Extreme limit distribution of truncated models for daily rainfall

Aracelis Hernández[*]

Facultad Experimental de Ciencias y Tecnología

Universidad de Carabobo, Naguanagua, Valencia Venezuela

Lelys Guenni[†]

Centro de Estadística y Software Matemático

Universidad Simón Bolívar

Apartado 89.000, Caracas 1080-A, Venezuela

Bruno Sansó[‡]

Department of Applied Mathematics and Statistics

University of California, Santa Cruz, CA, USA

Abstract

We propose truncated and power-transformed (TPT) models for daily rainfall and we derive the Generalized Extreme Value (GEV) limit distributions for these models. We find that these limit distributions belong to the domain of attraction of the Fréchet family when the parent distribution of the daily values is a TPT t-Student model. In this case the shape parameter of the limiting GEV model depends on the degrees of freedom and the power transformation parameter. When the parent distribution of the daily values is a TPT Normal model, the limiting GEV model is independent of the parameters of the parent model. We perform a detailed inference and predictive analysis to validate these theoretical results using a Bayesian approach. Markov Chain Monte Carlo methods (MCMC) were used to estimate the posterior distribution of the parameters of the t-Student model for daily rainfall on one hand, and to estimate the posterior distribution of the parameters of the GEV model for the annual maxima on the other hand. Numerical results are presented for two locations: Maiquetía (Vargas State), and La Mariposa (Miranda State), Venezuela. Simulations from the predictive distribution of the daily values suggest a good approximation between

[*]e-mail:arhernan@uc.ve

[†]Corresponding author: e-mail:lbravo@cesma.usb.ve

[‡]e-mail:bruno@ams.ucsc.edu

the extreme distribution of the TPT t-Student model and the Fréchet model found by standard extreme value limit theory.

1 Introduction

Rainfall modeling has been a topic of high interest for many years in the hydrological science literature from both a deterministic and a stochastic perspective. From the stochastic point of view, rainfall models are developed with the motivation to understand the probabilistic structure of the physical phenomenon mostly to simulate rainfall data to be used as inputs to hydrological, agricultural and environmental models. Cox and Isham (1984) identified three types of rainfall modeling approaches: Empirical Statistical models, with no consideration of the physical dynamics of the process (Gabriel and Neumann, 1962; Stern and Coe, 1984; Woolhiser, 1992); purely deterministic approaches resembling the physical rainfall dynamic through the solution of complex partial differential equation systems. One kind of these models are the Regional Climate Models (RCMs) which produce detailed meteorological data (including rainfall) by simulating atmosphere and land-surface processes (see for example Liston and Pielke (2000); Richard et al. (2002)). A third kind of models are intermediate models including in their definitions physical components of the rainfall process as rain cells, rain bands, fronts and storm nuclei. Pioneer work on this topic was developed by LeCam (1961); Waymire et al. (1984); Cox and Isham (1988) and Phelan and Goodall (1990), with space-time extension by Rodríguez-Iturbe et al. (1987, 1988) and Cowpertwait (1994).

Within the context of the empirical statistical models, truncated and power transformed (TPT) rainfall models are a good modeling option since they naturally include the non-zero probability mass point of zero values by the truncation of the latent distribution function. If X is a random variable measuring the accumulated rainfall on a particular time scale, the TPT normal model is defined as:

$$X = \begin{cases} W^{\beta} & \text{if } W > 0 \\ 0 & \text{if } W \leq 0 \end{cases} \tag{1}$$

where $W \sim N(\mu, \sigma^2)$, and $\beta > 0$ is the unknown power transformation parameter.

The power transformation accommodates the lack of symmetry of rainfall data especially for finer time scales as daily and hourly; and the use of the normal distribution as a latent random variable is always convenient from the inference point of view. These models have been used by Stidd (1953, 1973); Richardson (1977)and Hutchinson et al.(1993). Bárdossy and Plate (1992) proposed a spatio-temporal extension of the TPT normal model; Glasbey and Nevison (1997) also considered a truncated normal distribution model with a different transformation family. Sansó and Guenni (1999) considered a spatio-temporal structure of the TPT model and used a Bayesian approach to estimate the model parameters and to make model predictions. This model was extended to a dynamic version in Sansó and Guenni (2000) to handle changes in seasonal patterns through time and to produce model predictions including all the uncertainty in the model parameters.

Extreme rainfall events are of great interest due to their potential societal and economic impact. TPT models have been shown to be a flexible tool for modeling rainfall. Thus, it is important to study the distribution of extreme values under a TPT model. The theory of extreme events is well developed both probabilistically and inferentially. It is based on limiting approximations to the distribution of the sequence $M_n = \max\{X_1, \ldots, X_n\}$ where X_i are independent and identically distributed random variables from the parental distribution. In our case, the parental distribution will be a TPT. Usually n represents the number of observations in a year, and M_n represents the annual maximum. After some normalization, the distribution of M_n converges to a member of the Generalized Extreme Value (GEV) distribution family (von Mises, 1936; Jenkinson, 1955). GEV distributions are of the form

$$G(z) = \exp\left\{ -\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi} \right\}, \tag{2}$$

for

$$\left\{ z : 1 + \xi\left(\frac{z-\mu}{\sigma}\right) > 0 \right\},$$

$-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$. The parameters μ, σ and ξ represent the location, scale and shape parameters respectively. The case $\xi \to 0$ leads to the Gumbel family; the case $\xi < 0$ leads to the Weibull family and the case $\xi > 0$ leads to the Fréchet family. More precisely, a distribution function F, discrete or absolutely continuous, belongs to the maximum domain of attraction of a non-degenerated distribution function G, if there are sequences $\{a_n > 0\}$ and $\{b_n\}$ such that:

$$\lim_{n\to\infty} F^n(a_n z + b_n) = G(z) \tag{3}$$

and G belongs to the (GEV) distribution family. This type of convergence is called weak convergence. Galambos (1978) determined the necessary conditions for F in order that the normalizing sequences $\{a_n > 0\}$ and $\{b_n\}$ exist and (3) holds.

The results for the GEV encompass those in von Mises (1936). According to the earlier work, it is possible to establish sufficiency conditions for F to belong to the domain of attraction of either the Gumbel, Fréchet or Weibull families. Such results are used in this work to determine the domain of attraction of the TPT model. We consider two cases: when the underlying TPT distribution is normally distributed and when it is a t-Student distribution. These results are presented in section 2.

The paper continues with an extensive data analysis to validate our theoretical results. The rationale of the validation is as follows:

- A TPT model is fitted to the daily data values using a Bayesian approach.

- We calculate the annual maxima of the simulations of daily values obtained from the predictive TPT distribution
- The theoretical extreme limit distribution corresponding to the domain of attraction of the TPT model is fitted to the observed annual maxima using a Bayesian approach.
- Simulations from the predictive distribution of the extreme value model are compared to the annual maxima simulated from the predictive distribution obtained from the TPT model for daily values.

These results are presented in sections 3.1 to 3.4 and finally, the discussion and conclusions of this analysis are presented in section 4.

