

Alexandre Henrique Quadros Gramosa

**Modelos bayesianos para análise de dados extremos
inflados de zeros**

Recife

27 de fevereiro de 2024



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

Modelos bayesianos para análise de dados extremos inflados de zeros

Trabalho apresentado ao Programa de Pós-Graduação em Biometria e Estatística Aplicada da Universidade Federal Rural de Pernambuco, em cumprimento com as exigências legais para defesa do trabalho de Tese de Doutorado.

Área de concentração: Biometria e Estatística Aplicada

Orientador: Dr. Frank Sinatra Gomes da Silva

Coorientador: Dr. Antonio Samuel da Silva

Coorientador: Dr. Fernando Ferraz do Nascimento

Recife

27 de fevereiro de 2024

Alexandre Henrique Quadros Gramosa

Modelos bayesianos para análise de dados extremos inflados de zeros/ Alexandre Henrique Quadros Gramosa. – Recife, 27 de fevereiro de 2024-
68 p. : il. (algumas color.) ; 30 cm.

Orientador: Dr. Frank Sinatra Gomes da Silva

– UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA, 27 de fevereiro de 2024.

1. Eventos extremos. 2. Inflados de zeros. 3. Análise Bayesiana. 4. Precipitação.
5. MCMC. 6. Sazonalidade. I. Frank Sinatra Gomes da Silva. II. Universidade Federal Rural do Piauí. III. Programa de Pós-Graduação em Biometria e Estatística Aplicada. IV. Doutorado em Biometria e Estatística Aplicada

CDU 02:141:005.7

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA**

Modelos bayesianos para análise de dados extremos inflados de zeros

Alexandre Henrique Quadros Gramosa

Trabalho apresentado ao Programa de Pós-Graduação em Biometria e Estatística Aplicada da Universidade Federal Rural de Pernambuco, em cumprimento com as exigências legais para defesa do trabalho de Tese de Doutorado.

Orientador:

Dr. Frank Sinatra Gomes da Silva
Orientador

Banca examinadora:

Dr. Antonio Samuel da Silva
UFRPE.

Dr. Fernando Ferraz do Nascimento
UFPI.

Dr. Josimar Mendes de Vasconcelos
UFRPE.

Dr. Fidel Ernesto Castro Morales
UFRN.

*Este trabalho é dedicado a toda minha família,
em especial a minha esposa e as minhas três filhas, aos meus pais e irmãos,
como também aos meus professores e amigos.*

Agradecimentos

Agradeço a Deus, por tudo que Ele fez, tem feito e fará. Agradeço-O pela minha vida, pela minha família, pelas conquistas alcançadas até aqui, alegrias, consolo e conforto em momentos difíceis, e esperança de dias melhores. Agradeço a minha amada família: minha esposa Eulenir e minhas filhas Geovana, Júlia e Luna, como também aos meus pais João e Socorro Gramosa e aos meus irmãos Daniel e Nayara, por serem sempre um porto seguro, dando-me forças para prosseguir, apoiando e incentivando-me em todos os momentos. Agradeço aos meus professores orientadores, professor Frank e Professor Samuel, que acolheram-me desde quando entrei para o PPGBEA/UFRPE, vindo de um outro estado, a saber de Teresina-PI, conduzindo-me até este momento da realização de um sonho. De maneira especial, agradeço também ao professor Fernando pela caminhada desde a graduação pela UFPI, no mestrado pela UFRN, e até aqui no Doutorado pela UFRPE. Agradeço também dos meus amigos mestrandos, doutorandos e doutores do PPGBEA/UFRPE, aos queridos professores, coordenadores, secretário e toda equipe da direção deste programa, que prontamente me receberam e me apoiaram. Agradeço de maneira geral a UFRPE pela disponibilidade de suas instalações. Agradeço a CAPES, pela bolsa de doutorado, que tem sido um apoio importantíssimo para produção das pesquisas. Que Deus continue abençoando a todos.

*“Mas, buscai primeiro o reino de Deus, e a sua justiça,
e todas estas coisas vos serão acrescentadas”.*
(Bíblia Sagrada, Mateus 6, 33)

Resumo

É de fundamental importância ter o conhecimento do resultado limite para a modelagem de máximos de blocos de tamanho n , conhecido como distribuição Generalizada de Valores Extremos (GEV) que é utilizada na modelagem de eventos extremos. Entretanto, pode ocorrer em que nesses dados extremos possa conter um número excessivo de zeros, o que dificulta as análises e estimações feitas pela distribuição GEV. A distribuição Generalizada de Valores Extremos Inflada de Zeros (ZIGEV) foi criada, recentemente, para solucionar esta problemática com o auxílio do seu parâmetro inflador ω . Um dos objetivos deste trabalho é fazer uma aplicação da nova distribuição ZIGEV, em dados de precipitação diária de chuva, transformados em blocos de máximos mensais. Nesses dados podem ocorrer meses em que não houve precipitação, sendo estes computados como zero. Foram analisadas séries temporais das mesorregiões do estado de Pernambuco, região nordeste do Brasil. Algumas delas com predominância de meses não chuvosos. É também objetivo desse trabalho a criação da distribuição Generalizada de Valores Extremos Inflada de Zeros Sazonais (SZIGEV), um modelo sazonal para analisar a sazonalidade de dados que tenham essas características, extremos e inflados de zeros. Neste caso, utiliza-se os dados de precipitação de chuva das cidades de Recife e Petrolina do estado de Pernambuco e das cidades de São João do Piauí e Teresina do estado do Piauí. Em ambas análises, as inferências foram feitas sobre o paradigma Bayesiano, com as estimações dos parâmetros sendo feitas por aproximações numéricas da distribuição *a posteriori* utilizando o método de Monte Carlo Markov Chain (MCMC). Os resultados dessas aplicações, em consonância com o trabalho de Gramosa et al. (2019), reforçaram que as análises e estimações feitas pela distribuição ZIGEV, em comparação com a distribuição GEV, foram mais precisas e com melhor qualidade de ajuste reforçando a importância de se utilizar a ZIGEV para modelar dados extremos, em especial quando esses estiverem inflados de zeros. Entretanto, comparando a aplicação da ZIGEV com a distribuição SZIGEV, percebemos que, pelo comportamento de sazonalidade que os dados em estudo apresentam, os resultados obtidos pelo modelo sazonal (SZIGEV) foram melhores e mais precisos, enfatizando a relevância do uso da distribuição SZIGEV para a modelagem de dados extremos inflados de zeros e sazonais.

Palavras-chave: eventos extremos, inflados de zeros, análise bayesiana, precipitação, MCMC, sazonalidade.

Abstract

It is of fundamental importance to have knowledge of the limit result for modeling maximum blocks of size n , known as Generalized distribution of Extreme Values (GEV) which is used in modeling extreme events. However, may occur that these extreme data may contain an excessive number of zeros, which complicates the analyzes and estimates made by the GEV distribution. The Zero-Inflated Generalized Distribution of Extreme Values (ZIGEV) was raised, recently, to solve this problem with the help of its inflator parameter ω . One of the objectives of this work is to make an application of the new ZIGEV distribution, in daily rainfall data, transformed into blocks of monthly highs. In these data there may be months in which there was no precipitation, these being computed as zero. Time series of the mesoregions of the state of Pernambuco were analyzed, northeast region of Brazil. Some of them with a predominance of non-rainy months. But, the main objective of this work is the creation of the Generalized distribution of Extreme Values Inflated by Seasonal Zeros (SZIGEV), a seasonal model to analyze the seasonality of data that have these characteristics, extremes and inflated zeros. In this case, rain precipitation data from the cities of Recife was used and Petrolina from the state of Pernambuco and from the cities of São João do Piauí and Teresina from the state of Piauí. In both analyzes, inferences were made on the Bayesian paradigm, with parameter estimates being made by numerical approximations of the distribution *a posteriori* using the Monte Carlo Markov Chain (MCMC) method. The results of these applications, in line with the work of Gramosa et. al (2019), reinforced that the analyses and estimates made by the ZIGEV distribution, compared to the GEV distribution, were more accurate and with better quality of adjustment, reinforcing the importance of using the ZIGEV to model extreme data, especially when these are inflated from zeros. However, comparing the application of the ZIGEV with the SZIGEV distribution, we noticed that, due to the seasonality behavior that the data under study present, the results obtained by the seasonal model (SZIGEV) were better and more accurate, emphasizing the relevance of the use of the SZIGEV distribution for the modeling of extreme data inflated from zeros and seasonals. However, comparing the application of the ZIGEV with the SZIGEV distribution, we noticed that, due to the seasonality behavior that the data under study present, the results obtained by the seasonal model (SZIGEV) were better and more accurate, emphasizing the relevance of the use of the SZIGEV distribution for the modeling of extreme data inflated from zeros and seasonals.

Key-words: extreme events, zero inflation, bayesian analysis, precipitation, MCMC, seasonality.

Lista de Figuras

| | |
|--|----|
| Figura 1 – Nove cidades de cinco mesorregiões do estado de Pernambuco | 11 |
| Figura 2 – Gráfico de série para os dados de máximos mensais da cidade de Petrolina | 17 |
| Figura 3 – Gráficos de séries de máximos mensais das cidades de Recife, Barreiros, Surubim, Garanhuns, Triunfo, Arcoverde, Ouricuri e Cabrobó | 18 |
| Figura 4 – Gráfico dos estimadores ω e S da distribuição ZIGEV para a cidade de Petrolina | 21 |
| Figura 5 – Gráfico dos níveis de retorno da GEV (vermelho) e ZIGEV (azul) para Recife, Barreiros, Surubim, Garanhuns, Triunfo, Arcoverde, Ouricuri e Cabrobó, com seus respectivos intervalos de confiança representados pelas linhas tracejadas | 26 |
| Figura 6 – Gráfico do nível de retorno pela ZIGEV para a cidade de Petrolina . . . | 27 |
| Figura 7 – Gráfico dos pontos dos máximos dos meses do ano nas cidades em estudo | 41 |
| Figura 8 – Gráfico dos zeros (mm) ao longo dos meses no ano para as cidades em estudo | 42 |
| Figura 9 – Histogramas dos betas, para os parâmetros μ , σ e ξ da SZIGEV, para a cidade de Recife | 45 |
| Figura 10 – Gráficos de séries para dos betas, para os parâmetros μ , σ e ξ da SZIGEV, para a cidade de Teresina | 47 |
| Figura 11 – Gráficos de séries para os betas, para os parâmetros μ , σ e ξ da SZIGEV, para a cidade de Petrolina | 49 |
| Figura 12 – Histogramas para os betas, para os parâmetros μ , σ e ξ da SZIGEV, para a cidade de São João do Piauí | 51 |
| Figura 13 – gráficos das estimativas de μ , σ e ξ para as cidade de Recife e Teresina | 55 |
| Figura 14 – gráficos das estimativas de μ , σ e ξ das cidades em estudo | 56 |
| Figura 15 – Quantis extremos de 90%, 95% e 99% da SZIGEV, para as cidades de Recife e Teresina | 59 |
| Figura 16 – Quantis extremos 90%, 95% e 99% da SZIGEV, para as cidade de Petrolina e São João do Piauí | 60 |
| Figura 17 – Quantis extremos 90%, 95% e 99% da SZIGEV, com as observações da amostra, para as cidades de Recife e Teresina | 61 |
| Figura 18 – Quantis extremos 90%, 95% e 99% da SZIGEV, com as observações da amostra, para as cidades de Petrolina e São João do Piauí | 62 |

Lista de tabelas

| | |
|--|----|
| Tabela 1 – Quantidade (S) e proporção de zeros (ω) nos dados pluviométricos das mesorregiões do estado de Pernambuco | 19 |
| Tabela 2 – Estimativas para os parâmetros das distribuições GEV e ZIGEV para as cidades em estudo | 20 |
| Tabela 3 – Comparação das medidas de ajuste para as distribuições GEV e ZIGEV nas cidades de Pernambuco | 22 |
| Tabela 4 – Comparação dos níveis de retornos correspondentes aos quantis extremos do GEV, ZIGEV e Empírico das cidades em estudo | 24 |
| Tabela 5 – Quantidade (S) e proporção de zeros (ω) nos dados de máximos de chuva dos estados de Pernambuco e do Piauí | 39 |
| Tabela 6 – Quantidade (S) e proporção de zeros, das cidades em estudo, observados em cada mês do ano | 40 |
| Tabela 7 – Estimativas para os parâmetros ω e S do modelo SZIGEV respectivo a cada mês do ano nos dados de Recife | 43 |
| Tabela 8 – Estimativas para os β 's (0,1,2, e 3) referente a cada parâmetro μ , σ e ξ do modelo SZIGEV, para os dados de Recife, com seus respectivos intervalos de confiança | 44 |
| Tabela 9 – Estimativas para os parâmetros ω e S do modelo SZIGEV respectivo a cada mês do ano nos dados de Teresina | 46 |
| Tabela 10 – Estimativas para os parâmetros μ , σ e ξ do modelo SZIGEV, nos dados de Teresina, com seus respectivos intervalos de confiança | 46 |
| Tabela 11 – Estimativas para os parâmetros ω e S do modelo SZIGEV respectivo a cada mês do ano nos dados de Petrolina | 48 |
| Tabela 12 – Estimativas para os parâmetros μ , σ e ξ do modelo SZIGEV, nos dados de Petrolina, com seus respectivos intervalos de confiança | 48 |
| Tabela 13 – Estimativas para os parâmetros ω e S do modelo SZIGEV respectivo a cada mês do ano nos dados de São João do Piauí | 50 |
| Tabela 14 – Estimativas para os parâmetros μ , σ e ξ do modelo SZIGEV, nos dados de São João do Piauí, com seus respectivos intervalos de confiança | 50 |
| Tabela 15 – Medidas de qualidade de ajustes para Recife, Teresina, Petrolina e São João do Piauí, para comparação entre os modelos ZIGEV e SZIGEV | 52 |

| | |
|---|----|
| Tabela 16 – Estimativas para os estimadores $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\xi}$ em cada mês do ano, para as cidades de Recife, Teresina, Petrolina e São João do Piauí | 54 |
| Tabela 17 – Níveis de retornos correspondentes aos quantis extremos da SZIGEV para as cidades em estudo | 57 |

Sumário

| | | |
|----------|---|-----------|
| 1 | Introdução | 1 |
| 2 | Objetivos | 3 |
| 2.1 | Objetivo Geral | 3 |
| 2.2 | Objetivos específicos | 3 |
| 3 | Revisão de Literatura | 4 |
| 3.1 | Análise bayesiana | 4 |
| 3.1.1 | Distribuição a priori | 4 |
| 3.1.2 | Distribuição a posteriori | 4 |
| 3.2 | Eventos extremos | 5 |
| 3.2.1 | Distribuição dos máximos e mínimos | 6 |
| 3.2.2 | A distribuição de Valores Extremos Generalizada (GEV) | 7 |
| 3.3 | Dados inflados de zeros | 7 |
| 3.3.1 | Distribuição Generalizada de Valores Extremos Inflada de Zeros (ZIGEV) | 8 |
| 3.4 | Próximos capítulos | 8 |
| 4 | Aplicação das distribuições GEV e ZIGEV a dados de chuva do estado de Pernambuco | 9 |
| 4.1 | Introdução | 9 |
| 4.2 | Material e métodos | 10 |
| 4.2.1 | Área de estudo e dados | 10 |
| 4.3 | Distribuição Generalizada de Valores Extremos (GEV) | 12 |
| 4.3.1 | Inferência bayesiana para GEV | 12 |
| 4.3.2 | Estimação dos níveis de retorno para GEV | 13 |
| 4.4 | Distribuição Generalizada de Valores Extremos Inflada de Zeros (ZIGEV) | 14 |
| 4.4.1 | Inferência bayesiana para ZIGEV | 14 |
| 4.4.2 | Estimação dos níveis de retorno para ZIGEV | 16 |
| 4.5 | Resultados e Discussões | 17 |
| 4.6 | Conclusão | 27 |
| 5 | Distribuição generalizada de valores extremos inflada de zeros sazonais (SZIGEV) | 29 |
| 5.1 | Introdução | 29 |
| 5.2 | Material e métodos | 30 |

| | | |
|-------|---|----|
| 5.2.1 | Área de estudo e dados | 30 |
| 5.3 | Distribuição generalizada de valores extremos inflada de zeros sazonais (SZIGEV) | 31 |
| 5.3.1 | Função de verossimilhança | 32 |
| 5.3.2 | Parametrização do ω em função dos meses do ano | 33 |
| 5.3.3 | Distribuição a priori | 35 |
| 5.3.4 | Distribuição a posteriori | 36 |
| 5.3.5 | Função log-posteriori | 37 |
| 5.4 | Aplicação da SZIGEV em dados de chuvas em algumas cidades dos estados de Pernambuco e Piauí | 38 |
| 5.4.1 | Cidade de Recife | 43 |
| 5.4.2 | Cidade de Teresina | 45 |
| 5.4.3 | Cidade de Petrolina | 48 |
| 5.4.4 | Cidade de São João do Piauí | 50 |
| 5.5 | Medidas de ajustes | 51 |
| 5.6 | Níveis de retornos | 53 |
| 5.6.1 | Estimadores $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\xi}$ | 53 |
| 5.6.2 | Quantis extremos (90, 95 e 99) | 57 |
| 5.6.3 | Níveis de retorno com observações da amostra | 61 |
| 5.7 | Conclusão | 63 |
| 6 | Considerações Finais | 64 |
| 7 | Referências | 65 |

1 Introdução

A sociedade civil tem enfrentado enormes perdas, tanto humanas como materiais e financeiras, devido às mudanças climáticas que tem potencializado a ocorrência de eventos extremos, causando destruição por onde quer que esses eventos passem. O poder público e a população em geral necessitam de ferramentas que os auxiliem na compreensão do padrão dessas ocorrências e efeitos, para mensurarem o impacto que esses eventos podem causar a sociedade (NASCIMENTO et al., 2011).

Os eventos ocasionados por chuvas extremas são as principais causas de inundações repentinas e deslizamentos de terra pelo mundo, dos quais em condições extremas de chuvas durante curtos períodos de tempo causam desastres (MODRICKA, et al., 2015; STALEY et al., 2015). Ressalta-se que os eventos de chuvas extremas sempre existiram, porém se constata que eles aumentaram em frequência e intensidade no contexto das mudanças climáticas, conforme pesquisas, como as de Jost et al. (2012).

Faz-se necessário a existência de um mecanismo, como um modelo probabilístico, que proporcione as autoridades governamentais e a população em geral, informações sobre as ocorrências desses eventos extremos que tanto prejuízos promovem a sociedade, para que de posse dessas informações possam traçar estratégias de combate e prevenção (GAMERMAN et al., 2006).

A teoria de valores extremos (TVE) é aplicada com a finalidade de descrever o comportamento de eventos raros e tem sido aplicada com sucesso no tratamento estatístico de dados meteorológicos, tais como precipitações máximas, temperaturas mínimas, ventos máximos, entre outros (Beijo & Avelar, 2011).

Os primeiros trabalhos foram com Fisher-Tippett (1928) e Gnedenko (1943), que contribuíram muito para o avanço do estudo da Teoria de Valores Extremos (TVE), tanto com as distribuições limites para o máximo, Gumbel, Fréchet e Weibull, na qual tem-se as condições necessárias para as distribuições assintóticas. A TVE teve um outro grande avanço com os trabalhos de von Mises (1954) e Jenkinson (1955), com o surgimento da distribuição Generalizada de Valores Extremos (GEV). Uma única distribuição, que engloba as três distribuições limites citadas acima, para modelar o máximo coletado em blocos de tamanho n .

No entanto, os trabalhos de Ottero et al. (2018), Gonçalves e Gutter (2019) e Gramosa et al. (2019), utilizando a distribuição GEV para modelar seus dados de máximos, obtiveram dificuldades em suas análises, pois, os dados estudados continham uma grande proporção de zeros. É comum encontrar uma porcentagem de valores zeros nos conjuntos de dados extremos. Esses zeros em excesso dificultam, e impedem a elaboração de uma análise estatística precisa para o problema. Os modelos usuais desenvolvidos não ajustam bem os dados nessa situação (MARTIN, et al., 2005).

A precipitação é, dentre os parâmetros meteorológicos, a que apresenta a maior variabilidade temporal e espacial sendo, portanto, a de maior dificuldade de estimativa. Com isto, tornam-se fundamentais a caracterização do seu padrão local e suas variações (ALMEIDA et al., 2011).

Recentemente, Gramosa et al. (2019), desenvolveram a distribuição Generalizada de Valores Extremos Inflada de Zeros (ZIGEV), com o objetivo principal de solucionar essa grande problemática na modelagem de dados de valores extremos inflados de zeros. A distribuição mostrou resultados bastante expressivos, desde as simulações, como também nas aplicações a dados reais e testes de medidas de qualidade de ajustes, alcançando precisão em suas estimativas, mesmo em situações extremas com 80% de zeros nos dados. Uma característica importante é que, quando a proporção de zeros nos dados de máximos mensais for aproximadamente zero, $\omega \rightarrow 0$, a distribuição ZIGEV converge para GEV, sendo assim a GEV é um caso particular da ZIGEV (GRAMOSA et al., 2019).

Os dados utilizados neste trabalho referem-se as séries de precipitações diárias de chuvas, coletadas nas estações de Recife, Barreiros, Surubim, Garanhuns, Triunfo, Arcoverde, Ouricuri, Cabrobó e Petrolina do estado de Pernambuco e das estações de São João do Piauí e Teresina do estado do Piauí, ambos os estados pertencentes a região nordeste do Brasil, região semiárida, com predominância de meses não chuvosos. Esses dados foram distribuídos em blocos mensais e então separado o maior valor registrado de cada bloco, resultando em dados de máximos mensais, assim utilizados nas análises.

2 Objetivos

2.1 Objetivo Geral

É objetivo geral deste trabalho, apresentar novos modelos para analisar dados de chuvas intensas e extremas. Nesta proposta, são generalizados os modelos propostos por NASCIMENTO et al. (2011) e GRAMOSA et al. (2019).

2.2 Objetivos específicos

Aplicar a distribuição ZIGEV, modelo recém criado, em dados de precipitação de chuvas de Pernambuco. Na aplicação será utilizada também a distribuição GEV, e então, faremos uma comparação com os resultados obtidos pelos dois modelos. A motivação em utilizar o modelo ZIGEV deve-se ao fato de que existe um longo período sem chuva em regiões secas e por consequência, o máximo em um determinado bloco tem uma probabilidade razoável de ser zero.

Criar uma distribuição que modele bem os dados de valores extremos inflados de zeros e que possa também modelar a sazonalidade existente nesses dados, tornando assim suas estimações mais precisas e com melhor qualidade de ajuste. Então, aplicar esse modelo proposto a dados de chuvas de Pernambuco e Piauí e comparar com os resultados obtidos na primeira aplicação dessa Tese, como também os obtidos em GRAMOSA et al. (2019).

3 Revisão de Literatura

3.1 Análise bayesiana

O contexto geral no qual a inferência bayesiana funciona é semelhante ao da inferência clássica: há um parâmetro populacional θ a respeito do qual deseja-se fazer inferências, e um mecanismo probabilístico $f(y|\theta)$ que determina a probabilidade de observar cada valor diferente de y , sob diferentes valores de θ . A diferença essencial entretanto é a de que θ é tratado como um quantidade aleatória. Isto pode parecer bastante inócuo, mas de fato leva a abordagens substancialmente diferentes à modelagem e inferência. Essencialmente, a inferência vai ser baseada em $f(\theta|y)$ ao invés de $f(y|\theta)$; isto é, a probabilidade do parâmetro condicional aos dados obtidos, ao invés da dos dados, condicional o valor do parâmetro. Em muitas formas tal abordagem leva a inferências muito mais naturais, mas para atingir isto será necessário especificar, adicionando ao modelo, uma distribuição a priori de probabilidades, $f(\theta)$, que representa crenças sobre a distribuição de θ antes de se considerar qualquer informação proveniente dos dados. Esta noção de distribuição a priori para o parâmetro θ está no cerne do pensamento bayesiano. Dependendo, discute-se com um defensor ou oponente da metodologia, tal fato é ou sua principal vantagem sobre a teoria clássica, ou sua maior armadilha (COLES e RIBEIRO, 2022).

3.1.1 Distribuição a priori

Em quase todas as situações, quando se tenta estimar um parâmetro θ , tem-se algum conhecimento ou crença sobre o valor de θ antes de considerar os dados (O'HAGAN, 1994). Invariavelmente, temos algum conhecimento sobre o processo sendo investigado antes de obter os dados. É essencial que as inferências devem ser baseadas na informação combinada que o conhecimento a priori e os dados representam.

3.1.2 Distribuição a posteriori

O Teorema de Bayes expresso em termos de variáveis aleatórias, com densidades denotadas genericamente por $f(\cdot)$, tem a forma:

$$f(\theta|y) = \frac{f(\theta)f(y|\theta)}{f(y)} = \frac{f(\theta)f(y|\theta)}{\int f(\theta)f(y|\theta)d\theta}.$$

Esta notação será usada em ambos os casos, com Y discreta ou contínua. No caso contínuo f é a função de densidade de probabilidade como usual, mas no caso discreto, f é a função de massa de probabilidade de Y [$P(Y = y)$]. De forma similar, θ pode ser discreto ou contínuo mas, caso seja discreto, $\int f(\theta)f(y|\theta)d\theta$ deve ser interpretado como $\sum_j f(\theta_j)f(y|\theta_j)$.

Pode-se dizer que:

$$\text{posteriori} = \frac{\text{priori} \times \text{verossimilhança}}{\text{marginal}}.$$

Note que o denominador no Teorema de Bayes é uma função apenas de y , resultante de uma integração em θ (ou seja, θ foi “integrado fora”). Desta forma, uma outra maneira de escrever o Teorema de Bayes é: $f(\theta|y) \propto f(\theta)f(y|\theta)$ ou, em palavras, “a posteriori é a proporcional da priori vezes a verossimilhança”.

3.2 Eventos extremos

Em estudos de temperatura, o aquecimento global tem sido um assunto em destaque nos últimos anos devido as rápidas alterações climáticas que o planeta vem sofrendo nas últimas décadas. Estas mudanças implicam em alterações no número de eventos extremos de temperatura, seja de máximos ou mínimos (NASCIMENTO, 2012). Eventos de temperatura extrema são mais responsáveis por mudanças na natureza do que mudanças na temperatura média (PARMESAN et al., 2000).

Para analisar a ocorrência de grandes catástrofes e perdas, a Teoria de Valores Extremos (TVE) tem auxiliado na predição de valores extremos. Através da TVE, pode-se formular um modelo específico para estimar dados extremos e suas probabilidades.

Quando falamos de valores extremos, máximos, lembramos de valores que estão próximos ao limite superior da distribuição. Sendo o máximo, a maior observação de uma amostra de valores x_1, x_2, \dots, x_n de uma variável X , denotada por:

$$M_n = \max(x_1, x_2, \dots, x_n).$$

Em uma amostra em que todos os indivíduos são independentes, e identicamente distribuídos (iid), se conhecermos a distribuição desta variável F_X , então é possível encontrar a distribuição do máximo, dado por

$$\begin{aligned} P(M_n < y) &= P[\max(X_1, X_2, \dots, X_n) < y] = P(X_1 < y, X_2 < y, \dots, X_n < y) \\ &= \prod_{i=1}^n P(X_i < y) = P(X < y)^n \\ &= F_X(y)^n. \end{aligned}$$

A distribuição do mínimo $M_1 = \min(X_1, X_2, \dots, X_n)$, é encontrada por

$$\begin{aligned} P(M_1 < y) &= P[\min(X_1, X_2, \dots, X_n) < y] = 1 - P[\min(X_1, X_2, \dots, X_n) > y] \\ &= 1 - P(X_1 > y, X_2 > y, \dots, X_n > y) \\ &= 1 - \prod_{i=1}^n P(X_i > y) \\ &= 1 - [1 - F_X(y)]^n, \end{aligned}$$

outra maneira de encontrar o mínimo é utilizando a relação

$$M_1 = -\max(-x_1, -x_2, \dots, -x_n).$$

3.2.1 Distribuição dos máximos e mínimos

A origem da caracterização assintótica de máximo amostral tem início em Fisher e Tippet (1928), cujos argumentos foram complementados e formalizados por Gnedenko (1943), através da normalização linear da variável M_n , como vemos a seguir

$$M_n^* = \frac{M_n - b_n}{a_n}.$$

Escolhas apropriadas de a_n e b_n estabilizam a posição e a escala de M_n quando n cresce, evitando as dificuldades que aparecem com a variável M_n . Procura-se, portanto, distribuições limite para M_n^* , ao invés de M_n .

