MODELING THE DISTRIBUTION OF MAXIMUM RAINFALL IN URUGUAY
Juan Kalemkerian, Carolina Ábalo, Silvia Carámbula y Matías Quiñones
Meteorologica, Vol. 48 Nº1 (2023), e019, enero-junio 2023
ISSN 1850-468X — https://doi.org/10.24215/1850468Xe019
http://www.meteorologica.org.ar
Centro Argentino de Meteorólogos
Buenos Aires - Argentina

MODELING THE DISTRIBUTION OF MAXIMUM RAINFALL IN URUGUAY

Juan Kalemkerian[1], Carolina Ábalo[2], Silvia Carámbula[2] y Matías Quiñones[2]

[1]1Centro de Matemática, Facultad de Ciencias, UdelaR
[2]Dirección General de Educación Secundaria, ANEP

ABSTRACT

This paper shows, based on daily records, the modeling of maximum precipitations in each quarter of eighteen meteorological stations located in different parts of Uruguay. We compared the performance of the classic likelihood ratio test with one of the truncated Crámer-von Mises type. Most of the stations did adjust under the Gumbel distribution with few Fréchet and Weibull cases, obtaining a most appropriate truncated Crámer-von Mises test performance. From the adjustment in each of the stations and the combination of three statistical techniques (k-means, Kolomgorov–Smirnov test of equality of distributions and test of independence) we concluded that the maximum rainfall throughout the Uruguayan territory is homogeneous with a slight difference between the southern and northern regions.
Keywords: extreme rainfall, GEV distribution, Gumbel distribution, geostatistics.

MODELACIÓN DE LA DISTRIBUCIÓN DE PRECIPITACIONES MÁXIMAS EN URUGUAY

RESUMEN

En el presente trabajo, a partir de registros diarios, se modelan las precipitaciones máximas en cada trimestre de 18 estaciones meteorológicas ubicadas en distintos puntos de Uruguay. Se comparó la performance del clásico test de la razón de verosimilitud contra uno del tipo de Crámer—von Mises recortado. La mayoría de las estaciones ajustaron según la distribución Gumbel existiendo pocos casos de Fréchet y de Weibull y se obtuvo una performance más apropiada del test de Crámer—von Mises recortado. A partir del ajuste en cada una de las estaciones, combinando tres técnicas estadísticas (k-means, test de igualdad de distribuciones de Kolmogorov—Smirnov y test de independencia) se concluyó que las precipitaciones máximas a lo largo del territorio uruguayo son homogéneas existiendo una leve diferencia entre la región sur y la norte.
Palabras clave: precipitaciones extremas, distribuciones GEV, distribución Gumbel, geoestadística.

Dirección Electrónica: jkalem@cmat.edu.uy

1. INTRODUCTION

The importance of the study of extreme events is well-known in various areas as food production, economics, energy planning among many others. In the particular case of extreme rainfall events, both floods and severe droughts can bring great economic, resource and human losses. Therefore, governments should have precise models to better understand the phenomenon and use it to estimate both the probability of events not yet observed and the probability of return of the ones occurred already. On the one hand, there are several works about extreme precipitation in South America focused in physical and statistical aspects, see for example Bettolli et al (2021), Calvacanti (2012), Calvacanti et al (2015), Carril et al (2016).

On the other hand, its spatial study is also of vital importance since its both occurrence and modeling can radically change from one region to another. For instance, Hernández et al (2011) extreme rainfalls in different locations in Venezuela were modeled using Bayesian methods. In small and geographically homogeneous countries such as Uruguay, it is expected to have no major changes in modeling the different regions although no previous clustering work has been found with the maximums. Some Brazilian papers (Medeiros et al. 2019) presented a modeling for the maximum daily rainfall in the municipality of Jataí, Goiás, adjusted for Gumbel, to estimate the return levels up to 100 years. In some, (Anderson et al. 2020) the maximum rainfall in 12 municipalities in the northeast of Rio Grande do Sul were modeled by Gumbel with the objective of designing hydraulic structures. In others, (Silva et al. 2019) Gumbel models were adjusted to estimate the maximum intensity of the rains. In Argentina, (Vich et al. 2014) the generalized distributions of extreme values were used in order to find the magnitude of the annual flow for return. In the work of Santiñaque et al. (2021), can be found (through spatial clustering techniques applied to the annual maximums recorded in 20 meteorological stations distributed throughout the entire Uruguayan territory) the expected homogeneity among the stations considered with an exception (Mercedes). In this article, we will delve into what it has been already found (Santiñaque et al. 2021) by working with quarterly data, that is, quadruple the information by taking four values corresponding to the maximum in each of the quarters of each year and through a precise modeling of each station in each quarter, apply the classic k-means method to deepen the conclusion at the spatial level obtained in it. Section 2 describes the data which the investigation was carried out with and the objectives it pursues. Section 3 describes the mathematical-statistical methods, including references. Section 4 describes the results gathered with their preliminary conclusions. Last but not least, section 5 describes the fundamental conclusions of the investigation as well as possible line of work to be developed within the statistics field both at a theoretical and practical level.