2 Domain of attraction of TPT models

The results presented in this section follow the methods in von Mises (1936). We use the the sufficiency conditions for weak convergence of the distribution of maxima used to determine the domain of attraction of the truncated Normal model and the truncated t-Student model. Additionally, we establish the relationship between the parameters of the limiting and the parental distributions.

2.1 Truncated Normal model

Let X be a random variable that corresponds to the accumulated rainfall values for a given time scale. Assume that X follows the model described by (1). The distribution function and probability density function of X are given respectively by:

$$\begin{aligned} F_X(x) &= \Phi\left(-\frac{\mu}{\sigma}\right) I_{\{x=0\}} + \Phi\left(\frac{x^{1/\beta}-\mu}{\sigma}\right) I_{\{x>0\}} \\ f(x) &= \frac{1}{\beta\sigma} x^{1/\beta-1}\varphi\left(\frac{x^{1/\beta}-\mu}{\sigma}\right) I_{\{x>0\}} + \delta_0(x)\Phi(-\mu/\sigma) \end{aligned}$$

where Φ and φ are, respectively, the distribution and density functions of a standard normal, I_A is the indicator function of the set A and δ_0 is the Dirac delta. function at 0, such that $lim_{\Delta\to 0}\int_{-\Delta}^{\Delta}\delta_0(x)dx = 1$

According to von Misses (1939), if F is an absolutely continuous and $h(z)$ is defined as

$$h(z) = \frac{f(z)}{1-F(z)}, \tag{4}$$

then

1. If $h(z) > 0$ and, for some $\delta > 0$

$$\lim_{z\to\infty} z h(z) = \delta, \tag{5}$$

then F belongs to domain of attraction of a Fréchet.

2. If $F^{-1}(1) < \infty$ and, for some $\delta > 0$

$$\lim_{z\to F^{-1}(1)} (F^{-1}(1) - z)h(z) = \delta,$$

then F belongs to the domain of attraction of a Weibull.

3. If $h(z)$ is non-zero and differentiable for z close to $F^{-1}(1)$, then F belongs to domain of attraction of a Gumbel if

$$\lim_{z\to F^{-1}(1)} \frac{d}{dz}\left\{\frac{1}{h(z)}\right\} = 0. \tag{6}$$

For Fréchet and Weibull distributions, δ is the shape parameter of the GEV. We have that $\delta = 1/\xi$ if $\xi > 0$ (Fréchet case) and $\delta = -1/\xi$ if $\xi < 0$ (Weibull case).

We apply the previous results to model (1). We notice that the corresponding distribution function is absolutely continuous in $(0, \infty)$ and, following (4), the risk function $h(x)$ is

$$\frac{\frac{1}{\beta\sigma}\, x^{1/\beta-1}\, \varphi\left(\frac{x^{1/\beta}-\mu}{\sigma}\right)\, I_{\{x>0\}} + \delta_0(x)\Phi(-\mu/\sigma)}{1 - \Phi\left(-\frac{\mu}{\sigma}\right)\, I_{\{x=0\}} - \Phi\left(\frac{x^{1/\beta}-\mu}{\sigma}\right)\, I_{\{x>0\}}}\ .$$

Since we are interested in the limit for $x \to \infty$, the term $\delta_0(x)\Phi(-\mu/\sigma)$ can be dropped from the analysis. So, it can be seen that the conditions for (6) hold and

$$\begin{aligned}\frac{d}{dx}\left(\frac{1}{h(x)}\right) \quad &= \quad -I_{\{x>0\}} - \frac{\left[1 - \Phi\left(-\frac{\mu}{\sigma}\right)\, I_{\{x=0\}} - \Phi\left(\frac{x^{1/\beta}-\mu}{\sigma}\right)\, I_{\{x>0\}}\right]}{\varphi\left(\frac{x^{1/\beta}-\mu}{\sigma}\right)\, I_{\{x>0\}}} \\ &\cdot \frac{\left[\beta\sigma\left(\frac{1}{\beta} - 1\right)\, x^{-1/\beta} - \left(\frac{x^{1/\beta}-\mu}{\sigma}\right)\right]\, I_{\{x>0\}}}{\varphi\left(\frac{x^{1/\beta}-\mu}{\sigma}\right)\, I_{\{x>0\}}}\ .\end{aligned}$$

The limit of this expression, as $x \to \infty$ is equal to 0. Therefore the distribution of the TPT normal model belongs to the domain of attraction of the Gumbel distribution. This result is not surprising given the relationship between the parent distribution and the normal model. It also follows that the limit distribution does not depend on the power parameter β.

2.2 Truncated t-Student model

We consider now a modification of model (1) where $W \sim Student(\mu, \sigma^2, \alpha)$. Again, as in the TPT normal model, $\beta > 0$ is an unknown parameter. We shall refer to this model as the truncated student model (TSM). The distribution and density functions of a random variable X following a truncated student model are given by

$$\begin{aligned} F_X(x) &= F_W(0)\, I_{\{x=0\}} + F_W(x^{1/\beta})\, I_{\{x>0\}} \\ f(x) &= \frac{1}{\beta}\, x^{1/\beta-1} f_W(x^{1/\beta})\, I_{\{x>0\}} + \delta_0(x) F_W(0) \end{aligned} \tag{7}$$

where F_W and f_W are the t-Student distribution and density functions with location μ, scale σ^2 and α degrees of freedom. Dropping the density term at 0, the risk function for this model is given by

$$h(x) = \frac{\frac{1}{\beta}\, x^{1/\beta-1}\, f_w(x^{1/\beta})\, I_{\{x>0\}}}{1 - F_w(0)\, I_{\{x=0\}} - F_w(x^{1/\beta})\, I_{\{x>0\}}} \ .$$

To see that (5) holds we need to calculate $\lim_{x\to\infty^+} xh(x)$. We observe that

$$\begin{aligned} xh(x) &= \frac{\frac{1}{\beta}\, x^{1/\beta}\, f_w(x^{1/\beta})\, I_{\{x>0\}}}{1 - F_w(0)\, I_{\{x=0\}} - F_w(x^{1/\beta})\, I_{\{x>0\}}} \\ &= \frac{\frac{c}{\beta}\left(\frac{x^{2/\beta(\alpha+1)}}{1+\frac{\tau}{\alpha}(x^{1/\beta}-\mu)^2}\right)^{\frac{\alpha+1}{2}} I_{\{x>0\}}}{1 - F_w(0)\, I_{\{x=0\}} - F_w(x^{1/\beta})\, I_{\{x>0\}}} \ , \end{aligned}$$

so, after some calculations we have that

$$\begin{aligned} \lim_{x\to\infty^+} xh(x) &= -\frac{1}{\beta}\, I_{\{x>0\}} + \lim_{x\to\infty^+} \frac{(\alpha+1)}{2\beta}\, \frac{(2x^{1/\beta}-\mu)}{(x^{1/\beta}-\mu)}\, I_{\{x>0\}} \\ &= \frac{\alpha}{\beta}\, I_{\{x>0\}} > 0 \ . \end{aligned}$$

Therefore the extreme values of the TPT TSM tend to a limit distribution belonging to the Fréchet family with shape parameter given by α/β, where α are the degrees of freedom of the underlying t-Student distribution and β is the power transformation parameter. The former implies that the truncated t-Student model provides much larger flexibility to model the behavior of the tails of rainfall distribution. Even though the Gumbel is the default choice for maxima of rainfall, it has been observed that a Fréchet model can be more appropriate in some cases (see, for example, Coles and Pericchi, 2003). This fact will be corroborated in the section that follows. It should be noticed that, as $\alpha \to \infty$ the TPT TSM tends to the TPT normal model and the corresponding extreme limiting distribution approaches the Gumbel.