Teorema 2.3.1 [Fisher-Tippet (1928) e Gnedenko (1943)]: Se existem seqüências de constantes $a_n > 0$ e b_n tais que

$$Pr \left\{ \frac{M_n - b_n}{a_n} \leq x \right\} \rightarrow G(x) \text{ quando } n \rightarrow \infty,$$

em que G é uma função de distribuição não-degenerada, então G pertence a uma das seguintes famílias:

$$\text{Gumbel} : H_I(y) = \left\{ \exp \left\{ -\exp \left[-\left(\frac{y-b}{a} \right) \right] \right\}, \quad -\infty < y < \infty, \right.$$

$$\text{Fréchet} : H_{II}(y) = \begin{cases} 0, & \text{se } y \leq b \\ \exp \left\{ -\exp \left[-\left(\frac{y-b}{\alpha} \right) \right] \right\}, & \text{se } y > b \end{cases},$$

$$\text{Weibull} : H_{III}(x) = \begin{cases} \exp \left\{ -\exp \left[-\left(\frac{y-b}{\alpha} \right) \right] \right\}, & \text{se } y \leq b \\ 1, & \text{se } y > b \end{cases},$$

para parâmetros $a > 0$, b , e no caso das famílias II e III, $\alpha > 0$.

3.2.2 A distribuição de Valores Extremos Generalizada (GEV)

Ao invés de trabalhar com três distribuições distintas descritas em Fisher e Tippett (1928), von Mises (1954) e Jenkinson (1955) propuseram a distribuição GEV, que engloba as três distribuições do Teorema de Fisher-Tippett. Esta função é denotada por H e possui a seguinte função de distribuição:

$$H(y|\xi, \sigma, \mu) = \begin{cases} \exp \left\{ -\left(1 + \xi \left(\frac{y - \mu}{\sigma}\right)^{-1/\xi}\right) \right\}, & \text{se } \xi \neq 0 \\ \exp \left\{ -\exp \left\{ -\left(\frac{y - \mu}{\sigma}\right) \right\} \right\}, & \text{se } \xi = 0 \end{cases}$$

definida em $\{y : 1 - \xi(y - \mu)/\sigma > 0\}$.

Estimativas de quantis extremos dos máximos são obtidas invertendo a equação acima, fazendo $z_p = H^{-1}(1 - p)$, obtendo assim

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi} \left[1 - (-\log(1 - p))^{-\xi}\right], & \text{se } \xi \neq 0 \\ \mu - \sigma \log[-\log(1 - p)]\{1 - [-\log(1 - p)]\}, & \text{se } \xi = 0 \end{cases}$$

em que $H(z_p) = 1 - p$. O termo z_p é conhecido como nível de retorno associado ao período $1/p$, ou seja, é o valor esperado que os dados originais ultrapassem o valor z_p pelo menos uma vez a cada $1/p$ períodos de tempo.

3.3 Dados inflados de zeros

Em muitos estudos, na análise de dados, é possível encontrar uma grande quantidade de zeros, dificultando assim a elaboração de uma análise estatística precisa para o problema, já que os modelos usuais desenvolvidos não ajustam bem tal situação.

Em aplicações de valores extremos pode-se encontrar uma inflação de zeros nos dados, surgindo a necessidade de uma distribuição específica que modele bem os dados de valores extremos com a presença excessiva de zeros, buscando uma maior precisão na estimação dos parâmetros. Assim, respeitando todos os respectivos pressupostos das distribuições, através de mistura da distribuição de Valores Extremos Generalizada e da distribuição Inflada de Zeros, desenvolve-se uma nova Distribuição denominada de Distribuição de Valores Extremos Generalizada Inflada de Zeros (ZIGEV) (GRAMOSA, 2019).

3.3.1 Distribuição Generalizada de Valores Extremos Inflada de Zeros (ZIGEV)

A distribuição ZIGEV tem como característica analisar dados de máximos, em especial quando estes estiverem inflados de zeros. Sua função densidade é dada por

$$h[Y = y | \Theta = (\theta, \omega)] = \omega I_{\{0\}}(y) + (1 - \omega) \frac{1}{\sigma} \left[1 + \frac{\xi}{\sigma}(y - \mu) \right]^{-\frac{1}{\xi} - 1} \\ \times \exp \left\{ - \left[1 + \frac{\xi}{\sigma}(y - \mu) \right]^{-\frac{1}{\xi}} \right\},$$

para $y > 0$ e $0 \leq \omega \leq 1$, em que a função indicadora $I_{\{0\}}(y)$ é uma distribuição que é degenerada no ponto zero, definida em

$$\left\{ y : 1 + \frac{\xi}{\sigma}(y - \mu) \right\},$$

em que $-\infty < \mu < \infty$, $\sigma > 0$ e $-\infty < \xi < \infty$, $\xi \neq 0$, são parâmetros referente a locação, escala e forma, respectivamente.

3.4 Próximos capítulos

No capítulo 3, faremos uma aplicação nos dados de máximos referente as 5 mesorregiões: Metropolitana do Recife, Mata Pernambucana, Agreste Pernambucano, Sertão Pernambucano e São Francisco Pernambucano, do estado de Pernambuco, utilizando a distribuição GEV e a recente distribuição ZIGEV para a realização das análises e modelagens. No capítulo 4, apresentamos uma nova distribuição, a distribuição Generalizada de Valores Extremos Inflada de Zeros Sazonais (SZIGEV), para analisar dados de máximos com comportamentos sazonais, em especial para as situações em que os dados estiverem inflados de zeros. Ainda nesse capítulo faremos uma aplicação com os dados das estações pluviométricas nas cidades de Recife e Petrolina do estado de Pernambuco e nas cidades de São João do Piauí e Teresina do estado do Piauí, comparando seus resultados com os obtidos pela distribuição ZIGEV. No capítulo 5, apresentamos as considerações finais resultantes do trabalho proposto.

4 Aplicação das distribuições GEV e ZIGEV a dados de chuva do estado de Pernambuco

4.1 Introdução

Análise de dados extremos tem sido uma grande ferramenta nas mais diversas áreas do conhecimento nas últimas décadas. Duas áreas onde esta análise se destaca são as áreas ambientais e econômicas (NASCIMENTO, 2012). De acordo com os dados do International Disaster Database (EM-DAT), entre 1990 e 2014, a ocorrência de inundações representou a maior parte dos desastres brasileiros reportados internacionalmente. Em segundo lugar, estão os desastres relacionados aos deslizamentos (CRED, 2016). O Brasil encontra-se entre os países do mundo mais atingidos por inundações e enchentes, tendo registrado 94 desastres cadastrados no EM-DAT no período de 1960 a 2008, com 5.720 mortes e mais de 15 milhões de pessoas afetadas (desabrigados e/ou desalojados) (TOMINAGA et al., 2009). No ano de 2008 o Brasil ocupou a 10^a posição em número de vítimas de desastres naturais, com 1,8 milhões de pessoas afetadas (OFDA/CRED, 2009).

O Brasil experimentou durante os anos de 2010 e 2011 o seu pior momento em relação a desastres naturais quando deslizamentos e enchentes geraram centenas de fatalidades e deixaram milhares de pessoas desabrigadas e desalojadas. Esses desastres colocaram o Brasil no 3^o lugar dos países mais afetados por catástrofes no ano de 2011 (GUHA-SAPIR et al., 2012). Estes eventos extremos têm afetado diversas cidades brasileiras causando múltiplos e constantes impactos, tais como: elevado número de mortes, feridos, desabrigados, proliferação de doenças, perdas econômicas, impactos ao meio ambiente, dentre outros (SENA et al., 2017).

Segundo um levantamento produzido pela Confederação Nacional de Municípios (CNM), a cada desastre natural ocorrido no Brasil em 2022, cerca de 3,4 mil pessoas foram afetadas diretamente. Os prejuízos financeiros causados por extremos ambientais, em 2022, somam mais de R\$ 72 bilhões de reais.

As chuvas intensas e extremas mais recentes que ocorreram em 2022, em Pernambuco, causaram deslizamentos de barreiras e enxurradas, provocando a morte de 126 pessoas, tornando a maior catástrofe dos últimos 50 anos do estado. Superando a cheia de 1975, que naquele ano resultou na morte de 107 pessoas. Poucos estudos referentes a índices de extremos climáticos foram desenvolvidos para alguns postos ou regiões específicas do

estado de Pernambuco (ASSIS et al., 2012, MARINHO et al., 2023).

Para se fazer a previsão desses eventos extremos, faz-se necessário o conhecimento da Teoria de Valores Extremos (TVE), pois, ela fornece o conhecimento sobre as distribuições assintóticas para as análises dos máximos por blocos de tamanho n . A TVE garante que, para $n \rightarrow \infty$, a distribuição dos máximos converge para a distribuição Generalizada de Valores Extremos (GEV) (NASCIMENTO, 2012). No entanto, nos trabalhos de Ottero et al. (2018), Gonçalves e Gutter (2019) e Gramosa et al. (2019) utilizando a distribuição GEV, tiveram dificuldades nas análises dos dados de máximos, pelo fato deles terem grandes proporções de zeros.

É comum encontrar uma porcentagem de zeros nos conjuntos de dados, especialmente em situações práticas que envolvem valores extremos. Esses zeros em excessos dificultam, e muitas vezes impedem, a elaboração de uma análise estatística precisa para o problema. Recentemente, Gramosa et al. (2019), desenvolveram a distribuição Generalizada de Valores Extremos Inflada de Zeros (ZIGEV), com o objetivo principal de solucionar essa grande problemática na modelagem de dados de valores extremos inflados de zeros.

A distribuição mostrou resultados bastante expressivos, nas simulações, nas aplicações a dados reais e nas medidas de ajustes, alcançando precisão em suas estimativas, mesmo em situações extremas com 80% de zeros nos dados. Uma característica importante é que, quando a proporção de zeros nos dados de máximos mensais for aproximadamente zero, $\omega \rightarrow 0$, a distribuição ZIGEV converge para GEV, sendo assim a GEV um caso particular da ZIGEV (GRAMOSA et al., 2019).

Neste contexto, um dos objetivos é a aplicação da distribuição ZIGEV a dados de precipitação de chuvas de Pernambuco. Na aplicação utilizaremos também a distribuição GEV, e então, faremos uma comparação com os resultados obtidos pelos dois modelos.

4.2 Material e métodos

4.2.1 Área de estudo e dados

Essa região do estado de Pernambuco é considerada seca e perpassa por vários períodos sem chuvas, e por consequência o banco de dados composto pelos máximos mensais fica inflado de zeros, o que dificulta e impede uma modelagem mais precisa.

Para realizar as análises nos dados, como as estimativas dos parâmetros, fez-se necessário uma transformação nos dados de máximos mensais, com a multiplicação da série dos dados por 10. Porque os valores dos máximos mensais não-zeros estavam muito próximos do valor zero, como por exemplo 0,1; 0,3; 0,5, dificultando a análise pelo modelo,

pois no algoritmo esses valores próximos de zero eram tidos como zeros contínuos. Por isso, multiplicamos a série dos dados por 10, para que o programa entendesse melhor o que era zero e o que não era. Contudo, nos cálculos dos retornos e dos gráficos de retornos dividimos todos os valores dos quantis extremos por 10, fazendo a transformação inversa para que ficasse tudo na escala original dos dados de chuvas.

Os dados utilizados neste trabalho referem-se as séries de precipitações diárias de chuvas, para 9 estações pluviométricas do estado de Pernambuco, que foram distribuídos em blocos mensais e separado o maior valor registrado de cada bloco, resultando nos dados de máximos mensais. As estações estão espalhadas (não uniformemente) pelas quatro mesorregiões do estado de Pernambuco, região nordeste do Brasil, descritas na Figura 4, são elas as cidades de Recife, Barreiros, Surubim, Garanhuns, Triunfo, Arcoverde, Ouricuri, Cabrobó e Petrolina.

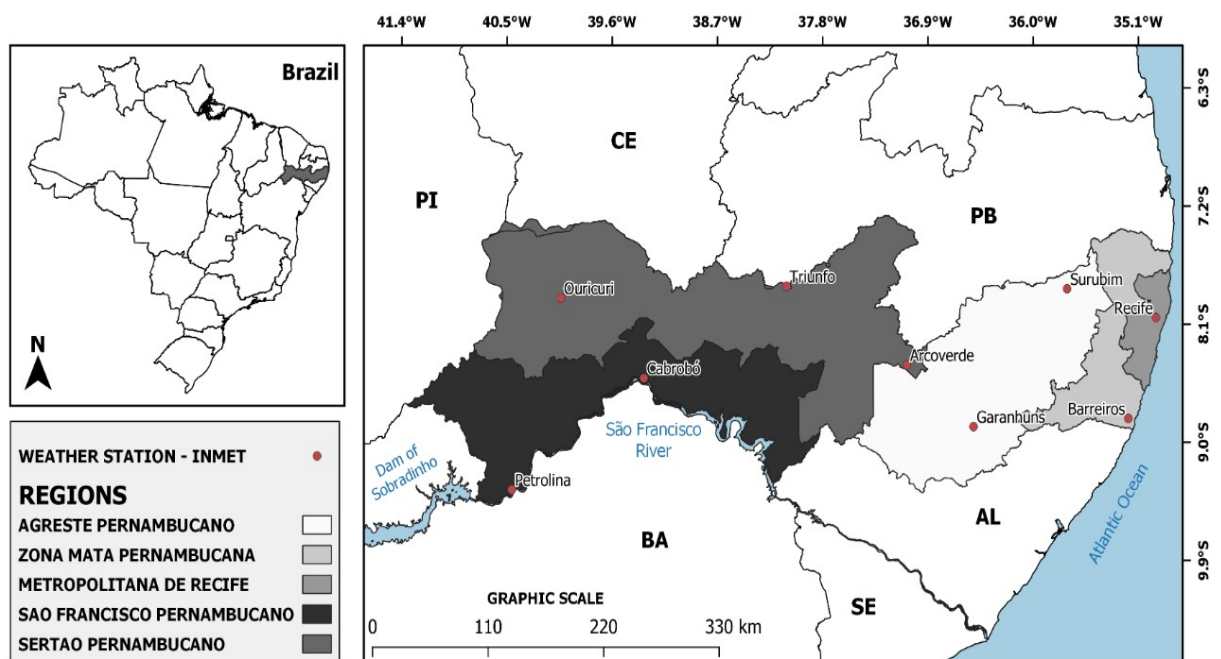


Figura 1 – Nove cidades de cinco mesorregiões do estado de Pernambuco

Esses dados foram coletados na Agência Pernambucana de Águas e Clima (APAC), e no Instituto Nacional de Meteorologia (INMET), e são resultantes de todos os dados históricos, do INMET e da APAC, durante o período do dia de 02/03/1961 a 15/07/2019 respectivo a cidade de Recife, a cidade de Barreiros do dia 01/01/1970 à 31/12/2003, Garanhuns do dia 02/12/1963 à 12/07/2019, Triunfo do dia 02/01/1961 à 31/03/2018, Surubim do dia 02/01/1961 à 15/07/2019, Arcoverde do dia 02/02/1973 à 20/06/2019, Ouricuri do dia 02/10/1975 à 15/07/2019, Cabrobó do dia 02/01/1961 à 15/07/2019 e

Petrolina do dia 04/12/1962 à 15/07/2019.

4.3 Distribuição Generalizada de Valores Extremos (GEV)

A distribuição de probabilidade generalizada de valores extremos (GEV) surgiu após a metade do século XX, com os trabalhos de von Mises (1954) e Jenkinson (1955) que realizaram a unificação das distribuições Fréchet, Gumbel e Weibull. A distribuição GEV possui a seguinte função de distribuição:

$$G(y|\xi, \sigma, \mu) = \begin{cases} \exp \left\{ - \left[1 + \xi \left(\frac{y - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}, & \text{se } \xi \neq 0; \\ \exp \left\{ -\exp \left\{ - \left(\frac{y - \mu}{\sigma} \right) \right\} \right\}, & \text{se } \xi = 0, \end{cases} \quad (4.1)$$

definida em $1 + \xi \left(\frac{y - \mu}{\sigma} \right) > 0$, em que $\mu \in R$ e $y \in R$.

Esse modelo possui três parâmetros: um parâmetro de locação (μ), um de escala (σ) e um de forma (ξ). Os casos em que $\xi < 0$ ou $\xi > 0$ correspondem respectivamente as distribuições Weibull ou Fréchet. Quando $\xi = 0$, ou quando no caso limite $\xi \rightarrow 0$, temos a distribuição Gumbel.

4.3.1 Inferência bayesiana para GEV

Segundo Nascimento (2012), na inferência Bayesiana, temos dois ingredientes para estimar o comportamento dos parâmetros, que é a função de verossimilhança $L(\theta) = f(y|\theta)$ e a distribuição *a priori* $p(\theta)$. A pergunta que temos que responder é: Qual seria a distribuição do vetor de parâmetros, após inserirmos a informação contida nos dados observados, e a informação a priori do pesquisador? Esta distribuição é conhecida como a posteriori, sendo obtida a partir do teorema de Bayes

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

em que, $f(y) = \int f(y|\theta)p(\theta)d\theta$.

Gamerman (2006) apresenta uma maneira compacta de escrever a função a posteriori dada por

$$\pi(\theta) \propto L(\theta)p(\theta), \quad (4.2)$$

em que $\theta = (\mu, \sigma, \xi)$.

Logo, sua função de verossimilhança é expressa por

$$L(\mu, \sigma, \xi) = \sigma^{-n} \exp \left\{ - \sum_{i=1}^n \left[1 + \xi \left(\frac{y_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \times \prod_{i=1}^n \left[1 + \xi \left(\frac{y_i - \mu}{\sigma} \right) \right]^{-(1+\frac{1}{\xi})}, \quad (4.3)$$

em que, $1 + \xi \left(\frac{y_i - \mu}{\sigma} \right) > 0$.

Assim, informando uma distribuição a priori para o vetor paramétrico (μ, σ, ξ) pode-se encontrar a distribuição a posteriori para os parâmetros. Para σ , que é necessariamente positivo, um candidato a priori é a distribuição *Gama*(a, b). Para μ e ξ , como podem ser negativos também, atribuímos prioris Normais $N(\mu_0, \sigma_\mu^2)$ e $N(\xi_0, \sigma_\xi^2)$ (NASCIMENTO, 2012). Dessa maneira, a distribuição conjunta a priori dos parâmetros é expressa por

$$p(\mu, \sigma, \xi) \propto \sigma^{a-1} \exp(-b\sigma) \exp \left[-\frac{(\mu - \mu_0)^2}{2\sigma_\mu^2} \right] \exp \left[-\frac{(\xi - \xi_0)^2}{2\sigma_\xi^2} \right]. \quad (4.4)$$

Portanto, substituindo as equações (4.3) e (4.4) na equação (4.2), encontra-se a distribuição a posteriori da GEV, que é representada da seguinte maneira

$$\begin{aligned} \pi(\mu, \sigma, \xi) \propto & \sigma^{a-n-1} \exp(-b\sigma) \exp \left[-\frac{(\mu - \mu_0)^2}{2\sigma_\mu^2} \right] \exp \left[-\frac{(\xi - \xi_0)^2}{2\sigma_\xi^2} \right] \times \\ & \exp \left\{ - \sum_{i=1}^n \left[1 + \xi \left(\frac{y_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \times \prod_{i=1}^n \left[1 + \xi \left(\frac{y_i - \mu}{\sigma} \right) \right]^{-(1+\frac{1}{\xi})}. \end{aligned} \quad (4.5)$$

4.3.2 Estimação dos níveis de retorno para GEV

Estimativas de quantis extremos dos máximos são obtidas invertendo a equação (4.1), fazendo $z_p = G^{-1}(1 - p)$, obtendo assim

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi} \left\{ 1 - [-\log(1 - p)]^{-\xi} \right\}, & \text{se } \xi \neq 0; \\ \mu - \sigma \log [-\log(1 - p)], & \text{se } \xi = 0, \end{cases}$$

em que $G(z_p) = 1 - p$. O termo z_p é conhecido como nível de retorno associado ao período $1/p$, ou seja, é o valor esperado que os dados originais ultrapassem o valor z_p pelo menos uma vez a cada $t = 1/p$ períodos de tempo.

4.4 Distribuição Generalizada de Valores Extremos Inflada de Zeros (ZIGEV)

A distribuição de probabilidade generalizada de valores extremos inflada de zeros ZIGEV surgiu recentemente, com o trabalho de Gramosa et al. (2019), que propuseram a mistura das distribuições GEV e ZID, com o propósito de resolver o problema na análise de dados de máximos inflados de zeros. A distribuição ZIGEV possui função de densidade dada por

$$h[X = x | \Theta = (\theta, \omega)] = \omega I_{\{0\}}(x) + (1 - \omega) \frac{1}{\sigma} \left[1 + \frac{\xi}{\sigma}(x - \mu) \right]^{-\frac{1}{\xi} - 1} \times \exp \left\{ - \left[1 + \frac{\xi}{\sigma}(x - \mu) \right]^{-\frac{1}{\xi}} \right\}, \quad (4.6)$$

para $x > 0$ e $0 \leq \omega \leq 1$, em que a função indicadora $I_{\{0\}}(x)$ é uma distribuição que é degenerada no ponto zero, definida em $z : 1 + \frac{\xi}{\sigma}(x - \mu)$, em que $-\infty < \mu < \infty$, $r > 0$ e $-\infty < \xi < \infty$, com $\xi \neq 0$.

Gramosa et al. (2019) mostraram algumas propriedades para a distribuição ZIGEV. Por exemplo, para encontrar a mediana da ZIGEV, basta calcular o quantil $p^* = \frac{0,50 - \omega}{1 - \omega}$.

4.4.1 Inferência bayesiana para ZIGEV

Apresentamos a seguir uma abordagem Bayesiana para a distribuição ZIGEV. Conforme já visto, para se fazer a estimação Bayesiana é necessário encontrar a distribuição a posteriori, que para a ZIGEV é apresentada da seguinte maneira

$$\pi(\theta, \omega | D) = L(\theta, \omega | D) \pi(\theta, \omega), \quad (4.7)$$

em que, $L(\theta, \omega | D)$ é a função de verossimilhança conjunta para os dados ampliados, expressa por

$$\begin{aligned}
L(\theta, \omega | D) &= \omega^S (1 - \omega)^{n-S} \left(\frac{1}{\sigma} \right)^{m-S} \left[\left(1 - \frac{\xi \mu}{\sigma} \right)^{-\frac{1}{\xi} - 1} \right]^{m-S} \times \\
&\prod_{i=1}^{n-m} \left\{ \frac{1}{\sigma} \left[1 + \frac{\xi}{\sigma} (x_i - \mu) \right]^{-\frac{1}{\xi} - 1} \exp \left\{ - \left[1 + \frac{\xi}{\sigma} (x_i - \mu) \right]^{-\frac{1}{\xi}} \right\} \right\} \times (4.8) \\
&\left[\exp \left\{ - \left(1 - \frac{\xi \mu}{\sigma} \right)^{-\frac{1}{\xi}} \right\} \right]^{m-S},
\end{aligned}$$

e $\pi(\theta, \omega)$ é a priori conjunta dos dados que é representada por

$$\pi(\theta, \omega) \propto \omega^{c-1} (1 - \omega)^{d-1} \exp \left\{ -\frac{\mu^2}{2\sigma_{0\mu}^2} \right\} \exp \left\{ -\frac{\xi^2}{2\sigma_{0\xi}^2} \right\} \sigma^{a_0-1} \exp\{-\sigma b_0\}, \quad (4.9)$$

em que c e d são os hiperparâmetros da distribuição Beta. As distribuições dos parâmetros $(\omega, S, \mu, \sigma, \xi)$ são dadas por

$$\begin{aligned}
\mu &\sim N(0, \sigma_{0\mu}^2) \sim \exp \left\{ -\frac{\mu^2}{2\sigma_{0\mu}^2} \right\}, \\
\xi &\sim N(0, \sigma_{0\xi}^2) \sim \exp \left\{ -\frac{\xi^2}{2\sigma_{0\xi}^2} \right\}, \\
\sigma &\sim G(a_0; b_0) \sim \sigma^{a_0-1} \exp\{-\sigma b_0\}.
\end{aligned}$$

$$\begin{aligned}
\omega &\sim \text{Beta}(S + c, n - S + d) \sim \omega^{S+c-1} (1 - \omega)^{n-S+d-1} \\
S &\sim \text{Bin}(m, p(\theta, \omega)) \sim \binom{m}{S} p(\theta, \omega)^S [1 - p(\theta, \omega)]^{m-S}.
\end{aligned}$$

Portanto, substituindo as equações (4.8) e (4.9) na equação (4.7), encontra-se a distribuição conjunta a posteriori da ZIGEV, que é expressa por

$$\begin{aligned}
\pi(\theta, w|D) &\propto w^{S+c-1}(1-w)^{n-S+d-1} \left(\frac{1}{\sigma}\right)^{m-s} \left[\left(1 - \frac{\xi\mu}{\sigma}\right)^{-\frac{1}{\xi} - 1} \right]^{m-S} \times \\
&\prod_{i=1}^{n-m} \left\{ \frac{1}{\sigma} \left[1 + \frac{\xi}{\sigma}(x_i - \mu)\right]^{-\frac{1}{\xi} - 1} \exp \left\{ - \left[1 + \frac{\xi}{\sigma}(x_i - \mu)\right]^{-\frac{1}{\xi}} \right\} \right\} \times \\
&\left[\exp \left\{ - \left(1 - \frac{\xi\mu}{\sigma}\right)^{-\frac{1}{\xi}} \right\} \right]^{m-S} \exp \left\{ -\frac{\mu^2}{\sigma_{0\mu}^2} - \frac{\xi^2}{\sigma_{0\xi}^2} \right\} \sigma^{a-1} \exp\{-\sigma b_0\}.
\end{aligned}$$

Uma vez que a posteriori encontrada não tem uma distribuição conhecida, faz-se necessário obter aproximações numéricas através das cadeias de Markov (MCMC), Gamerman (2006).

4.4.2 Estimação dos níveis de retorno para ZIGEV

Os níveis de retornos (z_{p^*}) para $1/p^*$ períodos, $0 < p^* < 1$ são baseados na estimação dos p^* quantis da ZIGEV, e obtidos pelas estimativas de máxima verossimilhança dos parâmetros da GEV. Segundo Gramosa et al. (2019) o quantil p^* é encontrado na expressão $\omega + (1 - \omega) = 1$, em que ω é a proporção de zeros, e $(1 - \omega)$ o complementar, sendo apresentado por $(p - \omega) + (1 - p)$, em que o quantil $p = 1 - 1/t$ é o nível de retorno esperado em média a cada $1/p^*$ períodos. Logo,

$$\begin{aligned}
p^* &= 1 - \frac{(1-p)}{(1-\omega)} = 1 - \frac{(1-1+1/t)}{(1-\omega)} = 1 - \frac{1/t}{(1-\omega)} \\
&= 1 - \frac{1}{t(1-\omega)},
\end{aligned}$$

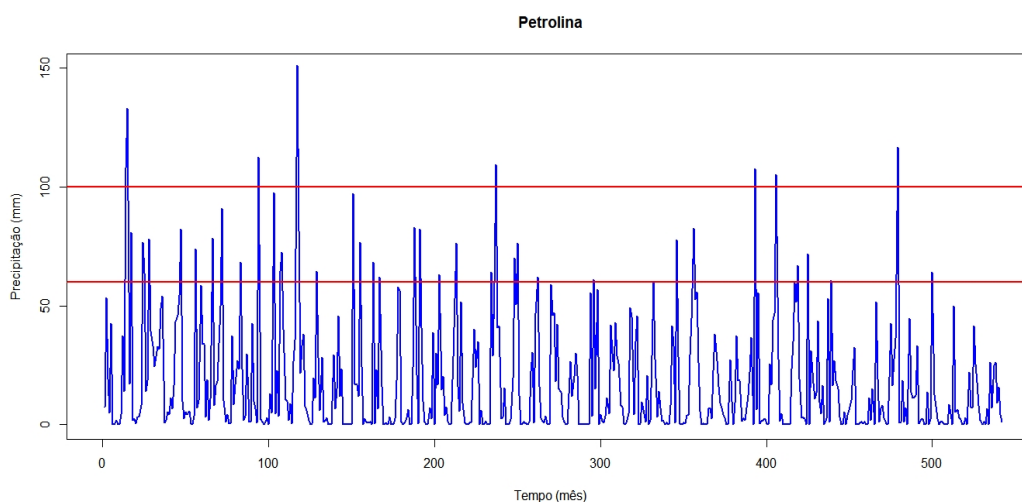
em que t é o número de períodos ao longo do tempo, normalmente sendo meses ou anos.

