5,-5) | 27,5 | 0,85 | -0,15 |
| (-42.5,-10) | 29 | 0,47 | -0,08 | (-32.5,-7.5) | 27,5 | 0,73 | -0,17 |
| (-42.5,-12.5) | 29,1 | 0,77 | -0,08 | (-32.5,-10) | 29 | 0,41 | -0,08 |
| (-42.5,-15) | 33 | 1,07 | -0,12 | (-32.5,-12.5) | 28,5 | 0,59 | -0,14 |
| (-42.5,-17.5) | 38 | 1,15 | -0,16 | (-32.5,-15) | 27,5 | 0,92 | -0,02 |
| (-42.5,-20) | 41 | 1,04 | -0,14 | (-32.5,-17.5) | 30 | 2,58 | -0,23 |

variância (Figura 4.4a) apresentam menor intensidade na região Norte. Enquanto que as estimativas do ξ (Figura 4.4b) apresentam um comportamento mais descontínuo.

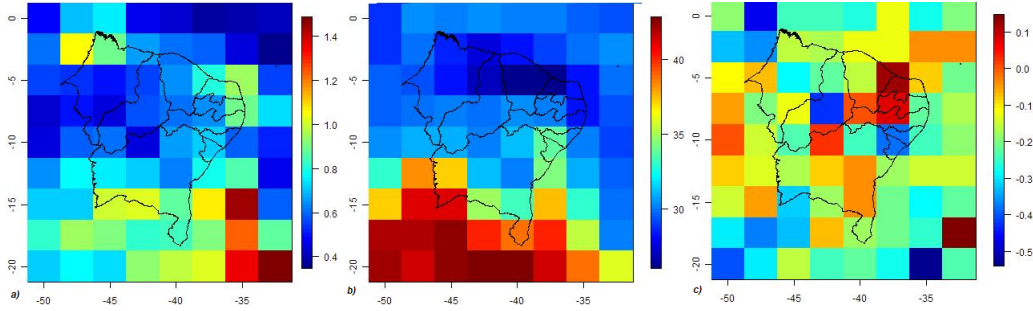


Figura 4.3: Mapa Coroplético das Estimativas de σ a), μ b) e ξ c) do modelo GEV

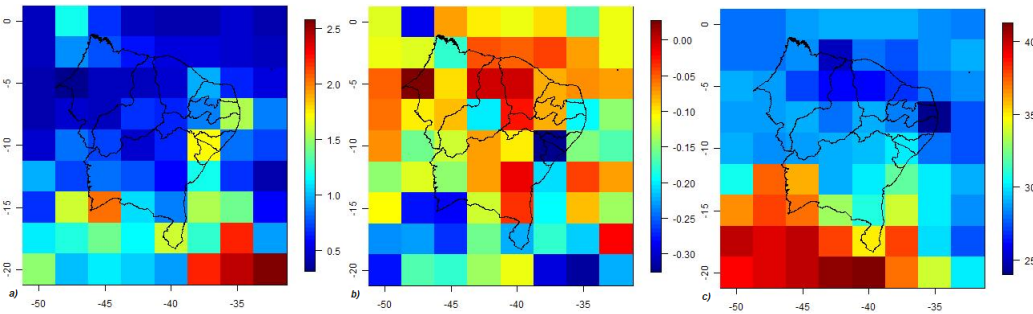


Figura 4.4: Mapa Coroplético das Estimativas de σ a), ξ b) e Limiar c) do modelo GPD.

Para verificar o comportamento da distribuição do processo espacial dos parâmetros estimados, foram utilizadas uma classe de transformações denominadas família Box-Cox (Box e Cox, 1964). As Figuras 4.5 e 4.6 mostram o gráfico de logverossimilhança para as possíveis transformações da família Box-Cox para a média (GEV) e o limiar da GPD, indicando as transformações mais eficientes para $\lambda = -3,27$ e $\lambda = -3,34$ para a média e o limiar respectivamente. Note aqui que apesar de ser um parâmetro, estamos considerando as observações como sendo variáveis aleatórias. Tal suposição pode ser considerada razoável devido a sua variabilidade sobre o espaço.

Outro importante aspecto na construção das superfícies é o conhecimento da tendência dos dados sobre o espaço. Assim, foram gerados gráficos de dispersão das coordenadas Longitudes e Latitudes com os dados. Nestes gráficos pode-se verificar tendências

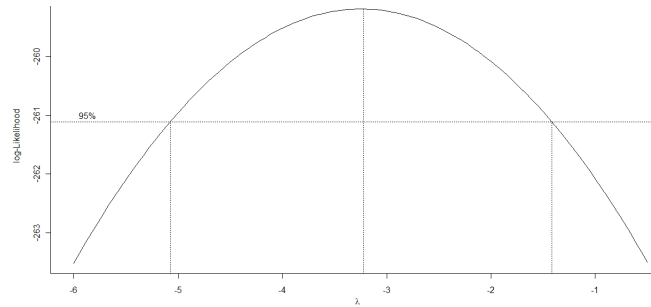


Figura 4.5: Gráfico da Logverossimilhança para as Transformações da família BoxCox, μ - GEV

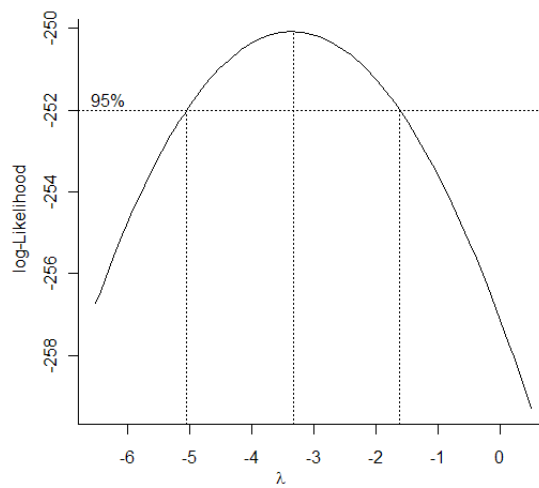


Figura 4.6: Gráfico da Logverossimilhança para as Transformações da família BoxCox, limiar - GPD

direcionais o que pode ser determinante na qualidade da superfície interpolada e na determinação do modelo de *krigagem* a ser utilizado. O gráfico de dispersão da Dados x Latitude (Figuras 4.7 e 4.8) mostra uma tendência decrescente de Norte a Sul para os parâmetros de forma (ξ) para o modelo GPD e mantêm-se constante para o modelo GEV. Para o restante dos parâmetros o comportamento é crescente do Norte para o Sul. Para a dispersão Longitude x Dados (Figuras 4.7 e 4.8), o gráfico de dispersão apresenta uma tendência de valores mais intensos na região Sul. Estes resultados indicam que a média poderá ser explicada com a utilização das coordenadas geográficas com covariáveis explicativas.

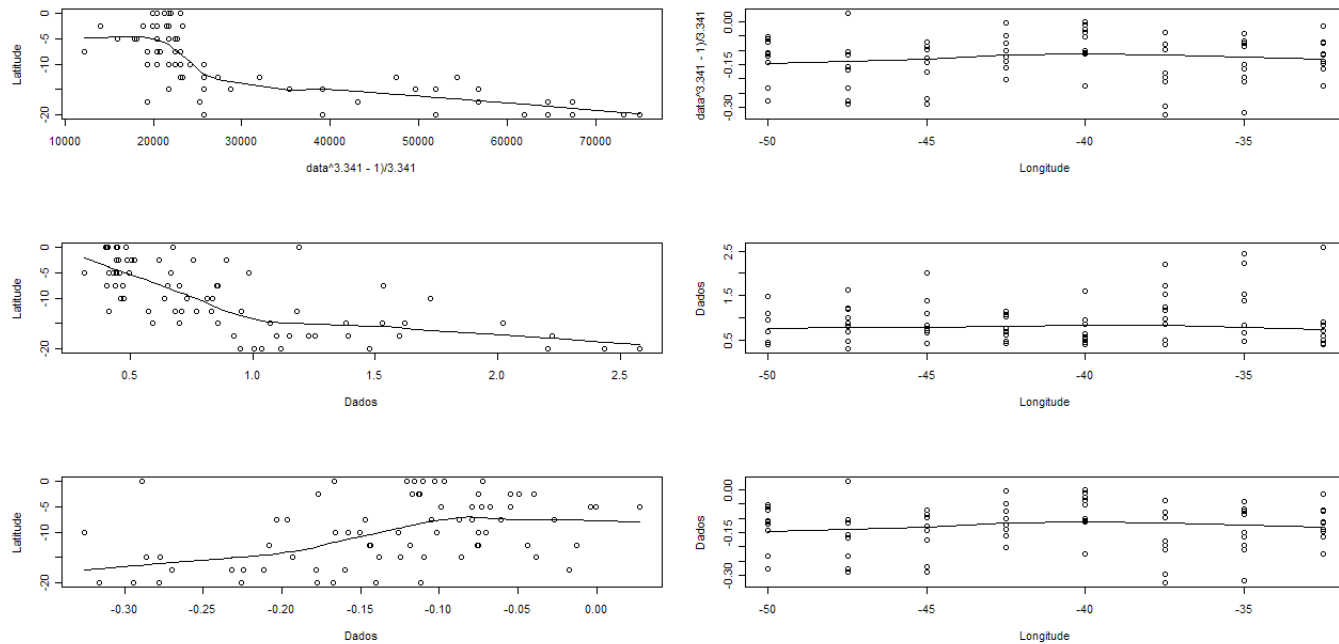


Figura 4.7: Gráficos de Dispersão dos dados com as coordenadas geográficas para o Limiar (linha - a), σ (linha - b) e ξ (linha - c) - GPD

4.3.2 Estimação dos Parâmetros

Após o estudo visual da variabilidade espacial das estimativas, foi verificado que a estrutura da média no espaço será melhor explicada com a inserção das covariáveis Latitude e Longitude. O próximo passo será o estudo da correlação espacial das estimativas.

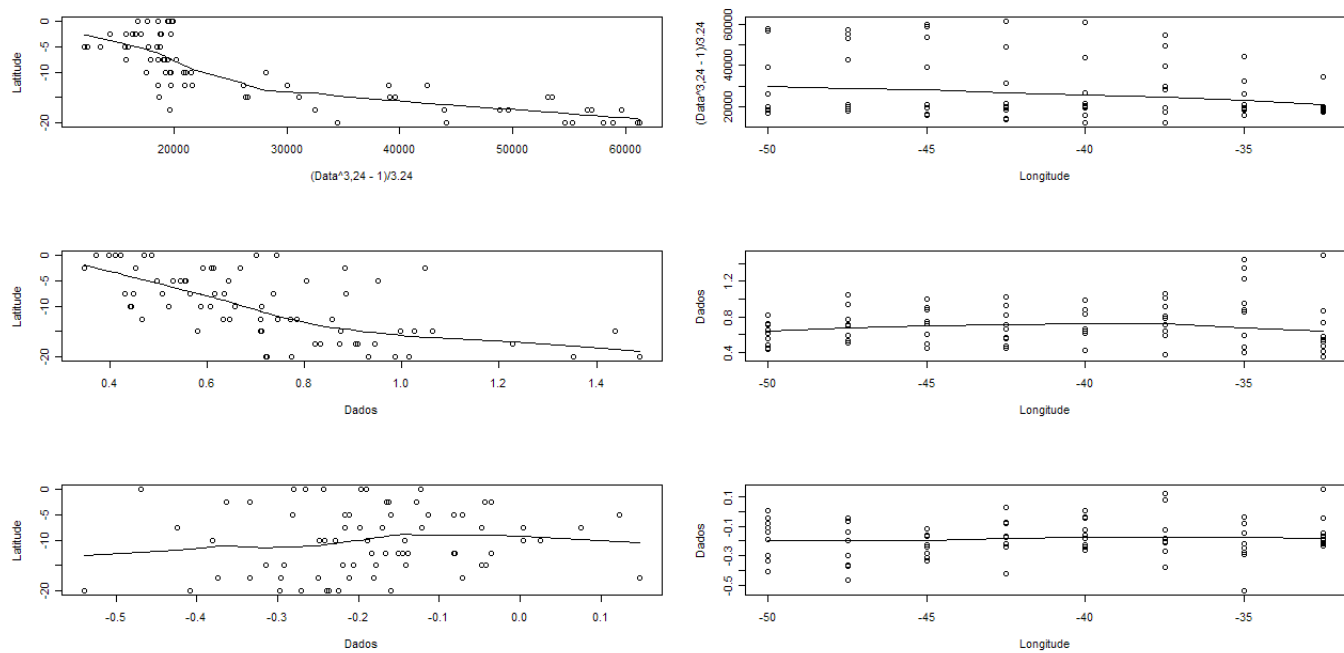


Figura 4.8: Gráficos de Dispersão dos dados com as coordenadas geográficas para a μ (linha - a), σ (linha - b) e ξ (linha - c) - GEV

Como visto em seções anteriores muitos trabalhos utilizaram a abordagem “a olho”, ou seja, o pesquisador gera o variograma empírico e a olho deduz qual a melhor função modelo das observações. Assim, os parâmetros da função teórica seriam as estimativas do modelo ajustado a olho. Devido a sua subjetividade, esta abordagem pode não ser muito precisa, pois os valores dos semivariogramas empíricos podem se afastar muito do semivariograma real devido ao tamanho, acaso amostral, variabilidade entre outros. Segundo Fonseca (2008) assumindo que, o *processo espacial gaussiano* possui estacionariedade forte, pode-se utilizar estimadores baseados na máxima verossimilhança. Sendo assim, seguindo a recomendação de Diggle e Ribeiro Jr. (2007), que discutem com maiores detalhes este assunto, será utilizado aqui a abordagem “a olho” para obtenção dos parâmetros iniciais para entrada em métodos iterativos de estimação. As Figuras 4.9 e 4.10 mostram os semivariograma empíricos para as 6 amostras das estimativas dos parâmetros.