3 Validation results for the TPT TSM

von Misses's theory provides asymptotic approximations to the distribution of extremes. It is based on the assumption that the observations are independent. The number of available observations is usually limited and it is seldom the case that they are really independent. In this section we validate the results obtained in the previous section with the analysis of some rainfall data. A GEV model was fitted to the annual maxima of several locations in Venezuela. We selected two locations whose estimated limiting extreme distribution belonged to the Fréchet family. Additionally, one of the locations is relevant to the study of catastrophic extreme rainfall events (Maiquetía, Vargas State) and the other to the monitoring of water supply (La Mariposa, Capital district).

3.1 Data description

We consider daily precipitation data collected at Maiquetía, Vargas State, from year 1961 until 1999. This station is located in the central coast of Venezuela at the sea level. Although data from this location represent a dry climate, this area is directly under the topographic effects of the north-central mountain chain running parallel to the coast. This effect played an important role in the extreme events occurred during December 1999. The daily time series is shown in Figure 1 where the 1999 peaks stands out from the data values. Data for the annual maxima was available for a longer period of records than the daily time series. Forty nine years were available since 1951 until 1999. This data set is presented in Figure 2.

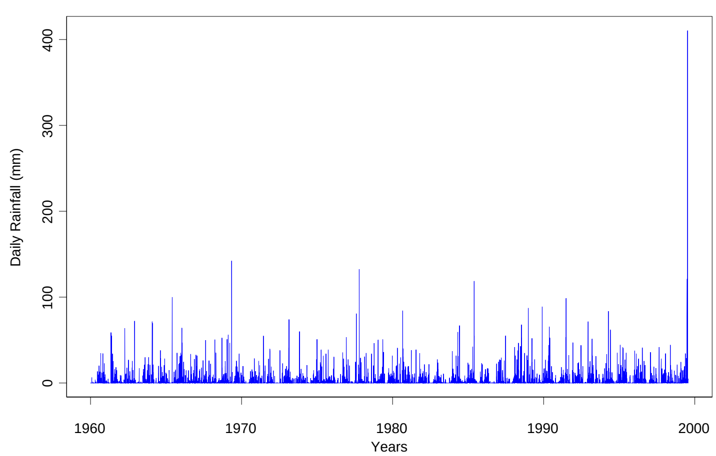

Figure 1: Daily Rainfall for Maiquetía Meteorological station

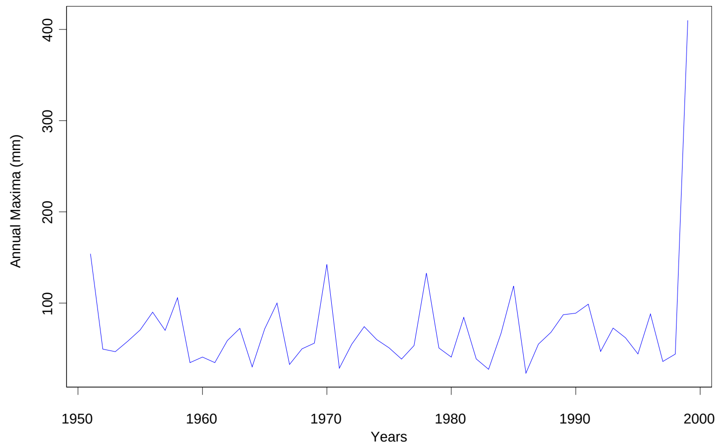

Figure 2: Annual Maxima for Maiquetía Meteorological station

The second location used in this analysis is La Mariposa, located in the Capital District of Venezuela. This station monitors the rainfall regimen affecting La Mariposa dam, which serves 25% of the population of Caracas, the capital of the country. The daily time series is available for the period 1949-1996. This data set is shown in Figure 3 and the annual maxima is presented in Figure 4.

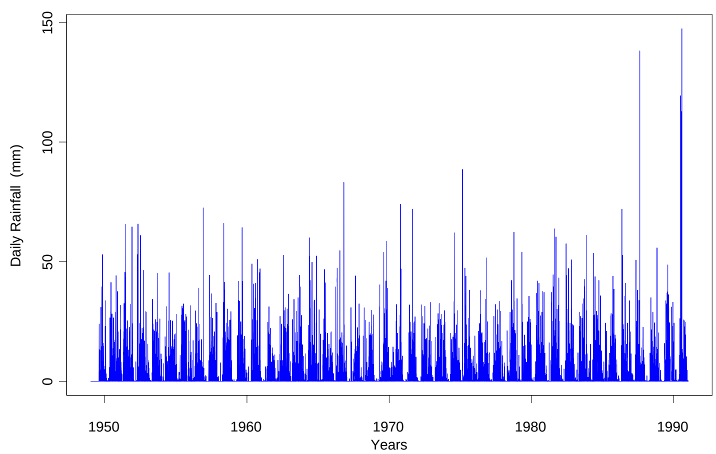

Figure 3: Daily Rainfall for La Mariposa Meteorological station

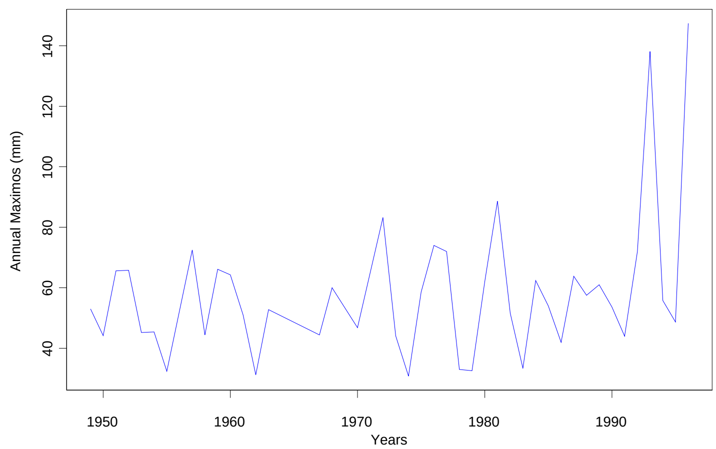

Figure 4: Annual Maxima for La Mariposa Meteorological station

3.2 Fitting the GEV distribution to the annual maxima

The parameters of the GEV model corresponding to Equation (2) were fitted using a Bayesian approach. Let $z_1, \dots, z_T$ represent the annual maxima, for T years, the likelihood of the GEV model can be expressed in the form:

$$g(z_1, \dots, z_T|\xi, \mu, \sigma) = \frac{1}{\sigma^T} \exp\left(\left(-\frac{1}{\xi} - 1\right) \sum_{i=1}^{T} \log\left(1 + \xi\left(\frac{z_i - \mu}{\sigma}\right)\right)\right) \\ \times \exp\left(-\sum_{i=1}^{T}\left(1 + \xi\left(\frac{z_i - \mu}{\sigma}\right)\right)^{-1/\xi}\right)$$