Assim, quando o $\omega = 0 \Rightarrow p^* = p$, ou seja, quando a proporção de zeros dos dados (ω) for 0%, a estimação dos quantis pela GEV será idêntica a da ZIGEV, portanto, quanto mais próximo de 0% estiver a proporção de zeros dos dados, mais semelhantes serão as estimativas de retorno destas distribuições. Entretanto, quanto mais distante de 0% estiver a proporção de zeros maior será a diferença entre as suas estimativas (GRAMOSA et al., 2019).

4.5 Resultados e Discussões

Apresentamos os resultados das análises dos dados de precipitação diária de chuvas das nove cidades em estudo, que compõem as quatro mesorregiões do estado de Pernambuco, na forma de tabelas e gráficos, juntamente com as estimações e medidas de ajustes realizadas nos modelos GEV e ZIGEV.

Figura 2 – Gráfico de série para os dados de máximos mensais da cidade de Petrolina

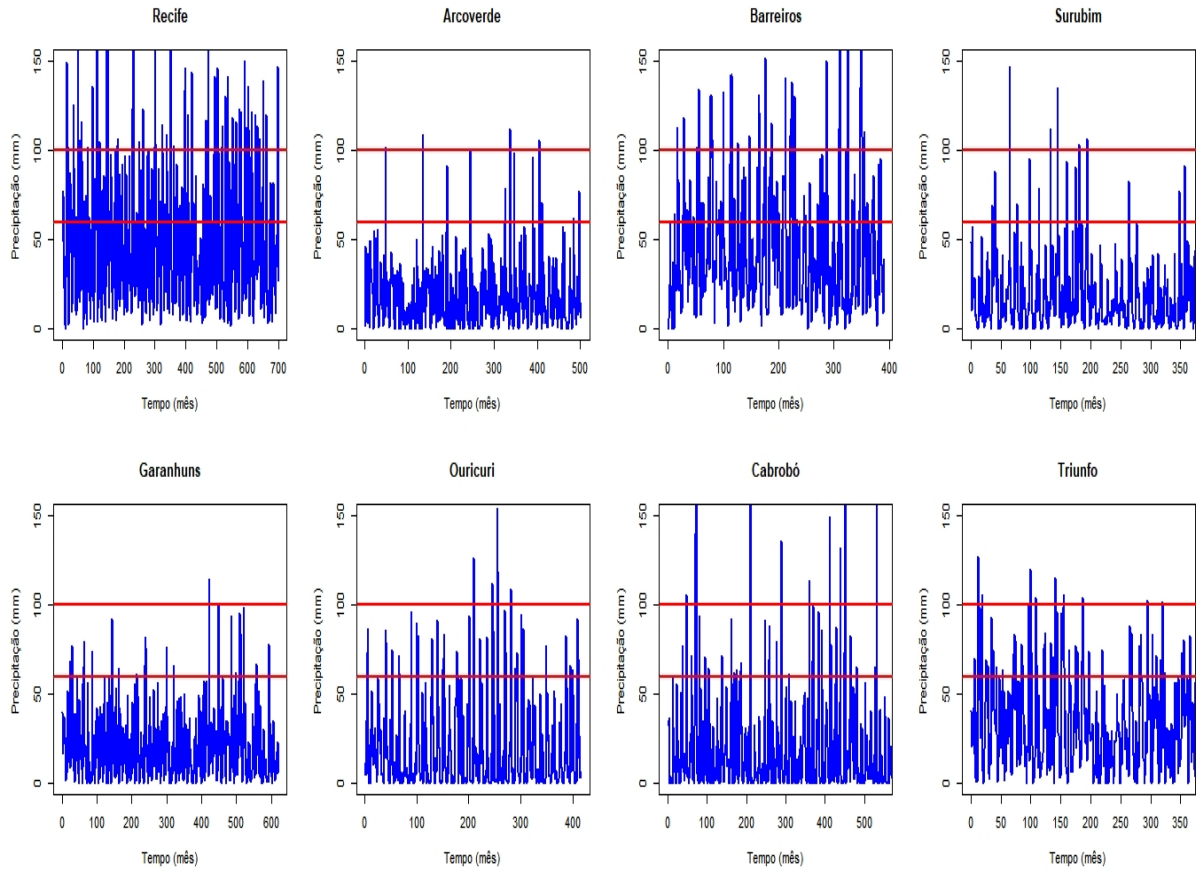


A Figura 5 mostra a série histórica da cidade de Petrolina-PE, com mais de 500 meses de observações em que cada um destes são as precipitações de chuvas máximas referentes a cada mês. Percebe-se que a cidade teve algumas chuvas intensas, que variam entre 60 mm e 100 mm, e que poucas vezes foi surpreendida com chuvas extremas, acima de 100 mm, chegando a uma precipitação de chuva até 150 mm.

Na Figura 6, referente as séries para os dados de máximos mensais das cidades de Barreiros, Surubim, Garanhuns, Triunfo, Arcoverde, Ouricuri, Petrolina e Cabrobó, nota-se que na maioria das cidades houver muitas chuvas intensas, entre 60 mm e 100 mm, e que não raramente foram surpreendidas com chuvas extremas, acima de 100 mm. Percebe-se que em muitas vezes as chuvas chegam a 150 mm, como nas cidades de Arcoverde, Garanhuns, Barreiros, Surubim, Triunfo, Ouricuri, e Cabrobó, havendo registros de chuvas maiores que 300 mm, ocorrido em um único dia, como aconteceu na cidade de Recife, em que provocaram enchentes, enxurradas e desastres geológicos, ocasionando diversas perdas humanas, materiais e financeiras.

Nota-se na Tabela 1 que as cidades de Recife e Barreiros, das mesorregiões Metropolitanas do Recife e Mata Pernambucana, respectivamente, possuem as menores proporções de zeros nos dados de máximos mensais das séries em estudo com 0,28% e 1,53%, respecti-

Figura 3 – Gráficos de séries de máximos mensais das cidades de Recife, Barreiros, Surubim, Garanhuns, Triunfo, Arcoverde, Ouricuri e Cabrobó



vamente, ou seja, em todo o histórico de dados mensais destas cidades, somente em alguns meses não choveu. Em Recife dos 700 meses registrados apenas 2 foram sem chuvas e em Barreiros dos 390 meses em apenas 6 não choveu. Em contraste, as cidades de Cabrobó e Petrolina, da mesorregião de São Francisco, possuem as maiores proporções de zeros nos dados de máximos mensais das séries em estudo com 11,75% e 16,60% de zeros nos dados, respectivamente, com 67 e 90 meses registrados sem chuvas.

Percebe-se que as mesorregiões Metropolitana, do Agreste e uma parte do Sertão Pernambucano possuem menor proporção de zeros nos seus respectivos dados de máximos mensais, indicando a existência de uma maior frequência de chuvas nesta região. Entretanto uma parte da mesorregião do Sertão Pernambucano e a mesorregião do São Francisco são as que mais registram zeros nos seus dados de máximos mensais, indicando que são as mesorregiões que mais passam por períodos sem chuvas.

Na Tabela 2, foram realizadas as estimativas dos parâmetros em estudo com os modelos GEV e ZIGEV, e verificou-se que, para a cidade de Recife, com 0,28% de zeros

Tabela 1 – Quantidade (S) e proporção de zeros (ω) nos dados pluviométricos das mesorregiões do estado de Pernambuco

| Mesorregiões | Cidade | Amostra (n) | Zeros | |
|---------------|-----------|-----------------|--------------------|----------------------------|
| | | | Quantidade (S) | Proporção (ω) (%) |
| Metropolitana | Recife | 700 | 2 | 0,00286 (0,28%) |
| Mata | Barreiros | 390 | 6 | 0,01538 (1,53%) |
| Agreste | Surubim | 592 | 23 | 0,03885 (3,88%) |
| | Garanhuns | 620 | 20 | 0,03226 (3,22%) |
| Sertão | Triunfo | 565 | 19 | 0,03363 (3,36%) |
| | Arcoverde | 502 | 36 | 0,07173 (7,17%) |
| | Ouricuri | 416 | 41 | 0,09856 (9,85%) |
| São Francisco | Cabrobó | 570 | 67 | 0,11754 (11,75%) |
| | Petrolina | 542 | 90 | 0,16605 (16,60%) |

nos dados, as estimativas dos parâmetros μ , σ e ξ , referente a locação, escala e forma, respectivamente, obtiveram resultados bem semelhantes [$\hat{\mu} = (251,1886; 251,3571)$, $\hat{\sigma} = (212,4715; 212,8001)$, $\hat{\xi} = (0,32846; 0,32950)$], respectivamente, como já era esperado, tendo em vista que em toda a série de dados de Recife, apenas dois máximos mensais são zeros. Estes resultados indicam que o modelo ZIGEV está bem ajustado, pois, quando não existe zeros de máximos nos dados ou existe, mas a proporção de zeros é próximo de zero, $\omega \rightarrow 0$, ele converge ao modelo GEV.

Para as cidades de Barreiros, Surubim, Garanhuns, Triunfo e Arcoverde, com proporções a partir de 1,53% até 7,17% de zeros nos dados, existe uma diferença não tão grande nas estimativas dos parâmetros μ , σ e ξ , entre os modelos ZIGEV e GEV, sendo que aparentemente os resultados obtidos pela ZIGEV são mais precisos, como acontece com as estimativas dos parâmetros ω e S , que representam a proporção e a quantidade de zeros nas séries dos dados, respectivamente. Na cidade de Triunfo, em que as estimativas obtidas em $\hat{\omega} = 3,26\%$, foram bem precisas, com relação a proporção de zeros desta série

de dados que é igual a 3,36%, o que aconteceu também com as estimativas obtidas em $\hat{S} = 35,32$, que são muito próximas da quantidade de 36 zeros descrita nos dados.

Tabela 2 – Estimativas para os parâmetros das distribuições GEV e ZIGEV para as cidades em estudo

| Cidades | Modelos | Estimativas | | | | |
|-----------|---------|----------------|-----------|-------------|----------------|-------------|
| | | $\hat{\omega}$ | \hat{S} | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\xi}$ |
| Recife | ZIGEV | 5,16e-05 | 1,0332 | 251,3571 | 212,8001 | 0,32950 |
| | GEV | | | 251,1886 | 212,4715 | 0,32846 |
| Barreiros | ZIGEV | 0,01502 | 5,8236 | 240,2976 | 196,9377 | 0,31135 |
| | GEV | | | 234,3691 | 203,5819 | 0,28383 |
| Surubim | ZIGEV | 0,03726 | 22,1095 | 86,9025 | 84,4410 | 0,49799 |
| | GEV | | | 79,8593 | 84,4598 | 0,49600 |
| Garanhuns | ZIGEV | 0,03010 | 19,1650 | 111,9388 | 103,7401 | 0,29898 |
| | GEV | | | 104,5986 | 103,8782 | 0,30639 |
| Triunfo | ZIGEV | 0,03263 | 18,4796 | 183,3413 | 171,5136 | 0,20645 |
| | GEV | | | 170,7257 | 171,5286 | 0,22148 |
| Arcoverde | ZIGEV | 0,07046 | 35,3238 | 106,2129 | 99,2791 | 0,32554 |
| | GEV | | | 89,6824 | 97,5092 | 0,36790 |
| Ouricuri | ZIGEV | 0,09451 | 39,3063 | 56,4642 | 79,5135 | 1,08662 |
| | GEV | | | 31,4210 | 56,6414 | 1,52705 |
| Cabrobó | ZIGEV | 0,11353 | 64,8129 | 52,5047 | 72,1119 | 1,07187 |
| | GEV | | | 25,6782 | 47,9616 | 1,60312 |
| Petrolina | ZIGEV | 0,16201 | 87,8794 | 50,1508 | 71,6142 | 1,15300 |
| | GEV | | | — | — | — |

Já a partir de 9,85% de zeros, referente a cidade de Ouricuri, as estimativas obtidas pela distribuição GEV tiveram resultados bem diferentes das estimativas encontradas pela distribuição ZIGEV, $[\hat{\mu} = (31,4210; 56,4642)$, $\hat{\sigma} = (56,6414; 79,5135)$, $\hat{\xi} = (1,52705; 1,08662)]$, respectivamente. Estes resultados mostram indícios de dificuldades da distribuição GEV em analisar dados na presença excessiva de zeros de máximos, pois, aparentemente os resultados obtidos pela ZIGEV foram mais precisos, como aconteceu com as estimativas dos parâmetros ω e S , em relação aos registros nas séries dos dados.

Para a série de dados da cidade de Petrolina, o modelo GEV não conseguiu estimar os parâmetros. Uma proporção de 16,60% de zeros foi o suficiente para que o modelo, com o algoritmo baseado nas cadeias MCMC, não convergisse para a distribuição de equilíbrio, mesmo quando alterado o número de interações de 5.000 para 50.000, não conseguindo gerar as médias a posteriori da distribuição referentes aos parâmetros do modelo. Já o modelo ZIGEV foi bastante preciso com as estimações, para a mesma série de dados da cidade de Petrolina, em relação aos parâmetros ($\omega = 16,60\%$, $\hat{\omega} = 16,20\%$ e $S = 90$; $\hat{S} = 87,87$), como observa-se na Figura 7 e, aparentemente, muito preciso também em relação aos parâmetros μ , σ e ξ , mostrando a importância de uma distribuição que modele bem dados com estas características, dados de máximos e inflados de zeros.

Este comportamento da distribuição GEV com os dados da cidade de Petrolina foi semelhante ao ocorrido em Gramosa et al. (2019), quando a distribuição não conseguiu estimar as séries de dados, das cidades de Picos (PI), Paulistana (PI) e São João do Piauí (PI), com uma proporção de 17,77%, 25,81% e 25,30% de zeros nos dados, respectivamente. Nesta situação a ZIGEV obteve resultado satisfatório tanto para a cidade de Petrolina, quanto para as demais em estudo, como é observado na Figura 7, estimando seus parâmetros com muita precisão, e mostrando ganho desta distribuição para a modelagem de dados extremos com excessos de zeros.

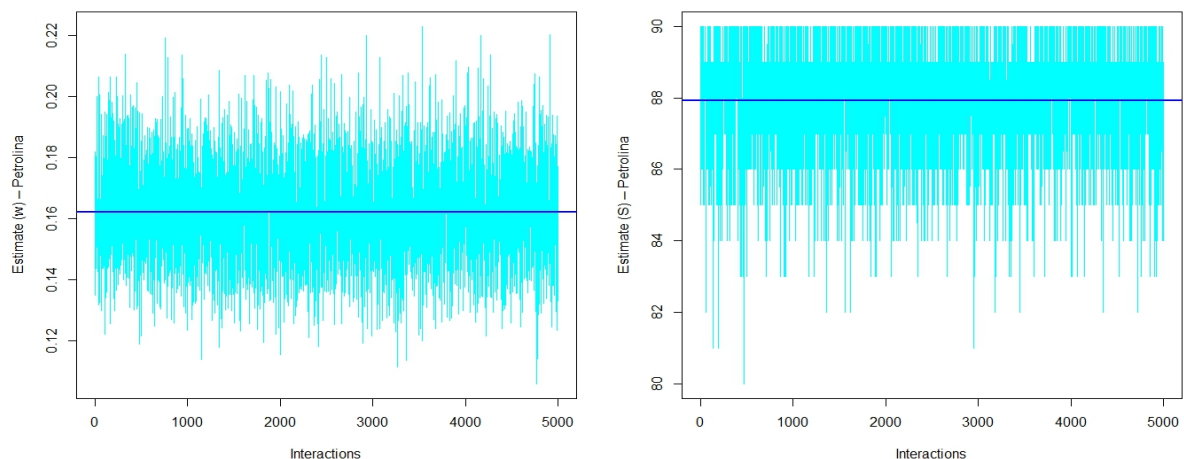


Figura 4 – Gráfico dos estimadores ω e S da distribuição ZIGEV para a cidade de Petrolina

Na Tabela 3 é apresentada a comparação entre as medidas de qualidade de ajuste AIC (Akaike's Information Criterion), BIC (Bayesian Information Criterion, também conhecido como Schwarz Criterion) e DIC (Deviance Information Criterion), para as distribuições GEV e ZIGEV nas cidades de Pernambuco. DIC é uma generalização dos

critérios AIC e BIC, e foi proposto por Spiegelhalter et al., (2002). Valores pequenos de DIC indicam um modelo de melhor ajuste. O DIC consiste em dois termos, um representando a qualidade do ajuste e o outro uma penalidade por aumentar a complexidade do modelo.

Tabela 3 – Comparação das medidas de ajuste para as distribuições GEV e ZIGEV nas cidades de Pernambuco

| Modelos | Medidas de Ajuste | | | | | |
|---------|-------------------|-----------|-----------|-----------|----------|----------|
| | Recife | | | Barreiros | | |
| | AIC | BIC | DIC | AIC | BIC | DIC |
| GEV | 10032,300 | 10054,500 | 10031,130 | 5551,651 | 5571,515 | 5550,548 |
| ZIGEV | 10038,130 | 10060,880 | 10030,770 | 5525,933 | 5545,764 | 5519,803 |
| Modelos | Surubim | | | Triunfo | | |
| | AIC | BIC | DIC | AIC | BIC | DIC |
| GEV | 7460,076 | 7481,610 | 7458,978 | 7766,964 | 7788,311 | 7765,862 |
| ZIGEV | 7376,284 | 7398,201 | 7370,204 | 7675,488 | 7697,172 | 7669,336 |
| Modelos | Garanhuns | | | Arcoverde | | |
| | AIC | BIC | DIC | AIC | BIC | DIC |
| GEV | 7939,604 | 7961,323 | 7938,597 | 6399,472 | 6420,346 | 6398,341 |
| ZIGEV | 7865,992 | 7888,141 | 7859,868 | 6211,428 | 6232,521 | 6205,402 |
| Modelos | Ouricuri | | | Cabrobó | | |
| | AIC | BIC | DIC | AIC | BIC | DIC |
| GEV | 5310,300 | 5330,423 | 5309,299 | 7124,306 | 7145,689 | 7123,231 |
| ZIGEV | 5188,310 | 5208,463 | 5182,032 | 6908,782 | 6930,511 | 6902,597 |

Nesta tabela observa-se que, para a cidade de Recife, com 0,28% de zeros nos dados, a distribuição GEV obteve melhores resultados nas medidas de ajuste AIC e BIC em comparação com a distribuição ZIGEV. Pois, mesmo que a ZIGEV tenha obtido resultados muito semelhantes nas estimativas dos seus parâmetros em relação as estimativas da GEV, para esta série de dados, estas medidas de ajuste aplicam uma penalização em relação a quantidade de parâmetros, e, como o modelo ZIGEV possui cinco parâmetros, dois a mais que o modelo GEV, sofre uma penalização maior. Entretanto, ainda para a cidade de Recife, com relação a medida de ajuste DIC, a ZIGEV obteve melhor resultado que a GEV

(10030,770; 10031,130), respectivamente, pois, mesmo com resultados semelhantes, seus procedimentos são relacionados ao método Monte Carlo em Cadeias de Markov (MCMC) que se mostram flexíveis e poderosos para análises em modelos Bayesianos.

Outro resultado muito importante é que todas as medidas de ajustes realizados nas demais cidades apresentadas na Tabela 3, comprovaram que o modelo ZIGEV obteve uma melhor qualidade no ajuste em comparação com o modelo GEV. Inclusive para a cidade de Barreiros, com uma proporção de apenas 1,53% de zeros nos dados de máximos, nas três principais medidas de ajuste selecionadas [AIC=(5551,651; 5525,933), BIC=(5571,515; 5545,764) e DIC=(5550,548; 5519,803)], respectivamente, em comparação aos modelos GEV e ZIGEV.

De posse destes resultados, pode-se afirmar que a partir de uma pequena proporção de zeros nas séries dos dados de máximos, no caso a partir de 1,53%, as estimativas da distribuição ZIGEV são mais precisas que as estimativas da distribuição GEV, e que, quanto maior for a proporção de zeros nos dados de máximos, maior será a imprecisão dos estimadores da GEV em relação a ZIGEV. Portanto, é notável que para esse tipo de aplicação em que, existe uma quantidade excessiva de zeros nos dados de máximos e consecutivamente uma dificuldade inerente nas análises desses dados por parte da GEV, a distribuição ZIGEV surge como uma possibilidade para se realizar uma modelagem estatística precisa.

Da Tabela 4 observa-se que os níveis de retornos dos períodos 5, 10, 20 e 100 das distribuições GEV e ZIGEV, respectivos aos seus quantis altos 80%, 90%, 95% e 99%, obtiveram resultados semelhantes para a cidade de Recife, como também já era esperado, tendo em vista que a proporção de zeros dos seus dados de máximo mensais era bem próximo de zero (0,28%), ficando bastante próximo dos respectivos quantis da distribuição empírica.

Percebe-se também que os quantis 80% e 90%, referentes aos níveis de retorno dos períodos 5 e 10, estimados pela GEV e pela ZIGEV para as cidades de Barreiros (1,53%), Garanhuns (3,22%), Triunfo (3,36%) e Surubim com 3,88% de proporção de zeros nos dados, obtiveram resultados semelhantes, ficando também próximos aos respectivos quantis da distribuição empírica. Entretanto nota-se que para estas mesmas cidades, os níveis de retorno dos períodos 20 e 100, calculados pela distribuição ZIGEV, correspondentes aos quantis altos de 95% e 99%, que estão mais para o final da cauda da distribuição, mostraram-se mais precisos e mais próximos dos resultados empíricos do que os respectivos quantis calculados pela GEV. Estes resultados indicam que, mesmo quando a inflação de zeros nos dados é pequena, principalmente nas estimações dos quantis altos, as predições feitas pelos níveis de retorno do modelo ZIGEV são mais precisas que as predições feitas

Tabela 4 – Comparação dos níveis de retornos correspondentes aos quantis extremos do GEV, ZIGEV e Empírico das cidades em estudo

| Cidades | Modelos | Níveis de Retornos | | | | Retornos | | | | Cidades |
|----------|----------|--------------------|--------|--------|---------|----------|--------|--------|---------|-----------|
| | | 80% | 90% | 95% | 99% | 80% | 90% | 95% | 99% | |
| Recife | GEV | 66,24 | 95,77 | 131,74 | 252,75 | 61,40 | 87,16 | 117,88 | 214,58 | Barreiros |
| | ZIGEV | 66,36 | 95,69 | 131,53 | 251,73 | 61,00 | 87,36 | 119,11 | 224,24 | |
| | Empírico | 75,16 | 98,08 | 118,08 | 162,00 | 69,68 | 90,92 | 116,54 | 152,77 | |
| Surubim | GEV | 26,78 | 42,86 | 65,05 | 156,65 | 47,63 | 67,07 | 88,93 | 153,43 | Garanhuns |
| | ZIGEV | 26,76 | 42,60 | 64,39 | 154,75 | 47,55 | 66,31 | 87,30 | 147,71 | |
| | Empírico | 28,76 | 42,36 | 54,27 | 95,72 | 53,68 | 69,32 | 81,72 | 105,54 | |
| Triunfo | GEV | 30,19 | 43,99 | 60,58 | 114,73 | 28,49 | 43,11 | 61,41 | 126,60 | Arcoverde |
| | ZIGEV | 30,15 | 43,64 | 59,76 | 111,83 | 28,46 | 41,86 | 58,17 | 112,74 | |
| | Empírico | 32,12 | 43,40 | 56,02 | 83,72 | 31,68 | 41,18 | 51,39 | 98,57 | |
| Ouricuri | GEV | 35,48 | 111,46 | 332,51 | 3953,47 | 32,19 | 107,26 | 340,84 | 4577,83 | Cabrobó |
| | ZIGEV | 31,35 | 73,58 | 163,70 | 978,02 | 27,58 | 64,24 | 141,67 | 829,92 | |
| | Empírico | 38,40 | 60,95 | 82,00 | 109,08 | 34,52 | 51,05 | 70,01 | 136,82 | |

pela GEV.

Para as cidades como Arcoverde, Ouricuri e Cabrobó, que possuem uma maior proporção de zeros iguais a 7,17%, 9,85% e 11,75% apresentados na Tabela 4, respectivamente, são perceptíveis que os quantis extremos 90%, 95% e 99% calculados pela distribuição ZIGEV são mais próximos dos resultados empíricos e mais precisos do que os respectivos quantis calculados pela GEV, com a observação de que quanto maior a proporção de zeros, maior é a imprecisão do modelo GEV, principalmente nas predições dos quantis extremos 95% e 99%, como visto na cidade de Cabrobó, na mesorregião do São Francisco, no qual o quantil da ZIGEV de 95%, que separa os 95% valores menores aos 5% maiores, teve uma precipitação de 141,67 mm. Quer dizer que este nível de chuva é muito raro, e é esperado que ocorra em 5% das vezes, logo, espera-se que a cada 20 períodos de tempo, neste caso que a cada 20 meses, se tenha pelo menos uma vez uma chuva igual ou maior a 141,67 mm. Como os quantis da GEV e o empírico de 95% foram estimados em 340,84 mm e 70,01 mm cada, temos que, por estes modelos, é esperado que pelo menos uma vez a cada período de 20 meses tenha-se uma chuva maior ou igual a 340,84 mm e 70,01 mm, respectivamente.

Assim, mais claramente, percebe-se que quanto maior a proporção de zeros de máximos nos dados, que neste caso é de 11,75%, maior será a imprecisão nas predições

feita pela distribuição GEV, pois, a precipitação predita (340,84 mm) pelo seu quantil 95% é mais que o dobro da precipitação estimada (141,67 mm) pelo quantil da distribuição ZIGEV, que é confirmada através dos resultados obtidos nas medidas de qualidade de ajuste. A precipitação estimada (70,01 mm) pelo quantil 95% da distribuição empírica, pode-se afirmar que é baseada na amostra ordenada dos dados e possui uma grande dificuldade em estimar os quantis extremos a medida que vão se aproximando do final da cauda, em que se tem poucas observações para calcular. Logo, o quantil empírico acaba ficando muito impreciso, e nestes casos o quantil estimado da distribuição ZIGEV obtém um melhor resultado. É exatamente o que esperamos que aconteça, até pela própria TVE, que é exatamente para detectar o comportamento preciso da cauda.