Devido a grande quantidade de gráficos e informações produzidas será apresentado

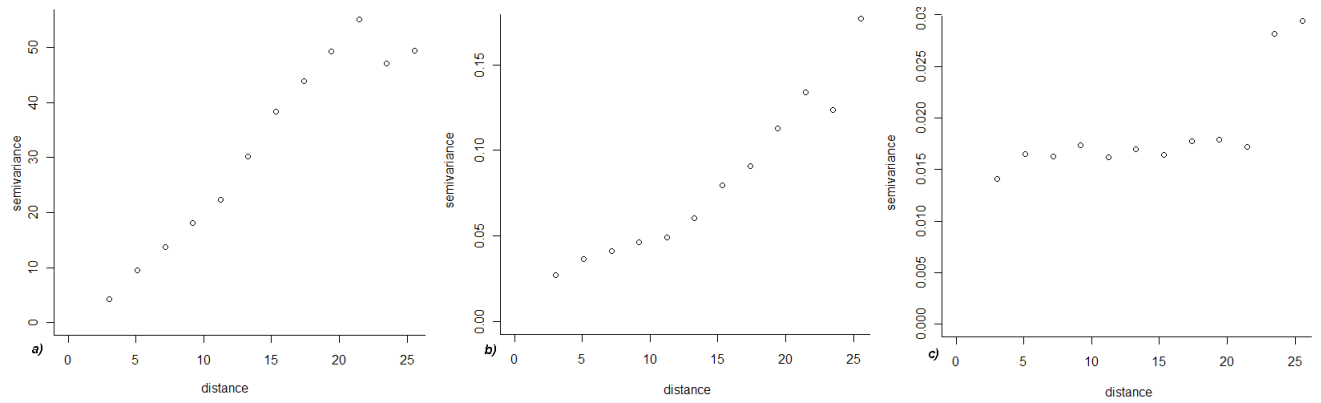


Figura 4.9: Variograma Empírico para as Estimativas dos Parâmetros μ transformado a), σ b) e ξ GEV

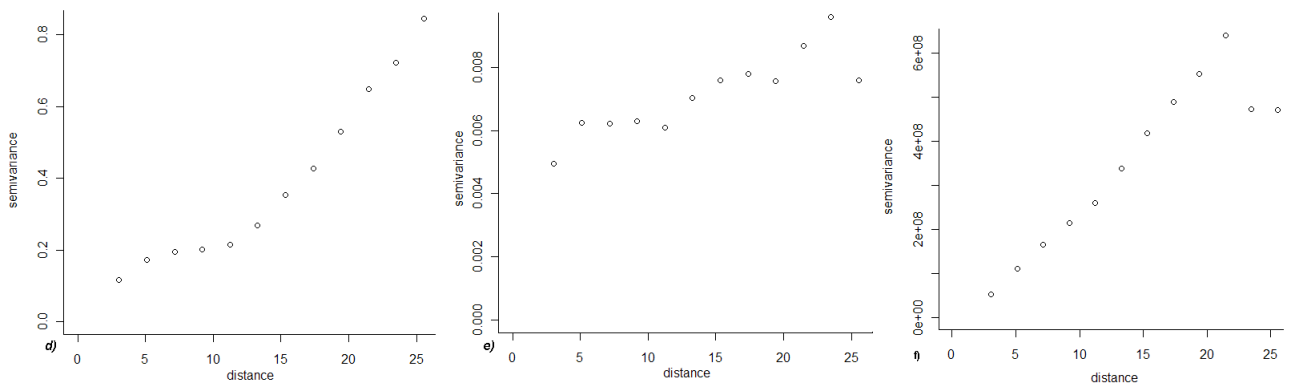


Figura 4.10: Variograma Empírica para as Estimativas dos Parâmetros σ a), ξ b) e Limiar GPD

aqui apenas o processo de construção para o parâmetro de forma ξ . As estimativas de máxima verossimilhança com tendências na média induzidas pelas coordenadas encontram-se nas Tabelas 4.3 a 4.5. Utilizando as diferentes abordagens para a média, os modelos com tendência constante apresentam menores valores para o critério Akaike. Analisando os valores maximizados do logaritmo da função de verossimilhança e os seus respectivos AIC, foi selecionado o modelo *Matèrn* com $\kappa = 0,68$. Este ajuste gerou o valor máximo da função logverossimilhança e o menor valor do critério de penalidade por número de parâmetros associados ao modelo.

Tabela 4.3: Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para o ξ com média constante

| Model | β | τ | σ | ϕ | AIC | Log |
|----------------------------|---------|--------|----------|--------|-----------|-------|
| <i>powered.exponential</i> | -0,1938 | 0 | 0,0167 | 0,4505 | -85,31103 | 46,66 |
| <i>powered.exponential</i> | -0,1883 | 0 | 0,0165 | 1,4237 | -86,91086 | 47,46 |
| <i>Matern</i> | -0,1883 | 0 | 0,0165 | 1,4237 | -86,91086 | 47,46 |
| <i>Matern</i> | -0,1895 | 0 | 0,0165 | 2,0314 | -86,7218 | 47,36 |
| <i>Matern</i> | -0,1878 | 0 | 0,0165 | 1,1980 | -86,94499 | 47,47 |
| <i>Matern</i> | -0,1895 | 0 | 0,0165 | 2,0314 | -86,7218 | 47,36 |

Tabela 4.4: Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para o ξ com tendencia na média induzida pelas coordenadas Oeste - Leste

| Model | β_0 | β_1 | τ | σ | ϕ | AIC | Log |
|----------------------------|-----------|-----------|--------|----------|--------|-----------|-------|
| <i>Exponencial</i> | -0,0859 | 0,0025 | 0 | 0,4 | 1,3690 | -85,40997 | 47,70 |
| <i>Matern</i> | -0,0859 | 0,0025 | 0 | 0,163 | 1,3127 | -85,42363 | 47,71 |
| <i>Powered Exponential</i> | -0,0839 | 0,0026 | 0 | 0,164 | 0,3684 | -83,8566 | 46,93 |
| <i>Matern</i> | -0,0859 | 0,0025 | 0 | 0,163 | 1,3127 | -83,8566 | 47,71 |
| <i>Matern</i> | -0,0852 | 0,0025 | 0 | 0,162 | 2,7225 | -84,81707 | 47,41 |
| <i>Powered Exponential</i> | -0,0839 | 0,0026 | 0 | 0,164 | 0,3684 | -83,8566 | 46,93 |

4.3.3 Krigagem

Com os modelos finais estabelecidos, foram calculadas as predições espaciais para as 6 estimativas dos modelos GEV e GPD em uma malha de 1000 localizações, tendo

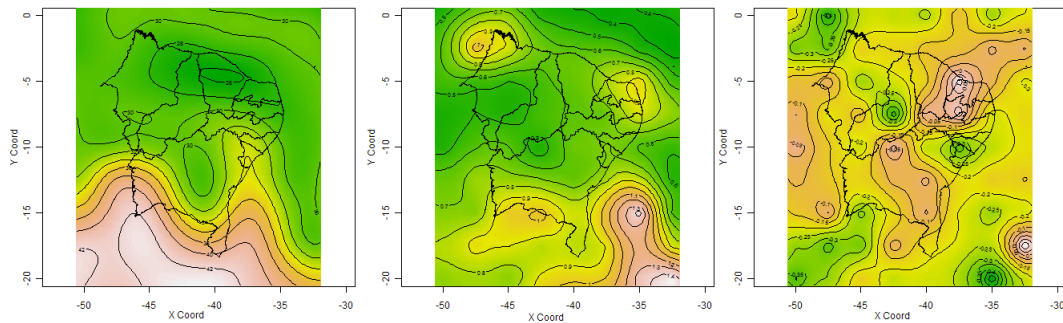
Tabela 4.5: Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para o ξ com tendencia na média induzida pelas coordenadas Norte-Sul

| Model | β_0 | β_1 | τ | σ | ϕ | AIC | Log |
|----------------------------|-----------|-----------|--------|----------|--------|-----------|-------|
| <i>Exponencial</i> | -0,1665 | 0,0025 | 0 | 0,0163 | 1,3679 | -85,43033 | 47,72 |
| <i>Matern</i> | -0,1665 | 0,0022 | 0 | 0,163 | 1,9127 | -85,2173 | 47,61 |
| <i>Powered Exponential</i> | -0,1693 | 0,0023 | 0 | 0,164 | 0,3558 | -83,8087 | 46,9 |

Tabela 4.6: Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para o ξ com tendencia na média induzida pela área

| Model | β_0 | β_1 | β_2 | τ | σ | ϕ | AIC | Log |
|----------------------------|-----------|-----------|-----------|--------|----------|--------|-----------|-------|
| <i>Exponencial</i> | -0,06226 | 0,0025 | 0,0023 | 0 | 0,0161 | 1,3104 | -83,9606 | 47,98 |
| <i>Matern</i> | -0,0627 | 0,0025 | 0,0023 | 0 | 0,0161 | 1,4524 | -83,90715 | 47,95 |
| <i>Powered Exponential</i> | -0,0627 | 0,0025 | 0,0023 | 0 | 0,0161 | 1,4524 | -83,90715 | 47,95 |

como base a estimativa paramétricas dos modelos estimados na seção anterior. A Figura 4.12 mostra os mapas preditivos dos processos estocásticos. As imagens estão condizentes com os dados observados, ou seja, localizações com menores valores para a coordenada possuem menores valores em sua vizinhança. Por fim, a Tabela 4.7 apresenta os intervalos de variação das variâncias preditivas.

Figura 4.11: Krigagem para os parâmetros do modelo GEV ajustados (a) - σ , (b) - ξ e (c) - μ .