Assuming a priori parameter independence with marginal distributions $p(\mu), p(\sigma)$ and $p(\xi)$, the joint posterior distribution is proportional to $g(z_1, \dots, z_T|\mu, \sigma, \xi)\, p(\mu)\, p(\sigma)\, p(\xi)$. A MCMC method was implemented to obtain samples from this joint posterior distribution. The prior distributions were selected as $p(\mu) \propto 1$; $p(\sigma) \propto \frac{1}{\sigma}$ and $p(\xi) \propto 1$. The full conditional distributions of the parameters μ, σ and ξ were used to simulate the Markov chains. These are given by:

$$p(\mu|z_1, \dots, z_T, \sigma, \xi) \propto \exp\left(\left(-\frac{1}{\xi} - 1\right) \sum_{i=1}^{T} \log\left(1 + \xi\left(\frac{z_i - \mu}{\sigma}\right)\right)\right) \\ \times \exp\left(-\sum_{i=1}^{T}\left(1 + \xi\left(\frac{z_i - \mu}{\sigma}\right)\right)^{-1/\xi}\right)$$

$$p(\sigma|z_1,\ldots,z_T,\mu,\xi) \propto \sigma^{\frac{T}{\xi}-1} \exp\left(\left(-\frac{1}{\xi}-1\right)\sum_{i=1}^{T}\log(\xi(z_i-\mu)+\sigma)\right)$$
$$\times \exp\left(-\sum_{i=1}^{T}\left(1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right)^{-1/\xi}\right)$$

$$p(\xi|z_1,\ldots,z_T,\mu,\sigma) \propto \exp\left(\left(-\frac{1}{\xi}-1\right)\sum_{i=1}^{T}\log\left(1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right)\right)$$
$$\times \exp\left(-\sum_{i=1}^{T}\left(1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right)^{-1/\xi}\right)$$

We used Metropolis-Hasting steps to obtain samples from each conditional distributions given their non-standard forms. We run the MCMC for 20,000 iterations. Geweke (1972) and Heidelberger-Welch (1983) tests were performed to check the convergence of the chain. Output results of the Running means of the Bayesian Output Analysis (BOA) Program for S-plus indicate convergence. The estimated values of the posterior means are presented in Table 1 for Maiquetía station and Table 2 for La Mariposa station.

Parameter	**Mean**	**SD**	**MSE**	**q0.025**	**q0.5**	**q0.975**
μ	48.9792	0.6679	0.0114	47.7059	48.9874	50.2713
σ	21.1406	2.1960	0.0434	17.4865	20.9179	26.0737
ξ	0.2999	0.0100	0.000066	0.2801	0.2999	0.3195

Table 1: **Statistics summary for the GEV model at Maiquetía station**

Parameter	**Mean**	**SD**	**MSE**	**q0.025**	**q0.5**	**q0.975**
μ	45.9994	0.0101	0.00007	45.9794	45.9994	46.0190
σ	13.7843	1.6983	0.0579	10.9307	13.6235	17.6196
ξ	0.1203	0.0099	0.00006	0.1011	0.1202	0.1401

Table 2: **Statistics summary for the GEV model at La Mariposa station**

The posterior densities for the simulated values are presented in Figure 5 for Maiquetía and 6 for La Mariposa. It is important to note that the estimated mean value of the parameter ξ is 0.2998 in Maiquetía , and 0.1203 in La Mariposa. The values corresponding to 95% posterior probability intervals are all positive, indicating that the extreme limit distributions belong to the Fréchet family. Therefore, these data sets can be used to validate the theoretical results found in section 2 about the domain of attraction of the TPT TSM. We expect that the distribution of rainfall for these stations will correspond to a truncated t-Student.

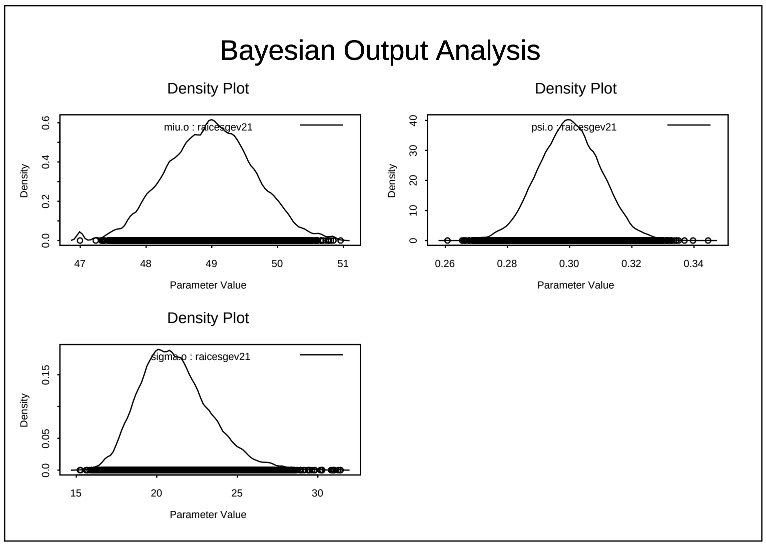

Figure 5: Posterior densities of the parameters of the GEV model at Maiquetía station

3.3 Predictive inference for the GEV model

In this section we present a validation of the results from the GEV model. Our diagnostics are based on comparing observed values with simulations from the predictive distribution of the annual maxima. The comparison is conducted as follows: i) Choose one sample for every 200 iterations of the MCMC. This produces 100 simulated values of the parameters. ii) Use each simulated value, say, $\mu^{(i)}$, $\sigma^{(i)}$ and $\xi^{(i)}$, $i = 1, \ldots, 100$ to sample annual maxima from the GEV model that corresponds to such values. The resulting sample corresponds to the posterior predictive distribution of the annual maxima.

The posterior predictive median and the 2.5% and 97.5% quantiles from the posterior predictive distribution are presented in Figure 7 for Maiquetía station and Figure 8 for La Mariposa station. For comparison, in each case the observed maxima have been superimposed. In most cases the observed values are located within the 95% probability intervals and are close to the median of the predictive distribution. We observe that the value for year 1999 is out of the probability intervals in Maiquetía station. This is not surprising, since this is an extremely unlikely event. In fact, from simulations of the predictive distribution we can estimate that $P(z > 350\ mm) \cong 0.006$. The return levels z_p vs. return period $1/p$ and their corresponding probability intervals are presented in Figure 9 for Maiquetía station and in Figure 10 for La Mariposa station.