Na Figura 8 é perceptível que, para as cidades de Recife (0,28%), Barreiros (1,53%), Garanhuns (3,22%), Triunfo (3,36%) e Surubim (3,88%), com proporções de zeros nos dados de máximos mensais entre 0,28% e 3,88%, o comportamento das estimativas dos níveis de retornos dos dois modelos são praticamente idênticas, pois, as linhas do gráfico da GEV (*em vermelho*) são todas sobrepostas pelas linhas da ZIGEV (*em azul*), inclusive na variação dos seus limites inferiores e superiores. Percebe-se que, para a cidade de Recife (0,28%), as linhas contínuas em vermelho e em azul entre as linhas tracejadas sobre o ponto de 60 meses (5 anos), apontam para níveis de retornos muito semelhantes, acima de 200 mm, 207,59 mm pela GEV e 207,30 mm pela ZIGEV, permitindo-nos afirmar que, segundo o modelo, é esperado que a cada 5 anos ocorra, pelo menos uma vez, uma chuva extrema igual ou maior ao valor estimado (207 mm). Estes resultados implicam que, quando a proporção de zeros nos dados são próximos de zero, a distribuição ZIGEV converge para a GEV, tornando a GEV um caso particular da ZIGEV.

Observa-se que para a cidade de Arcoverde, com uma proporção de 7,17% de zeros nos dados, existe uma visível diferença nas predições realizadas pelos modelos GEV (vermelho) e ZIGEV (azul) da Figura 8, representadas pelas linhas contínuas, inclusive com um distanciamento entre seus respectivos limites inferiores e superiores, representadas pelas linhas tracejadas. Gramosa (2019) verifica que, com uma proporção de 8,60% de zeros nos dados de máximos mensais, a distribuição GEV superestima a distribuição ZIGEV em quase duas vezes mais.

Já nos gráficos de retornos para as cidades de Ouricuri e Cabrobó da Figura 8, que possuem uma proporção de 9,85% e 11,75% de zeros nos dados de máximos, respectivamente, nota-se uma grande diferença nas curvas de retorno dos modelos GEV (vermelho) e ZIGEV (azul), nas predições ao longo do tempo, e também nos intervalos entre seus respectivos limites inferiores e superiores. Pois, como a distribuição GEV não considera os zeros, ela tem o vício para cima, já a distribuição ZIGEV, que analisa e

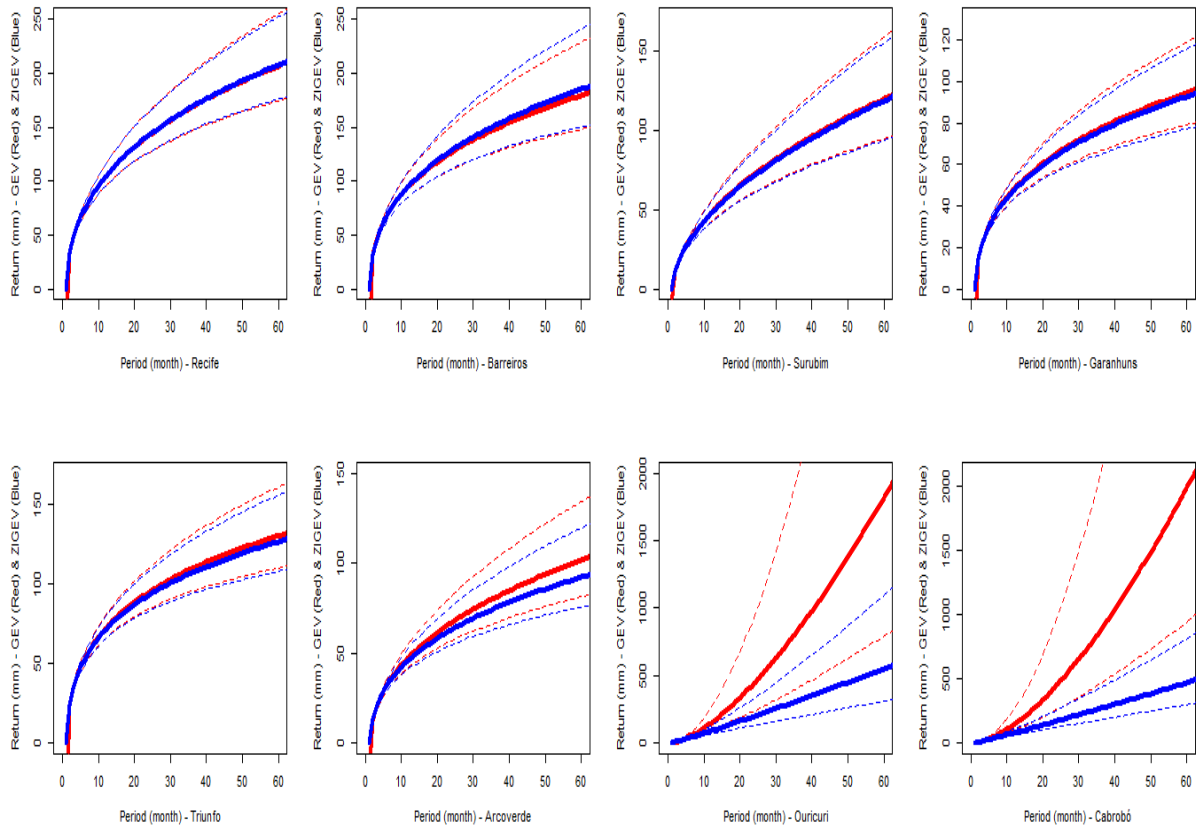


Figura 5 – Gráfico dos níveis de retorno da GEV (vermelho) e ZIGEV (azul) para Recife, Barreiros, Surubim, Garanhuns, Triunfo, Arcoverde, Ouricuri e Cabrobó, com seus respectivos intervalos de confiança representados pelas linhas tracejadas

modela os zeros, tem o seu retorno mais para baixo. Percebe-se também que a estimação realizada com os dados da cidade de Ouricuri (9,85%), no ponto de 36 meses (3 anos), de acordo com a distribuição GEV (vermelho), é aguardado um retorno de 824,84 mm, enquanto que a estimação realizada pela distribuição ZIGEV (azul), no mesmo ponto de 36 meses, é aguardado um retorno de 304,28 mm.

Estes resultados nos mostram que uma proporção de 9,85% de zeros nos dados de máximos mensais, é suficiente para tornar as previsões feitas pelo modelo GEV muito imprecisas, com uma superestimação de quase três vezes maior (em milímetros) em comparação com o modelo ZIGEV. Tornando bastante claro que a medida que aumenta a proporção de zeros nos dados de máximos mensais, aumenta também a imprecisão do modelo GEV na análise destes dados. Mostrando a importância e a necessidade de se utilizar a distribuição ZIGEV para modelar dados extremos e inflados de zeros.

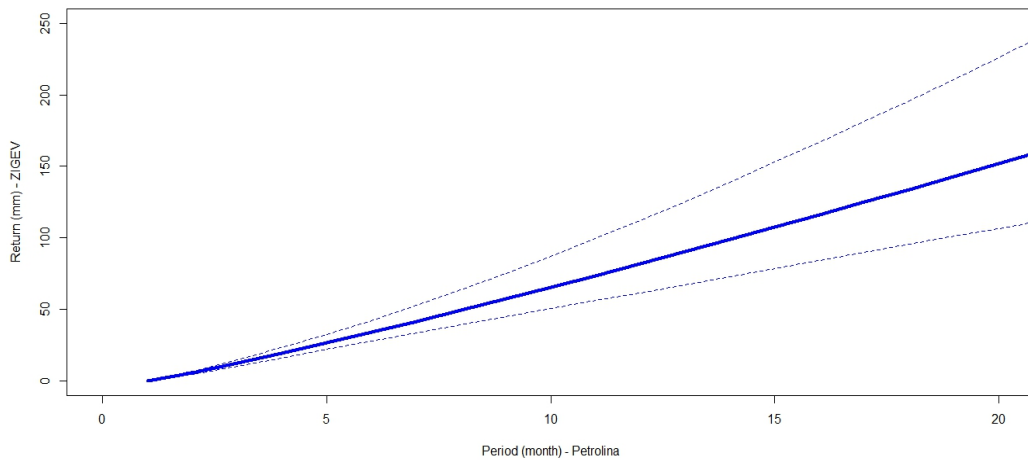


Figura 6 – Gráfico do nível de retorno pela ZIGEV para a cidade de Petrolina

Como vimos, muitas vezes a cidade de Petrolina foi afetada com chuvas intensas e até chuvas extremas (150 mm), indicando a importância de se fazer estudos de previsão, que mostrem quando acontecerá novamente e com qual intensidade virá. No entanto o modelo GEV não conseguiu estimar os níveis de retornos para esta cidade, por haver uma proporção de 16,60% de zeros nos dados de máximos. Já o modelo ZIGEV, foi bastante preciso em suas previsões, conseguindo estimar os níveis de retorno para a mesma série de dados, como observado na Figura 9, mostrando-nos que a linha central sobre o ponto de 12 meses (1 ano), indica um retorno de 81,96 mm, o que nos permite afirmar, segundo o modelo, que é esperado a cada ano que ocorra, pelo menos uma vez, uma chuva intensa igual ou maior ao valor estimado. Revelando, dessa maneira, como é recomendável a utilização da distribuição ZIGEV para a análise de dados extremos, principalmente quando eles estiverem, inflados de zeros.

4.6 Conclusão

Este trabalho fez uma aplicação com a distribuição Generalizada de Valores Extremos Inflada de Zeros, ZIGEV, recomendada para conjuntos de dados de máximos, com excesso de zeros. Utilizou-se dados de precipitação diárias de chuvas do estado de Pernambuco, região nordeste do Brasil, para as cidades de Recife da mesorregião Metropolitana do Recife, Barreiros da mesorregião Mata Pernambucana, Surubim e Garanhuns do Agreste Pernambucano, Triunfo, Arcoverde e Ouricuri do Sertão Pernambucano, Cabrobó e Petrolina da mesorregião do São Francisco Pernambucano. Nesta aplicação, o modelo ZIGEV foi excelente na qualidade do ajuste, conseguindo modelar muito bem os dados de

máximos, em todas as diferentes proporções de zeros, variando entre 0,28%, nos dados da cidade de Recife, a 16,60% nos dados de Petrolina.

Para se fazer comparação com os resultados obtidos pela ZIGEV, foi aplicado também aos mesmos dados a distribuição GEV, que usualmente é utilizada para modelar dados de valores extremos. Observou-se que a partir de uma mínima proporção de zeros nos dados, fica claro a existência de uma variação entre as estimativas das duas distribuições, com uma forte indicação para a utilização de uma distribuição que consiga modelar também os zeros máximos presentes nos dados. A diferença entre as estimativas dos modelos GEV e ZIGEV cresceu a medida que a proporção de zeros aumentou, como aconteceu em todas as cidades com proporção de zeros maior e igual a 1,53%.

A partir de uma proporção de 9,85% de zeros nos dados, o modelo GEV se apresentou muito impreciso em suas estimativas, e para a cidade de Petrolina com 16,60% de zeros, o modelo não convergiu pelo método de inferência proposto, mostrando a necessidade e a importância da distribuição ZIGEV para análises de dados de máximos, e inflados de zeros. Também, a medida que a proporção de zeros diminui e vai a zero, $\omega \rightarrow 0$, a distribuição ZIGEV converge para GEV, como aconteceu no caso dos dados da cidade de Recife, com 0,28% de zeros, em que os resultados obtidos pelos dois modelos foram praticamente idênticos. Confirmando o que aconteceu em Gramosa (2019), para a cidade de Natal-RN com 0,20% de zeros nos dados de máximos. Isto corrobora com o importante fato da teoria de que a distribuição GEV é um caso particular da distribuição ZIGEV.

Estes resultados foram confirmados através das medidas da qualidade de ajuste (AIC, BIC e DIC), em que todas as medidas de ajustes para a distribuição ZIGEV foram melhores que para a GEV, com a exceção das medidas AIC e BIC para os dados de Recife com 0,28% de zeros, e que apesar dos resultados dos modelos terem sido bastante semelhantes, estas duas medidas aplicam uma penalização pela quantidade de parâmetros,

Espera-se que esta nova distribuição ZIGEV venha a ser aplicada em qualquer tipo de dado de valores extremos, principalmente com excesso de zeros, auxiliando tanto as autoridades competentes como a sociedade civil de maneira geral, fazendo uso desta ferramenta, para minimizar as perdas humanas, materiais e financeiras, que as chuvas intensas e extremas possam afetar, como também utilizá-la para potencializar ganhos.

5 Distribuição generalizada de valores extremos inflada de zeros sazonais (SZIGEV)

5.1 Introdução

Nos últimos anos, o mundo vem enfrentando grandes mudanças ambientais. Problemas como inundações, devido às precipitações de chuvas intensas e extremas, preocupam as autoridades pelos enormes prejuízos humanos, materiais e financeiros (GRAMOSA et al., 2019). A possibilidade de minimizar os riscos provenientes de eventos extremos com maior segurança, é de extrema importância para o auxílio do planejamento urbano e rural. Abordagens estatísticas focadas em valores extremos mostram resultados promissores na previsão de eventos incomuns em ciências da terra, genética e finanças (THOMAS et al., 2016).

As chuvas intensas são caracterizadas por apresentar grandes lâminas d'água precipitada durante um pequeno intervalo de tempo, e que, frequentemente, geram volumes significativos de escoamento superficial e podem provocar danos socioeconômicos, prejuízos ambientais, além de riscos a vida (SANTOS et al., 2010; PEREIRA et al., 2017).

A caracterização da variabilidade temporal da precipitação pluvial máxima permite estimar a intensidade máxima de chuva esperada em um determinado período de tempo, e possui inúmeras aplicações, como no dimensionamento de obras hidráulicas urbanas e/ou rurais, tais quais como vertedouros de barragens, canais de terraços, drenagem agrícola, urbana e rodoviária, desvio de cursos d'água, práticas mecânicas de controle de erosão hídrica, dentre outros (GARCIA et al., 2011; QUADROS et al., 2011).

Nesse contexto, a Teoria de Valores Extremos (TVE) tem ganhado cada vez mais espaço, tornando possível resultados mais precisos e, com um maior respaldo nas análises e previsões de eventos extremos. Em meio a uma enorme problemática, de analisar dados de valores extremos com excessos de zeros, que geram resultados imprecisos e que as vezes até impedem as previsões de eventos extremos, Gramosa et al. (2019) desenvolveram a distribuição ZIGEV. Contudo, muitos eventos extremos são caracterizados por apresentarem uma alternância no comportamento em algum momento do tempo, seja de forma contínua ou repentina. Situações assim são bem comuns em dados ambientais e financeiros.

Analisando as precipitações de chuvas, em especial as ocorridas no nordeste do Brasil, encontramos uma sazonalidade nos dados. De acordo com Olímpio et al. (2013) e Cavalcanti (2009), no nordeste brasileiro, os desastres causados por enxurradas, inundações bruscas ou graduais e alagamentos, deflagrados por eventos de chuvas *intensas* ($\geq 60,0$ mm/dia) e chuvas *extremas* ($\geq 100,0$ mm/dia), já foram registrados em cidades litorâneas, a exemplo de Fortaleza, bem como em João Pessoa, Maceió, Recife e Olinda.

Temos por principal objetivo nesse trabalho, propôr uma a distribuição Generalizada de Valores Extremos Inflada de Zeros Sazonais (SZIGEV) que permita melhorar a modelagem em dados de máximos inflados de zeros em especial com as características da sazonalidade, proporcionando mais segurança e maior precisão nas previsões desses eventos, prevenindo as autoridades competentes e a sociedade quanto aos seus efeitos e impactos que tantos prejuízos acarretam a sociedade, permitindo assim que mais vidas sejam poupadas. Para tanto, sugerimos fazer com que o parâmetro ω , parâmetro que controla a porcentagem de zeros nos dados, varie de acordo com cada mês do ano. Assim, haverá mês em que a chance de ter um máximo zero será maior, pois obteremos um parâmetro para cada mês. Logo, identificaremos os meses em que são mais propícios ao acontecimento de eventos extremos, no caso, de chuvas intensas e extremas.

A inferência utilizada nos modelos em estudo é baseada numa abordagem bayesiana, em que é assumido que tanto os parâmetros da ZIGEV quanto os parâmetros da SZIGEV possuem uma distribuição de probabilidade, que pode ser descrita também pela verossimilhança, dados os valores observados. A estimação é realizada com o uso do algoritmo de Metrópolis-Hastings, em que as amostras serão feitas por blocos, pelo Método de Monte Carlo via Cadeias de Markov (MCMC), encontradas em Murteira et al.(2018).

5.2 Material e métodos

5.2.1 Área de estudo e dados

Os dados em estudo são de precipitação diária de chuva, coletados na Agência Pernambucana de Águas e Clima (APAC), e no Instituto Nacional de Meteorologia (INMET), sendo resultantes de todos os dados históricos das determinadas estações a seguir, no INMET e na APAC, que foram transformados em blocos de máximos mensais de tamanho n . As estações em análise apresentadas são provenientes das cidades de Recife, Teresina, Petrolina e São João do Piauí dos estados de Pernambuco e do Piauí. Todas estas cidades são localizadas na região nordeste do Brasil, região semiárida, sendo que algumas delas com predominância de meses não chuvosos. Por isso, na série histórica dos dados podem aparecer meses em que não tenha ocorrido nenhuma precipitação de chuva,

sendo estes computados como zero.

A presença excessiva desses zeros, dificulta a modelagem estatística, promove insegurança e imprecisão nos resultados, proporcionam as vezes até o impedimento das análises e previsões, permitindo que inúmeras cidades e regiões fiquem sem o aviso de advertência para as ocorrências desses eventos extremos, que tantos prejuízos e calamidades acarretam na sociedade.

Os valores não-zeros dos dados de máximos mensais em estudo, estavam muito próximos de valores zeros, como 0,1; 0,3; 0,5, dificultando as modelagens realizadas pelo modelo, pois eram analisados zeros contínuos. Então, para solução, foi preciso realizarmos uma transformação, multiplicando as séries dos dados por 10, permitindo o programa analisar corretamente o que era zero e o que não era. Entretanto, fizemos a transformação inversa, dividindo todos os valores dos quantis extremos por 10, para os cálculos e gráficos dos níveis de retornos, para que os dados estivessem na sua escala original.

Gramosa et al. (2019) apresentam resultados de aplicações com algumas destas cidades, e com outras cidades com essas mesmas características, dados de máximos inflados de zeros, que mostram a importância e necessidade de se utilizar a distribuição ZIGEV, em comparação com a distribuição GEV, pois, obteve uma melhor qualidade no ajuste e maior precisão na estimação. As estimações realizadas para os níveis de retornos (5, 10, 20 e 100), correspondentes aos respectivos quantis altos (80%, 90%, 95% e 99%) da ZIGEV, são esperados que ocorram pelo menos uma vez a cada período de tempo (mês).

5.3 Distribuição generalizada de valores extremos inflada de zeros sazonais (SZIGEV)

Considere uma sequência de variáveis aleatórias iid, Y_1, \dots, Y_r que possam ser ordenados em tamanhos de bloco n . Seja X_i o de máximo de cada bloco, em que $i = 1, \dots, n$. Então, para a modelagem dos máximos, temos a seguinte densidade proposta para a distribuição Generalizada de Valores Extremos Inflada de Zeros Sazonais (SZIGEV):

$$h(X|\omega, \mu, \sigma, \xi) = \omega_i I_{\{0\}}(x) + (1 - \omega_i) \frac{1}{\sigma_i} \left[1 + \frac{\xi_i}{\sigma_i} (x - \mu_i) \right]^{-\frac{1}{\xi_i} - 1} \times \exp \left\{ - \left[1 + \frac{\xi_i}{\sigma_i} (x - \mu_i) \right]^{-\frac{1}{\xi_i}} \right\}, \quad (5.1)$$

em que os parâmetros μ_i , σ_i , ξ_i , de acordo com Nascimento (2011), apresentam as seguintes funções de ligação

$$\begin{cases} \mu_i = \beta'_\mu z_k \\ \sigma_i = \exp(\beta'_\sigma z_k) \\ \xi_i = \exp(\beta'_\xi z_k) - 1, \end{cases}$$

em que z_k são as covariáveis dos meses do ano, com $k = 1, \dots, 12$.

Para o parâmetro ω_i , que é responsável pela proporção de zeros no modelo ZIGEV, que possui valores entre 0 e 1, inicialmente propusemos uma função de ligação logística, dada por

$$\omega_i = \frac{\exp(\beta'_\omega z_k)}{1 + (\exp \beta'_\omega z_k)},$$

em que $1 + \exp(\beta'_\omega z_k) > 0$.

Substituindo os parâmetros, na expressão (4.1), por suas respectivas funções de ligação, encontramos a seguinte expressão da função densidade proposta para a SZIGEV

$$\begin{aligned} h(X|\omega, \mu, \sigma, \xi) &= \frac{\exp(\beta'_\omega z_k)}{1 + \exp(\beta'_\omega z_k)} I_{\{0\}}(x) + \left[1 - \frac{\exp(\beta'_\omega z_k)}{1 + \exp(\beta'_\omega z_k)} \right] \frac{1}{\exp(\beta'_\sigma z_k)} \times \\ &\quad \left[1 + \frac{\exp(\beta'_\xi z_k) - 1}{\exp(\beta'_\sigma z_k)} (x - \beta'_\mu z_k) \right]^{-\frac{1}{\exp(\beta'_\xi z_k) - 1} - 1} \times \quad (5.2) \\ &\quad \exp \left\{ - \left[1 + \frac{\exp(\beta'_\xi z_k) - 1}{\exp(\beta'_\sigma z_k)} (x - \beta'_\mu z_k) \right]^{-\frac{1}{\exp(\beta'_\xi z_k) - 1}} \right\}. \end{aligned}$$

5.3.1 Função de verossimilhança

Considerando uma amostra de tamanho m da densidade descrita na equação (4.2), obtemos a seguinte função de verossimilhança:

$$\begin{aligned}
L(\omega, \mu, \sigma, \xi) &= \prod_{x_i \in A} \left\{ \frac{\exp(\beta'_\omega z_k)}{1 + \exp(\beta'_\omega z_k)} + \left[1 - \frac{\exp(\beta'_\omega z_k)}{1 + \exp(\beta'_\omega z_k)} \right] \right. \\
&\quad \times \frac{1}{\exp(\beta'_\sigma z_k)} \left\{ 1 - \frac{\beta'_\mu z_k}{\exp(\beta'_\sigma z_k)} \left[\exp(\beta'_\xi z_k) - 1 \right] \right\}^{-\frac{1}{\exp(\beta'_\xi z_k) - 1} - 1} \\
&\quad \times \exp \left\{ - \left\{ 1 - \frac{\beta'_\mu z_k}{\exp(\beta'_\sigma z_k)} \left[\exp(\beta'_\xi z_k) - 1 \right] \right\}^{-\frac{1}{\exp(\beta'_\xi z_k) - 1}} \right\} \\
&\quad \times \prod_{x_i \notin A} \left\{ \left[1 - \frac{\exp(\beta'_\omega z_k)}{1 + \exp(\beta'_\omega z_k)} \right] \frac{1}{\exp(\beta'_\sigma z_k)} \right. \\
&\quad \times \left[1 + \frac{\exp(\beta'_\xi z_k) - 1}{\exp(\beta'_\sigma z_k)} (x_i - \beta'_\mu z_k) \right]^{-\frac{1}{\exp(\beta'_\xi z_k) - 1} - 1} \\
&\quad \times \exp \left\{ - \left[1 + \frac{\exp(\beta'_\xi z_k) - 1}{\exp(\beta'_\sigma z_k)} (x_i - \beta'_\mu z_k) \right]^{-\frac{1}{\exp(\beta'_\xi z_k) - 1}} \right\} \left. \right\}, \tag{5.3}
\end{aligned}$$

em que os elementos de A provém da distribuição degenerada de zero ou de $p(0)$.

5.3.2 Parametrização do ω em função dos meses do ano

No trabalho proposto, inicialmente foi pensado em fazer uma reparametrização nos parâmetros μ , σ , ξ e ω , da distribuição ZIGEV, para que variassem de acordo com suas covariáveis, num modelo de regressão. Para os parâmetros μ , σ e ξ , foi feita uma reparametrização similar ao realizado no trabalho de Nascimento (2011), utilizando as funções de ligações nele proposto. Já para o parâmetro ω , o natural seria fazer uma função de ligação logística, pois ω é um parâmetro entre 0 e 1. Ele é o peso, por isso, está entre 0 e 1.

Assim, com todos os parâmetros dependendo de suas covariáveis, teríamos a função de verossimilhança apresentada na expressão (4.3). Chegamos a encontrar a distribuição a posteriori, porém ao fazer a estimação por MCMC, os parâmetros associados a regressão logística do ω não tiveram uma forma conhecida, não sendo possível o mecanismo que Rodrigues (2003) e Gramosa et al.,(2019) fizeram. Tentamos fazer a estimação pelo MCMC, direto, calculando a priori, porém, como os parâmetros associados ao ω são os betas e não mais simplesmente os ω 's não é possível fazer a estimação desses parâmetros pelo

Metrópolis Hastings como foi feito pelo Rodrigues (2003) e Gramosa et al.,(2019). Então dessa maneira, a parametrização proposta não foi possível ser feita, a estimação do ω e os valores encontrados pela técnica de Metrópolis Hastings não coincidem com os valores reais de proporção de zeros no banco de dados trabalhado.

Gostaríamos de ter a mesma reparametrização feita no trabalho de Rodrigues (2003), a mesma que foi realizada no trabalho de Gramosa et. al. (2019). A diferença é que não foi possível propor uma priori com distribuição Beta, como Rodrigues fez em 2003 e dessa maneira não conseguimos encontrar uma posteriori fechada, ao contrário do que foi obtido em Gramosa (2019). Mas, agora, quando utilizamos a função de ligação logística, os betas perdem essa propriedade de ter uma posteriori "fechada", como acontece nos trabalhos de Rodrigues e Gramosa. Por isso, foi necessário pensar em trabalharmos ω de uma outra maneira, mas no sentido de ainda indicar a sazonalidade dos dados.

Com relação aos parâmetros μ , σ e ξ seguimos com a proposta inicial, sendo os seus respectivos parâmetros o β_μ , β_σ e β_ξ . Logo, sabendo os valores dos betas, saberemos os valores de μ , σ e ξ . Por exemplo, se sabemos quais são os betas do μ , fazemos $\beta'_\mu \times z$ e encontramos μ . Se conseguimos estimar os betas do σ , fazemos $\exp(\beta'_\sigma z)$ e encontramos o sigma. Da mesma forma é encontrado para o ξ , através da sua função de ligação.

Portanto, para manter ainda um padrão de sazonalidade do ω e ao mesmo tempo, utilizar os mecanismos de posteriori utilizados em Rodrigues (2003) e Gramosa (2019), considerando que as covariáveis são o seno do mês, o cosseno do mês e a multiplicação entre eles consideramos um ω diferente para cada mês do ano, sendo um vetor $(\omega_1, \dots, \omega_{12})$. Logo, consideramos que sua função densidade em um mês i pode ser expressa da seguinte maneira

$$\begin{aligned}
 h(x_i|\omega, \mu, \sigma, \xi) &= \omega_{mes_j} I_{\{0\}}(x_i) + (1 - \omega_{mes_j}) \frac{1}{\exp(\beta'_\sigma z_k)} \\
 &\times \left[1 + \frac{\exp(\beta'_\xi z_k) - 1}{\exp(\beta'_\sigma z_k)} (x_i - \beta'_\mu z_k) \right]^{-\frac{1}{\exp(\beta'_\xi z_k) - 1} - 1} \\
 &\times \exp \left\{ - \left[1 + \frac{\exp(\beta'_\xi z_k) - 1}{\exp(\beta'_\sigma z_k)} (x_i - \beta'_\mu z_k) \right]^{-\frac{1}{\exp(\beta'_\xi z_k) - 1}} \right\},
 \end{aligned}$$

em que $i = 1, \dots, n$ e $k = 1, \dots, 12$.