4.3.4 Análise de Resíduos

As Figuras 4.13 a 4.18 mostram o comportamento dos resíduos do modelo ajustado para as estimativas dos parâmetros da GEV e GPD como para o limiar utilizado. Para

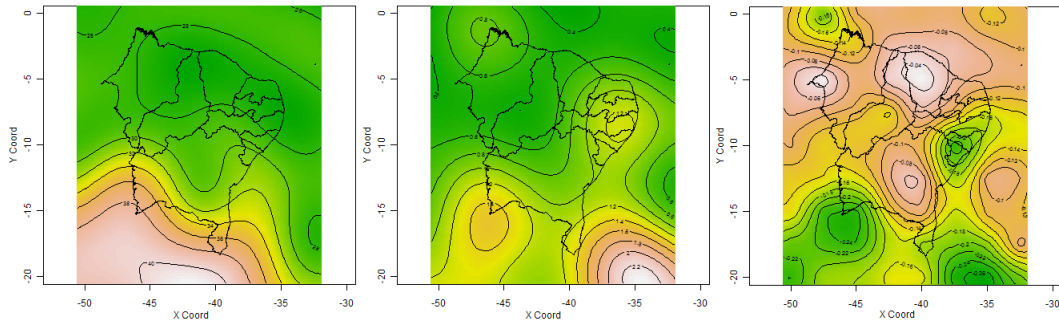


Figura 4.12: Krigagem para os parâmetros do modelo GPD ajustados (a) - *Limiar*, (b) - σ e (c) - ξ

Tabela 4.7: Intervalos de variação das variâncias preditivas para os modelos de *krigagem*

| Parâmetros | Variância Mínima | Variância Máxima |
|----------------|------------------|------------------|
| μ - GEV | 0,08412866 | 0,11505003 |
| σ - GEV | 0,000317161 | 0,012734518 |
| ξ - GEV | 0,0001274949 | 0,0122182960 |
| σ - GPD | 0,000317161 | 0,012734518 |
| ξ - GPD | 0,003329692 | 0,004260491 |

o limiar o comportamento dos resíduos indica aparentemente um fuga da suposição de normalidade, para o restante das estimativas dos parâmetros da GPD, o ajuste mostrou-se satisfatório. Para as estimativas dos parâmetros da GEV os gráficos dos quantis mostraram-se mais problemáticos com desvios mais fortes nas extremidades da distribuição. Apesar da necessidade da suposição de normalidade o método apresentado aqui será aplicado tanto para os parâmetros da GPD quanto para os estimativas dos parâmetros da GEV.

4.4 Regressão Quantílica

O modelo de regressão quantílica determina relações entre variáveis para qualquer quantil da distribuição. O nosso interesse está em determinar os extremos de qualquer Modelo de Circulação Global em regiões onde o mesmo não fornece informação relevante. Para isto, iremos obter uma relação entre um série da localização de interesse e a série do ponto da grade mais próximo da localização de interesse. Esta relação per-

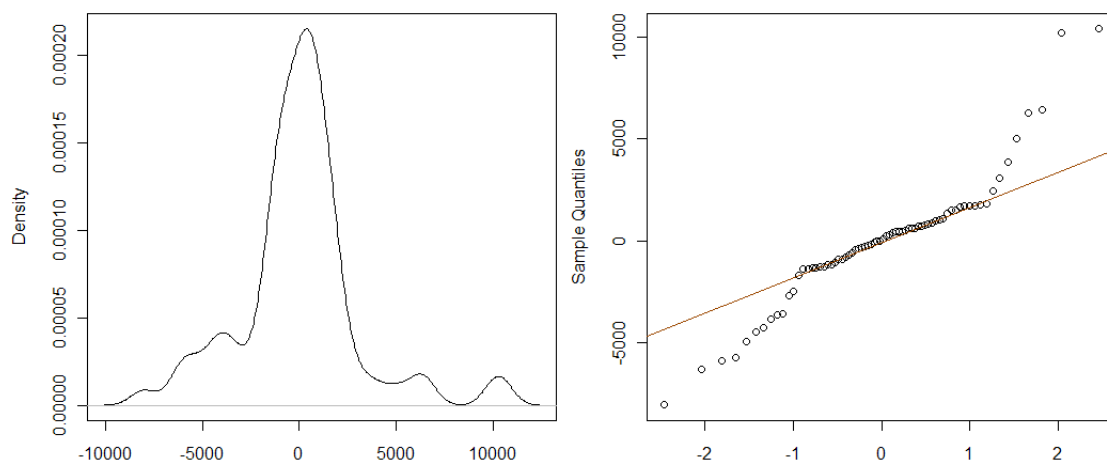


Figura 4.13: Densidade e Gráfico de quantis dos resíduos para o modelo final do limiar

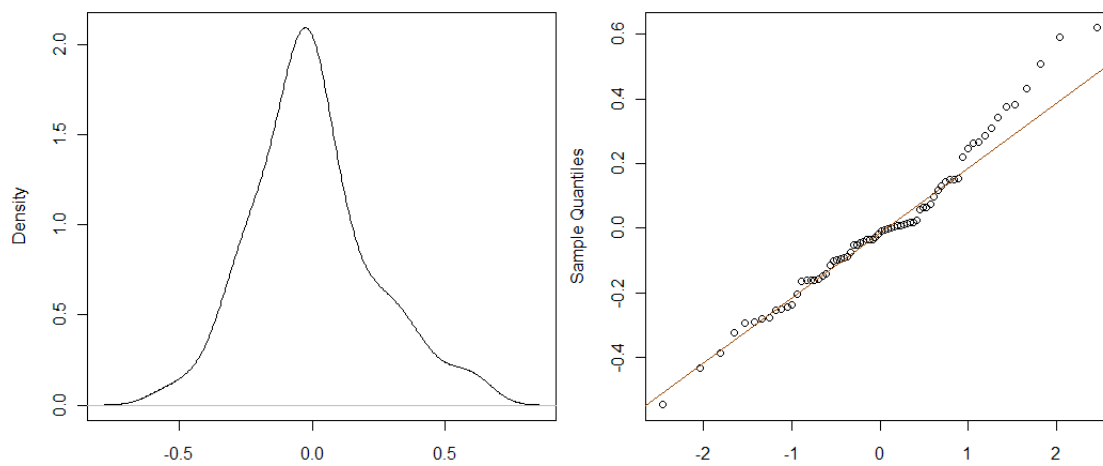


Figura 4.14: Densidade e Gráfico de quantis dos resíduos para o modelo final σ GPD

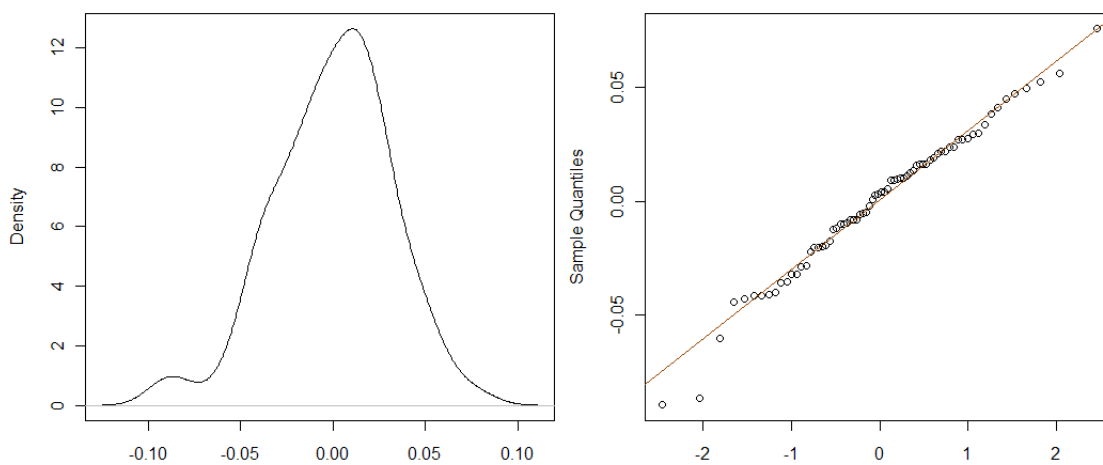


Figura 4.15: Densidade e Gráfico de quantis dos resíduos para o modelo final ξ GPD

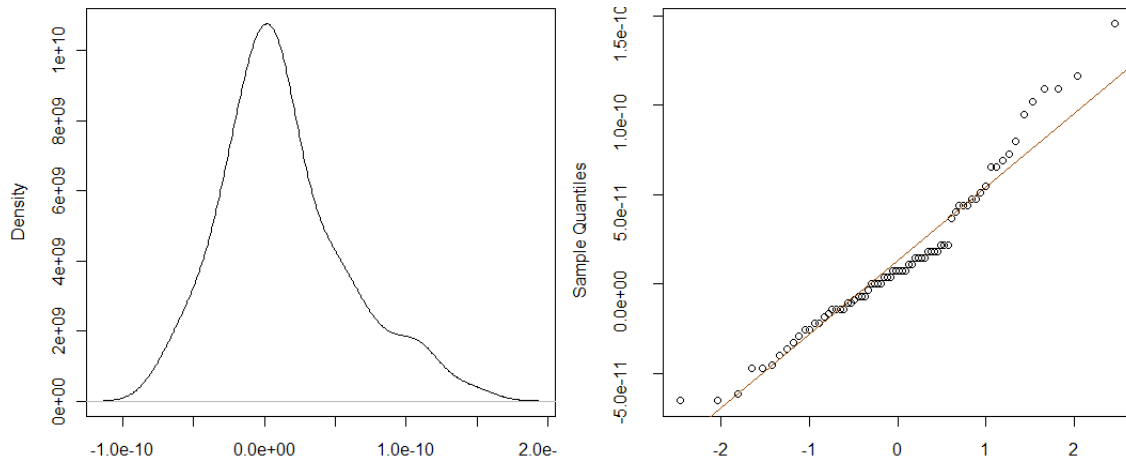


Figura 4.16: Densidade e Gráfico de quantis dos resíduos para o modelo final μ GEV

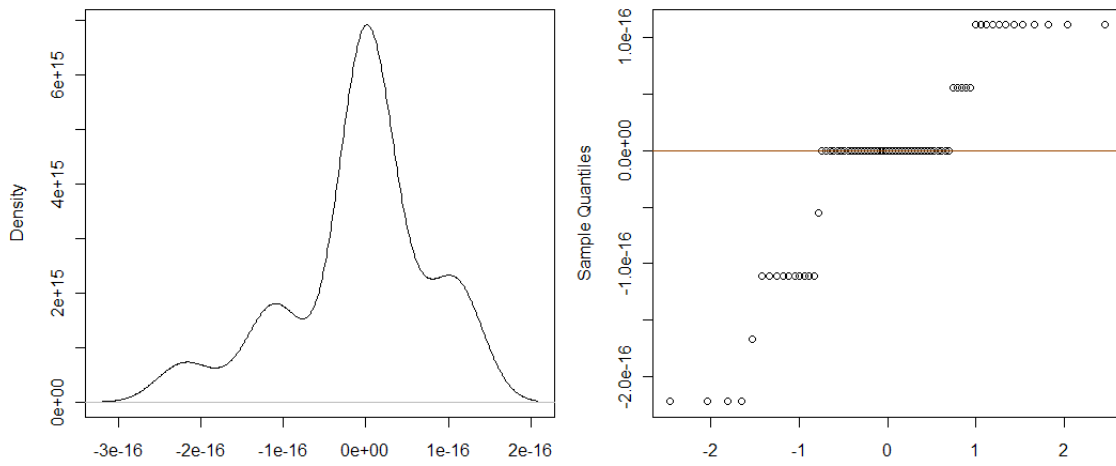


Figura 4.17: Densidade e Gráfico de quantis dos resíduos para o modelo final σ GEV

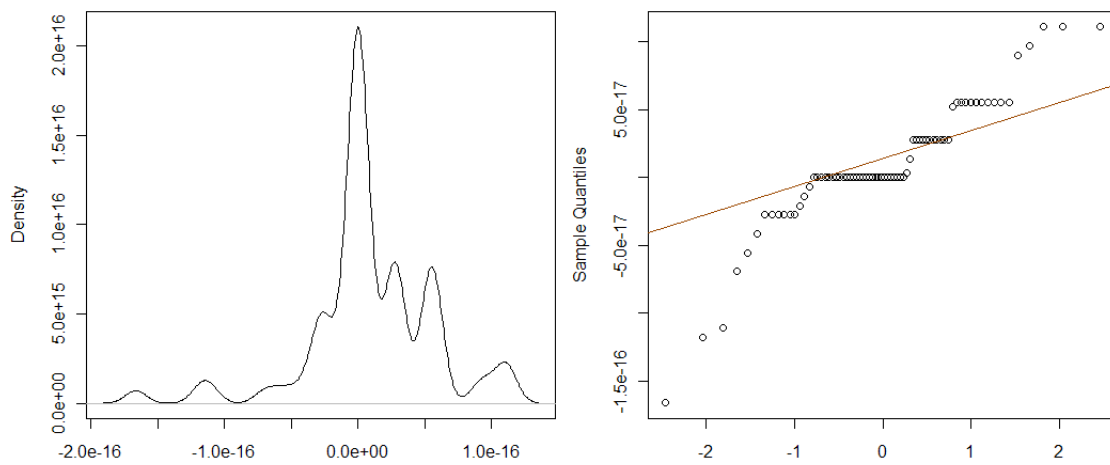


Figura 4.18: Densidade e Gráfico de quantis dos resíduos para o modelo final ξ GEV

mitirá a obtenção dos quantis para qualquer modelo de circulação global em qualquer localização, proporcionando o ajuste da distribuição de eventos extremos.