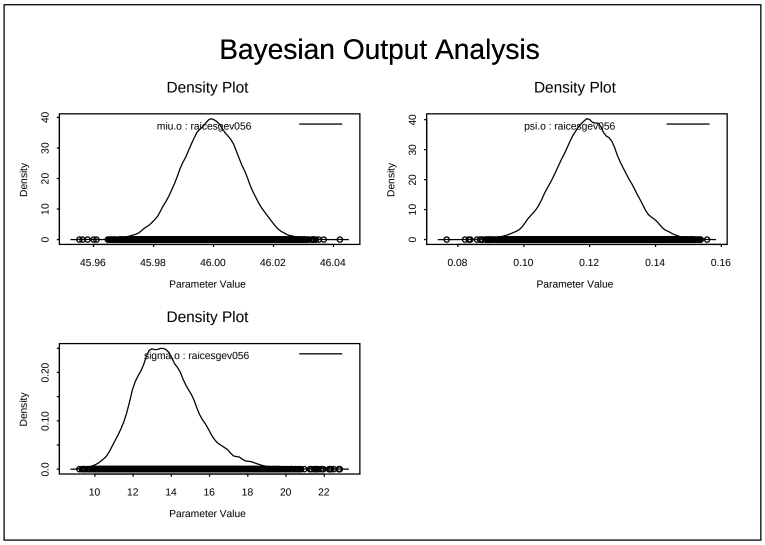

Figure 6: Posterior densities of the parameters of the GEV model at La Mariposa station

These values are calculated by plugging in the posterior median and the 2.5% and 97.5% quantiles of μ, σ and ξ into the GEV model. The observed maximum in Maiquetía close to 400 mm, causing the fatal events of year 1999 has an estimated return period of 500 years with a high uncertainty level.

3.4 Fitting the TPT TSM to the daily time series

The TPT TSM model was described in (1) where $W \sim Student(\mu, \sigma^2, \alpha)$ with parameter space given by the set $\Theta = \{(\mu, \sigma^2, \alpha, \beta) : -\infty < \mu < \infty, \sigma^2 > 0, \alpha > 0, \beta > 0\}$ The t-Student distribution with parameters μ, σ^2, α can be modeled through the hierarchical structure (Dickey, 1968) $W \sim N(\mu, \sigma^2/\eta),\ \eta \sim Gamma(\alpha/2, \alpha/2)$. In order to fit the truncated model we consider latent variables $v_i < 0$ that correspond to the dry periods. Within the MCMC, these are treated as additional model parameters.

Thus, letting Y_t be the total daily precipitation for $t = 1, \ldots, n$ at a particular location, then

$$Y_t = \begin{cases} W_t^{\beta} & \text{if } W_t > 0 \\ 0 & \text{if } W_t \leq 0 \end{cases}$$

where

$$\begin{aligned} W_t &\sim N\left(\mu_t, \frac{\sigma^2}{\eta_t}\right) \\ \eta_t &\sim Gamma\left(\frac{\alpha}{2}, \frac{\alpha}{2}\right). \end{aligned} \tag{8}$$

We observed that the seasonal variation of precipitation in the region under study may have a bimodal behavior during the year. Thus, a two-harmonics Fourier component

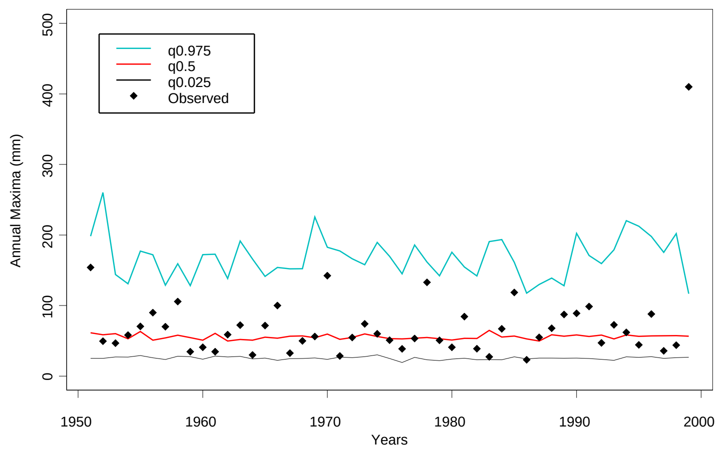

Figure 7: Quantiles for the simulated annual maxima in Maiquetía station

was added to the model. The resulting mean is of the form

$$\mu_t = a_0 + \sum_{k=1}^{2}\left[a_k\, cos\left(\frac{2k\pi t}{365}\right) + b_k\, sin\left(\frac{2k\pi t}{365}\right)\right]$$

for $t = 1, 2, \ldots, n$, where n is the number of daily records. In matrix form $\overrightarrow{\mu} = H\overrightarrow{a}$, where $\overrightarrow{\mu}$ is a vector of dimension n, H is a $n\times 5$ matrix of sines and cosines and $\overrightarrow{a}$ is the vector of the unknown harmonics coefficients of dimension 5 ($\overrightarrow{a} = (a_0, a_1, b_1, a_2, b_2)$). Given observations y_t for $t = 1, \ldots, n$, we can write the model as

$$w_t = \begin{cases} y_t^{1/\beta} & \text{if } t \in T \\ v_t & \text{if } t \in T_l \end{cases},$$

where $T_l = \{t : y_t = 0\}$ and $T = \{t : y_t \neq 0\}$ correspond to the time indices for dry and non-dry periods respectively.

Given the above notation, the model likelihood can be written as:

$$f(w_1, \ldots, w_n | \overrightarrow{a}, \sigma^2, \alpha, \beta, \overrightarrow{\eta}, \overrightarrow{v}) \propto \frac{(\sigma^2)^{-(n/2)}}{\beta^k}\left(\prod_{t=1}^{n}\eta_t\right)^{1/2}\left(\prod_{t\in T} w_t\right)^{1/\beta - 1}$$
$$\times \exp\left\{-\frac{1}{2\sigma^2}(\overrightarrow{w} - H\,\overrightarrow{a})' D\,(\overrightarrow{w} - H\,\overrightarrow{a})\right\} \qquad (9)$$

where $D = diag(\eta_1, \ldots, \eta_n)$. To complete the model we assume that the priors are proportional to $f(\overrightarrow{w}|\overrightarrow{a}, \sigma^2, \alpha, \beta, \overrightarrow{\eta}, \overrightarrow{v})p(\overrightarrow{a})p(\sigma^2)p(\alpha)p(\beta)p(\overrightarrow{\eta})p(\overrightarrow{v})$. MCMC methods were implemented to obtain samples from the joint posterior distribution. We