Considerando essa função densidade, na expressão anterior, temos que a sua função de verossimilhança será expressa por

$$\begin{aligned}
L(\omega, \mu, \sigma, \xi) = & \prod_{x_i \in A} \left\{ \omega_{mes_j} + (1 - \omega_{mes_j}) \frac{1}{\exp(\beta'_\sigma z_k)} \right. \\
& \times \left\{ 1 - \frac{\beta'_\mu z_k}{\exp(\beta'_\sigma z_k)} \left[\exp(\beta'_\xi z_k) - 1 \right] \right\}^{-\frac{1}{\exp(\beta'_\xi z_k) - 1} - 1} \\
& \times \exp \left\{ - \left\{ 1 - \frac{\beta'_\mu z_k}{\exp(\beta'_\sigma z_k)} \left[\exp(\beta'_\xi z_k) - 1 \right] \right\}^{-\frac{1}{\exp(\beta'_\xi z_k) - 1}} \right\} \left. \right\} \\
& \times \prod_{x_i \notin A} \left\{ (1 - \omega_{mes_j}) \frac{1}{\exp(\beta'_\sigma z_k)} \right. \\
& \times \left[1 + \frac{\exp(\beta'_\xi z_k) - 1}{\exp(\beta'_\sigma z_k)} (x_i - \beta'_\mu z_k) \right]^{-\frac{1}{\exp(\beta'_\xi z_k) - 1} - 1} \\
& \times \exp \left\{ - \left[1 + \frac{\exp(\beta'_\xi z_k) - 1}{\exp(\beta'_\sigma z_k)} (x_i - \beta'_\mu z_k) \right]^{-\frac{1}{\exp(\beta'_\xi z_k) - 1}} \right\} \left. \right\}, \tag{5.4}
\end{aligned}$$

em que os elementos de A provém da distribuição degenerada de zero ou de $p(0)$. A é o conjunto de valores em que x_i são iguais a zero.

5.3.3 Distribuição a priori

Propomos para os parâmetros da SZIGEV, referentes ao intercepto e demais parâmetros, as seguintes distribuições a priori:

$$\mu : \beta_{\mu_0} \sim N(0, V_{\beta_{\mu_0}}) \text{ e } \beta_{\mu_i} \sim N(0, V_{\beta_{\mu_i}})$$

$$\sigma : \beta_{\sigma_0} \sim N(0, V_{\beta_{\sigma_0}}) \text{ e } \beta_{\sigma_i} \sim N(0, V_{\beta_{\sigma_i}})$$

$$\xi : \beta_{\xi_0} \sim N(0, V_{\beta_{\xi_0}}) \text{ e } \beta_{\xi_i} \sim N(0, V_{\beta_{\xi_i}}).$$

Como temos um ω e um S para cada mês do ano, iremos considerar as prioris da seguinte maneira

$$\omega_j \sim \text{Beta}(S_j + c, n_j - S_j + d) \sim \omega_j^{S_j+c-1} (1 - \omega_j)^{n_j-S_j+d-1}$$

$$S_j \sim \text{Bin}[m_j, p(\theta, \omega_j)] \sim \binom{m_j}{S_j} p(\theta, \omega_j)^{S_j} [1 - p(\theta, \omega_j)]^{m_j-S_j},$$

em que $j = 1, \dots, 12$.

Assim, a proporcional da distribuição a priori conjunta da SZIGEV é expressa da seguinte maneira

$$p(\boldsymbol{\omega}, \beta_\mu, \beta_\sigma, \beta_\xi) \propto \prod_{j=1}^{12} \omega_j^{S_j+c-1} (1 - \omega_j)^{n_j-S_j+d-1}$$

$$\times \exp \left[\frac{\beta_{\mu_0}^2}{2V_{\beta_{\mu_0}}} + \sum_{i=1}^p \left(\frac{\beta_{\mu_i}^2}{2V_{\beta_{\mu_i}}} \right) \right]$$

$$\times \exp \left[\frac{\beta_{\sigma_0}^2}{2V_{\beta_{\sigma_0}}} + \sum_{i=1}^p \left(\frac{\beta_{\sigma_i}^2}{2V_{\beta_{\sigma_i}}} \right) \right]$$

$$\times \exp \left[\frac{\beta_{\xi_0}^2}{2V_{\beta_{\xi_0}}} + \sum_{i=1}^p \left(\frac{\beta_{\xi_i}^2}{2V_{\beta_{\xi_i}}} \right) \right],$$
(5.5)

com $\boldsymbol{\omega}$ sendo o vetor de ω em todos os meses $\omega_1, \dots, \omega_{12}$ e $i = 1, \dots, n$.

A estimação dos parâmetros é feita pelo algoritmo de Metrópolis Hastings para os Betas do ξ , σ e μ , e por amostrador de Gibbs para ω e S .

5.3.4 Distribuição a posteriori

A distribuição a posteriori conjunta da SZIGEV, representada por $\pi(\boldsymbol{\omega}, \mu, \sigma, \xi)$, é encontrada através da seguinte expressão

$$\pi(\boldsymbol{\omega}, \mu, \sigma, \xi) \propto L(\boldsymbol{\omega}, \mu, \sigma, \xi) p(\boldsymbol{\omega}, \beta_\mu, \beta_\sigma, \beta_\xi).$$
(5.6)

Portanto, para encontrarmos a função de distribuição a posteriori, consideremos a função de verossimilhança em (5.4) e a função de distribuição a priori descrita em (5.5), substituindo-as na equação (5.6) obtendo a seguinte expressão:

$$\begin{aligned}
\pi(\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}) &\propto \prod_{j=1}^{12} \left\{ \omega_j^{S_j+c-1} (1-\omega_j)^{n_j-S_j+d-1} \right\} \\
&\times \prod_{x_i \in A} \left\{ \frac{1}{\exp(\beta'_\sigma z_k)} \left\{ 1 - \frac{\beta'_\mu z_k}{\exp(\beta'_\sigma z_k)} [\exp(\beta'_\xi z_k) - 1] \right\}^{-\frac{1}{\exp(\beta'_\xi z_k)-1}-1} \right. \\
&\times \exp \left\{ - \left\{ 1 - \frac{\beta'_\mu z_k}{\exp(\beta'_\sigma z_k)} [\exp(\beta'_\xi z_k) - 1] \right\}^{-\frac{1}{\exp(\beta'_\xi z_k)-1}} \right\} \left. \right\} \\
&\times \prod_{x_i \notin A} \left\{ \left[1 - \frac{\exp(\beta'_\omega z_k)}{1 + \exp(\beta'_\omega z_k)} \right] \frac{1}{\exp(\beta'_\sigma z_k)} \right. \\
&\times \left[1 + \frac{\exp(\beta'_\xi z_k) - 1}{\exp(\beta'_\sigma z_k)} (x_i - \beta'_\mu z_k) \right]^{-\frac{1}{\exp(\beta'_\xi z_k)-1}-1} \\
&\times \exp \left\{ - \left[1 + \frac{\exp(\beta'_\xi z_k) - 1}{\exp(\beta'_\sigma z_k)} (x_i - \beta'_\mu z_k) \right]^{-\frac{1}{\exp(\beta'_\xi z_k)-1}} \right\} \left. \right\} \\
&\times \exp \left\{ \left[\frac{\beta_{\mu_0}^2}{2V_{\beta_{\mu_0}}} + \sum_{i=1}^n \left(\frac{\beta_{\mu_i}^2}{2V_{\beta_{\mu_i}}} \right) \right] + \left[\frac{\beta_{\sigma_0}^2}{2V_{\beta_{\sigma_0}}} + \sum_{i=1}^n \left(\frac{\beta_{\sigma_i}^2}{2V_{\beta_{\sigma_i}}} \right) \right] \right\} \\
&\times \exp \left\{ \left[\frac{\beta_{\xi_0}^2}{2V_{\beta_{\xi_0}}} + \sum_{i=1}^n \left(\frac{\beta_{\xi_i}^2}{2V_{\beta_{\xi_i}}} \right) \right] \right\}.
\end{aligned}$$

Como a posteriori não tem distribuição conhecida, obtemos aproximações numéricas através das cadeias de Markov (MCMC).

5.3.5 Função log-posteriori

Para uma melhor implantação do algoritmo de Metrópolis Hastings, aplicamos o logaritmo na equação anterior. Portanto, a função log-posteriori é apresentada por

$$\begin{aligned}
\log \pi(\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}) &\propto \sum_{j=1}^{12} [(S_j + c - 1) \log(w_j) + (n_j - S_j + d - 1)(\log(1 - w_j))] \\
&\sum_{x_i \in A} \log \left\{ \frac{1}{\exp(\beta'_\sigma z_k)} \left\{ 1 - \frac{\beta'_\mu z_k}{\exp(\beta'_\sigma z_k)} [\exp(\beta'_\xi z_k) - 1] \right\}^{-\frac{1}{\exp(\beta'_\xi z_k) - 1} - 1} \right. \\
&\quad \times \frac{1}{\exp(\beta'_\sigma z_k)} \left\{ 1 - \frac{\beta'_\mu z_k}{\exp(\beta'_\sigma z_k)} [\exp(\beta'_\xi z_k) - 1] \right\}^{-\frac{1}{\exp(\beta'_\xi z_k) - 1} - 1} \\
&\quad \times \exp \left\{ - \left\{ 1 - \frac{\beta'_\mu z_k}{\exp(\beta'_\sigma z_k)} [\exp(\beta'_\xi z_k) - 1] \right\}^{-\frac{1}{\exp(\beta'_\xi z_k) - 1} - 1} \right\} \left. \right\} \\
&+ \sum_{x_i \notin A} \left\{ \log \left[1 - \frac{\exp(\beta'_\omega z_k)}{1 + \exp(\beta'_\omega z_k)} \right] + \log \left\{ \frac{1}{\exp(\beta'_\sigma z_k)} \right. \right. \\
&\quad \times \left[1 + \frac{\exp(\beta'_\xi z_k) - 1}{\exp(\beta'_\sigma z_k)} (x_i - \beta'_\mu z_k) \right]^{-\frac{1}{\exp(\beta'_\xi z_k) - 1} - 1} \\
&\quad \times \exp \left\{ - \left[1 + \frac{\exp(\beta'_\xi z_k) - 1}{\exp(\beta'_\sigma z_k)} (x_i - \beta'_\mu z_k) \right]^{-\frac{1}{\exp(\beta'_\xi z_k) - 1} - 1} \right\} \left. \right\} \\
&+ \left[\frac{\beta_{\mu_0}^2}{2V_{\beta_{\mu_0}}} + \sum_{i=1}^p \left(\frac{\beta_{\mu_i}^2}{2V_{\beta_{\mu_i}}} \right) \right] + \left[\frac{\beta_{\sigma_0}^2}{2V_{\beta_{\sigma_0}}} + \sum_{i=1}^p \left(\frac{\beta_{\sigma_i}^2}{2V_{\beta_{\sigma_i}}} \right) \right] \\
&+ \left[\frac{\beta_{\xi_0}^2}{2V_{\beta_{\xi_0}}} + \sum_{i=1}^p \left(\frac{\beta_{\xi_i}^2}{2V_{\beta_{\xi_i}}} \right) \right].
\end{aligned}$$

5.4 Aplicação da SZIGEV em dados de chuvas em algumas cidades dos estados de Pernambuco e Piauí

Nesta seção foi feita uma aplicação com o modelo proposto para os dados de precipitação diárias de chuvas, transformados em máximos mensais, referentes as cidades de Recife e Petrolina dos estado de Pernambuco e das cidades de São João do Piauí e Teresina do estado do Piauí. Os resultados das análises são apresentados a seguir.

Analisando a Tabela 5, percebemos que dos 700 meses observados na cidade de Recife, apenas 2 meses foram registrados 0 mm de chuvas, ou seja, segundo o histórico

Tabela 5 – Quantidade (S) e proporção de zeros (ω) nos dados de máximos de chuva dos estados de Pernambuco e do Piauí

| Cidades | Amostras (n) | Zeros | | |
|-------------------|------------------|-----------------|--------------------------|---------|
| | | Quantidades (S) | Proporçõess (ω) | (%) |
| Recife | 700 | 2 | 0,0028 | (0,28) |
| Teresina | 462 | 40 | 0,0865 | (8,65) |
| Petrolina | 542 | 90 | 0,1660 | (16,60) |
| São João do Piauí | 359 | 101 | 0,2813 | (28,13) |

dos dados, praticamente todos os meses registrados tiveram chuvas, obtendo a menor proporção de zeros nos dados (0,28%) com relação as demais cidades em estudo. Notamos também que na cidade de São João do Piauí, dos 359 meses em estudos, 101 meses foram sem chuvas, correspondendo a 28,13% de 0 (mm) nos dados, de máximos mensais, ou seja, mais de 1/4 (um quarto) do período histórico foi sem chuvas, a maior porcentagem entre as cidades em análise.

Propomos aplicar a SZIGEV nessas quatro cidades, apresentadas na Tabela 5, por apresentarem algumas das porcentagens de zeros que mais tiveram destaque nos trabalhos em comparação com as distribuições GEV e ZIGEV, para dados de máximos mensais de chuva. Pois, com a proporção de 0,28% de zeros nos dados, os resultados obtidos pelas duas distribuições foram idênticos, já com 8,65% de zeros nos dados as análises feitas pela GEV são totalmente distorcidas, a partir de 16,60% de zeros ela não consegue mais modelar os dados, sendo 28,13% de zeros um caso ainda mais extremo. Também por ter provável sazonalidade em seus dados.

Temos na Tabela 6, a quantidade e proporção de zeros máximos mensais obtidas em cada cidade analisada. Os meses que sempre chovem, registram valores máximos diferentes de zeros (mm). Portanto, de acordo com a série histórica dos dados, os meses que não tiveram valores extremos sendo zeros, não tiveram contagens na quantidade e proporção de zero, estando representados pelo número zero (0) na tabela, indicando que naquele mês choveu, logo seu máximo mensal foi diferente de zero mm. Temos também os meses da série que pelo menos uma vez ficaram sem chuvas, no qual estão sendo informados a quantidade e a proporção de vezes que ficaram sem chuvas, ou seja, quando os valores dos máximos mensais foram zeros (mm).

Na Tabela 6, notamos que na cidade de Recife há chuva praticamente em todos os meses do ano, exceto outubro e novembro - meses esses em que os zeros (mm) foram quantificados. Em Teresina o período chuvoso vai de dezembro a meados de maio, nos demais meses pelo menos uma vez houve registro que passou sem chuvas, especialmente nos meses de julho a setembro. Os dados da cidade de Petrolina mostram que todos os meses do ano já ficaram sem chuvas, pelo menos uma vez, sendo os meses de maio a

Tabela 6 – Quantidade (S) e proporção de zeros, das cidades em estudo, observados em cada mês do ano

| Mês | Cidades | | | | | | | |
|---------------|---------|---------------|----------|---------------|-----------|---------------|-------------------|---------------|
| | Recife | | Teresina | | Petrolina | | São João do Piauí | |
| | (S) | (ω %) | (S) | (ω %) | (S) | (ω %) | (S) | (ω %) |
| Janeiro | 0 | 0 | 0 | 0 | 2 | 4,35 | 0 | 0 |
| Fevereiro | 0 | 0 | 1 | 2,56 | 1 | 2,17 | 1 | 3,23 |
| Março | 0 | 0 | 0 | 0 | 1 | 2,17 | 1 | 3,13 |
| Abril | 0 | 0 | 0 | 0 | 3 | 6,52 | 0 | 0 |
| Maio | 0 | 0 | 1 | 2,56 | 9 | 19,57 | 9 | 25,81 |
| Junho | 0 | 0 | 3 | 7,69 | 7 | 15,22 | 18 | 58,06 |
| Julho | 0 | 0 | 12 | 30,77 | 6 | 13,04 | 23 | 74,19 |
| Agosto | 0 | 0 | 11 | 28,21 | 18 | 39,13 | 23 | 74,19 |
| Setembro | 0 | 0 | 10 | 25,64 | 20 | 43,48 | 18 | 58,06 |
| Outubro | 1 | 1,69 | 1 | 2,56 | 17 | 36,96 | 8 | 25,81 |
| Novembro | 1 | 1,69 | 1 | 2,56 | 5 | 10,87 | 1 | 3,23 |
| Dezembro | 0 | 0 | 0 | 0 | 1 | 2,17 | 0 | 0 |
| Total (S) | 2 | | 40 | | 90 | | 101 | |

novembro aparentemente os mais críticos. Na cidade de São João do Piauí os meses de dezembro a abril, são os que chovem constantemente, nos demais, com exceção de janeiro e dezembro, pelo menos uma vez houve registro que passou sem chuvas, indicando serem os mais críticos os meses de maio a outubro.

Na Tabela 6, além de observarmos quais são os meses que sempre chovem, e os meses que ficaram sem chover, de acordo com a série histórica dos dados, percebemos também a proporção do período sem chuvas em cada mês. Quanto mais perto de zero estiver o valor da proporção no mês, maior é o indicativo de chuva naquele mês, sendo que quanto mais distante do valor zero estiver a proporção de zero (mm), menor é o indicativo de chuva para aquele mês, ou seja, maior é o período de seca.

Identificamos ainda na Tabela 6 que nos dados de máximos da cidade de Recife, a maioria dos meses não tiveram zeros (mm) como seus valores máximos, fazendo com que a proporção de zeros (mm) registrados nesses meses sejam zeros. Indicando constantes chuvas durante este período. Os meses de outubro e novembro tiveram 1,69 como proporção de zero (mm), um valor próximo de zero (≈ 0), mostrando que, mesmo passando por um pequeno período sem chuvas, são meses chuvosos. Na cidade de Teresina, nos possíveis meses críticos, de julho a setembro, foram registrados proporções de zeros respectivamente iguais a 30,8%, 28,2% e 25,6% indicando possível período de seca. Em Petrolina, no indicado período crítico, de maio a novembro, tiveram proporções de 10,9% a 43,5% e em São João do Piauí durante seus possíveis meses críticos, de maio a outubro, tiveram proporções de 25,8% a 74,2% de zeros nos dados, mostrando maior intensidade sem chuvas

durante esse período.

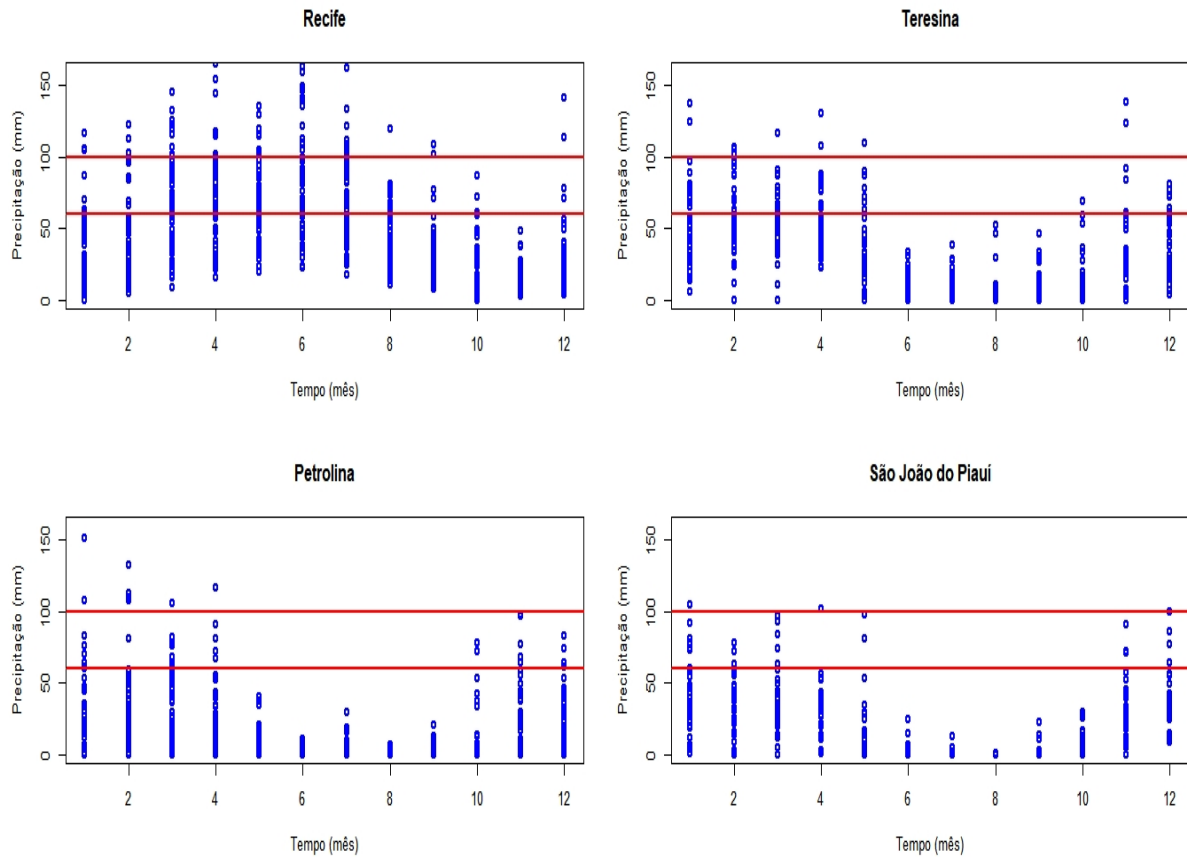


Figura 7 – Gráfico dos pontos dos máximos dos meses do ano nas cidades em estudo

Na Figura 7, percebemos que a sazonalidade é diferente para as cidades ao longo do ano. Na cidade de Recife, as maiores chances de ter chuvas intensas (entre 60 mm e 100 mm) e extremas (acima de 100 mm) são nos meses de abril a julho, enquanto que na cidade de Teresina as chances são maiores nos meses de janeiro, de março a maio, e em novembro. Para as outras cidades tiveram comportamentos diferentes, por exemplo para Petrolina as chuvas ocorrem no começo do ano, assim também como em São João do Piauí.

A Figura 8 é o contrário do gráfico anterior, pois apresenta o período sem chuvas, ou seja, o período de seca ao longo do ano, para cada cidade em estudo. A seca é um fenômeno natural que não possui uma definição rigorosa e universal. Pode ser entendida como deficiência em precipitação (chuva) por um extenso período de tempo, resultando em escassez hídrica com repercussões negativas significativas nos ecossistemas e nas atividades socioeconômicas.

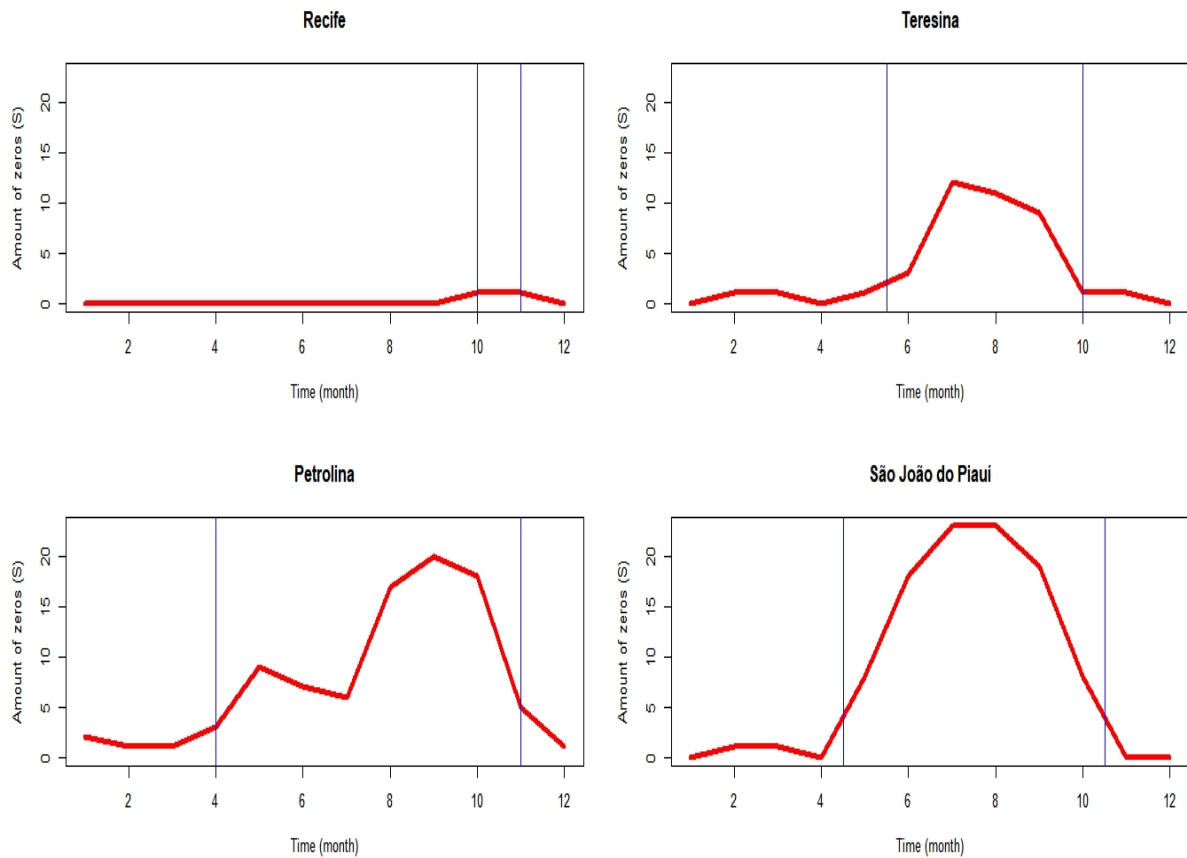


Figura 8 – Gráfico dos zeros (mm) ao longo dos meses no ano para as cidades em estudo

Percebemos, na Figura 8, que a cidade de Recife praticamente não apresenta períodos de seca, como é visto também na tabela 6, pois, praticamente, não há mês do ano em que não chove, com exceção de uma vez em outubro e em novembro. Já na cidade de Teresina, notamos que o período de junho até meados de outubro perpassa sem chuvas, sendo que elas acontecem mais no final de outubro a maio, mostrando bem essas características. Tendo o mês de julho como o pior mês de seca. As cidades de Petrolina e São João do Piauí possuem comportamentos bem diferentes na seca, apesar de serem bem parecidas com relação as chuvas. Podemos ver que os dados de São João do Piauí tem um movimento muito sazonal, quase que perfeito, começa em maio e vai subindo até o mês de julho, mantendo a altura constante até agosto e depois vai caindo de maneira quase que simétrica. Essa curva chega a um nível muito alto, apontando mais de 20 anos sem chuvas nesses meses de julho a agosto. Se observarmos, de junho a setembro temos quatro meses em que suas quantidades de zeros são acima de 15, ou seja, mais de 15 anos sem chuvas durante esse período. São quatro meses bem delicados. Enquanto que na cidade de Petrolina a sua curva mais alta são nos meses de agosto a outubro, em que apontam uma

quantidade acima de 15 zeros (mm), ou seja, são nesses três meses que a cidade tem o seu maior período de seca.

5.4.1 Cidade de Recife

Foi aplicado nos dados de precipitação de chuva da cidade de Recife o modelo proposto, a SZIGEV, com o intuito de analisar o seu comportamento em dados que possuem 0% de zeros máximos.

Os resultados a seguir na Tabela 7 apresentam as estimativas realizadas pelos estimadores $\hat{\omega}$ e \hat{S} , referindo-se a proporção e a quantidades de zeros respectivamente, do modelo SZIGEV para os dados da cidade de Recife.

Tabela 7 – Estimativas para os parâmetros ω e S do modelo SZIGEV respectivo a cada mês do ano nos dados de Recife

| Mês | Estimadores | |
|-----------|--------------------------|-------------------------------|
| | Quantidade (\hat{S}) | Proporção ($\hat{\omega}$ %) |
| Janeiro | 0 | 0 |
| Fevereiro | 0 | 0 |
| Março | 0 | 0 |
| Abril | 0 | 0 |
| Mai | 0 | 0 |
| Junho | 0 | 0 |
| Julho | 0 | 0 |
| Agosto | 0 | 0 |
| Setembro | 0 | 0 |
| Outubro | 0,20 | 2,00 |
| Novembro | 0,17 | 1,70 |
| Dezembro | 0 | 0 |

Observamos na Tabela 7, que as estimativas obtidas pelos estimadores $\hat{\omega}$ e \hat{S} do modelo SZIGEV, para a cidade de Recife, foram bem precisas em cada mês do ano, tanto com relação a quantidade de zeros ($\hat{S}_{mês}$) quanto com a proporção de zeros ($\hat{\omega}_{mês}$) dos dados, sendo confirmados pelos valores descritos e apresentados nas Tabelas 6. Por exemplo, na comparação dos valores de $S_{novembro}$ com os de $\hat{S}_{novembro}$ e na comparação dos valores de $\omega_{novembro}$ com os de $\hat{\omega}_{novembro}$ [(1,00 e 0,17) e (1,69 e 1,70)], respectivamente.