A Figura 4.19 mostra a gráfico de dispersão conjuntamente com os modelos de regressão ajustados para os quantis de 95% e 99% e o modelo linear. Nota-se claramente um certo desvio entre as retas do modelo linear e o modelo de regressão quantílica utilizando a mediana como quantil. A representação do modelo de regressão quantílica é apresentada de forma linear conforme a equação 3.18. A Tabela 4.8 mostra os parâmetros estimados e como comparação as estimativas dos modelos linear e quantil 50%.

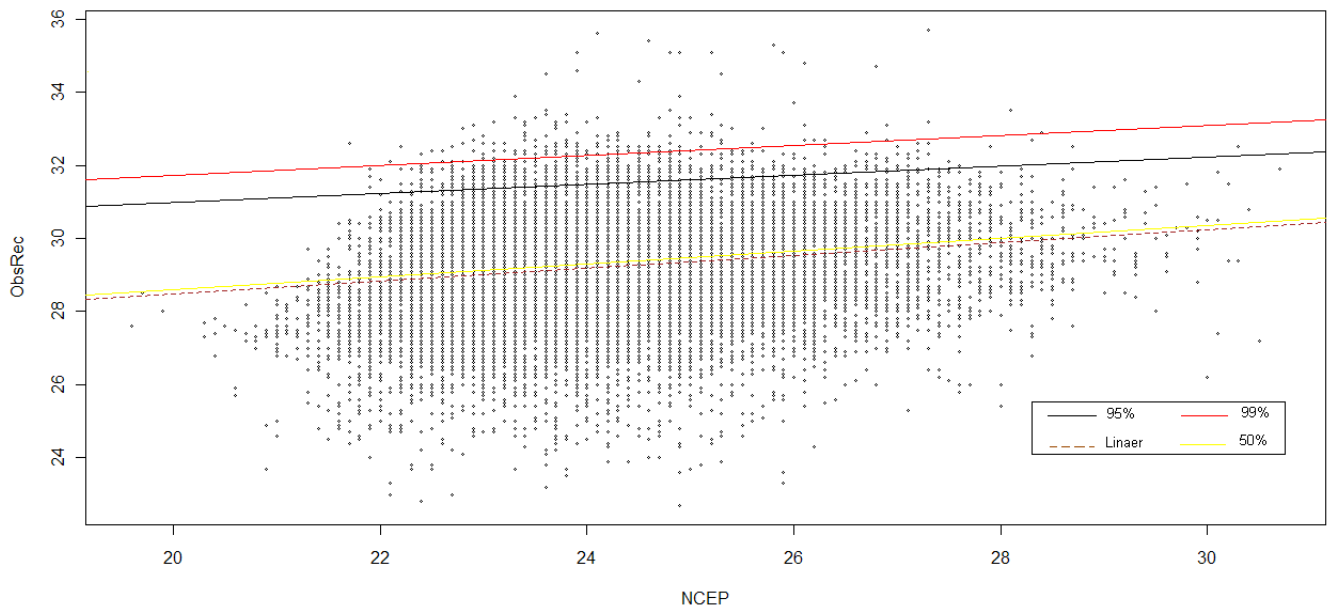


Figura 4.19: Modelos de Regressão Quantílicos os quantis 99% (vermelho), 90% (preto), 50% e modelo de regressão linear (amarela) .

4.4.1 TVE Aplicada aos Quantis 99% e 95%

Ao gerar os quantis dos dados do NCEP/NCAR sobre a estação localizada em Recife é possível realizar o ajuste da distribuição dos eventos extremos. A Figura 4.20 mostra as densidades estimadas dos quantis do NCEP/NCAR sobre a estação do Recife. Em

Tabela 4.8: Estimativas dos parâmetros do modelo de regressão quantílica para o quantis 95% e 99%

| Quantil | β_1 | β_2 |
|---------|-----------|-----------|
| 99% | 28.5 | 0.16 |
| 95% | 28.84 | 0.11 |
| 50% | 26.03 | 0.14 |
| Linear | 26.33 | 0.12 |

ambas as distribuições possuem uma certa assimetria com valores de mínimo de 31,0° e 31,7° e valores máximos 32,2° e 33,0° para os quantis de 95% e 99%, respectivamente.

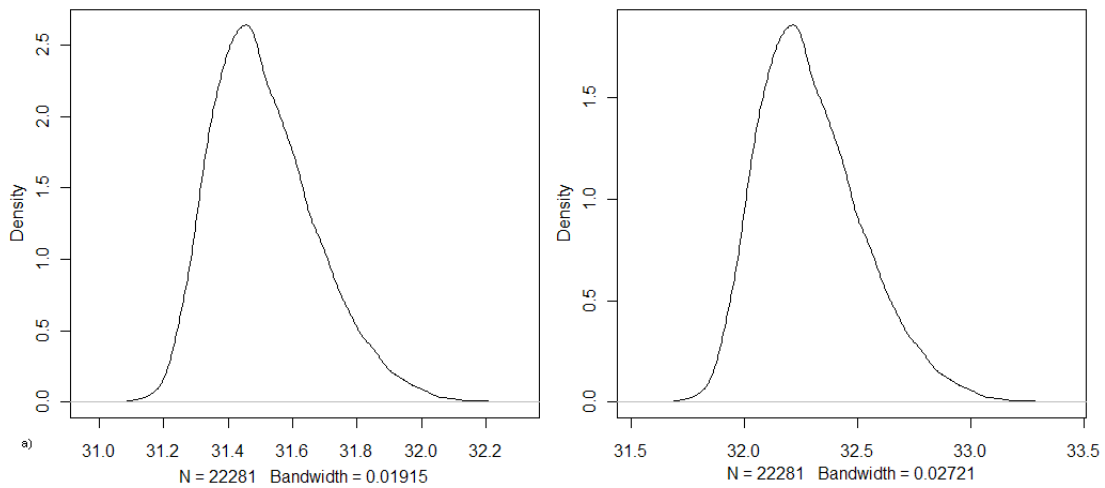


Figura 4.20: Densidades Estimadas dos Quantis de 95% e 99%

Análise GEV

A abordagem GEV necessita da criação de máximos em blocos. A série dos quantis possui o mesmo período dos dados originais do NCEP/NCAR, ou seja, 1 de janeiro de 1948 a 31 de dezembro de 2008. Atribuindo os anos como sendo os blocos teremos 61 valores máximos, podendo assim os eventos extremos serem ajustados a distribuição GEV. Utilizando o método da máxima verossimilhança os ajustes foram realizados produzindo os resultados contidos nas Tabelas 4.9 e 4.10. Os resultados mostram maior variância e média para a GEV do quantil 99% enquanto que o parâmetro de forma apresenta-se praticamente o mesmo, devido ao comportamento dos eventos extremos

ser o mesmo.

Tabela 4.9: Estimativas e Variância dos parâmetros GEV para o quantil 95%

| Modelo | μ | σ | ξ |
|-----------|---------|----------|---------|
| GEV | 31,9831 | 0,0983 | -0,2169 |
| Variância | 0,0143 | 0,0103 | 0,1033 |

Tabela 4.10: Estimativas e Variância dos parâmetros GEV para o quantil 99%

| Modelo | μ | σ | ξ |
|-----------|---------|----------|---------|
| GEV | 32,9656 | 0,1396 | -0,2168 |
| Variância | 0,0203 | 0,0146 | 0,1032 |

A qualidade do ajuste foi avaliada através da produção de gráficos de probabilidade, quantis, e estimação da densidade dos eventos extremos. As Figuras 4.21 e 4.22 apresentam estes resultados. Em ambos os casos o ajuste foi satisfatório. Para ambos, os gráfico de probabilidade e quantis, existe um tendência linear sobre a reta dos erros da distribuição GEV. O intervalo de confiança da estimativa do parâmetro de forma para ambos os quantis contêm o zero (Figuras 4.23 e 4.24) o que indica um distribuição Gumbel para os eventos extremos de Temperatura do Ar.

Análise GPD

A primeira etapa para realizar o ajuste é a definição do limiar para a obtenção dos valores que são extremos. Na literatura o método mais utilizado para a seleção é o Gráfico da Média dos Excessos. O comportamento dos Gráficos da Média dos Excessos (Figuras 4.25 e 4.26) mostram uma tendência linear e um melhor ajuste para os limiares $31,8^\circ$ e $32,5^\circ$ para os quantis de 95% e 99%, respectivamente. Com estes limiares o ajuste a GPD foi realizado utilizando o método da máxima verossimilhança, estimando os valores das Tabelas 4.11 e 4.12. Novamente a variância do modelo gerado a partir dos quantil 99% é maior enquanto que o parâmetro de forma são particamente semelhantes.