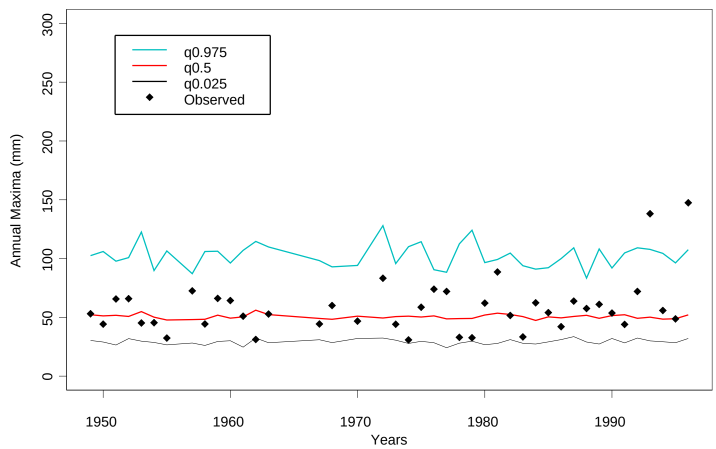

Figure 8: Quantiles for the simulated annual maxima in La Mariposa station

assume that $p(\eta_t)\text{Gamma}(\alpha/2, \alpha/2), p(\overrightarrow{a}) \propto 1, p(\sigma^2) = \text{IGamma}(a_0, b_0), p(\alpha) \propto 1/\alpha$. The resulting full conditionals are given by

$$p(\overrightarrow{a}|\sigma^2, \alpha, \beta, \overrightarrow{\eta}, \overrightarrow{v}, \overrightarrow{w}) \propto N_5(\overrightarrow{\hat{a}}, \sigma^2(H'DH)^{-1})$$

$$p(\eta_t|\overrightarrow{a}, \sigma^2, \alpha, \beta, \overrightarrow{v}, \overrightarrow{w}) \propto \text{Gamma}\left(\frac{\alpha}{2} + \frac{1}{2}, \frac{\alpha}{2} + \frac{(w_t - (H\overrightarrow{a})_t)^2}{2\sigma^2}\right) \quad t = 1, \ldots, n$$

$$p(\sigma^2|\overrightarrow{a}, \alpha, \beta, \overrightarrow{\eta}, \overrightarrow{v}, \overrightarrow{w}) \propto \text{IGamma}\left(a_0 + \frac{n}{2}, b_0 + \frac{1}{2}(\overrightarrow{w} - H\overrightarrow{a})'D(\overrightarrow{w} - H\overrightarrow{a})\right)$$

$$p(\alpha|\overrightarrow{a}, \sigma^2, \beta, \overrightarrow{\eta}, \overrightarrow{v}, \overrightarrow{w}) \propto \frac{1}{\alpha}\left(\prod_{t=1}^{n} \eta_t\right)^{\alpha/2-1} \exp\left\{-\frac{\alpha}{2}\sum_{t=1}^{n} \eta_t\right\}$$

and

$$p(\beta|\overrightarrow{a}, \sigma^2, \alpha, \overrightarrow{\eta}, \overrightarrow{v}, \overrightarrow{w}) \propto \frac{1}{\beta^k}\left(\prod_{t\in T} w_t\right)^{1/\beta-1} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t\in T} \eta_t(w_t - (H_1\overrightarrow{a})_t)^2\right\} \times p(\beta),$$

where H_1 is the sub-matrix of H corresponding to $t \in T$, and $(H_1\overrightarrow{a})_t$ is the t-th coordinate of that vector. A prior distribution for β was chosen such that its distribution

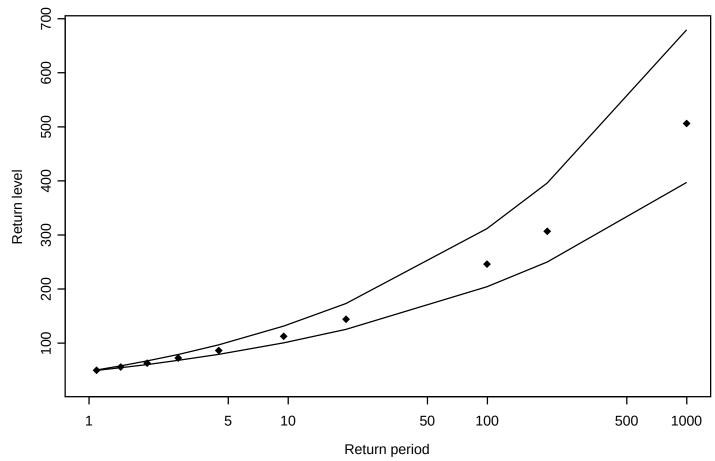

Figure 9: Return periods vs. return levels at Maiquetía station

was highly concentrated around 3, according to preliminary information available on this parameters and previous experience with this type of models. We used

$$p(\beta) \propto \beta^{9001-1} \exp\{-3000\beta\} \quad \text{for } \beta > 0 \; .$$

Finally, the posterior conditional distribution for $\overrightarrow{v}$ is given by

$$p(\overrightarrow{v}\,|\,\overrightarrow{a}, \sigma^2, \alpha, \beta, \overrightarrow{\eta}, \overrightarrow{w}) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{t\in T_l}\eta_t(v_t - (H_2\overrightarrow{a})_t)^2\right\} p(\overrightarrow{v}) \prod_{t\in T_l} I_{\{v_t\leq 0\}}(v_t)$$

where H_2 is the sub-matrix of H which corresponds to $t \in T_l$, and $(H_2\overrightarrow{a})_t$ is the t-th coordinate of the respective vector. By assuming prior independence of vector $\overrightarrow{v}$ and $p(v_t) \propto 1$ we get:

$$p(v_t|\overrightarrow{a}, \sigma^2, \alpha, \beta, \overrightarrow{\eta}, \overrightarrow{w}) \propto N\left((H_2\overrightarrow{a})_t, \frac{\sigma^2}{\eta_t}\right) I_{v_t(v_t)\leq 0} \quad \forall t \in T_l \; .$$

Sampling from this distribution implies simulating values from the negative side of a normal distribution. The Acceptance-Rejection algorithm of Devroye (1985) was used for this purpose.

A Gibbs sampler can be used to get samples from the conditional posterior distributions of $\overrightarrow{a}$, $\overrightarrow{\eta}$, $\overrightarrow{v}$ and σ, while a Metropolis-Hastings step is used for α and β. This MCMC was implemented for the Maiquetía and La Mariposa data. We used 5,000 iterations to perform Geweke (1972) and Heidelberger-Welch (1983) tests to check the convergence of each chain. A summary of the statistics for the model parameters is presented in Table 3 for Maiquetía station and Table 4 for La Mariposa station. Summary statistics for the harmonic coefficients are presented in Tables 5 and 6 respectively.