2. MATERIAL AND METHODS

2.1. Data description and objectives

The main objective of this investigation is to obtain the distribution of the variable defined as the maximum quarterly precipitation from daily recorded in 18 stations located across Uruguay. On the one hand, we will deep dive into Santiñaque et al (2021), founding since we got the information quadrupled, meaning that we contemplated each quarters maximums for each year considered. Taking into account the 18 stations' geographical distribution and each of their adjustments, on the other hand, we will apply k-means clustering to obtain results at spatial level as well. The data set consist of daily rainfall records from January 1st, 1981 to December 31st, 2013 in millimeters, in each of the 18 meteorological stations shown in Figure 1. Data were provided by INUMET (Instituto uruguayo de meteorología): www.inumet.gub.uy.

Figure 1: Geographical distribution of the 18 meteorological stations considered in this work The map was obtained from Google Maps.

Each year was split into four quarters as follows: from January 1st to March 31st (quarter 1), from April 1st to June 30th (quarter 2), from July 1st to September 30th (quarter 3) and from 1st October to December 31st (quarter 4). Due to the goal is modeling the quarterly maximums, only four values were considered per year: the maximum values of each the quarters, discarding all the rest of the data. Figure 1 shows the geographic distribution of the 18 stations across Uruguay.

2.2. Estimation of the distribution of the quarterly maximums in each station

If $D_1, D_2, \ldots, D_n$ are n independent and identically distributed (i.i.d.) observations of certain variable D, the Fisher—Tippett theorem (Fisher and Tippett, 1928; Gnedenko, 1943) assures that as n grows, $M_n = \max\{D_1, D_2, \ldots, D_n\}$ approximates to a Gumbel, Fréchet or Weibull distribution defined as $H_1(x;\mu;\sigma) = e^{-(\mu-x)/\sigma}$ where $\sigma > 0$, $H_2(x;\mu;\sigma;\xi) = e^{-(\frac{x-\mu}{\sigma})^{-1/\xi}}$ where $x < \mu$, σ, $\xi > 0$, and $H_3(x;\mu;\sigma;\xi) = e^{-(\frac{\mu-x}{\sigma})^{-1/\xi}}$ where $x < \mu$, $\sigma > 0$, $\xi > 0$ respectively. The three distributions' family can be expressed in a single formula given by $H(x;\mu;\sigma;\xi) = e^{-1+(\frac{\xi(x-\mu)}{\sigma})^{-1/\xi}}$ where $\sigma > 0$, and where $x > \mu - \sigma/\xi$, for the $\xi > 0$ case, or $x < \mu - \sigma/\xi$ for the $\xi < 0$ case. H is Fréchet when $\xi > 0$, Weibull when $\xi < 0$, and if $\xi \to 0$, H tends to a Gumbel distribution. μ is called the location parameter, σ the scale parameter and ξ the shape parameter. H is called Generalized Extreme Value Distribution (GEV) and was proposed by Jenkinson (1955) and Von Mises (1936). Considering D_i as the accumulated precipitation on day i, in Santiñaque (2020) the adjustment was applied for the same set of annual maximum data, this means n = 365, providing the adjustment was accurate. In our work, we will apply the theorem for n = 90 since we will work with the maximums in each quarter. Simultaneously, we also worked with semester data (n = 183). Even though these values of n are notoriously lower than the ones used for annual maximums, we can fortunately prove that the theorem still gives us good results. Assuming that the values at each station follow a GEV distribution, the parameter estimation was carried out by applying the weighted moment method (Greenwood et al, 1979) (method specially designed for the study of extreme values) and the maximum likelihood giving similar results. The calculations were made using R's "extRemes" package, as well as the confidence intervals for them.