A seguir na Tabela 8, apresentamos os resultados obtidos pelos estimadores $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\xi}$, referentes aos parâmetros de locação, escala e forma, respectivamente, do modelo SZIGEV, para os dados da cidade de Recife.

Para utilizar o modelo de regressão nos dados das cidades, temos que considerar as seguintes covariáveis: X_1 = o cosseno do mês, X_2 = o seno do mês e X_3 = a multiplicação

Tabela 8 – Estimativas para os β 's (0,1,2, e 3) referente a cada parâmetro μ , σ e ξ do modelo SZIGEV, para os dados de Recife, com seus respectivos intervalos de confiança

| Estimadores | | | |
|----------------------------|-------------------------------|---------------------------|--------------------------|
| $\hat{\beta}_{0,\mu}$ | $\hat{\beta}_{1,\mu}$ | $\hat{\beta}_{2,\mu}$ | $\hat{\beta}_{3,\mu}$ |
| 314,71 (298,14; 331,59) | -183,45 (-202,68; -163,55) | 110,70 (89,10; 131,51) | 23,19 (-8,92; 58,09) |
| $\hat{\beta}_{0,\sigma}$ | $\hat{\beta}_{1,\sigma}$ | $\hat{\beta}_{2,\sigma}$ | $\hat{\beta}_{3,\sigma}$ |
| 5,15 (5,08; 5,22) | -0,45 (-0,55; -0,35) | 0,38 (0,28; 0,48) | 0,36 (0,17; 0,55) |
| $\hat{\beta}_{0,\xi}$ | $\hat{\beta}_{1,\xi}$ | $\hat{\beta}_{2,\xi}$ | $\hat{\beta}_{3,\xi}$ |
| 0,18 (0,12; 0,24) | 0,07 (-0,02; 0,15) | -0,09 (-0,17; -0,01) | 0,05 (-0,11; 0,20) |

entre o cosseno e o seno. O β_0 é o intercepto da variável de chuva, e o β_1 , β_2 e o β_3 são os efeitos nos meses ao longo do ano.

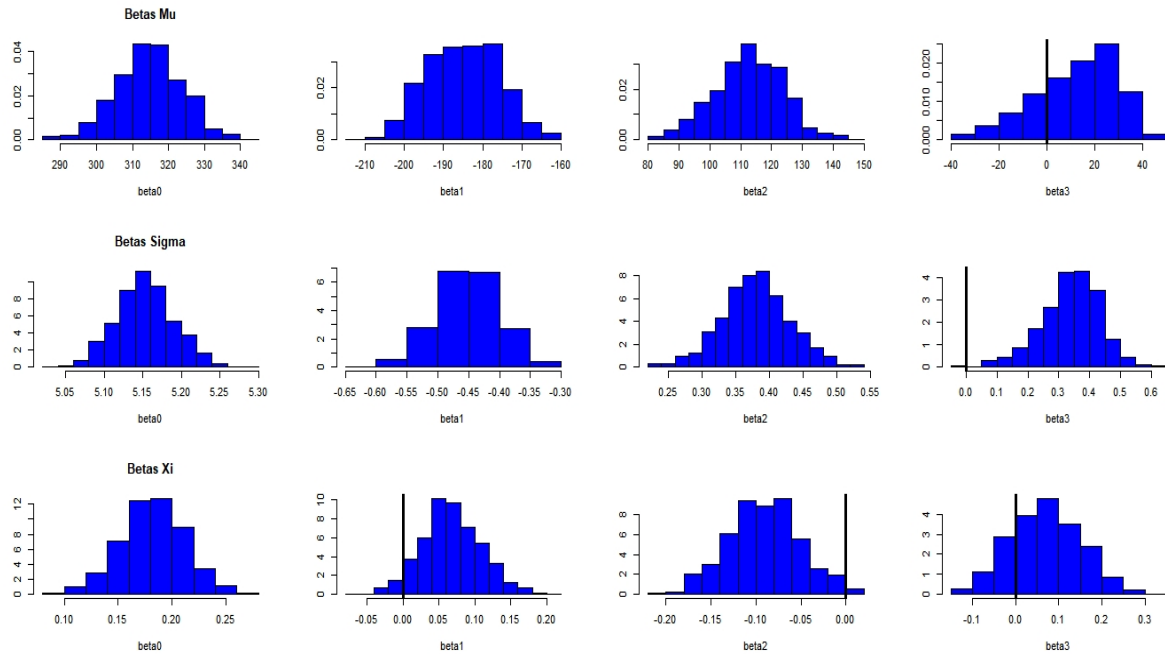
De acordo com os resultados apresentados na Tabela 8, tem-se que: para o parâmetro μ , tanto o β_1 quanto o β_2 são significativos. β_1 , tem o efeito significativo do cosseno. É uma covariável que está influenciando no parâmetro μ , e o β_2 , que é efeito do seno do mês, também está influenciando no parâmetro μ , já o β_3 não, pois o valor zero está dentro do seu intervalo $(-8,92; 58,09)$. Todos os parâmetros que tem o valor zero dentro dos seus intervalos de confiança, não estão influenciando. Para o parâmetro σ , todas as três covariáveis estão influenciando no modelo. Já com relação ao parâmetro ξ , percebemos que ele sofre pouco influência das covariáveis, tendo um comportamento similar para os meses do ano.

É interessante observamos que as precipitações de chuvas, com relação aos máximos mensais, são praticamente constantes a cada mês, sendo que em apenas dois meses os máximos foram zeros. Mesmo assim, existe uma sazonalidade na intensidade das chuvas, e essa intensidade é percebida também pelo modelo SZIGEV que nos mostra que as covariáveis selecionadas influenciam no tempo.

Na Figura 9, temos os histogramas dos betas (0, 1, 2 e 3) para os parâmetros μ , σ e ξ com dados de Recife.

Percebemos, na Figura 9, que o gráfico do β_3 , para o parâmetro μ , tem a linha do zero praticamente no meio do histograma, reforçando que este parâmetro não sofre

Figura 9 – Histogramas dos betas, para os parâmetros μ , σ e ξ da SZIGEV, para a cidade de Recife



influência dessa covariável, a saber a covariável mês. Para o parâmetro σ , apesar do gráfico do β_3 ter a linha do zero no histograma, o parâmetro sofre influência dessa covariável, pois o zero está bem na cauda, o que é confirmado na Tabela 8 em seu respectivo intervalo de confiança, que não contém o valor zero. O parâmetro ξ teve o comportamento do seu β_3 semelhante ao β_3 do parâmetro μ , e o comportamento do seu β_2 semelhante ao do β_3 do parâmetro σ .

Como os betas foram significativos para os parâmetros μ , σ e ξ , podemos afirmar que o modelo SZIGEV foi mais preciso que o modelo ZIGEV, para os dados de Recife.

5.4.2 Cidade de Teresina

O modelo proposto, SZIGEV, foi aplicado aos dados de precipitação de chuva de Teresina, com o objetivo de analisar o seu comportamento em dados que estejam inflados, com mais de 8%, de zeros máximos de maneira geral, mas que nos meses foram observados porcentagens bem maiores.

Os resultados a seguir na Tabela 9 representam as estimativas realizadas pelos estimadores $\hat{\omega}$ e \hat{S} do modelo SZIGEV nos dados da cidade de Teresina, referindo-se a proporção e a quantidades de zeros respectivamente.

Percebemos nessa Tabela 9, que a estimação realizada pela SZIGEV, através dos

Tabela 9 – Estimativas para os parâmetros ω e S do modelo SZIGEV respectivo a cada mês do ano nos dados de Teresina

| Mês | Estimadores | |
|-----------|--------------------------|-------------------------------|
| | Quantidade (\hat{S}) | Proporção ($\hat{\omega}$ %) |
| Janeiro | 0 | 0 |
| Fevereiro | 0,06 | 0,10 |
| Março | 0,05 | 0,02 |
| Abril | 0 | 0 |
| Mai | 0,03 | 0,10 |
| Junho | 2,27 | 5,00 |
| Julho | 11,73 | 30,10 |
| Agosto | 10,36 | 27,30 |
| Setembro | 8,66 | 22,30 |
| Outubro | 0,02 | 0,10 |
| Novembro | 0,07 | 0,20 |
| Dezembro | 0 | 0 |

estimadores \hat{S} e $\hat{\omega}$, também foi precisa para os dados de Teresina, comparando seus resultados com os dados apresentados na tabela 6, como por exemplo para o mês de julho segue que: $\hat{S} = 11,73$ e o $\hat{\omega} = 30,10$. Sendo que os valores observados nos dados, também para o mês de julho, foram: $S = 12$ e $\omega = 30,77$. Esses resultados nos indicam o quão bem ajustado está o modelo.

Tabela 10 – Estimativas para os parâmetros μ , σ e ξ do modelo SZIGEV, nos dados de Teresina, com seus respectivos intervalos de confiança

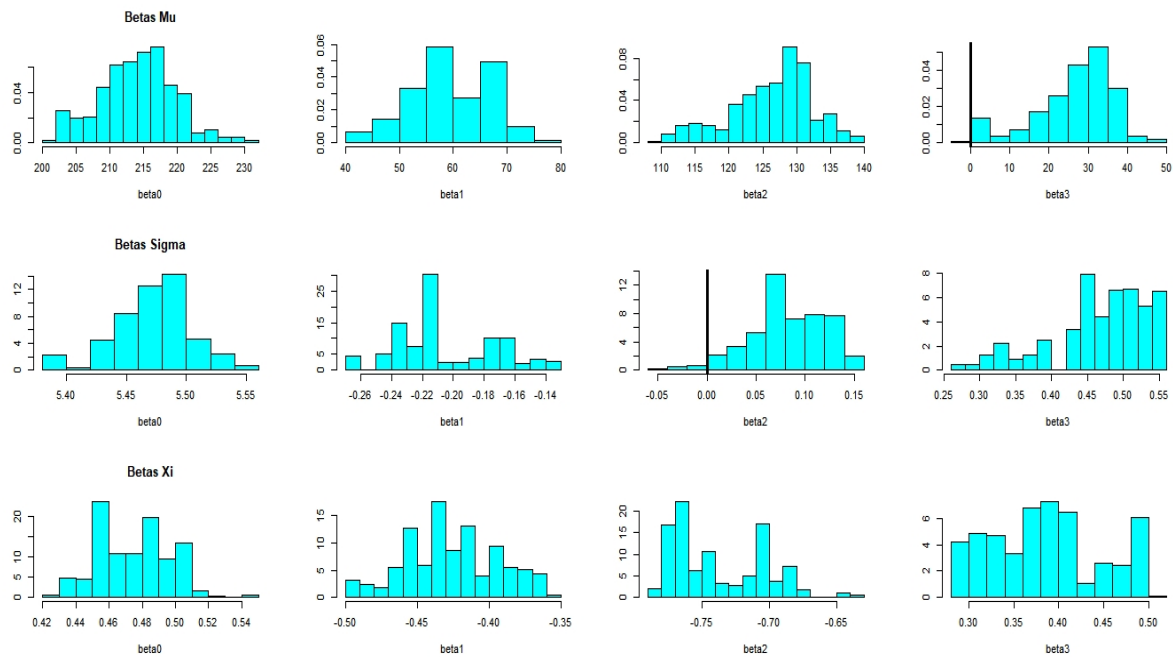
| Estimadores | | | |
|----------------------------|--------------------------|----------------------------|--------------------------|
| $\hat{\beta}_{0,\mu}$ | $\hat{\beta}_{1,\mu}$ | $\hat{\beta}_{2,\mu}$ | $\hat{\beta}_{3,\mu}$ |
| 214,06 (203,51; 225,64) | 59,62 (44,63; 72,98) | 126,36 (112,81; 136,40) | 26,43 (1,41; 42,23) |
| $\hat{\beta}_{0,\sigma}$ | $\hat{\beta}_{1,\sigma}$ | $\hat{\beta}_{2,\sigma}$ | $\hat{\beta}_{3,\sigma}$ |
| 5,47 (5,40; 5,53) | -0,20 (-0,26; -0,13) | 0,08 (0,01; 0,14) | 0,46 (0,31; 0,55) |
| $\hat{\beta}_{0,\xi}$ | $\hat{\beta}_{1,\xi}$ | $\hat{\beta}_{2,\xi}$ | $\hat{\beta}_{3,\xi}$ |
| 0,47 (0,43; 0,51) | -0,43 (-0,50; -0,37) | -0,74 (-0,78; -0,67) | 0,38 (0,29; 0,50) |

Esses resultados apresentados na Tabela 10, nos mostram que todos os betas, do β_1 ao β_3 , tem efeitos significativos no modelo. As 3 covariáveis, cosseno do mês, seno do

mês e a interação entre eles, cosseno e seno, estão influenciando em todos os parâmetros relacionados, μ , σ e ξ . Essa confirmação de que o modelo SZIGEV teria uma vantagem a mais, em relação com a ZIGEV, era aguardada desde a ideia da sua construção, também pelo comportamento dos dados ao longo dos meses do ano, como registros de chuvas extremas, intensas e moderadas, e também como um período sem chuvas, para uma mesma cidade, nesse caso Teresina.

Na Figura 10, temos os histogramas dos betas (0, 1, 2 e 3) para os parâmetros μ , σ e ξ com dados de Teresina.

Figura 10 – Gráficos de séries para dos betas, para os parâmetros μ , σ e ξ da SZIGEV, para a cidade de Teresina



Notem na Figura 10, que apesar dos gráficos do β_3 com relação ao parâmetro μ e do β_2 com relação ao parâmetro σ apresentarem as linhas do zero nos seus respectivos histogramas, temos que esses betas também estão produzindo efeitos significativos nos parâmetros μ e σ como no ξ , mostrando que as covariáveis estão influenciando no modelo, também para a cidade de Teresina.

Percebam que esses resultados significativos nos mostram que o modelo SZIGEV é melhor que o modelo ZIGEV, também para os dados de Teresina.

5.4.3 Cidade de Petrolina

Aplicamos nos dados de precipitação de chuva de Petrolina, a SZIGEV, para analisar o comportamento do modelo em dados inflados com mais de 16% de zeros máximos, junto com a sua provável sazonalidade.

Tabela 11 – Estimativas para os parâmetros ω e S do modelo SZIGEV respectivo a cada mês do ano nos dados de Petrolina

| Mês | Estimadores | |
|-----------|--------------------------|----------------------------------|
| | Quantidade (\hat{S}) | Proporção ($\hat{\omega}$) (%) |
| Janeiro | 0,33 | 0,70 |
| Fevereiro | 0,01 | 0 |
| Março | 0,01 | 0 |
| Abril | 0,63 | 1,40 |
| Mai | 8,50 | 18,70 |
| Junho | 6,45 | 15,60 |
| Julho | 5,77 | 11,80 |
| Agosto | 15,77 | 32,00 |
| Setembro | 19,00 | 40,40 |
| Outubro | 17,18 | 39,00 |
| Novembro | 4,64 | 9,70 |
| Dezembro | 0,01 | 0 |

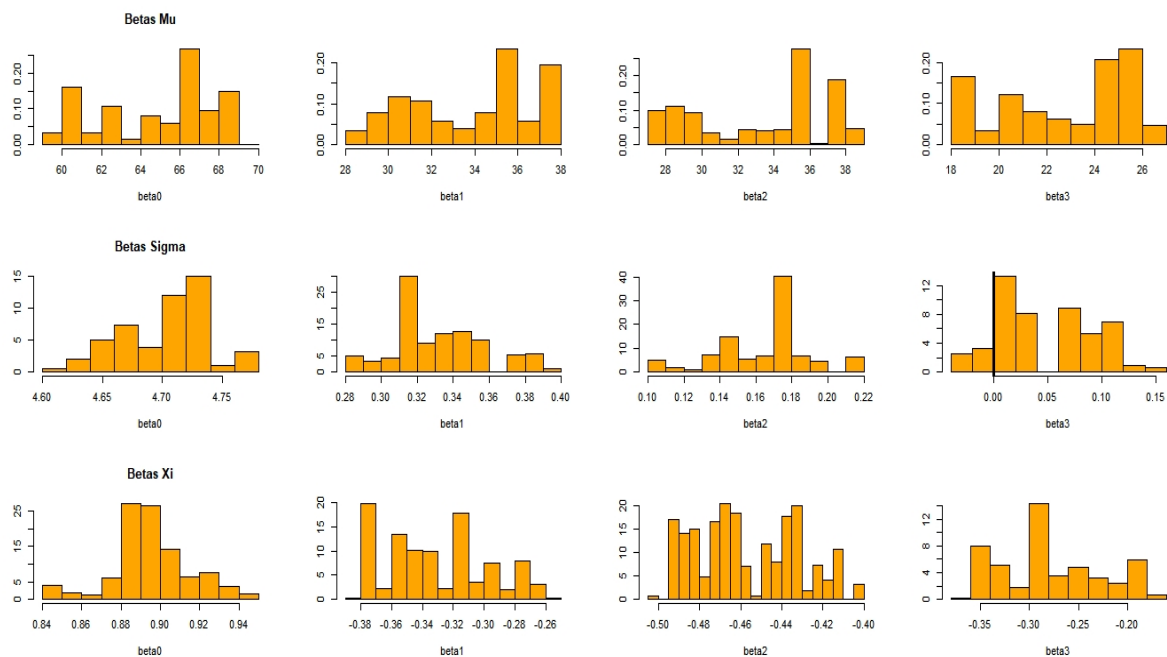
Identificamos na Tabela 11, que os valores obtidos pelos estimadores \hat{S} e $\hat{\omega}$ da SZIGEV foram muito precisos, para cada mês do ano, quando estes comparados com os valores que realmente ocorreram nos dados, apresentados anteriormente. Com base nesses resultados, temos indícios positivos sobre a qualidade do ajuste deste modelo.

Tabela 12 – Estimativas para os parâmetros μ , σ e ξ do modelo SZIGEV, nos dados de Petrolina, com seus respectivos intervalos de confiança

| Estimadores | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|
| $\hat{\beta}_{0,\mu}$ | $\hat{\beta}_{1,\mu}$ | $\hat{\beta}_{2,\mu}$ | $\hat{\beta}_{3,\mu}$ |
| 64,76 (59,75; 68,66) | 33,89 (28,62; 37,71) | 33,32 (27,27; 38,20) | 22,72 (18,25; 26,25) |
| $\hat{\beta}_{0,\sigma}$ | $\hat{\beta}_{1,\sigma}$ | $\hat{\beta}_{2,\sigma}$ | $\hat{\beta}_{3,\sigma}$ |
| 4,71 (4,60; 4,77) | 0,33 (0,28; 0,40) | 0,16 (0,10; 0,22) | 0,06 (-0,04; 0,16) |
| $\hat{\beta}_{0,\xi}$ | $\hat{\beta}_{1,\xi}$ | $\hat{\beta}_{2,\xi}$ | $\hat{\beta}_{3,\xi}$ |
| 0,89 (0,84; 0,95) | -0,33 (-0,39; -0,27) | -0,45 (-0,51; -0,40) | -0,27 (-0,38; -0,16) |

Temos que, dentre todos os betas apresentados na Tabela 12, somente o β_3 com relação ao parâmetro σ não foi significativo. Isso implica que a covariável 3, relativa a multiplicação do cosseno com o seno do mês, não influencia neste parâmetro. Em contrapartida todos os outros betas referentes aos parâmetros, tem efeitos significativos no modelo, tem suas covariáveis influenciando em seus respectivos parâmetros, μ , σ e ξ . Também era de se aguardar que, na utilização do modelo SZIGEV, para esses dados da cidade de Petrolina, tivéssemos uma maior vantagem em relação a ZIGEV, também pelo comportamento sazonal dos seus dados ao longo dos meses do ano. Esperamos também que, para os dados da cidade de São João do Piauí, que possui esta mesma característica sazonal, sejam obtidos resultados semelhantemente significativos.

Figura 11 – Gráficos de séries para os betas, para os parâmetros μ , σ e ξ da SZIGEV, para a cidade de Petrolina



Com base na Figura 11, temos que somente o gráfico do β_3 relacionado com o parâmetro σ apresentou as linha do zero, mostrando que a covariável, cosseno vezes o seno do mês, não tem efeito significativo nesse parâmetro, não influenciando no modelo. Já os gráficos dos outros betas, ainda para a cidade de Petrolina, mostram que são significativos aos seus respectivos parâmetros, sendo o modelo muito influenciado pelas 3 covariáveis em estudo.

Como os β_1 , β_2 e β_3 são significativos para os seus respectivos parâmetros μ , σ e ξ podemos afirmar que, para os dados de Petrolina, o modelo SZIGEV foi mais preciso que o modelo ZIGEV.

5.4.4 Cidade de São João do Piauí

Nesta seção, aplicamos a SZIGEV aos dados de São João do Piauí, com o objetivo de analisar o seu comportamento em dados sazonais que possuem mais de 28% de zeros.

Tabela 13 – Estimativas para os parâmetros ω e S do modelo SZIGEV respectivo a cada mês do ano nos dados de São João do Piauí

| Mês | Estimadores | |
|-----------|--------------------------|----------------------------------|
| | Quantidade (\hat{S}) | Proporção ($\hat{\omega}$) (%) |
| Janeiro | 0 | 0 |
| Fevereiro | 0,08 | 0,29 |
| Março | 0,06 | 0,20 |
| Abril | 0 | 0 |
| Mai | 7,92 | 25,40 |
| Junho | 17,90 | 64,00 |
| Julho | 22,89 | 78,70 |
| Agosto | 22,90 | 81,70 |
| Setembro | 17,88 | 63,80 |
| Outubro | 7,91 | 27,20 |
| Novembro | 0,03 | 0,10 |
| Dezembro | 0 | 0 |

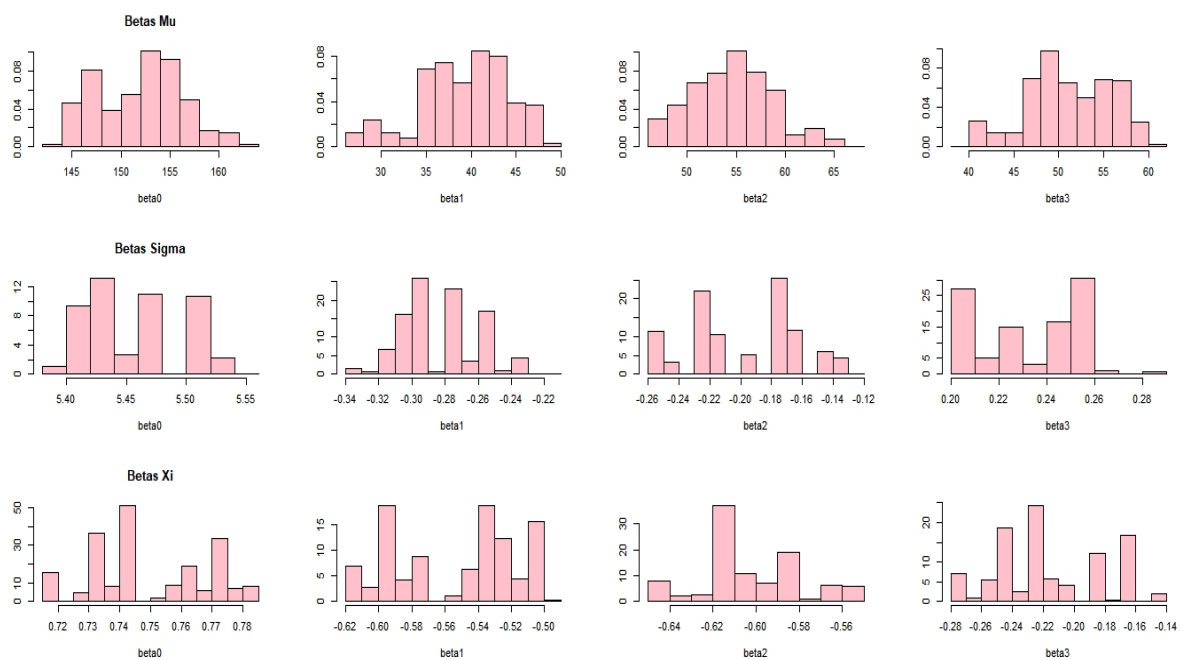
Mais uma vez percebemos, agora através dos resultados apresentados na Tabela 13, o quão aparentemente ajustado está o modelo SZIGEV, tendo em vista a precisão de suas estimativas, para cada mês do ano, feitas pelos seus estimadores \hat{S} e $\hat{\omega}$, em comparação com os reais valores encontrados nos dados da cidade de São João do Piauí.

Tabela 14 – Estimativas para os parâmetros μ , σ e ξ do modelo SZIGEV, nos dados de São João do Piauí, com seus respectivos intervalos de confiança

| Estimadores | | | |
|----------------------------|--------------------------|--------------------------|--------------------------|
| $\hat{\beta}_{0,\mu}$ | $\hat{\beta}_{1,\mu}$ | $\hat{\beta}_{2,\mu}$ | $\hat{\beta}_{3,\mu}$ |
| 151,98 (145,22; 160,70) | 39,11 (28,06; 46,98) | 54,72 (47,51; 63,74) | 51,06 (41,51; 58,77) |
| $\hat{\beta}_{0,\sigma}$ | $\hat{\beta}_{1,\sigma}$ | $\hat{\beta}_{2,\sigma}$ | $\hat{\beta}_{3,\sigma}$ |
| 5,46 (5,41; 5,53) | -0,28 (-0,31; -0,24) | -0,19 (-0,25; -0,13) | 0,23 (0,20; 0,26) |
| $\hat{\beta}_{0,\xi}$ | $\hat{\beta}_{1,\xi}$ | $\hat{\beta}_{2,\xi}$ | $\hat{\beta}_{3,\xi}$ |
| 0,75 (0,72; 0,78) | -0,55 (-0,61; -0,50) | -0,60 (-0,65; -0,56) | -0,22 (-0,28; -0,16) |

De acordo com os resultados obtidos, nenhum dos intervalos de confiança dos betas tiveram valores zeros entres eles, logo, todos os betas, do β_1 ao β_3 , tem efeitos significativos nos parâmetros relacionados na Tabela 14 (μ , σ e ξ), ou seja, as 3 covariáveis, influenciam muito no modelo. Percebam que o modelo sofre influência de praticamente todas as estações do ano, e que os parâmetros μ , σ e ξ sofrem muita influência das covariáveis, tendo um comportamento diferente para os meses.

Figura 12 – Histogramas para os betas, para os parâmetros μ , σ e ξ da SZIGEV, para a cidade de São João do Piauí



Notem na Figura 12, que em nenhum dos gráficos dos betas, para os dados de São João do Piauí, tem a linha do zero. Mostrando que todos eles são significativos para seus respectivos parâmetros e que as covariáveis selecionadas muito influenciam no modelo proposto. Assim, com os resultados significativos obtidos pelos betas, para a cidade de São João do Piauí, podemos afirmar que existe um forte indicativo de que a distribuição SZIGEV seja mais precisa que a distribuição ZIGEV.

5.5 Medidas de ajustes

Spiegelhalter et al. (2002) afirma que quanto menor for o valor das medidas AIC, BIC e DIC, melhor será o ajuste do modelo. Tendo em vista a busca pelo melhor ajuste, apresentamos na Tabela 15, com relação aos modelos ZIGEV e SZIGEV, a comparação entre as medidas de qualidade de ajuste AIC (Akaike's Information Criterion), BIC

(Bayesian Information Criterion, também conhecido como Schwarz Criterion) e DIC (Deviance Information Criterion), para as cidades de Recife, Teresina, Petrolina e São João do Piauí nos estados de Pernambuco e Piauí.