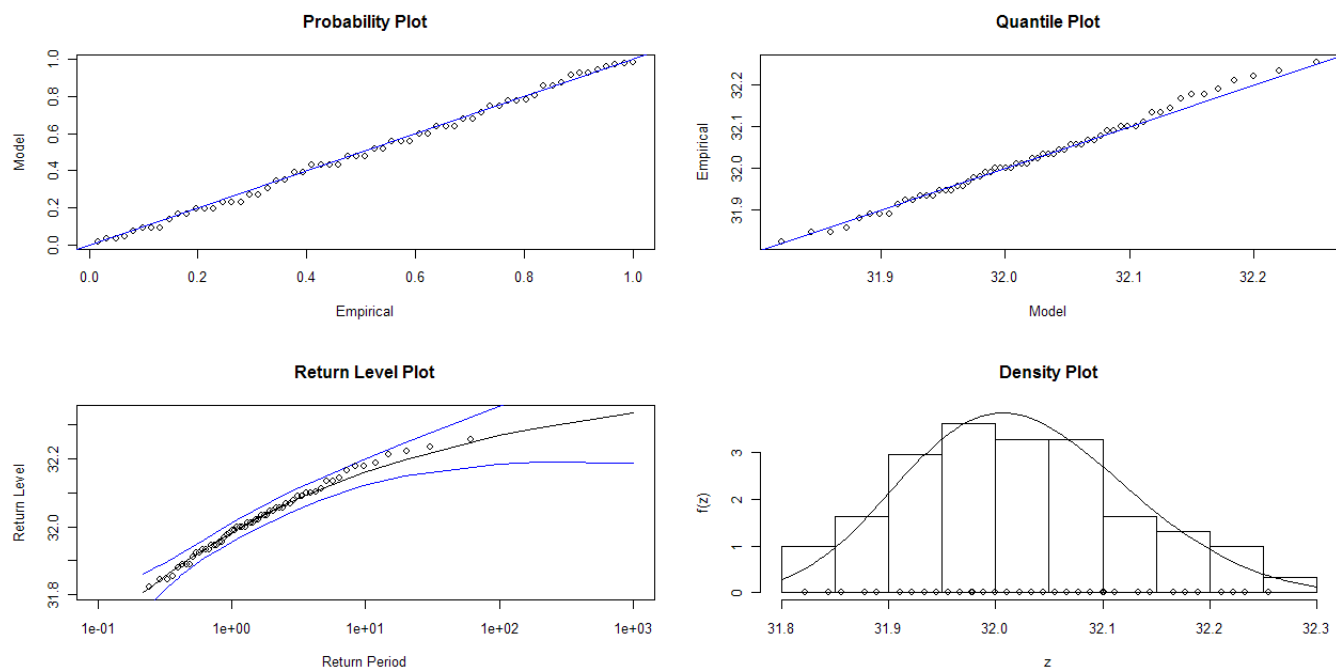


Figura 4.21: Gráficos de Diagnóstico para o ajuste aos quantis 95%

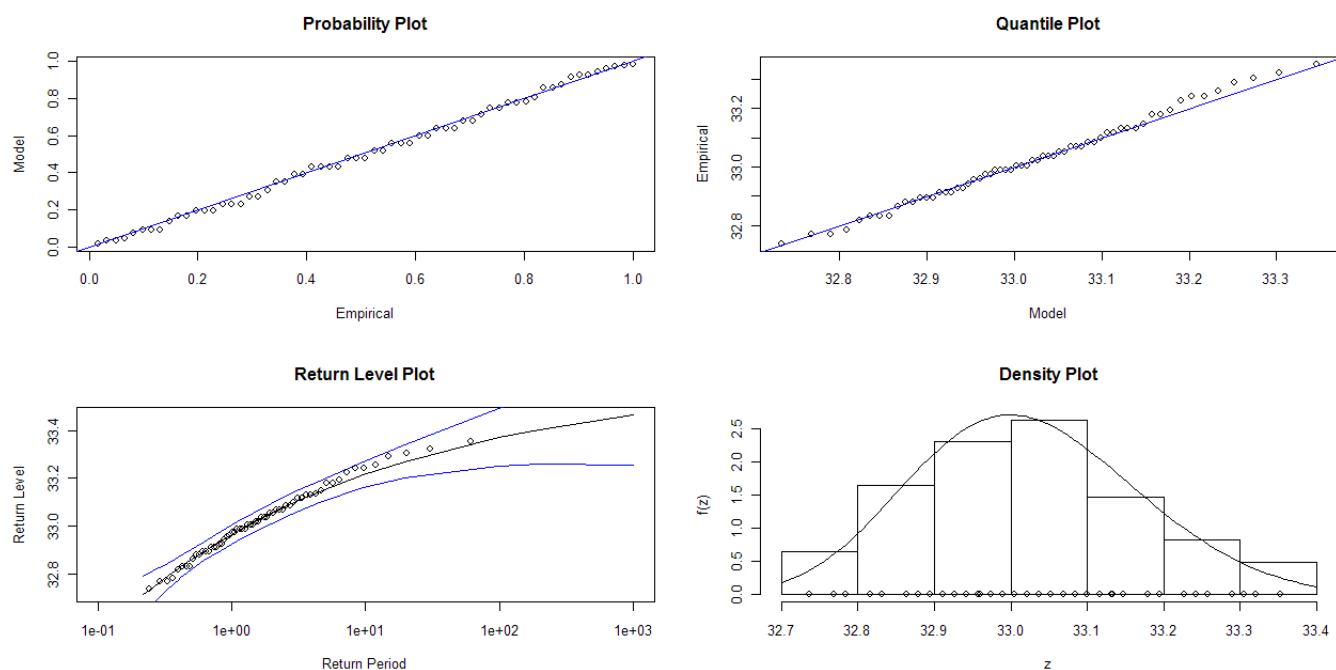


Figura 4.22: Gráficos de Diagnóstico para o ajuste ao quantis 99%

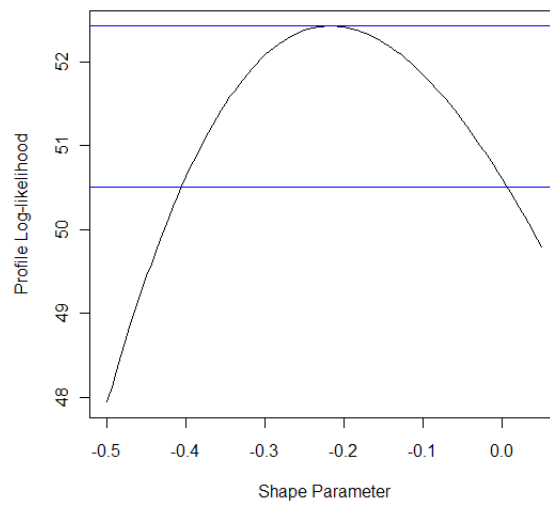


Figura 4.23: Intervalo de Confiança para a estimativa do parâmetros de forma quantil 95%

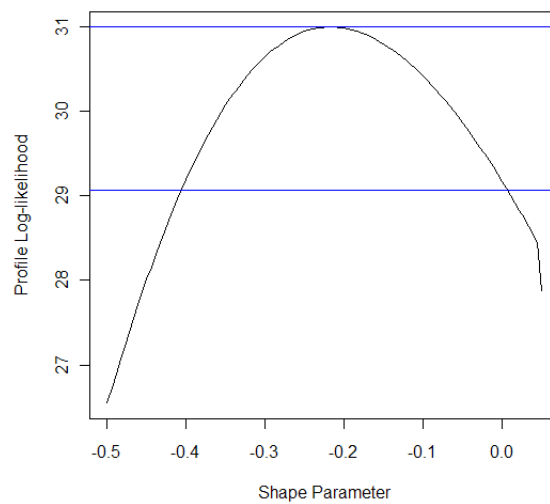


Figura 4.24: Intervalo de Confiança para a estimativa do parâmetros de forma quantil 99%

Tabela 4.11: Estimativas e Variância dos parâmetros da GPD para o quantil 95%

| Modelo | Limiar | σ | ξ |
|-----------|--------|----------|---------|
| GPD | 31,8 | 0,1190 | -0,2069 |
| Variância | | 0,00425 | 0,0221 |

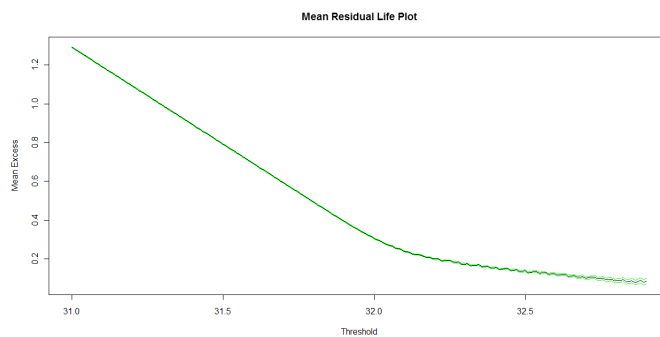


Figura 4.25: Gráficos da Médias dos Excessos para o quantil 99%

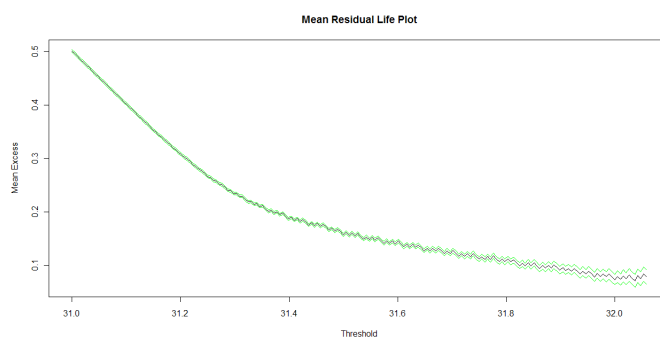


Figura 4.26: Gráficos da Médias dos Excessos para o quantil 95%

Tabela 4.12: Estimativas e Variância dos parâmetros da GPD para o quantil 99%

| Modelo | Limiar | σ | ξ |
|-----------|--------|----------|---------|
| GEV | 32,5 | 0,2049 | -0,2023 |
| Variância | | 0,0039 | 0,0116 |

4.5 Comparações dos Resultados

Como resumo dos resultados obtidos, a Tabela 4.13 apresenta as estimativas dos parâmetros da GEV e GPD produzidos por regressão quantílica (R. Q) por interpolação espacial (I. E.) e a partir dos dados observados. Os resultados obtidos para as estimativas da GPD mostram-se mais satisfatório principalmente em relação ao parâmetro de forma ξ , apresentando melhores resultados para o método de interpolação espacial, ocorrendo uma maior discrepância para o limiar dos dados observados. Enquanto que para a GEV, a estimativa do parâmetro de forma para os dados observados apresentou uma diferença significativa em relação a estimativa obtida a partir dos métodos de interpolação espacial e regressão quantílica.

Tabela 4.13: Resultados das estimativas dos parâmetros da GPD e GEV gerados pelos métodos apresentados

| | Limiar | σ - GPD | ξ - GPD | μ - GEV | σ - GEV | ξ - GEV |
|------------|---------|----------------|-------------|-------------|----------------|-------------|
| Dados Obs. | 30°C | 1,0 | -0,16 | 32,23°C | 0,92 | -0,04 |
| I. E. | 26,45°C | 1,16 | -0,17 | 28,57°C | 0,79 | -0,22 |
| R. Q. 95% | 31,8°C | 0,12 | -0,21 | 31,98°C | 0,10 | -0,22 |
| R. Q. 99% | 32,5°C | 0,21 | -0,20 | 32,97°C | 0,14 | -0,22 |

Capítulo 5

Considerações Finais

Este trabalho tem como objetivo a produção de modelos de eventos extremos de temperatura do ar para MGC, em localizações onde os Modelos Globais não fornecem informação. Como vimos, a primeira etapa consistiu em ajustar modelos GEV e GPD sobre os dados de reanálise de Temperatura do Ar do NCEP/NCAR, e a partir dos parâmetros ajustados obter superfícies interpoladas destes parâmetros. Com esta interpolação conheceremos a função distribuição GEV e GPD de localizações aonde não possuíamos informação, podendo assim, obter os níveis de retorno para qualquer Modelo Global de Circulação. A primeira parte do método consiste no ajuste do modelos extremos sobre as séries do MCG. Os ajustes dos modelos GEV e GPD apresentaram-se satisfatórios para as 72 séries dos dados do NCEP/NCAR. A obtenção da superfície interpolada das estimativas dos parâmetros dos modelos GEV e GPD, através da *krigagem*, necessita de um processo espacial gaussiano e a falha na suposição leva a inconsistência dos resultados. As estimativas produzidas por *krigagem* apresentaram um certa discrepância entre as estimativas obtidas através do ajuste aos dados observados principalmente para o parâmetro de forma do modelo GEV. Para o modelo GPD as estimativas são mais próximas da estimativas geradas pelos dados observados. O método de regressão quantílica, curiosamente, produziu a estimativa do parâmetro de forma igual ao estimado através da *krigagem* tanto para o quantil 95% quanto para o quantil 99%. Para os outros parâmetros a média e o limiar foram os que mais se distanciaram do valor

ajustado por dados observacionais com uma diferença de aproximadamente 4°C tanto para o limiar quanto para média. Esta diferença se deve ao fato da menor variabilidade espacial com um menor intensidade da média e limiar. Tais discrepâncias podem ser também devido a qualidade da informação fornecida pelos dados do NCEP/NCAR e a forma como os modelos produzem as observações. Os modelos de circulação globais produzem as informação baseada na média geral de cada pixel da grade, suavizando as variáveis climáticas sobre o espaço. Este forma de produção das informação pode ser notada comparando as estimativas produzidas pela interpolação espacial da média e o limiar. Uma consideração importante a ser realizada acerca da abordagem quantílica se deve ao fato desta abordagem necessitar de uma série na localização de interesse, o que inviabiliza o processo de obtenção de distribuições de eventos extremos em localizações onde os Modelos Globais de Circulação não fornecem informação relevante.

Referências

- [1] BAUTISTA, E. A. L. A Distribuição Generalizada de Valores Extremos no Estudo da Velocidade Máxima do Vento em Piracicaba, SP. 2002. 47f. Dissertação - Escola Superior de Agricultura Luiz de Queiroz. Abril/2002.
- [2] BEIJO, L. A. et al. Maximum Rainfall Return Period by Extreme Values Type I distribution in Larvas, Minas Gerais. Ciênc. agrotec., Lavras, v. 29, n. 3, p. 657-667, Junho/2005.
- [3] BENESTAD, R. E.; Chen, D.; Hanssen-Bauer, I. "Empirical - Statistical Down-scaling". Compendium, Disponível em: <http://home.broadpark.no/rbene>. Junho/2007.
- [4] CARVALHO, D. M., MORAES, R. M. Análise de conglomerados de homicídios por arma branca e arma de fogo na cidade de João Pessoa entre 2002 e 2005. In: VII Encontro Regional de Matemática Aplicada e Computação, Recife-PE, 2007.
- [5] CEREZER, S. M. Uso da Teoria de Valores Extremos para Estimar Valores de Pressões Hidrodinâmicas em um Ressalto Hidráulico Formado a Justante de um Vertedouro: O Caso de UHE Porto Colômbia. Abril/2008.
- [6] CHOWDHURY, J.U; J.R. STENDINGER; L.H. LU. Goodness of - Fit Test for Regional Generalized Extreme Value Flood Distributions. Water Resources Research, AGU, Vol.27(7), Julho/1991.
- [7] COLES, S.G. An Introduction to Statistical Modeling of Extreme Values. Springer-Verlag, London, 2001.