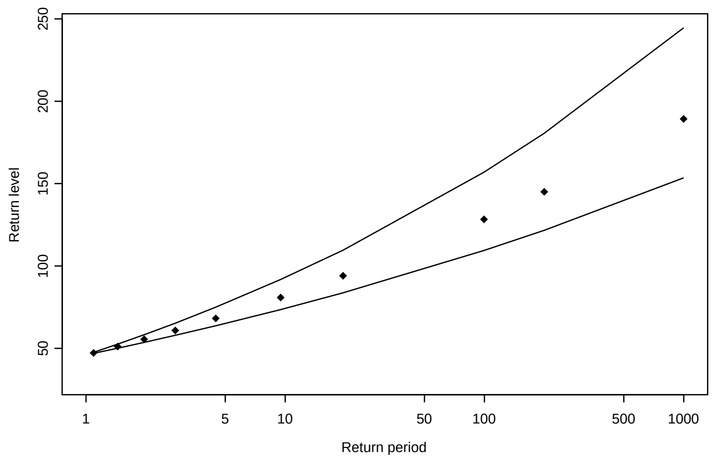

Figure 10: Return periods vs. return levels at La Mariposa station

Parameter	Mean	SD	MSE	q0.025	q0.5	q0.975
α	7.0340	0.0449	0.0005	6.9475	7.0332	7.1232
β	2.4068	0.0234	0.0008	2.3605	2.4061	2.4516
α/β	2.9227	0.0343	0.0010	2.8563	2.9228	2.9910
σ^2	3.9511	0.1326	0.0064	3.6992	3.9477	4.2132

Table 3: **Statistical summary of the parameters: α, β, the ratio α/β and σ^2 for Maiquetía station**

From these results it is important to notice that the value of the posterior mean of α/β is 2.9227 for Maiquetía station, and 8.3897 for La Mariposa station. These results are very close to the posterior mean of the inverse of the parameter ξ of the GEV distribution. This provides an empirical finite sample confirmation of the theoretical results found in section 2 about the relationship between the shape parameter of the GEV distribution and the parameters α and β. Notice that the value of β is very similar for both stations. This implies that the tail behavior of the extreme value distribution is solely determined by the degrees of freedom of the student distribution underlying the distribution of the daily records.

Parameter	Mean	SD	MSE	q0.025	q0.5	q0.975
α	19.2304	0.0728	0.0017	19.0918	19.2302	19.3690
β	2.2923	0.0199	0.0013	2.2558	2.2910	2.3300
α/β	8.3897	0.0783	0.0050	8.2385	8.3908	8.5382
σ^2	3.7362	0.1079	0.0074	3.5234	3.7398	3.9508

Table 4: **Statistical summary of the parameters: α, β, the ratio α/β and σ^2 for La Mariposa station**

Parameter	Mean	SD	MSE	q0.025	q0.5	q0.975
a_0	-1.4005	0.0410	0.0017	-1.4776	-1.4009	-1.3284
a_{11}	-0.5121	0.0367	0.0009	-0.5826	-0.5131	-0.4424
a_{12}	-0.8504	0.0410	0.0011	-0.9297	-0.8503	-0.7752
b_{11}	0.2778	0.0361	0.0009	0.2076	0.2771	0.3516
b_{12}	-0.1676	0.0366	0.0009	-0.2381	-0.1676	-0.0973

Table 5: **Statistical summary for simulated traces of harmonic coefficients in Maiquetía station**

4 Validation of theoretical results through predictive inference

Predictive inference was used to validate the theoretical results by simulating daily time series from the TPT TSM predictive distribution. The objective of this analysis is to check whether the annual maxima of the simulated series follow a Fréchet model with shape parameter $\delta = 1/\xi = \alpha/\beta$. While in the previous section we compared directly the posterior means of the parameters of the maxima and parental distributions, in this case the posterior predictive distribution is used to show that this relationship can be reproduced through predictive simulation.

Thirty time series of daily values of the same length as the observed time series were simulated by using the predictive distribution of the TSM. The annual maxima for each time series was calculated and their median and 95% quantile intervals are presented in Figures 11 and 12 jointly with the observed values in Maiquetía and La Mariposa station respectively.

To each of the thirty time series, a Fréchet model was fitted using the Bayesian approach described before. Twenty thousand samples of the posterior distribution of the model parameters were produced for each of the thirty time series. For Maiquetía the median of the simulated values is around 2.9. This is similar to the posterior median obtained by using the observed annual maxima as well as to the posterior mean of α/β in Table 3. For La Mariposa station the posterior median is around 8.3, which is also

Parameter	Mean	SD	MSE	q0.025	q0.5	q0.975
a_0	-0.4076	0.0271	0.0013	-0.4585	-0.4086	-0.3591
a_{11}	-0.6963	0.0296	0.0012	-0.7529	-0.6969	-0.6415
a_{12}	-1.0889	0.0350	0.0013	-1.1484	-1.0897	-1.0272
b_{11}	0.4654	0.0284	0.0010	0.4109	0.4649	0.5208
b_{12}	-0.2709	0.0286	0.0010	-0.3254	-0.2710	-0.2166

Table 6: **Statistical summary for simulated traces of harmonic coefficients in La Mariposa station**

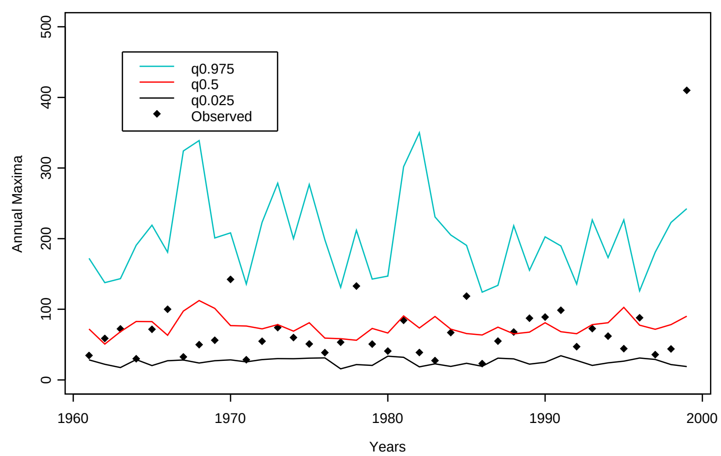

Figure 11: Quantiles from simulations of the predictive distribution at Maiquetía station

very close to the posterior median produced with the observed annual maxima and to the posterior mean of α/β in Table 4. This is an additional finite sample empirical validation of the theoretical results in Section 2.

5 Conclusions

By using extreme values classic convergence theory it was possible to demonstrate that the annual maxima for the truncated and power transformed Normal model follows a limit distribution from the Gumbel family. On the other hand, when the parent distribution follows a TPT TSM model, the extreme limit distribution belongs to the Fréchet family. This result is intuitively interesting, since taller tails from the t-Student model result in higher probabilities of risky events for a given location.