Tabela 15 – Medidas de qualidade de ajustes para Recife, Teresina, Petrolina e São João do Piauí, para comparação entre os modelos ZIGEV e SZIGEV

| Modelos | Medidas de Ajuste | | | | | |
|---------|-------------------|----------|----------|-------------------|---------|---------|
| | Recife | | | Teresina | | |
| | AIC | BIC | DIC | AIC | BIC | DIC |
| ZIGEV | 10032,30 | 10054,50 | 10031,13 | 6174,14 | 6194,81 | 6167,99 |
| SZIGEV | 9669,02 | 9832,85 | 9607,43 | 5834,35 | 5963,23 | 5762,62 |
| Modelos | Petrolina | | | São João do Piauí | | |
| | AIC | BIC | DIC | AIC | BIC | DIC |
| | ZIGEV | 6349,50 | 6370,97 | 6343,37 | 3941,66 | 3961,08 |
| SZIGEV | 5979,29 | 6113,92 | 5887,54 | 3222,58 | 3362,38 | 3161,79 |

As medidas de qualidade de ajuste constituem uma outra maneira de comparar qual dos modelos está melhor ajustado. Podemos ver que, para as cidades, a utilização do modelo SZIGEV proporciona um ganho real e substancial, obtendo uma melhor qualidade de ajuste e maior precisão em suas estimações. Pois, para as quatro cidades em estudo, todas as medidas de ajuste foram melhores (menores) com a distribuição SZIGEV. Apesar dos dados de Recife terem apenas 2 (dois) zeros máximos, correspondente a uma proporção de 0,28% de zeros, a sazonalidade presente foi determinante para que a SZIGEV obtivesse um melhor ajuste em relação a ZIGEV. Esses expressivos resultados nos mostram quão ajustado está o modelo para modelar dados de máximos inflados de zeros e sazonais e obter resultados mais confiáveis e fidedignos.

Quanto maior for a quantidade de zeros e a sazonalidade, no banco de dados, maior a diferença entre os modelos. As maiores diferenças entre a ZIGEV e SZIGEV são com relação a Petrolina e São João do Piauí, pois são essas as cidades que tem mais sazonalidades, tem as maiores proporções de zeros, sendo que elas variam muito ao longo dos meses do ano.

Vimos que tem mês em que a proporção de zeros, de máximos mensais, é de 0%, enquanto há mês que a proporção de zeros é de 70%. Essa discrepância faz com que se justifique ainda mais o modelo sazonal.

5.6 Níveis de retornos

O p_i^* -ésimo quantil da distribuição SZIGEV tem seguinte a expressão

$$z_{p_i^*} = \begin{cases} \mu_i + \frac{\sigma_i}{\xi_i} \left\{ [-\log(1 - p_i^*)]^{-\xi_i} - 1 \right\}, & \text{se } \xi_i \neq 0; \\ \mu_i - \sigma_i \log [-\log(1 - p_i^*)], & \text{se } \xi_i = 0, \end{cases} \quad (5.7)$$

em que os parâmetros tem as seguintes funções de ligação: $\mu_i = \beta'_\mu z_k$, $\sigma_i = \exp(\beta'_\sigma z_k)$, $\xi_i = \exp(\beta'_\xi z_k) - 1$, como apresenta Nascimento (2011).

Substituindo os parâmetros por suas respectivas funções de ligação, encontramos a expressão do p_i^* -ésimo quantil proposto da SZIGEV

$$z_{p_i^*} = \begin{cases} \beta'_\mu z_k + \frac{\exp(\beta'_\sigma z_k)}{\exp(\beta'_\xi z_k) - 1} \left\{ [-\log(1 - p_i^*)]^{-[\exp(\beta'_\xi z_k) - 1]} - 1 \right\}, & \text{se } \exp(\beta'_\xi z_k) \neq 1; \\ \beta'_\mu z_k - \exp(\beta'_\sigma z_k) \log [-\log(1 - p_i^*)], & \text{se } \exp(\beta'_\xi z_k) = 1, \end{cases}$$

em que $p_i^* = \frac{p - \omega_i}{1 - \omega_i}$, $i = 1, \dots, 12$ e $p_i^* \in (0, 1)$, sendo ω_i a proporção de zeros a cada mês do ano, e os z_k são as covariáveis dos meses do ano.

Na TVE, o p -ésimo quantil é usado para calcular o nível de retorno esperado em que ocorra uma vez a cada t períodos de tempo, sendo representado pelo quantil $p = 1 - 1/t$.

Assim, para obter o valor da mediana da distribuição SZIGEV, é preciso calcular o quantil $p = 0,5$ que é expresso por

$$Md(Z) = \begin{cases} \beta'_\mu z_k + \frac{\exp(\beta'_\sigma z_k)}{\exp(\beta'_\xi z_k) - 1} \left\{ [\log(2) + \log(1 - \omega_i)]^{-[\exp(\beta'_\xi z_k) - 1]} - 1 \right\}, & \exp(\beta'_\xi z_k) \neq 1; \\ \beta'_\mu z_k - \exp(\beta'_\sigma z_k) \log [-\log(2) + \log(1 - \omega_i)], & \exp(\beta'_\xi z_k) = 1. \end{cases}$$

É importante salientarmos que os quantis da SZIGEV (p_i^*), da ZIGEV (p_i^*) e GEV (p_i) serão idênticos, quando a proporção de zeros (ω_i) nos dados forem iguais a zeros, ou seja, quando $\omega_i = 0 \implies p_i^* = p_i^* = p_i$.

5.6.1 Estimadores $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\xi}$

Vimos na seção anterior que os parâmetros μ , σ e ξ estão ligados diretamente a construção dos quantis extremos. Portanto, ao sabermos seus valores e comportamento,

podemos ter um indicativo de como os níveis de retorno estarão desempenhando-se.

Na Tabela 16 temos os estimadores, $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\xi}$, com suas respectivas estimativas encontradas em cada mês do ano para as quatro cidades em estudo.

Tabela 16 – Estimativas para os estimadores $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\xi}$ em cada mês do ano, para as cidades de Recife, Teresina, Petrolina e São João do Piauí

| Cidade | Estim | Meses | | | | | | | | | | | |
|-----------|----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | | Jan | Fev | Mar | Abr | Mai | Jun | Jul | Ago | Set | Out | Nov | Dez |
| Recife | $\hat{\mu}$ | 221,6 | 329,2 | 425,1 | 491,0 | 517,8 | 497,9 | 428,0 | 320,7 | 203,9 | 116,8 | 90,3 | 131,3 |
| | $\hat{\sigma}$ | 165,7 | 224,7 | 252,7 | 256,9 | 263,4 | 270,7 | 246,9 | 182,4 | 118,0 | 84,6 | 82,5 | 110,1 |
| | $\hat{\xi}$ | 0,24 | 0,17 | 0,09 | 0,05 | 0,06 | 0,12 | 0,21 | 0,28 | 0,31 | 0,31 | 0,30 | 0,28 |
| Teresina | $\hat{\mu}$ | 220,3 | 232,8 | 224,2 | 192,5 | 142,4 | 87,3 | 46,0 | 33,2 | 52,4 | 94,8 | 144,9 | 189,4 |
| | $\hat{\sigma}$ | 213,4 | 244,1 | 267,4 | 242,5 | 167,4 | 97,7 | 61,8 | 53,9 | 67,5 | 101,4 | 146,7 | 184,8 |
| | $\hat{\xi}$ | 0,19 | 0,16 | 0,16 | 0,18 | 0,23 | 0,29 | 0,34 | 0,37 | 0,36 | 0,33 | 0,28 | 0,23 |
| Petrolina | $\hat{\mu}$ | 122,6 | 124,7 | 107,5 | 78,8 | 50,3 | 29,4 | 18,3 | 16,1 | 23,0 | 41,4 | 70,2 | 101,4 |
| | $\hat{\sigma}$ | 150,8 | 150,5 | 135,6 | 105,4 | 70,5 | 44,9 | 33,1 | 33,3 | 45,4 | 71,3 | 106,7 | 137,2 |
| | $\hat{\xi}$ | 0,22 | 0,22 | 0,22 | 0,22 | 0,23 | 0,23 | 0,23 | 0,23 | 0,23 | 0,22 | 0,22 | 0,22 |
| São João | $\hat{\mu}$ | 196,7 | 197,6 | 184,6 | 161,1 | 132,2 | 107,1 | 92,7 | 91,2 | 103,7 | 127,2 | 156,0 | 181,5 |
| | $\hat{\sigma}$ | 183,2 | 189,1 | 184,5 | 180,4 | 180,2 | 185,3 | 189,7 | 184,2 | 172,5 | 161,9 | 161,5 | 171,5 |
| | $\hat{\xi}$ | 0,05 | 0,04 | 0,05 | 0,05 | 0,04 | 0,05 | 0,04 | 0,05 | 0,05 | 0,04 | 0,05 | 0,05 |

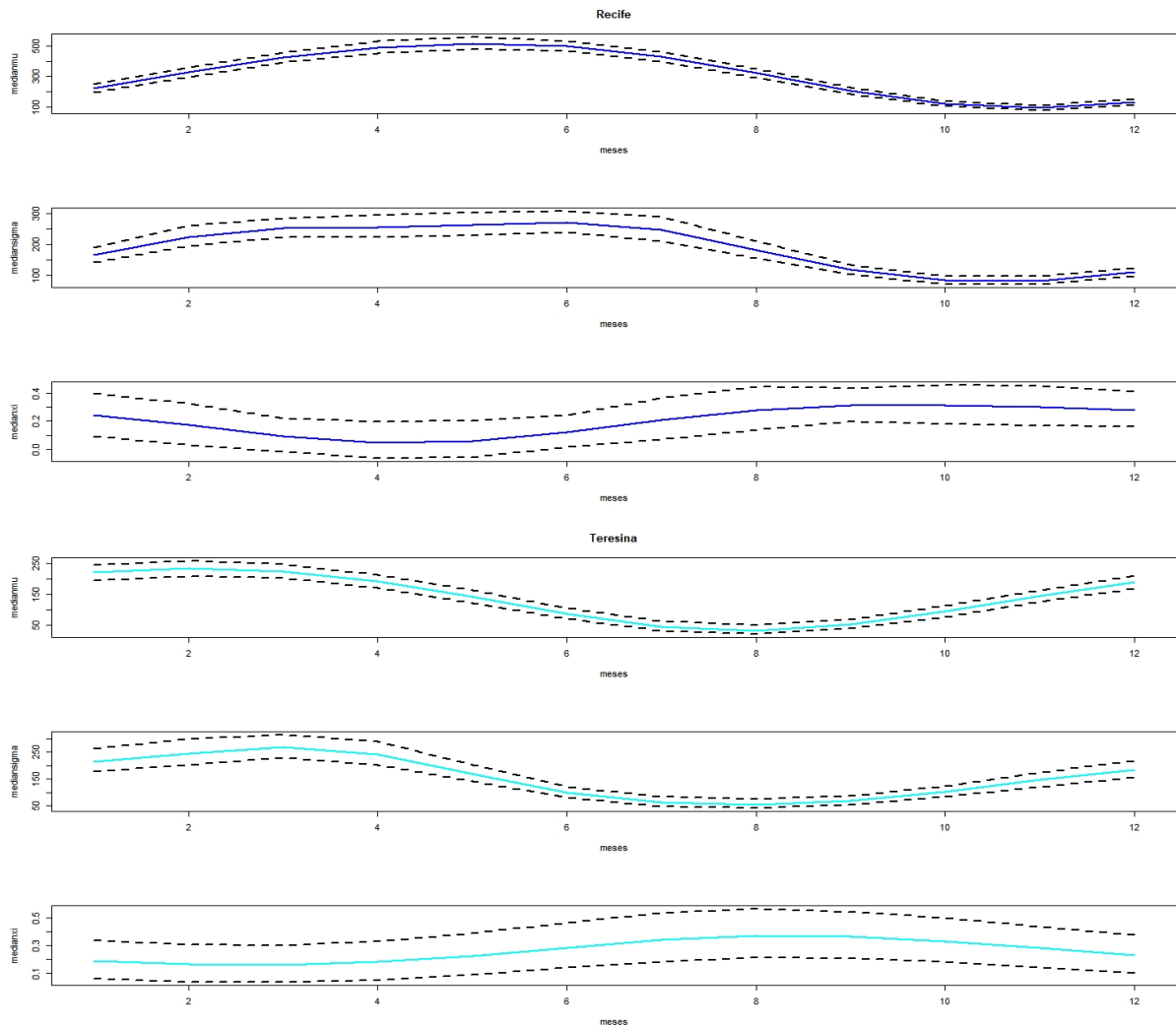
Analisando os dados das quatro cidades na Tabela 16, percebemos que os valores dos parâmetros aumentam até um determinado mês, depois começam a diminuir até voltarem a subir novamente ao longo dos meses do ano, apresentando forte indicação de características sazonais.

Assim, pelo conhecimento da ligação existente entre os parâmetros μ , σ e ξ e os níveis de retornos, podemos sugerir que os resultados encontrados pelos quantis extremos também terão desempenhos sazonais.

Diferente dos demais estimadores, as estimativas obtidas pelos $\hat{\xi}_i$ para as cidades de Petrolina e São João do Piauí foram muito semelhantes mês a mês dentro de cada uma dessas cidades, mostrando que não sofreram variação ao longo do ano.

Observamos na Figura 13 os gráficos das estimativas para os parâmetros μ , σ e ξ dos dados de Recife e Teresina.

Figura 13 – gráficos das estimativas de μ , σ e ξ para as cidade de Recife e Teresina



Observando os gráficos dos estimadores $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\xi}$, para a cidade de Recife ao longo dos meses, na Figura 13, notamos mais facilmente a presença da sazonalidade, mostrando seus respectivos comportamentos e variações a cada mês do ano. Ao observarmos os gráficos de Teresina vemos que possui comportamento semelhante aos estimadores de Recife.

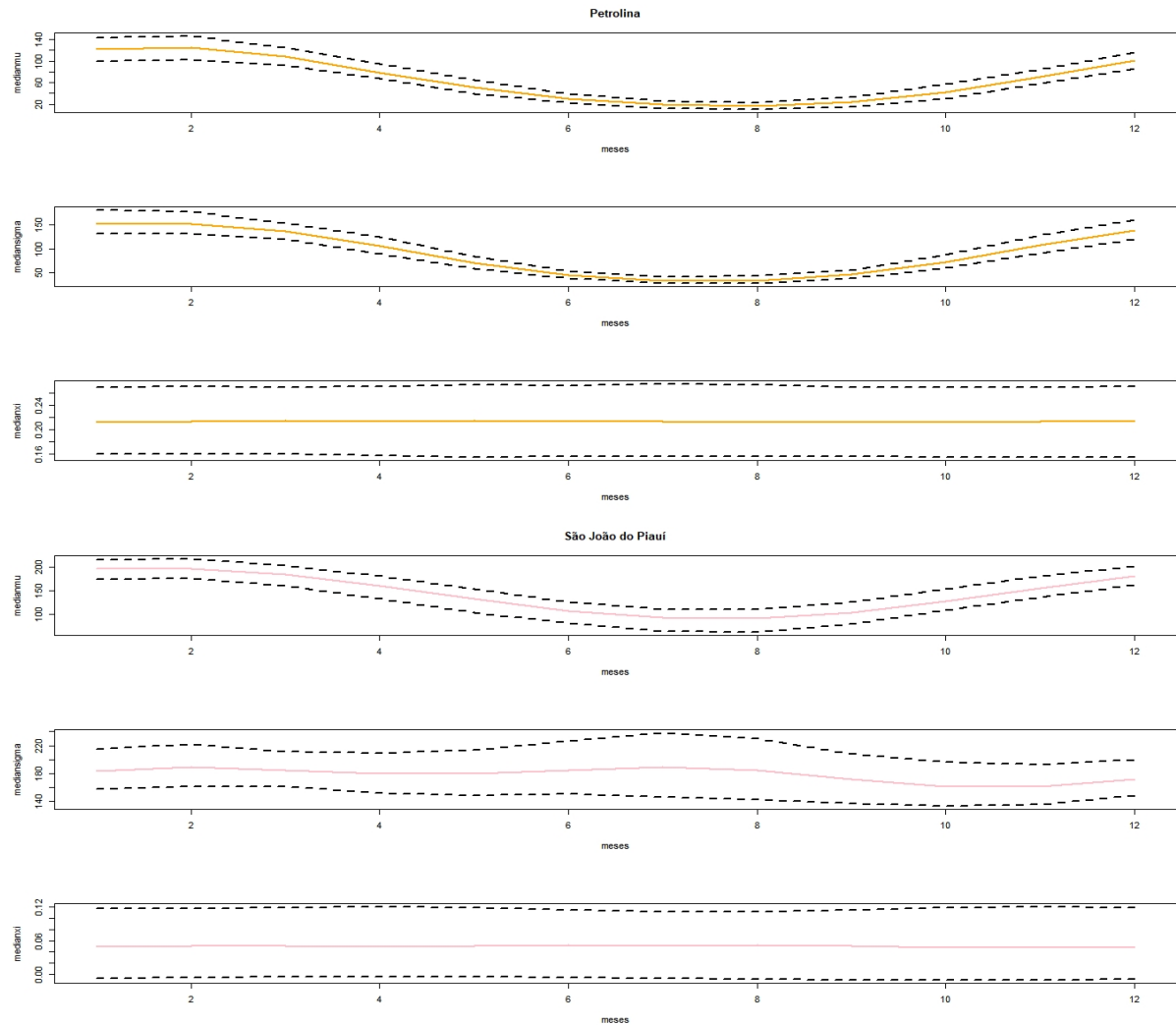
É interessante analisarmos que praticamente todos os meses chove em Recife, que tem a menor proporção de zeros (ω) das cidades, igual a 0,28%, mas mesmo assim, percebemos um forte índice de sazonalidade, Fica mais claro para o entendimento quando olhamos para o comportamento de seus dados de máximos mensais, em que existem meses de grande precipitação (mm) de chuvas como também há meses de baixa precipitação (mm). Essas oscilações nas intensidades das chuvas já indicam uma sazonalidade nos dados.

Os resultados observados nesse trabalho nos mostram a importância da construção

de uma distribuição que modele também a parte sazonal dos dados, buscando alcançar resultados mais precisos e significativos.

Na Figura 14, temos os gráficos das estimativas para os parâmetros μ , σ e ξ dos dados de Petrolina e São João do Piauí.

Figura 14 – gráficos das estimativas de μ , σ e ξ das cidades em estudo



Analisando a Figura 14, encontramos as mesmas características sazonais nas estimativas de $\hat{\mu}_i$ e $\hat{\sigma}_i$ ao longo dos meses do ano, como as que foram ressaltadas nos gráficos anteriores. Com exceção do estimador $\hat{\xi}_i$ para ambas cidades. Pois notamos que no decorrer do ano que ele tem pouca ou nenhuma variação, com resultados muito semelhantes dentro de cada uma das cidades.

As Figuras 13 e 14, indicam a existência de períodos cíclicos anuais, que começam e terminam em períodos mensais muito semelhantes. Esses indícios são confirmados pelo

bom desempenho nas medidas de qualidade de ajuste, que mostram a necessidade de trabalhar-se com um modelo sazonal.

5.6.2 Quantis extremos (90, 95 e 99)

Nesta seção, trabalharemos com os quantis altos de 90%, 95% e 99%, das distribuições GEV, ZIGEV e SZIGEV, respectivos aos níveis de retornos dos períodos 5, 10, 20 e 100. Por exemplo, o quantil 90% separa categoricamente os 90% de valores menores dos 10% de valores maiores dos dados em ordem. Essa parte final da cauda são referente aos máximos. Quando encontramos um valor estimado de precipitação de chuva para o quantil 90%, é esperado que esse valor ocorra 10% das vezes, ou seja, espera-se que aconteça pelo menos uma vez em cada 10 períodos (mês).

Na Tabela 17, encontramos os quantis altos, 90%, 95% e 99% para as cidades de Recife, Teresina, Petrolina e São João do Piauí dos estados de Pernambuco e Piauí. Esses resultados foram obtidos para cada mês do ano, baseados nas séries históricas dos respectivos dados.

Tabela 17 – Níveis de retornos correspondentes aos quantis extremos da SZIGEV para as cidades em estudo

| Quantis | Meses | | | | | | | | | | | |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | Jan | Fev | Mar | Abr | Mai | Jun | Jul | Ago | Set | Out | Nov | Dez |
| Recife | | | | | | | | | | | | |
| 90% | 71,3 | 94,3 | 105,9 | 110,3 | 115,2 | 119,8 | 113,5 | 89,1 | 58,8 | 39,1 | 35,5 | 47,6 |
| 95% | 93,2 | 119,3 | 129,1 | 131,1 | 137,2 | 146,3 | 143,9 | 116,0 | 77,9 | 52,8 | 48,4 | 63,9 |
| 99% | 160,0 | 188,7 | 187,7 | 181,2 | 190,1 | 215,3 | 231,6 | 201,3 | 140,5 | 97,6 | 90,3 | 115,6 |
| Teresina | | | | | | | | | | | | |
| 90% | 82,0 | 89,8 | 94,8 | 86,6 | 63,5 | 38,4 | 20,6 | 18,0 | 24,8 | 43,3 | 60,4 | 73,5 |
| 95% | 107,4 | 116,8 | 124,1 | 114,7 | 85,4 | 53,0 | 30,4 | 27,2 | 36,4 | 60,6 | 82,2 | 97,7 |
| 99% | 179,1 | 191,5 | 204,1 | 193,5 | 150,2 | 99,6 | 63,0 | 59,1 | 77,0 | 118,8 | 151,5 | 170,1 |
| Petrolina | | | | | | | | | | | | |
| 90% | 56,0 | 56,3 | 50,4 | 37,9 | 22,9 | 14,8 | 10,8 | 8,6 | 11,0 | 18,8 | 36,1 | 50,0 |
| 95% | 75,2 | 75,7 | 67,9 | 51,4 | 31,6 | 20,4 | 15,0 | 12,6 | 16,2 | 27,1 | 49,5 | 67,5 |
| 99% | 132,3 | 132,5 | 119,6 | 91,3 | 57,2 | 37,1 | 27,4 | 24,2 | 31,4 | 51,3 | 88,8 | 119,0 |
| São João | | | | | | | | | | | | |
| 90% | 63,5 | 64,9 | 62,7 | 59,0 | 49,6 | 31,5 | 14,9 | 11,2 | 29,4 | 45,1 | 54,1 | 59,1 |
| 95% | 78,7 | 80,6 | 78,0 | 74,0 | 64,5 | 46,9 | 32,2 | 28,5 | 43,9 | 58,5 | 67,5 | 73,2 |
| 99% | 115,4 | 118,3 | 115,1 | 109,4 | 100,0 | 83,1 | 68,8 | 64,3 | 77,0 | 90,7 | 99,9 | 107,2 |

Analisando os resultados dos níveis de retornos para a cidade de Recife, identificamos que todas as previsões até o mês de agosto são acima de 60 mm, sendo a maioria acima de 100 mm, anunciando muitas chuvas intensas e mais ainda de chuvas extremas. Por exemplo, olhando para o quantil 90%, temos uma previsão para o mês de janeiro em Recife de chuvas intensas de 71,3 mm, ou seja, pelo menos uma vez em 10 períodos (mês) acontecerá chuvas intensas maiores ou iguais a 71,3 mm. Para fevereiro, olhando para o quantil 95%, temos a previsão de que haja chuvas extremas de 129,1 mm, ou seja, de que pelo menos uma vez em 20 meses chegará chuvas extremas maiores ou iguais ao valor predito de 129,1 mm.

Analisando esses mesmos meses, de janeiro e fevereiro, e os quantis de 90% e 95% na cidade de Teresina, notamos que o comportamento com relação a previsão de chuvas intensas e extremas são também bem semelhantes nesse período. Mas, quando avançamos nos meses ao longo do ano percebemos que em Teresina, a partir do mês de maio, a quantidade das chuvas começam a diminuir, de acordo com as previsões, enquanto que em Recife as chuvas continuam intensas e extremas até os meses de julho, agosto e setembro.

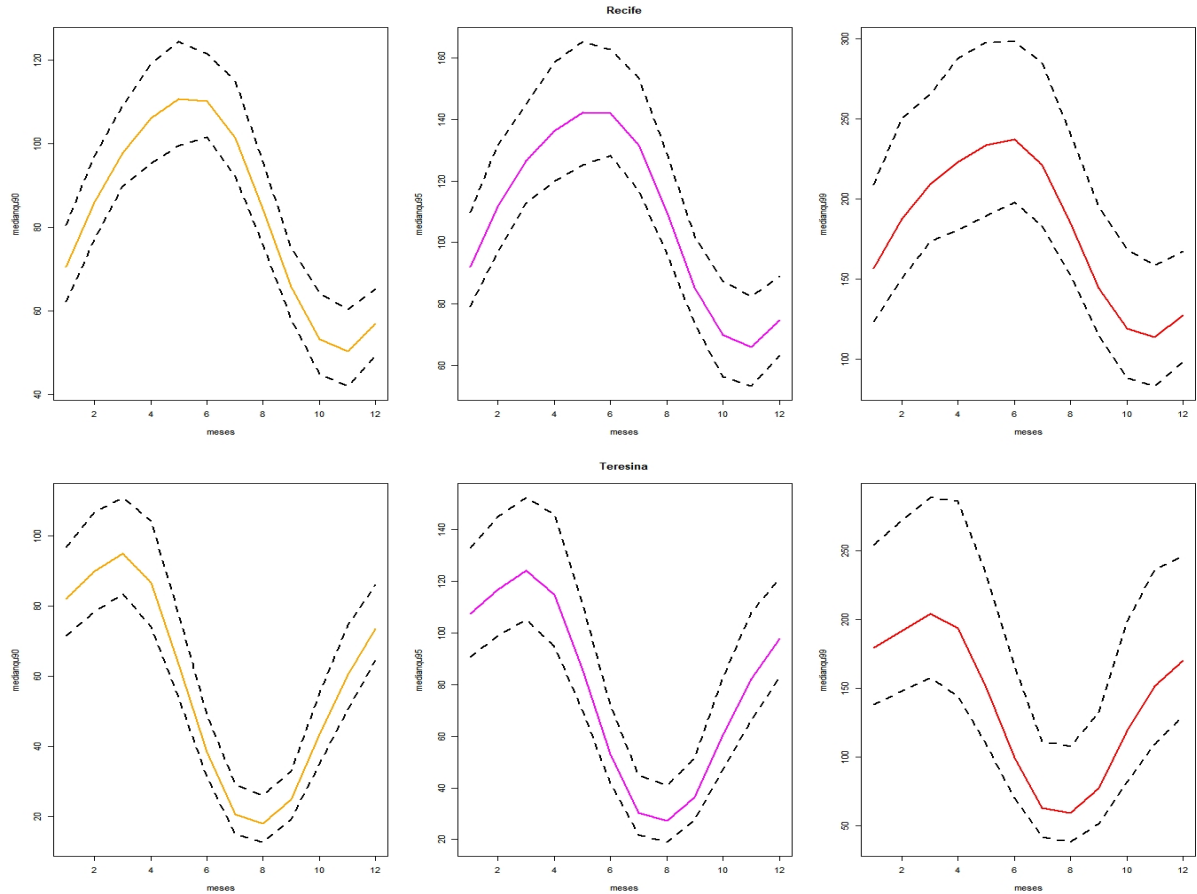
Essas previsões baseadas nas séries históricas dos dados nos traz um alerta, pois essas chuvas intensas e extremas são muito fortes e impetuosas que provocam estragos por onde passam, sejam danos materiais ou até mesmo com vidas. Daí a importância de tanto as autoridades competentes como a sociedade civil, terem em mãos essas informações com o máximo de precisão possível, para que sejam tomadas as devidas providências.

Na Figura 15, visualizamos as previsões de chuvas em cada mês do ano para as cidades de Recife e Teresina. Assim, conseguimos analisar o comportamento das previsões sobre as precipitações de chuvas para essas cidade ao longo dos meses do ano.