- [8] COLES, S. G.; PERICCHI, L. R. Anticipating catastrophes through extreme value modelling. *Applied Statistics*, London, v. 52, n. 3, p. 405-416, Out/2003.
- [9] COLES, S. R port and R documentation files by Alec Stephenson. ismev: An Introduction to Statistical Modeling of Extreme Values. R package version 1.3. 2006. <http://www.maths.lancs.ac.uk/stephena/>
- [10] DIGGLE, P. J.; RIBEIRO, Jr. P. J. *Model-Based geostatistics*. New York: Springer, 2007 230p.
- [11] DIGGLE, P. J.; RIBEIRO, Jr. P. J. geoR: a package for geostatistical analysis *R-NEWS*, 1(2):15-18. 2001
- [12] DRUCK, S.; CARVALHO, M.S.; CÂMARA, G.; MONTEIRO, A.V.M. (eds) "Análise Espacial de Dados Geográficos". Brasília, EMBRAPA, 2004 (ISBN: 85-383-260-6)
- [13] FISHER, R.A.; TIPPETT, L.H.C. Limiting forms of the frequency distribution of the largest or smallest number of the sample. *Proc. Cambridge Philos. Soc.* p180-190, 1928.
- [14] FONSECA, B. H. F. Um estudo sobre estimação e predição em modelos geoestatísticos bivariados. Escola Superior de Agricultura Luiz de Queiroz. Mestrado em Estatística e Experimentação Agronômica, 2009
- [15] FRIEDERICHS, P.; HENSE, A. Statistical Downscaling of Extreme Precipitation Events Using Censored Quantile Regression. *Monthly Weather Review*, 135 2365-2378, 2007
- [16] FRIEDERICHS, P. Statistical downscaling of extreme precipitation using extreme value theory. *Extremes* 13. 109-132. 2010.
- [17] FURRER, R.; NYCHKA, D.; SAIN, S. (2009). *fields: Tools for spatial data*. R package version 6.01. <http://CRAN.R-project.org/package=fields>
- [18] GALVÃO, C. O. e Trigo, R. M. "Downscaling da precipitação e temperatura de modelos de circulação global para a escala de bacia hidrográfica. In: IV SIM-

- PÓSIO DE HIDRÁULICA E RECURSOS HÍDRICOS DOS PAÍSES DE LÍNGUA OFICIAL PORTUGUESA. Coimbra: Portugal. Em CD-ROM, 1999.
- [19] GLAHN, H.R. and LOWRY D.A. "The use of model output statistics (MOS) in objective weather forecasting." *J.Appl.Meteor.*, 11, 1203-1211, 1972.
- [20] GNEDENKO, P. B. Sur la Distribution Limite du Terme Maximum d'une Série Aléatoire. *Annals of Mathematics*, v. 44, n. 3, p. 423-453, 1943.
- [21] GIORGI, F. "Simulation of regional climate using a limited area model in a general circulation model", *J. Clim.*, Vol3, pág. 942-963, 1990
- [22] HANN, Z. Actuarial Modelling of Extremal Events Using Transformed Generalized Extreme Value Distributions and Generalized Pareto Distributions., Tese 91f. Universidade do Estado de Ohio, 2003.
- [23] HANSSEN-BAUER, I.; ACHBERGER, C.; BENESTAD, R. E.; CHEN, D.; FORLAND, E. J."Statistical downscaling of climate scenarios over Scandinavia" *Clim. Res.* Vol. 29: 255-268, 2005
- [24] HEWISTON, B. C. ; CRANE, R. G. "Climate downscaling: techniques and application", *Clim. Res.*, Vol. 7, pág. 85-95, 1996.
- [25] HOSKING, J.R.M.; WALLIS, J.R. ; WOOD, E.F.. An appraisal of the regional flood frequency procedure in the UK flood studies report, *Hydrological Science Journal*, 30(1), 85-109. 1985.
- [26] HUIJBREGTS, C. J. Regionalized variables and quantitative analysis of spatial data. In: Davis, J. C. McCullagh, M. J. (ed) *Display and analysis of spatial data*. New York, John Wiley. p. 38-53. 1975.
- [27] IPCC-TGCIA, Guidelines on the Use of Scenario Data for Climate Impact and Adaptation Assessment. Version 1. In. Carter, T.R.; Hulme, M.; Lal, M. Intergovernmental Panel on Climate Change, Task Group on Scenarios for Climate Impact Assessment, p.66, 1999.

- [28] JENKINSON, A. F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly Journal of the Royal Meteorological Society, v. 81, p.158-171, Abril/1955.
- [29] JOYCE, P. et al. A General Extreme Value Theory Model for the Adaptation of DNA Sequences Under Strong Selection and Weak Mutation. Genetics 180: 1627-1643 Novembro/2008.
- [30] JOURNEL, A. G. Fundamentals of geostatistics in five lessons. California, Stanford Center for Reservoir Forecasting Applied Earth Sciences Department, 1988.
- [31] KLEIN, W.H., LEWIS, B.M., ENGER, I. Objective Prediction of Five Day Mean Temperatures during Winter. Journal of Meteorology, 16, 672–681, 1959.
- [32] KOENKER, R., BASSETT, G. Regression Quantiles. Econometrica, 46, 33-45, 1978
- [33] KOENKER, R., BASSETT, G. quantreg: Quantile Regression. R package version 4.36. <http://CRAN.R-project.org/package=quantreg> 2009.
- [34] MAO, Q.; McNIDER, R. T.; MUELLERE, S. F.; JUANG, H. An Optimal Model Output Calibration Algorithm Suitable for objective temperature forecasting. Analysis and Forecasting, v.14, p.190-202, 1998.
- [35] MATÈRN, B. Spatial variation. Verlag, Berlin: Springer, 365 p. 1986.
- [36] NICHOLAS, J. et. al. maptools: Tools for reading and handling spatial objects. R package version 0.7-13. 2008.
- PIERCE, D. ncdf: Interface to Unidata netCDF data files. R package version 1.6. <http://cirrus.ucsd.edu/pierce/ncdf>
- [37] PINTO, L. I. C. et. al. "Comparação entre os campos de precipitação de reanálise do CPTEC e NCEP/NCAR para América do Sul."Congresso Brasileiro de Meteorologia Edição XIV - Florianópolis - 2006

- [38] PIELKE, R. A., COTTON, W. R., WALKO, R. L., TREMBACK, C. J., LYONS, W. A., GRASSO, L. D., NOCHOLLS, M. E., MORAN, M. D., WESLEY, D. A., LEE, T. J. e COPELAND, J. H. "A comprehensive meteorological modeling system – RAMS". *Meteorology Atmospheric Physics*, v.49, p.69-91, 1992.
- [39] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, 2009. URL <http://www.R-project.org>
- [40] RAMIREZ, M. C. V.; "Previsão e Análise da Precipitação sobre as regiões sudeste e sul do brasil utilizando redes neurais artificiais". 12 de dezembro de 2003.
- [41] RAMOS, A. M. "Desagregação Espacial da Precipitação Simulada por Modelos Atmosféricos no Nordeste do Brasil". Dissertação apresentada ao Curso de Pós-Graduação em Engenharia Civil, na área de Engenharia de Recursos Hídricos, em cumprimento às exigências para obtenção do Grau de Mestre. Campina Grande, Dezembro/2000.
- [42] RIBATET, M. POT: Generalized Pareto Distribution and Peaks Over Threshold. R package version 1.0-7. 2008. <http://r-forge.r-project.org/projects/pot/>
- [43] SANSIGOLO, C. A.; NERY, J. T. Distribuições de extremos de temperatura mínima no Paraná. *Ver. Bras. Agromet*, v. 8, n. 2, p. 247-253,2000.
- [44] SCHANBENBERGER, O.; GOTWAY, C. A. Statistical methods for spatial data analysis. BPCA Raton: Chapman and Hall, 2005. 488p
- [45] SCHLATHER, M. An introduction to positive definite functions and to unconditional simulation of random fields. Technical Report ST 99-10, Lancaster University 1999.
- [46] SILVA, R. R. Distribuição generalizada de Pareto e mistura de distribuições de Gumbel no estudo da vazão e da velocidade máxima do vento em Piracicaba, SP. Escola Superior de Agricultura Luiz de Queiroz, Dissertação de Mestrado em Agronomia. Área de concentração: Estatística e Experimentação

- Agronômica, 2008.
- [47] SMITH, R.L; Extreme Value Theory Based On the r Largest Annual Evensts. Journal of Hydrology. Amsterdan, v. 86,n.1, p.27-43.1986.
- [48] SMITH, R.L; Extreme Value Analysis of Environmental Times Series: An Application to Trend Detection in Ground-Level ozone. Statistical Science, Hayward, v. 4, n. 4, p. 367-393,1989.
- [49] TABWIN 3.5. DataSus - Ministério da Saúde. <http://www.datasus.gov.br>. 2009.
- [50] VENABLES, W. N.; RIPLEY, B. D. Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0, 2002.
- [51] VIOLA, M.L.L.,Teoria de Valores Extremos e Cópulas: Distribuição Valor Extremo Generalizada e Cópulas Arquimedianas Generalizadas Trivariadas, Dissertação. 179f. Campinas/SP 2006.
- [52] WILBY, R. L., WILGEY, T. M. L., CONWAY, D., JONES, P. D., HEWISTON, B. C., MAIN, J. e WILKS, D. S. "Statistical downscaling of general circulation model output: a comparison of methods". Water Resources Research, Washington, v.34, n.11 p.2995-3008, 1998.

Apêndice A

Programas

Neste apêndice apresentamos alguns algoritmos produzidos no *Software R*.

Algoritmo para edição dos dados brutos do NCEP/NCAR,

```
Edit = function(cam = NULL)
{
  ## Lendo os dados do arquivo no formato NetCDF ##
  require(ncdf)
  nc <- open.ncdf(cam)
  Dados <- get.var.ncdf(nc)
  lat = get.var.ncdf(nc, "lat")
  lon = get.var.ncdf(nc, "lon")
  close.ncdf(nc)
  dados = Dados[c(53:60),c(37:45),]
  ##### Pacote de Mapas #####
  require(maptools)
  test = function(x)
  {
    Seq2 = seq(0,dim(dados)[3],4)
    Seq2 = Seq2[-1]
    Seq1 = seq(1,(dim(dados)[3] - 3),4)
    a = 0
    for(k in 1:366)
    {
      a = rbind(a,Seq1[k],Seq2[k])
    }
  }
}
```

```

    }
    return(a)
  }
  a = test(1)
  a = a[-1]
  A = matrix(0,dim(dados)[3]/4,72)
  for(k in 1:(dim(dados)[3]/4))
  {
    l = 0
    for(i in 1:8)
    {
      l = l + 1
      for(j in 1:9)
      {
        l = l + 1
        A[k,l-i] = max(dados[i,j,a[2*k-1]:a[2*k]])
      }
    }
  }
  return(A)
}

```

DADOS = DADOS - 273

Algoritmo para a edição dos máximos em blocos:

```
##### Modelagem GEV #####
```

```

library(ismev)
library(POT)
DADOS = read.csv2("")
Dados = DADOS - 273
jpeg("BoxPlot_Max_Diarioa.jpeg")
boxplot(as.data.frame(DADOS))
dev.off()