The connection between the parameters of the t-Student truncated model, as the

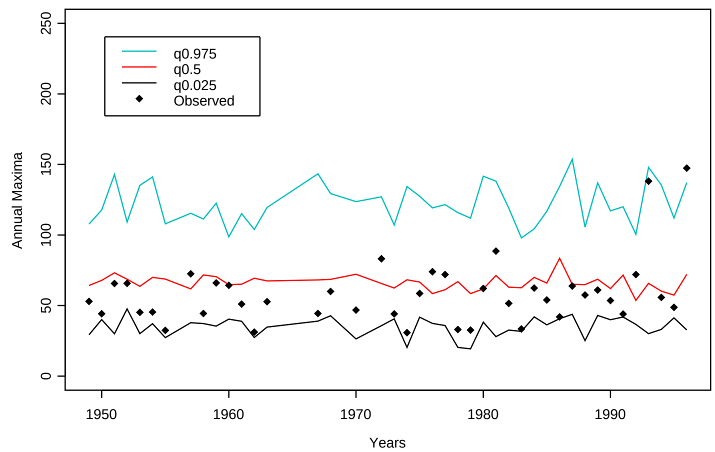

Figure 12: Quantiles from simulations of the predictive distribution at La Mariposa station

parent distribution of the daily rainfall values, and the Fréchet model is also an interesting. It is natural to connect the shape of the extreme value distribution with the tail behavior of the parent distribution. In the case of the TPT TSM model, two parameters are directly associated with the tails characteristics of the TSM: the number of degrees of freedom α and the power transformation parameter β. Nevertheless, empirical evidence shows that α is actually the most relevant one.

The theoretical results have been widely confirmed in this study from two points of view: first by comparing the posterior distribution of the extreme model parameters with the posterior distribution of the daily rainfall model parameters; second by sampling from the posterior predictive distributions of the daily rainfall and fitting the GEV extreme model to the annual maxima simulations. The posterior mean of the shape parameter δ from the Fréchet model was compared with the posterior mean of the ratio α/β resulting from the t-Student fit to the daily rainfall values. These comparisons were in good agreement and do validate the theoretical results presented in section 2. These theoretical results were validated for two important stations in Venezuela: Maiquetía, located in Vargas State and La Mariposa, located in Miranda State.

Further analysis on the extreme rainfall behavior of Venezuelan rainfall for several locations appears to indicate an interesting relationship according to the climatological nature of the daily rainfall values and their natural spatial variability. Further research is being carried out in this direction.

Acknowledgments

The third author was partially funded by the National Science Foundation grant DMS-0417753.

References

Bárdossy, A., Plate, E. J. (1992). Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resources Research*, **28**, $1247 - 1259$.

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics, London.

Coles S. and Pericchi L.R. (2003). Anticipating catastrophes through extreme value modeling. *Journal of the Royal Statistical Society, Series C: Applied Statistics*, **52**, 4, $405 - 416$.

Cox, D. R., Isham, V. S. (1994). Stochastic models of precipitation. *In Statistics for the Environment* 2*: Water Related Issues (eds. V, Barnett and K.F. Turkman), ch* 1*, Chichester: Wiley*. $3 - 18$.

Cowpertwait, P. S. P. (1994). A generalized point process model for rainfall. *Proceeding Royal Society London A*, **447**, $23 - 27$.

Devroye, L. (1985). *Non-uniform Random Variate Generation*. Springer-Verlag, New York.

Dickey, J.M. (1968).Three multidimensional integral identities with Bayesian applications. *Annals Mathematics Statistics*,**39**, $1615 - 1627$.

Gabriel, K. R., Neumann, J. (1962). A Markov Chain model for daily rainfall occurrence at Tel Aviv. *Journal Royal Meteorology Society*. **89**, $90 - 95$.

Galambos (1978). *The Asymptotic Theory of Extreme Order Statistics*. John Wiley and Sons, New York.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4*, eds. J.M. Bernardo, J.O. berger, A.P. Dawid and A.F.M. Smith. Oxford University Press.

Glasbey and Nevison. 1997. In: *Modelling Longitudinal and Spatially Correlated Data: Methods, Applications, and Future Directions* (T.G. Gregoire, D.R. Brillinger, P.J. Diggle, E. Russek-Cohen, W.G. Warren and R.D. Wolfinger, eds.). Lecture Notes in Statistics 122, Springer: New York, 233-242.

Gumbel, E. J. (1958). *S*tatistics of Extremes. Columbia University Press, New York.

Heidelberger, P., Welch, P. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, **31**, $1109 - 1144$.

Hutchinson, M. F., Richardson, C. W., Dyke, P. T (1993). Normalization of rainfall across different time steps. Management of irrigation and drainage systems, $21 - 23$ July 1993, Park City, U. T, Irrigation and Drainage Division ASCE, US Department of Agriculture. $432 - 439$.

Jenkinson, A.F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, **81**, $158 - 171$.

LeCam, L. (1961). A stochastic description of precipitation. *In proc, 4th Berkeley simp. Mathematical Statistics and Probability (ed. J. Newman.), University of California Press. Berkeley*, $165 - 186$.

Liston, G.E., Pielke, R.A. (2000). A climate version of the regional atmospheric modeling system. *Theory of Applied Climatology*, **66**, $29 - 47$.

Phelan, M. J., Goodall, C. R. (1990). An assessment of a generalized Waymire-Gupta- Rodríguez-Iturbe model for GARP Atlantic tropical experiment rainfall. *Journal Geoghysics Research*, **95**, $7603 - 7616$.

Richard, J., Kirtman, B., Laprise, R., von Storch, H. and Wergen, W. (2002). Atmospheric regional climate models (RCMs): A multiple purpose tool? Tech. rep., Report of the joint WGNE/WGCM ad hoc panel on Regional Climate Modelling *http://w3g.gkss.de/G/Mitarbeiter/Storch/pdf/RCM.report.04*

Richardson, C. W. (1977). A model of stochastic structure of daily precipitation over an area. Colo State Univ. Fort Collins Hydro. Pap. **91**

Rodríguez-Iturbe, I., Cox, D. R., Isham, V. S. (1987). Some models for rainfall based on stochastic point processes. *Proceeding Royal Society London A*, **410**, $269-288$.

Rodríguez-Iturbe, I., Cox, D. R., Isham, V. (1988). A point process model for rainfall: further developments. *Proceeding Royal Society London A*, **417**, $283 - 298$.

Sansó, B., Guenni, L. (1999). Venezuelan rainfall data analysis using a Bayesian space-time model. *Applied Statistical Journal of the Royal Statistical Society*, **48**, $345 - 362$.

Sansó, B., Guenni, L. (2000). A non-stationary Multi-site model for rainfall. *Journal of the American Statistical Association*, **95(452)**, $1089 - 1100$.

Stern, R., Coe, R. (1984). A model fitting analysis of rainfall data. *Journal of the Royal Statistical Society, series A*, **147**, $1 - 34$.

Stidd, C.K. (1953). The use of correlation fields in relating precipitation to circulation. *Journal of Meteorology*, **11**, 202 − 213.

Stidd, C. K. (1973). Estimating the precipitation climate. *Water Resources Research*, **9**, 1235 − 1241.

von Mises, R. (1936). La distribution de la plus grande de n valeurs. *Reprinted in Selected Papers Volumen II. American Mathematical Society, Providence, R. I.*, 1954, 271 − 294.

Waymire, E. D., Gupta, V. K., Rodríguez-Iturbe, I. (1984). Spectral theory of rainfall intensity at the meso-β scale. *Water Resources Research*, **20**, 1453 − 1465.

Woolhiser, D. A. (1992). Modelling daily precipitation-progress and problems. *In Statistics in the Environmental and Earth Sciences. New York: Halsted*, 20 − 22.