Ao observarmos os gráficos de retorno (Figura 15), logo encontramos a existência de um forte ciclo sazonal que acontece ao longo do ano e que vai se repetindo ao passar dos anos. Quando modelamos dados de máximos inflados de zeros sazonais, entendemos que os resultados obtidos pelos estimadores terão as mesmas características sazonais apresentados nesses dados. Percebam que os resultados das estimações, que são as estimativas, seguem a tendência dos máximos e da sazonalidade dos dados tanto em Recife como também em Teresina, adotando um comportamento natural.

Comparando os gráficos dos quantis 90%, 95% e 99% referente aos dados de uma mesma cidade, notamos que suas respectivas linhas contínuas ou tracejadas aparentemente perfazem o mesmo caminho, mas com valores diferentes para cada quantil. Analisando o gráfico do quantil 99% para a cidade de Recife, verificamos que a curva ascendente que começa em Janeiro já inicia com uma previsão de que aconteça pelo menos uma vez a cada 100 meses, chuvas extremas maiores ou iguais a aproximadamente 150 mm

Figura 15 – Quantis extremos de 90%, 95% e 99% da SZIGEV, para as cidades de Recife e Teresina

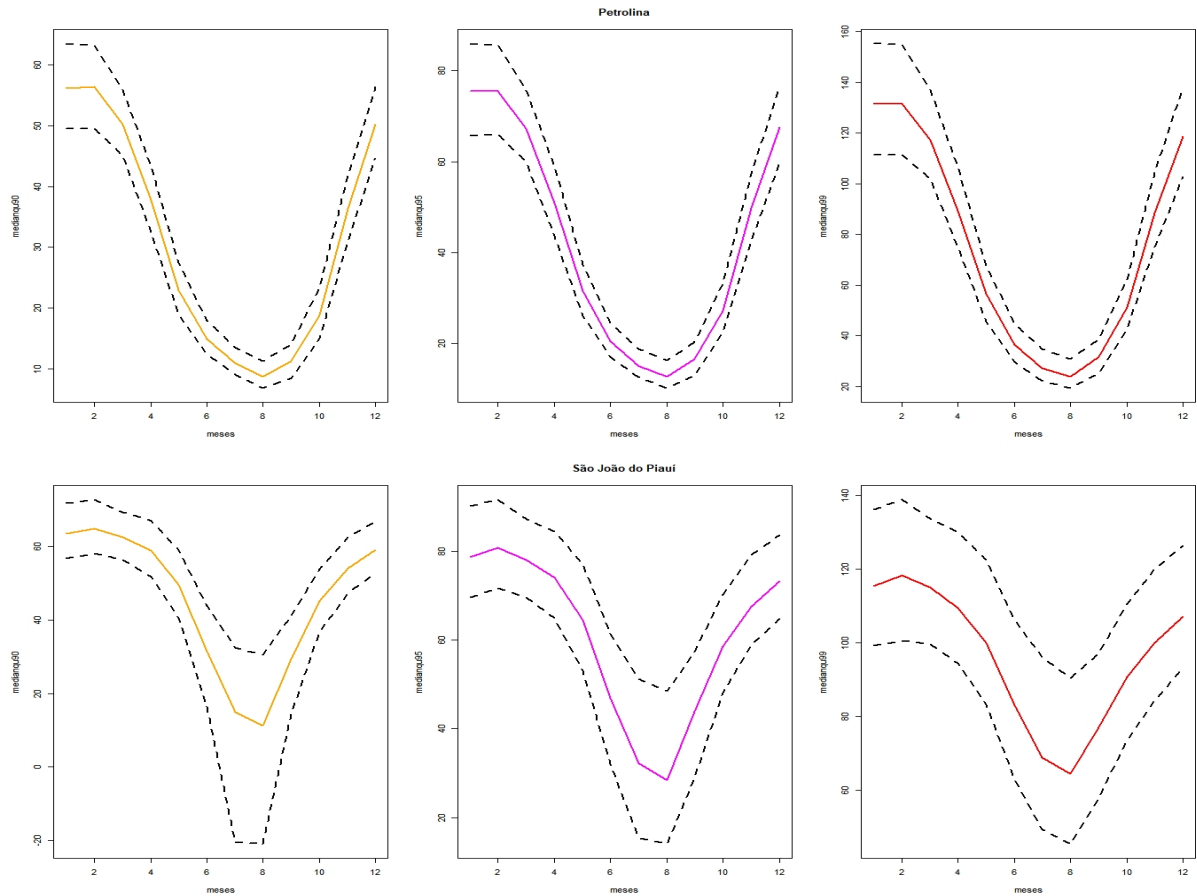


(valor na Tabela 17 é de 160 mm). Essa curva chega ao pico no final do mês de junho e início de julho, em que são esperados, que ocorram pelo menos uma vez a cada 100 meses, chuvas extremas de aproximadamente 250 mm (231,6 mm). A partir desse pico as chuvas começam a diminuir. Ainda observando o quantil 99%, vemos que no mês de novembro estão previstas a ocorrência das menores chuvas do ano. Nesse mês é esperado que haja, pelo menos uma vez a cada 100 meses, chuvas intensas maiores ou iguais a aproximadamente 100 mm (90,3 mm). Já a partir desse marco, percebemos o aumento novamente das chuvas. E assim, é o comportamento das séries analisadas.

Identificamos pela Figura 16, que os resultados alcançados com os dados da cidade de Teresina foram semelhantes ao comportamento dos resultados obtidos para a cidade de Recife.

A Figura 16 trata dos gráficos de retornos 10, 20 e 100, correspondentes aos quantis altos 90%, 95% e 99% para as cidades de Petrolina e São João do Piauí.

Figura 16 – Quantis extremos 90%, 95% e 99% da SZIGEV, para as cidade de Petrolina e São João do Piauí



Em harmonia com os gráficos anteriores, verificamos na Figura 16 a existência de características sazonais, de um ciclo anual que se repete ao longo do tempo. Na cidade de Petrolina, o ano já começa com a curva dos quantis praticamente no seu ápice. Analisando o quantil 95%, vemos que é esperado nesses meses de janeiro e fevereiro que aconteçam, pelo menos uma vez a cada 20 meses, chuvas intensas iguais a 75,7 mm, aproximadamente 80 mm. A partir desse pico as chuvas começam a diminuir, chegando a previsão de que ocorram no mês de agosto, valores iguais a 12,6 mm, aproximadamente 10 mm, pelo menos uma vez a cada 20 meses. Logo após, percebemos o aumento gradual das chuvas, culminando em dezembro com previsões de chuvas de até 67,5 mm, aproximadamente 70 mm.

Comparando as Figuras 15 e 16 com a Figura 7, percebemos o quão alinhadas estão as estimativas obtidas pelo modelo, em relação ao comportamento dos dados, em todas as cidades em estudo. Cada vez torna-se mais evidente a qualidade de ajuste e a precisão das estimações da distribuição SZIGEV, para a modelagem de dados de máximos

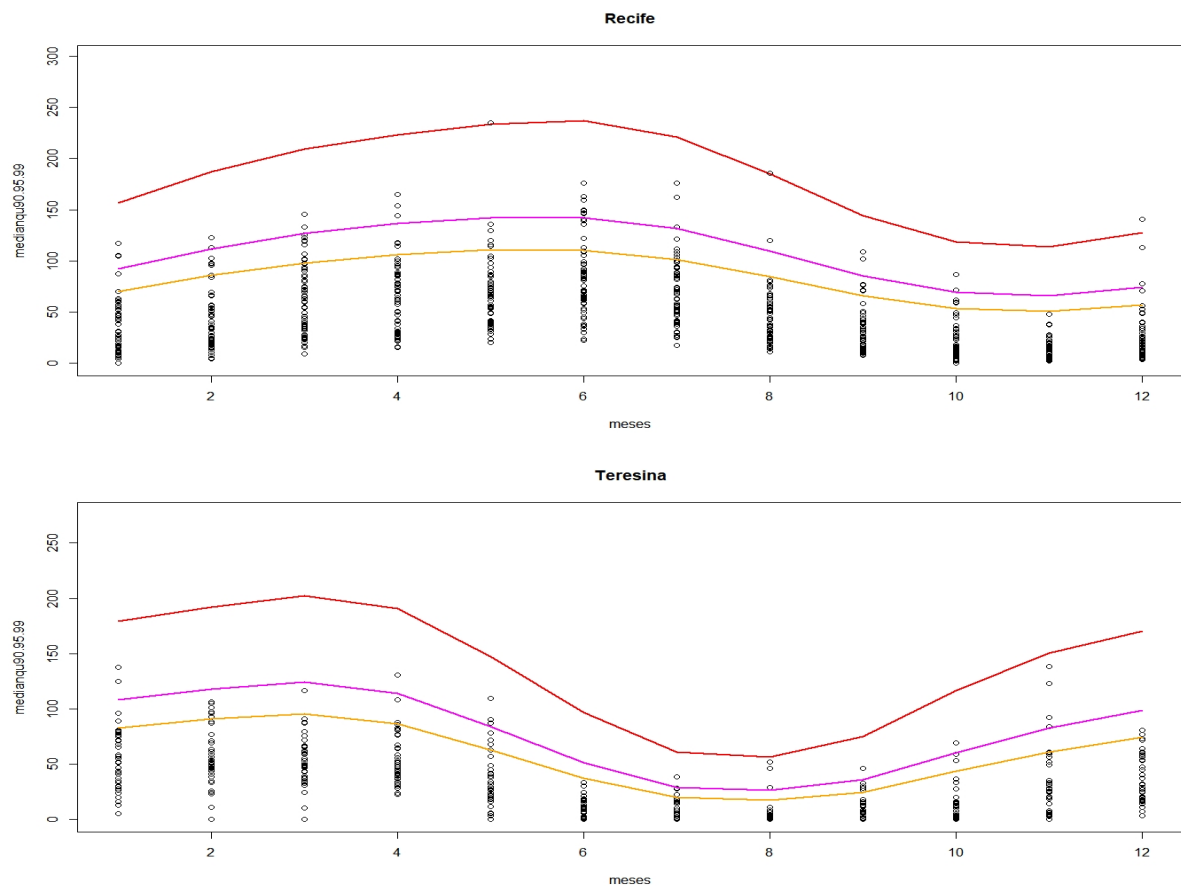
inflados de zeros sazonais.

Observamos que os resultados alcançados para a cidade de São João do Piauí, na Figura 16, foram semelhantes aqueles encontrados com os dados da cidade de Petrolina.

5.6.3 Níveis de retorno com observações da amostra

A Figura 17 apresenta os gráficos dos níveis de retorno de 10, 20 e 100 períodos referentes aos quantis altos de 90%, 95% e 99% da distribuição SZIGEV, aplicados aos dados de Recife e Teresina.

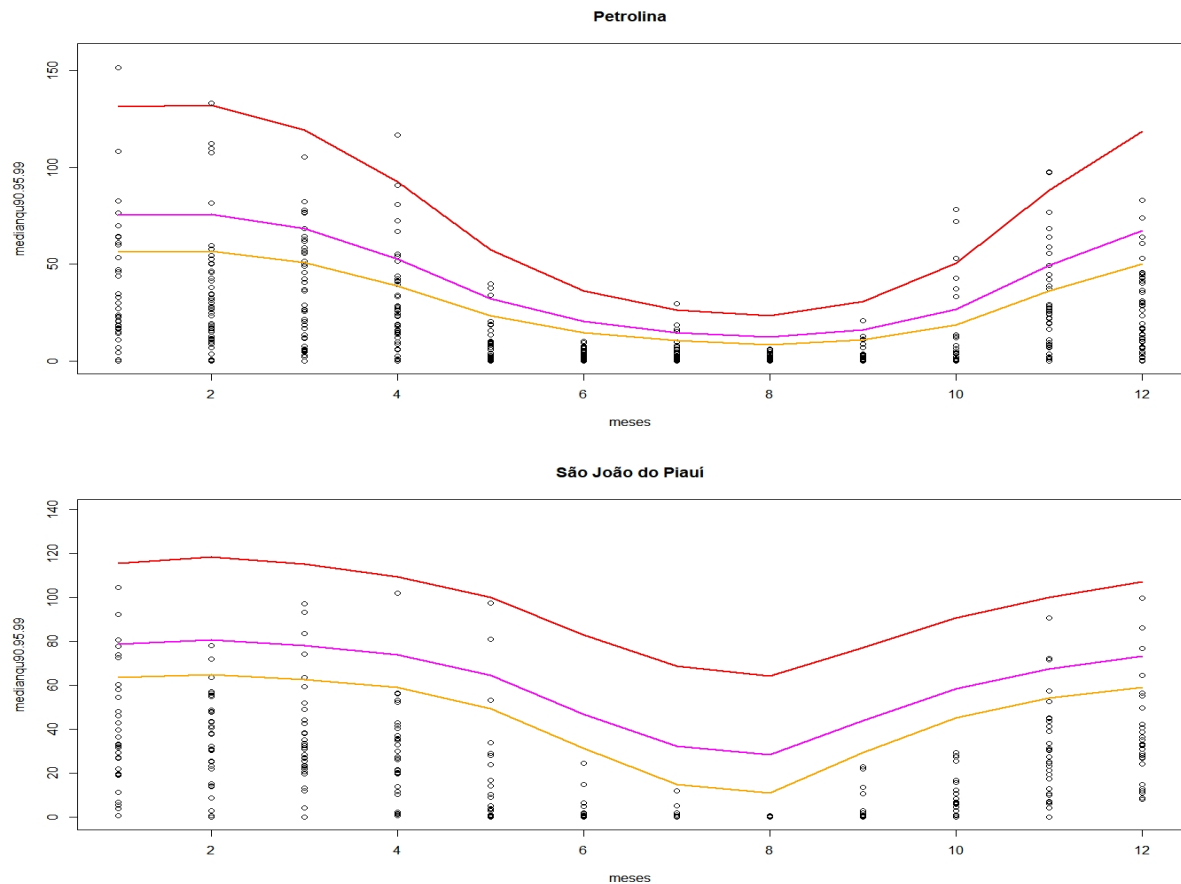
Figura 17 – Quantis extremos 90%, 95% e 99% da SZIGEV, com as observações da amostra, para as cidades de Recife e Teresina



Esses gráficos mostram o quanto foram precisas as estimativas feitas pela SZIGEV em relação ao desempenho das séries dos dados em análise. Verificamos que suas previsões foram muito assertivas. Enxergamos com muita nitidez a ocorrência das chuvas em cada mês, como também seu desempenho ao longo do ano, tanto para os dados da cidade de Recife quanto para os da cidade de Teresina.

Na Figura 18, analisamos os gráficos dos níveis de retornos de 10, 20 e 100 períodos referentes aos quantis altos de 90%, 95% e 99% da distribuição SZIGEV, aplicados nos dados de Petrolina e São João do Piauí.

Figura 18 – Quantis extremos 90%, 95% e 99% da SZIGEV, com as observações da amostra, para as cidades de Petrolina e São João do Piauí



Analisando o ciclo sazonal em cada mês ao longo do ano, para cada uma dessas duas cidade, percebemos que todos os níveis de retorno encontrados e que são esperados que aconteçam, foram muito precisos, em acordo com os registros das séries históricas dos dados. Enfatizando que esse modelo sazonal está bem ajustado, tornando-se ideal para a modelagem dos dados de valores extremos.

Esses resultados apresentam a necessidade e importância de utilizarmos a distribuição SZIGEV para fazerem-se as análises nos dados de máximos de chuvas, que geralmente apresentam ciclos sazonais. Pelos resultados alcançados nesse trabalho podemos afirmar que as estimações realizadas por esse modelo sazonal, estarão mais ajustadas e precisas a despeito dos acontecimentos das chuvas intensa e extremas para cada mês do ano.

5.7 Conclusão

De acordo com os resultados obtidos, podemos concluir que o modelo SZIGEV, que tem origem a partir da transformação sazonal na distribuição ZIGEV, realizada através do modelo de regressão e de suas funções de ligação, obteve resultados muito expressivos, tanto com relação aos seus estimadores, que foram bastante significativos e precisos para os parâmetros do modelo, que sofreram muita influência das 3 covariáveis em estudo, o cosseno do mês, o seno do mês e a interação entre o cosseno e o seno, respectivamente, como também em relação as medidas de qualidade de ajuste, AIC, BIC e DIC, que confirmaram um melhor ajuste para o modelo SZIGEV, em comparação com a ZIGEV, demonstrando ser razoável para analisar dados de valores extremos de precipitação de chuvas. Não é difícil que esses dados sejam sazonais, mas mesmo que isso aconteça e não apresente sazonalidade a distribuição SZIGEV converge a ZIGEV. Caso ainda os dados não tenham máximos de zeros em seus bancos de dados o modelo ZIGEV converge a distribuição GEV.

Surgindo assim como solução diante da problemática de se obter modelos que sejam precisos e acurados na análise de dados de precipitação de chuvas, na previsão de chuvas intensas e extremas, como também na previsão de períodos de seca. Auxiliando os governos e a sociedade civil tanto na prevenção de perdas de bens materiais e vidas humanas, quanto proporcionando ganhos.

6 Considerações Finais

Fizemos uma aplicação com a distribuição Generalizada de Valores Extremos Inflada de Zeros, ZIGEV, recomendada para conjuntos de dados de máximos, com excessos de zeros. Utilizou-se dados de precipitação diárias de chuvas do estado de Pernambuco, região nordeste do Brasil, para as cidades de, Recife, Barreiros e Surubim da mesorregião Metropolitana, Garanhuns do Agreste, Triunfo e Arcoverde do Sertão e Ouricuri, Cabrobó e Petrolina da mesorregião do São Francisco. Nesta aplicação, o modelo ZIGEV mostrou mais uma vez sua excelência na qualidade do ajuste, conseguindo modelar muito bem os dados de máximos, em todas as diferentes proporções de zeros, variando entre 0.28%, nos dados da cidade de Recife, a 16.60% nos dados de Petrolina. Estes resultados foram confirmados através das medidas da qualidade de ajuste (AIC, BIC e DIC), em que todas as medidas de ajustes para a distribuição ZIGEV foram melhores que para a GEV, com a exceção das medidas AIC e BIC para os dados de Recife com 0.28% de zeros, e que apesar dos resultados dos modelos terem sido bastante semelhantes, essas duas medidas aplicam uma penalização pela quantidade de parâmetros.

Propomos nesse trabalho a criação de um modelo sazonal, que além de modelar dados de máximos, com pouca ou muita presença de zeros nesses dados, mas que também conseguisse captar e modelar a presença sazonal em que muitos desses dados possuem. Então, aplicar esse novo modelo em dados de precipitação de chuvas e compará-lo com a recém criada distribuição ZIGEV para esses mesmos dados. Os resultados foram muito satisfatórios. Todos os dados estudados nesse trabalho, com presença de zeros infladas ou não, como foi o caso de Recife com aproximadamente $\omega = 0$ ($\omega \approx 0$), apresentaram sazonalidades e o modelo SZIGEV se sobressaiu melhor que o modelo ZIGEV em todas as situações. Esses resultados foram muito expressivos, tanto com relação aos seus estimadores, que foram mais precisos, quanto as medidas de qualidade de ajuste, AIC, BIC e DIC. Demonstrando está melhor ajustada para analisar dados de máximos mensais, inflados de zeros e sazonais.

Portanto, com os resultados obtidos e apresentados por meio de suas estimações e medidas de ajuste podemos afirmar que a distribuição SZIGEV é a ideal para a modelagem de dados de máximos de precipitação de chuvas, inflado de zeros e sazonais.

7 Referências

- Almeida, A.Q. de; Ribeiro, A.; Paiva, Y.G.; Rascon, N.J.L.; Lima, E.D.P. (2011). *Geoestatística no estudo de modelagem temporal da precipitação*. Revista Brasileira de Engenharia Agrícola e Ambiental, Campina Grande, v. 15, n. 4, p. 354-358. DOI: <http://dx.doi.org/10.1590/S1415-43662011000400004>.
- Assis, J.M.O. de, Sobral, M. do C.M., Souza, W.M. de (2012). *Análise de Detecção de Variabilidade Climáticas com Base na Precipitação nas Bacias Hidrográficas do Sertão de Pernambuco*. Revista Brasileira de Geografia Física, v. 5, n. 3, p. 630-645.
- Beijo, L.A. and Avelar, F.G. (2011). *Distribuição Generalizada de Valores Extremos no estudo de dados climáticos: uma breve revisão e aplicação*. Revista da Estatística da UFOP, Ouro Preto, pp. 10-16.
- Cavalcanti, A.S. (2009). *Avaliação de padrões atmosféricos associados à ocorrência de chuvas extremas no litoral da região Nordeste do Brasil: aspectos numéricos na previsão operacional do tempo*. Tese. (Doutorado em Engenharia Mecânica). Rio de Janeiro: Programa de Pós-Graduação em Engenharia Mecânica/ UFRJ.
- Coles, S., Ribeiro, P.J.Jr. (2022). *Introdução à Inferência Bayesiana*. Laboratório de Estatística e Geoinformação (LEG). UFPR.
- CRED. (2016). Centre for Research on the Epidemiology of Disasters. *Base de dados do EM-DAT - The International Disaster Database*. Disponível em: <http://www.emdat.be/>.
- Fisher, R.A. and Tippett, L.H.C. (1928). On the estimation of frequency distributions of the largest and smallest number of a sample, Proceedings of the Cambridge Philosophical Society, 24, 180.
- Gamerman, D. and Lopes, H.F. (2006) *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* 2nd ed. Baton Rouge: Chapman & Hall/CRC.
- Garcia, S.S., Amorim, R.S., COUTO, E.G., Stopa, W.H. (2011). *Determinação da equação intensidade-duração-frequência para três estações meteorológicas do Estado de Mato Grosso*. Revista Brasileira de Engenharia Agrícola e Ambiental, Campina Grande, v. 15, n. 6, p. 575-581.
- Gonçalves, R.C. and Guetter, A.K. (2019). *Análise de frequência regional de ventos*

- extremos no Paraná*. Revista Brasileira de Meteorologia v. 34, n. 2, 301-313.
DOI: <http://dx.doi.org/10.1590/0102-7786334003>
- Gnedenko, B.V. (1943). *Sur la distribution limite du terme maximum d'une serie aléatoire*, Annals of Mathematics, 44, 423-1943.
- Gramosa, A.H.Q., Nascimento, F.F. and Morales, F.E.C. (2019). *A Bayesian approach to zero-inflated data in extremes*. Communications in Statistics - Theory and Methods. doi: 10.1080/03610926.2019.1594305.
- Guha-Sapir, D.; Vos, F. and Below, R. (2011). *Annual Disaster Statistical Review the Numbers and Trends*. Bruxelles.
- Jenkinson, A.F. (1955). *The frequency distribution of the annual maximum (or minimum) values of meteorological events*. Quarterly Journal of the Royal Meteorology Society, 81, 158. doi:10.1002/qj.49708134804.
- Jost, G. et al. (2012). *A hillslope scale comparison of tree species influence on soil moisture dynamics and runoff processes during intense rainfall*. Journal of Hydrology, Amsterdam, v. 420, n. 2, p. 112-124.
- Marinho, C.S., Nascimento, A.C. do, Miranda, H.S., Costa, M.R. (2023). *Grupo Mulher Maravilha: a força das mulheres de Nova Descoberta / Recife em tempos de redemocratização (1975-1988)*. TEDE2 - UFRPE, Recife - (PE).
- Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J. e Possingham, H.P. (2005). *Zero Tolerance Ecology: Improving Ecological Inference by Modelling the source of Zero Observations*. Ecology Letters, 1235, 1246.
- Modricka, T.M. and Georgakakos, K.P. (2015). *O caráter e as causas da ocorrência de inundações repentinas mudam em pequenas bacias montanhosas do sul da Califórnia sob a mudança climática projetada*. Journal of Hydrology: Regional Studies, 3: 312-336.
- Murteira, B., Paulino, C.D., Amaral Turkman, M.A., Silva, G.L. (2018). *Estatística Bayesiana*, 2ª edição. Fundação Calouste Gulbenkian, Lisboa.
- Nascimento, F.F., Gamerman, D. and Lopes, H.F. (2011). *Regression models for exceedance data via the full likelihood*. Environmental and Ecological Statistics 18 495-512.
- Nascimento, F.F. (2012). *Modelos Probabilísticos para dados Extremos: Teoria e aplicações*. In: II COLÓQUIO DE MATEMÁTICA DA REGIÃO NORDESTE. UFPI. Teresina.
- OFDA/CRED. (2009). *The Office of US Foreign Disaster Assistance/Centre for Research*

- on the Epidemiology of Disasters* Université Catholique de Louvain. Annual Disaster Statistical Review 2008. The numbers and trends. Brussels, Belgium. Disponível em: www.emdat.be/Documents/Publications/ADSR.2008.pdf.
- O'Hagan, A. (1994). *Bayesian Inference*. Edward Arnold. Kendall's advanced theory of statistics.
- Olímpio, J.L.S, Vieira, P.M, Zanela, M.E and Sales, M.C.L. (2013). *Episódios Pluviais Extremos e a Vulnerabilidade Socioambiental do município de Fortaleza: o episódio do dia 27/03/2012*. Revista Geo UERJ, v. 1, p. 181-206.
- Ottero, C.R., Chargel, L.T. (2018). *Análise de frequência dos dados pluviométricos observados em 2011 e 2013 na Região Serrana*, Estado do Rio de Janeiro. Revista Brasileira de Meteorologia v. 33, n. 1, 131-139. DOI: <http://dx.doi.org/10.1590/0102-7786331007>.
- Parmesan, C., Root, T. and Willig, M.R. (2000). *Impacts of Extreme Weather and Climate on Terrestrial Biota*. Bulletin of the American Meteorological Society 81(3):443-450. DOI:10.1175/1520-0477(2000)081<0443:IOEWAC>2.3.CO;2.
- Pereira, D.C., Duarte, L.R., Sarmento, A.P. (2017). *Determinação da curva de intensidade, duração e frequência do município de Ipameri-Goiás*. Revista Eletrônica de Engenharia Civil, Goiânia, v.13, n.2, p.233-246.
- Quadros, L.E., De Queiroz, M.M.F., Antônio, M. Vilas Boas. (2011). Distribuição de frequência e temporal de chuvas intensas. Acta Scientiarum Agronomy, Maringá, v.33, n.3, p.401-410.
- Rodrigues, J. (2003). *Bayesian Analysis of Zero-Inflated Distributions*. Communications in Statistics-Theory and Methods 32 (2):2819. doi:10.1081/STA-120018186.
- Santos, G.G., Griebeler, N.P., De Oliveira, L.F.C. (2010). *Chuvas intensas relacionadas à erosão hídrica*. Revista Brasileira de Engenharia Agrícola e Ambiental, Campina Grande, v.14, n.2, p.115-123.
- Sena, J.P.O., Lucena, D.B. and Ribeiro, G.N. (2017). *Eventos extremos de precipitação no sertão paraibano: Variação espaço-temporal*. Revista Verde de Agroecologia e Desenvolvimento Sustentável, v. V.12, n.Nº 4, p. 748-755, DOI: <http://dx.doi.org/10.18378/rvads.v12i4.4938>.
- Spiegelhalter, D.J., Best, N.G. and Carlin, B.P. et al. (2002) *Bayesian measures of model complexity and fit*. Journal of the Royal Statistical Society, Series B - Statistical Methodology, v.64, p.583-616.
- Staley, M.D., Gartner, E.J., Kean, J.W. (2015). *Definição objetiva dos limites de in-*

- tensidade e duração das chuvas para inundações e fluxos de detritos repentinos na área queimada pelo incêndio do Waldo Canyon, Colorado, EUA.* Engineering Geology for Society and Territory, 2: 621-624. 10.1007 / 978-3-319-09057-3.103.
- Thomas, M. et al. (2016). *Applications of extreme value theory in public health.* PLoS ONE, p. 1 - 7.
- Tominaga, L.K., Santoro, J. and Amaral, R. (2009). *Desastres naturais: por que ocorrem.* IN.: Desastres Naturais Conhecer para prevenir. São Paulo.
- von Mises, R. (1954). *La distribution de la plus grande de n valeurs,* American Mathematical Society 2:271-94.