```

```
# Função para obter o máximo por anos, dos 48 anos de série.
```

```

MaxBlocos = function(Local)
{
  DADOS = read.csv2("")
  MaxTemp = as.vector(rep(0,61))
  dados = DADOS[,Local]
  MaxTemp[1] = max(dados[1:366])
  MaxTemp[61] = max(dados[21916:22281])
  for(i in 2:60)
    {
      s = seq(1,61,4)
      string = i == s
      bi = 0
      if(length(string[string == TRUE]) == 1){(bi = bi + 1) &
        (a = 366*i - 366 - 3*bi) & (b = 366*i - bi*3)}
      else{(a = 365*(i-1) + (2 + bi)) & (b = 365*i + (1 + bi))}
      MaxTemp[i] = max(dados[a:b])
    }
  return(MaxTemp)
}

Max = function(x)
{
  Maximos = matrix(c(rep(0,61*72)),61,72)
  for(i in 1:72)
    {
      Maximos[,i] = MaxBlocos(i)
    }
  return(Maximos)
}

Maxi = Max(NULL)
write.csv2(Maxi,"Max_Blocos_Temp.csv", row.names = FALSE)
boxplot(as.data.frame(Maxi))

```

Algoritmo para a modelagem de extremos via GEV

```

# Dados Máximo Anuais #
Dados = read.csv2("") # Os pontos da grade estão configurados da seguinte forma(1 a

##### Modelagem GEV #####

#####
# Algoritmo para Construção dos Gráficos de Diagnósticos e Estimativas #
#####

Diag_Gev = function(x)
  {
    require(ismev)
    require(POT)
    Dados = read.csv2("")
    names = read.csv2("")
    names = as.matrix(names)
    i = 1
    Max = gev.fit(Dados[,i])
    Est = Max$mle
    for(i in 2:72)
      {
        print("XXXXXXXXXXXXXXXXXXXXXXXXXXXX")
        Max = gev.fit(Dados[,i])
        Est = rbind(Est,Max$mle)
        jpeg(names[i])
        gev.diag(Max)
        dev.off()
      }
    write.csv2(Est, file = "Est.csv", sep = ";")
  }
Diag_Gev(x)

#####
# Gráfico do Intervalo de Confiança para o parâmetro de forma #
#####
Gev_profxi = function(x)

```

```

    {
        library(ismev)
        library(POT)
        Dados = read.csv2("")
        names = read.csv2("''')
        names = as.matrix(names)

        for(i in 1:72)
            {
                print("XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX")
                Max = gev.fit(Dados[,i])
                jpeg(names[i])
                gev.profxi(Max,-.5,0.5)
                dev.off()
            }
    }
Gev_profxi(x)

#####
# Algoritmo para construção do intervalo de confiança #
# para todos os pontos da grade do NCEP                #
#####
Intervalos_ZM = function(x)
{
    # Algoritmo para construção dos intervalos de confiança dos Níveis de Retorno #
    Zm = function(i,N,alfa) # para o parametro de fomra diferente de 0.
    {
        library(ismev)
        library(POT)
        Dados = read.csv2("")
        p = 1/N
        mle = gev.fit(Dados[,i])
        est = as.vector(mle$mle)
        zm = est[1] - (est[2]/est[3])*(1 - (-log(1-p))^(1-est[3]))
        V = mle$cov
        DeltaT = matrix(c(1,-est[3]^(-1)*(1-(-log(1-p))^(1-est[3]))),

```

```

        est[2]*est[3]^(-2)*(1-(-log(1-p))^-est[3])) -
        est[2]*est[3]^(-1)*(-log(1-p))^-est[3])*log(-log(1-p))),1,3)
VAR = DeltaT%*%V%*%t(DeltaT)
LI = zm - qnorm(1- alfa/2)*sqrt(VAR)
LS = zm + qnorm(1 - alfa/2)*sqrt(VAR)
return(i, N, alfa, sigma = est[2],
xi = est[3], u = est[1], LI, LS, zm)
}
i = 1
NR1 = as.numeric(Zm(i,10,0.05))
NR2 = as.numeric(Zm(i,50,0.05))
NR3 = as.numeric(Zm(i,100,0.05))
NR4 = as.numeric(Zm(i,200,0.05))
ZM_IC = rbind(NR1,NR2,NR3,NR4)
for(i in 1:72)
{
        NR1 = as.numeric(Zm(i,10,0.05))
        NR2 = as.numeric(Zm(i,50,0.05))
        NR3 = as.numeric(Zm(i,100,0.05))
        NR4 = as.numeric(Zm(i,200,0.05))
        ZM_IC = rbind(ZM_IC,NR1,NR2,NR3,NR4)
}
write.csv2(ZM_IC, file = "ZM.csv")
}

```

Comando utilizados na Krigagem do parâmetro de forma:

```

#####
# Modelagem Geo #
#####
library(geoR)
library(MASS)
library(maptools)

#### Mapa So com os limites ####
read.map = function(filename)

```

```
{
  zz=file(filename,"rb")
  #
  # header do .map
  #
  versao = readBin(zz,"integer",1,size=2) # 100 = versao 1.00
  #Bounding Box
  Leste = readBin(zz,"numeric",1,size=4)
  Norte = readBin(zz,"numeric",1,size=4)
  Oeste = readBin(zz,"numeric",1,size=4)
  Sul   = readBin(zz,"numeric",1,size=4)

  geocodigo = ""
  nome = ""
  xleg = 0
  yleg = 0
  sede = FALSE
  poli = list()
  i = 0

  #
  # repete para cada objeto no arquivo
  #
  repeat{
    tipoobj = readBin(zz,"integer",1,size=1)
    if (length(tipoobj) == 0) break
    i = i + 1

    Len = readBin(zz,"integer",1,size=1) # length byte da string Pascal
    geocodigo[i] = readChar(zz,10)
    Len = readBin(zz,"integer",1,size=1) # length byte da string Pascal
    nome[i] = substr(readChar(zz,25),1,Len)
    xleg[i] = readBin(zz,"numeric",1,size=4)
    yleg[i] = readBin(zz,"numeric",1,size=4)
    numPontos = readBin(zz,"integer",1,size=2)
```

```
sede = sede || (tipoobj = 1)

x=0
y=0
for (j in 1:numPontos){
  x[j] = readBin(zz,"numeric",1,size=4)
  y[j] = readBin(zz,"numeric",1,size=4)
}

# NAs separam vários polígonos no mesmo objeto
# BUG a corrigir: Assim como está o primeiro polígono não fecha e,
# em múltiplos polígonos, há um NA a mais no final
xInic = x[1]
yInic = y[1]
for (j in 2:numPontos){
  if (x[j] == xInic & y[j] == yInic) {x[j]=NA; y[j] = NA}
}

poli[[i]] = c(x,y)
dim(poli[[i]]) = c(numPontos,2)
}

class(poli) = "polylist"
attr(poli,"region.id") = geocodigo
attr(poli,"region.name") = nome
attr(poli,"centroid") = list(x=xleg,y=yleg)
attr(poli,"sede") = sede
attr(poli,"maplim") = list(x=c(Oeste,Leste),y=c(Sul,Norte))

close(zz)
return(poli)
}

# Dados #
Xi = read.csv2("C:/Daniel/Mestrado/Dissertação/Dados/NCEP/Kigagem/Coord_xi.csv")
```



```
attach(Xi)
Xi = cbind(Long,Lat,xi)
Xi = as.geodata(Xi)
attach(Xi)

# Análise Descritiva #
summary(Xi$data)
sd(Xi$data)
boxplot(Xi$data)
points(Xi)
plot(Xi)

# Estimação AD-HOC para os chutes iniciais com tend cte ##
va1 = variog(Xi)
plot(va1)
eyefit(va1)

# Estimação AD-HOC para os chutes iniciais com tend oeste-leste ##
va2 = variog(Xi,trend = ~ coords[,1])
plot(va2)
eyefit(va2)

# Estimação AD-HOC para os chutes iniciais com tend norte-sul ##
va3 = variog(Xi,trend = ~ coords[,2])
plot(va3)
eyefit(va3)

# Estimação AD-HOC para os chutes iniciais com tend coordenadas ##
va4 = variog(Xi,trend = ~ coords)
plot(va4)
eyefit(va4)

### Estimação de M. V. com tendencia constante ##
est1.1 = likfit(Xi, ini.cov.pars = c(0.02,8.517745042255328),nugget= 0.002936,
+ fix.kappa = T,kappa = 0.41, cov.model = "powered.exponential")
est1.2 = likfit(Xi, ini.cov.pars = c(0.02,4.83),nugget= 0.002936,
```

```

+ kappa = 1, cov.model = "powered.exponential")
est1.3 = likfit(Xi, ini.cov.pars = c(0.20,4.83),
+ nugget = 0.002936562449978015,fix.kappa = T,kappa = 0.5)
est1.4 = likfit(Xi, ini.cov.pars = c(0.01,17.27),
+ nugget = 0.01,fix.kappa = T,kappa = 0.27)
est1.5 = likfit(Xi, ini.cov.pars = c(0.13,9.74),
+ nugget = 0.01,fix.kappa = T,kappa = 0.68)
est1.6 = likfit(Xi, ini.cov.pars = c(0.02,8.29),
nugget = 0,fix.kappa = T,kappa = 0.27)

```

```

### Estimação de M. V. com tendencia oeste-leste ##

```

```

est2.1 = likfit(Xi, trend =~coords[,1], ini.cov.pars = c(0.01968045263839677,6.91),
+ nugget= 0.0024600565797995963,cov.model = "exponential")
est2.2 = likfit(Xi, trend =~coords[,1], ini.cov.pars = c(0.14137333430770968,
+ 10.50), nugget= 0.01, fix.kappa = T,kappa = 0.54)
est2.3 = likfit(Xi, trend =~coords[,1], ini.cov.pars = c(0.01,1.38),
+ nugget = 0.01, fix.kappa = T, kappa = 0.41, cov.model = "powered.exponential")
est2.4 = likfit(Xi, trend =~coords[,1], ini.cov.pars = c(0.02,5.53),
+ nugget = 0.01,fix.kappa = T,kappa = 0.54)
est2.5 = likfit(Xi, trend =~coords[,1], ini.cov.pars = c(0.02,15.89),
+ nugget = 0.002,fix.kappa = T,kappa = 0.14)
est2.6 = likfit(Xi, trend =~coords[,1], ini.cov.pars = c(0.01968045263839677,3.14),
+ nugget = 0.01, fix.kappa = T, kappa = 0.41, cov.model = "powered.exponential")

```

```

### Estimação de M. V. com tendencia norte-sul ###

```

```

est3.1 = likfit(Xi, trend =~coords[,2], ini.cov.pars = c(0.03840144596400271,8.5177
+ ,nugget= 0.004800180745500339,cov.model = "exponential" )
est3.2 = likfit(Xi, trend =~coords[,2], ini.cov.pars = c(0.03,3.45),
+ nugget= 0.01, cov.model = "exponential")
est3.3 = likfit(Xi, trend =~coords[,2], ini.cov.pars = c(0.06,6.14),
+ nugget = 0.01,cov.model = "exponential")
est3.4 = likfit(Xi, trend =~coords[,2], ini.cov.pars = c(0.03,6.22),
+ nugget = 0.01,fix.kappa = T,kappa = 0.27)
est3.5 = likfit(Xi, trend =~coords[,2], ini.cov.pars = c(0.01,1.38),
+ nugget = 0.01, fix.kappa = T, kappa = 0.41, cov.model = "powered.exponential")
est3.6 = likfit(Xi, trend =~coords[,2], ini.cov.pars = c(0.01968045263839677,4.14),

```

```
+ nugget = 0.01, fix.kappa = T, kappa = 0.41, cov.model = "powered.exponential")
```

```
### Estimação de M. V. com tendencia norte-sul ###
```

```
est4.1 = likfit(Xi, trend =~coords, ini.cov.pars = c(0.03840144596400271,7.51774504
```

```
+ nugget= 0.004800180745500339, cov.model = "exponential")
```

```
est4.2 = likfit(Xi, trend =~coords, ini.cov.pars = c(0.03,6.91),
```

```
+ nugget= 0.004673965251595824,cov.model = "exponential")
```

```
est4.3 = likfit(Xi, trend =~coords, ini.cov.pars = c(0.04,4.83),
```

```
+ nugget = 0.004673965251595824,fix.kappa = T,kappa = 0.41)
```

```
est4.4 = likfit(Xi, trend =~coords, ini.cov.pars = c(0.04,7.6),
```

```
+ nugget = 0.004673965251595824,fix.kappa = T,kappa = 0.41)
```

```
est4.5 = likfit(Xi, trend =~coords, ini.cov.pars = c(0.01,1.38),
```

```
+ nugget = 0.01, fix.kappa = T, kappa = 0.41, cov.model = "powered.exponential")
```

```
est4.6 = likfit(Xi, trend =~coords, ini.cov.pars = c(0.01968045263839677,1.14),
```

```
+ nugget = 0.01, fix.kappa = T, kappa = 0.41, cov.model = "powered.exponential")
```

```
#### Seleção do Modelo ####
```

```
Log = c(
```

```
est1.1$log,
```

```
est1.2$log,
```

```
est1.3$log,
```

```
est1.4$log,
```

```
est1.5$log,
```

```
est1.6$log,
```

```
est2.1$log,
```

```
est2.2$log,
```

```
est2.3$log,
```

```
est2.4$log,
```

```
est2.5$log,
```

```
est2.6$log,
```

```
est3.1$log,
```

```
est3.2$log,
```

```
est3.3$log,
```

```
est3.4$log,
```

```
est3.5$log,
```

```
est3.6$log,
```

```
est4.1$log,  
est4.2$log,  
est4.3$log,  
est4.4$log,  
est4.5$log,  
est4.6$log)
```

```
AIC = c(  
est1.1$AIC,  
est1.2$AIC,  
est1.3$AIC,  
est1.4$AIC,  
est1.5$AIC,  
est1.6$AIC,  
est2.1$AIC,  
est2.2$AIC,  
est2.3$AIC,  
est2.4$AIC,  
est2.5$AIC,  
est2.6$AIC,  
est3.1$AIC,  
est3.2$AIC,  
est3.3$AIC,  
est3.4$AIC,  
est3.5$AIC,  
est3.6$AIC,  
est4.1$AIC,  
est4.2$AIC,  
est4.3$AIC,  
est4.4$AIC,  
est4.5$AIC,  
est4.6$AIC)
```

```
names = c(  
"est1.1",  
"est1.2",
```

```
"est1.3",  
"est1.4",  
"est1.5",  
"est1.6",  
"est2.1",  
"est2.2",  
"est2.3",  
"est2.4",  
"est2.5",  
"est2.6",  
"est3.1",  
"est3.2",  
"est3.3",  
"est3.4",  
"est3.5",  
"est3.6",  
"est4.1",  
"est4.2",  
"est4.3",  
"est4.4",  
"est4.5",  
"est4.6")
```

```
Selecao_Xi = cbind(Log,AIC)  
row.names(Selecao_Xi) = names  
write.csv2(Selecao_Xi, file = "Selecao_Xi.csv")
```

```
# Modelo Selecionado #
```

```
est1.5 = likfit(Xi, ini.cov.pars = c(0.13,9.74),  
+ nugget = 0.01,fix.kappa = T,kappa = 0.68)
```

```
## Estabelecendo os extremos do grid para predição ####
```

```
y = seq(-20.5,.5, l = 100)  
x = seq(-50.5,-32, l = 100)  
p1 = expand.grid(x,y)
```

```
## Predição ##
mapa2 = read.map("C:/Daniel/Mestrado/Dissertação/TabWin/Mapas/MRNEUF.MAP")
predi1 = krige.conv(Xi, loc = p1, krige = krige.control(obj = est1.5))

jpeg("C:/Daniel/Mestrado/Dissertação/Resultados/NCEP/Krige/Krige_Sigma2.jpeg")
par(mfrow=c(2,2))
image(predi1, col = terrain.colors(100), xlim =c(-50.5,-32))
plotpolys(mapa2,add = TRUE)
points(Xi, add = T)
contour(predi1, add = T)
dev.off()

## Análise de Resíduos ##
library(nortest)
plot(density(est1.5$model.components[,3]))
qqnorm(est1.5$model.components[,3])
qqline(est1.5$model.components[,3], col = 2)
ks.test(est1.5$model.components[,3], "pnorm", mean(est1.5$model.components[,3]),
+ sd(est1.5$model.components[,3]))
shapiro.test(est1.5$model.components[,3])
```

Apêndice B

Gráficos

Neste apêndice apresentamos os gráficos mais relevantes.

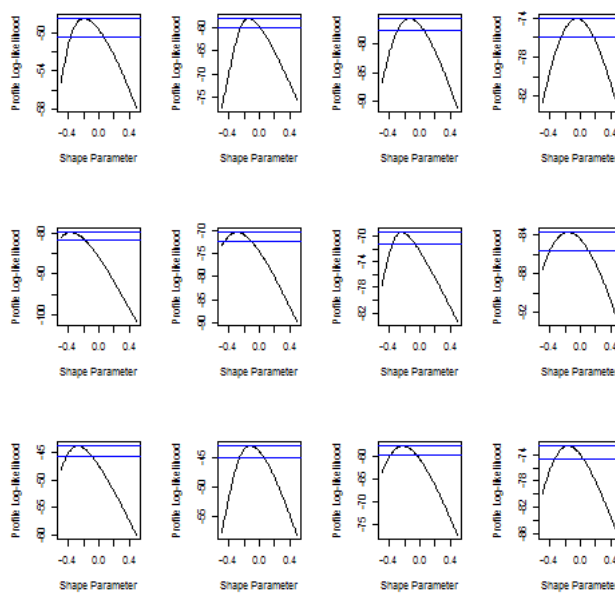


Figura B.1: Perfil de máxima verossimilhança das 12 séries de temperatura máxima do ar.

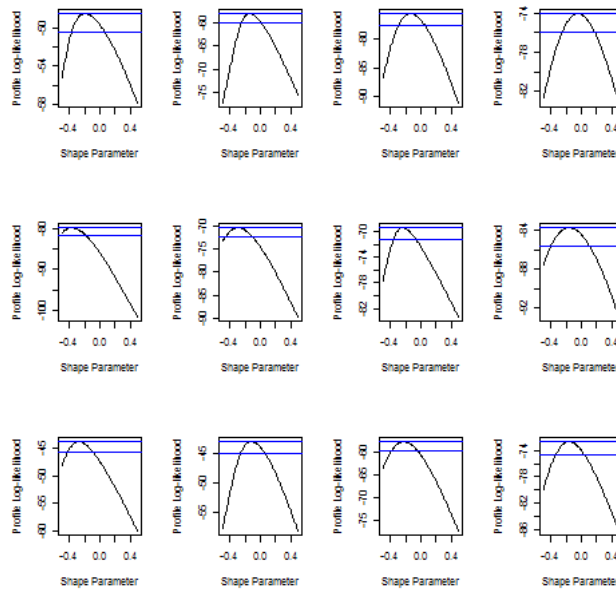


Figura B.2: Perfil de máxima verossimilhança das 12 séries de temperatura máxima do ar.

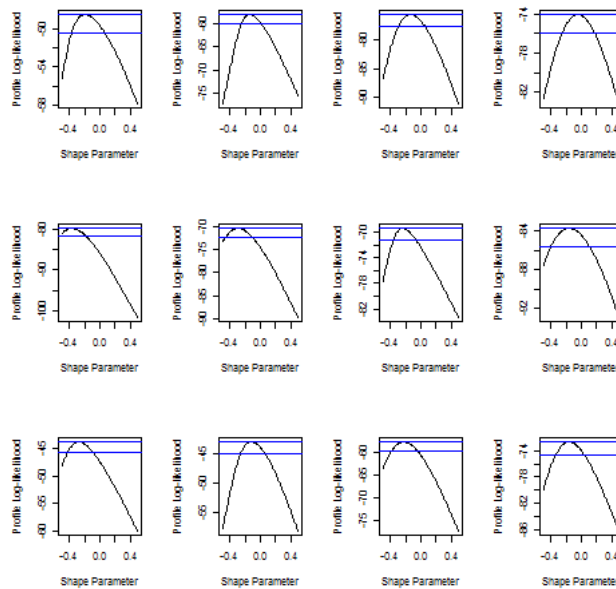


Figura B.3: Perfil de máxima verossimilhança das 12 séries de temperatura máxima do ar.

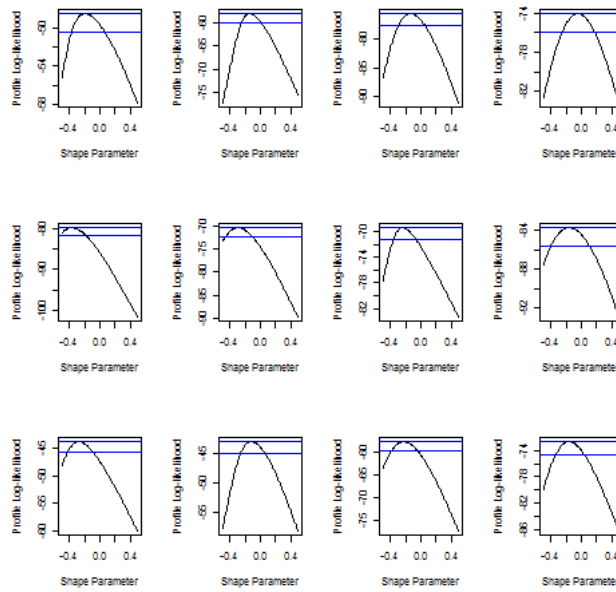


Figura B.4: Perfil de máxima verossimilhança das 12 séries de temperatura máxima do ar.

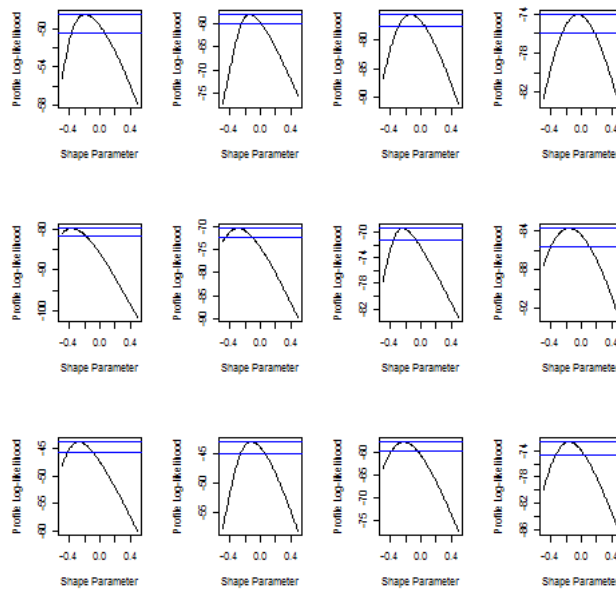


Figura B.5: Perfil de máxima verossimilhança das 12 séries de temperatura máxima do ar.

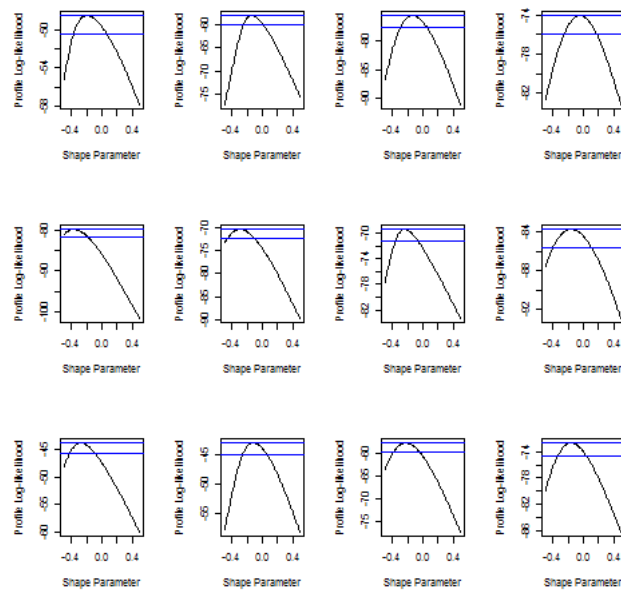


Figura B.6: Perfil de máxima verossimilhança das 12 séries de temperatura máxima do ar.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)