2.3. Model diagnosis

Once the GEV parameters were estimated for each station, the model was validated using the diagnostic graphs. The diagnostic graphs are a visual tool made up of four graphs where the adjusted distribution (GEV) is compared with the empirical one of the data observed through different measures. The first graph is the so-called PP-plot (represents the values of the adjusted cumulative distribution (GEV) versus the empirical one at different points); the closer to the diagonal, the better the fit of the model. The second graph is the so-called QQ-plot, which represents the quantile function of the adjusted GEV distribution versus the empirical quantile. Again the closer to the diagonal the points of this graph are seen, the better the model is. The third graph shows the empirical density versus the density of the fitted one. In this case, the more similar are the graphs one another, the better the fit. The fourth graph compares the return levels estimated by the adjusted GEV model with its confidence bands. If the values are within these bands, the fit is good. The closer the values to the straight line, the closer the distribution is to the Gumbel, if the points are drawn above (below) the diagonal using a convex (concave) graph, the more the distribution resembles a Fréchet (Weibull). Coles et al (2001) gives a more detailed explanation of the diagnostic graphics while Santiñaque (2020) only gives a synthesis of them. To have a more precise technique diagnostic model, two goodness-of-fit hypothesis tests were applied to the Gumbel distribution, which are the likelihood ratio test (LR) and the truncated Cramér —Von Mises test (TCVM). In this second case, when the Gumbel distribution hypothesis was rejected, the test was performed taking the Fréchet distribution (when the shape parameter estimate was positive) as the null hypothesis, or the Weibull distribution (when the shape parameter estimate was negative). TCVM is a test of the Crámer-von Mises type which truncates the integration region using a similar idea to the one applied in Kalemkerian (2019). Here, $H_0: X^{(i)} \equiv \text{Gumbel}(\mu,\sigma)$ it is posed versus H_1:H_0 does not hold, where $X^{(i)}$ is the maximum precipitation in the i station. If H_0 is rejected, the test is adapted to consider H_0:$X^{(i)} \equiv$ Fréchet(μ,σ,ξ) when the estimation of the shape parameter is positive or H_0:$X^{(i)} \equiv$ Weibull(μ,σ,ξ) when the estimation of the shape parameter is negative. In Santiñaque (2020) this adaptation it is explained in detail.

2.4. Clustering of estimated parameters

Once it was obtained a good fit in each of the stations, quarters and semesters, the k-means methodology was applied using the estimated parameters as indicators of the distribution. As it is well known, it is necessary to select the number of groups to apply k-means. In order to find the number of groups to be separated, it was calculated the Silhouette coefficient proposed in Rousseeuw (1986). This coefficient splits into k groups and calculates how well the elements are classified in the k groups, it takes values between -1 and 1 and the higher it is the coefficient, the better its elements are classified. This means that the highest k value the Silhouette coefficient takes it will the one suggested for applying clustering.

2.5. Kolmogorov—Smirnov test for equality of distributions

The classic Kolmogorov-Smirnov test was applied to test the equality or difference between the distributions of the maximum in the different stations. It is more explicitly stated H_0:$X^{(i)}$,$X^{(j)}$ have the same distribution versus H_1:H_0 does not hold, where $X^{(i)}$,$X^{(j)}$ are the maximum precipitations in the stations i and j respectively.

2.6. Independence test based on recurrence rates

Regarding the existence of associations or dependencies between the observations corresponding to the data observed in the stations, it was applied the recently proposed independence test based on recurrence percentages (Kalemkerian and Fernández, 2020a). This test aims to investigate if two variables X and Y are independent in a probabilistic sense. Then, starting from $(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$ sample of (X,Y) where X and Y can take values in any metric space (for example), we stated that H_0: X and Y are independent versus H_1:H_0 does not hold. We used this test where X and Y are the maximum values of all the pairs of stations considered in this work.

The theoretical details of the test are developed in Kalemkerian and Fernández (2020a) as well as its implementation and application to economic and meteorological data in Kalemkerian and Fernández (2020b).

3. RESULTS AND DISCUSSION

3.1. Estimation of the distribution parameters

Figures 2 and Figure 3 show the point estimates together with their 95 % confidence intervals for the parameters μ and σ respectively. Recall that μ and σ are not the mean and the deviation of a GEV distribution, but are called the location and scale parameters of the GEV distribution. In this investigation we are interested in the comparison between the distributions in each station. Except for Rocha station, a small difference can be observed between the stations in the south of the country (the 5 stations to the left of the graphs). Similarly, a small difference can be observed between the northern stations (the 4 stations to the right of the graphs). The differences are a little clearer with respect to the parameter μ than with respect to σ. Figure 4 and Figure 5 show the estimates of the shape parameter (ξ) for the 18 stations in each of the quarters and semesters respectively. It is observed that almost all the 95 % confidence intervals includes the zero value, so it is to be expected that most of the stations have a good fit to the Gumbel distribution, as will be seen in the next subsection. In addition to the comparison of the behavior of different stations,

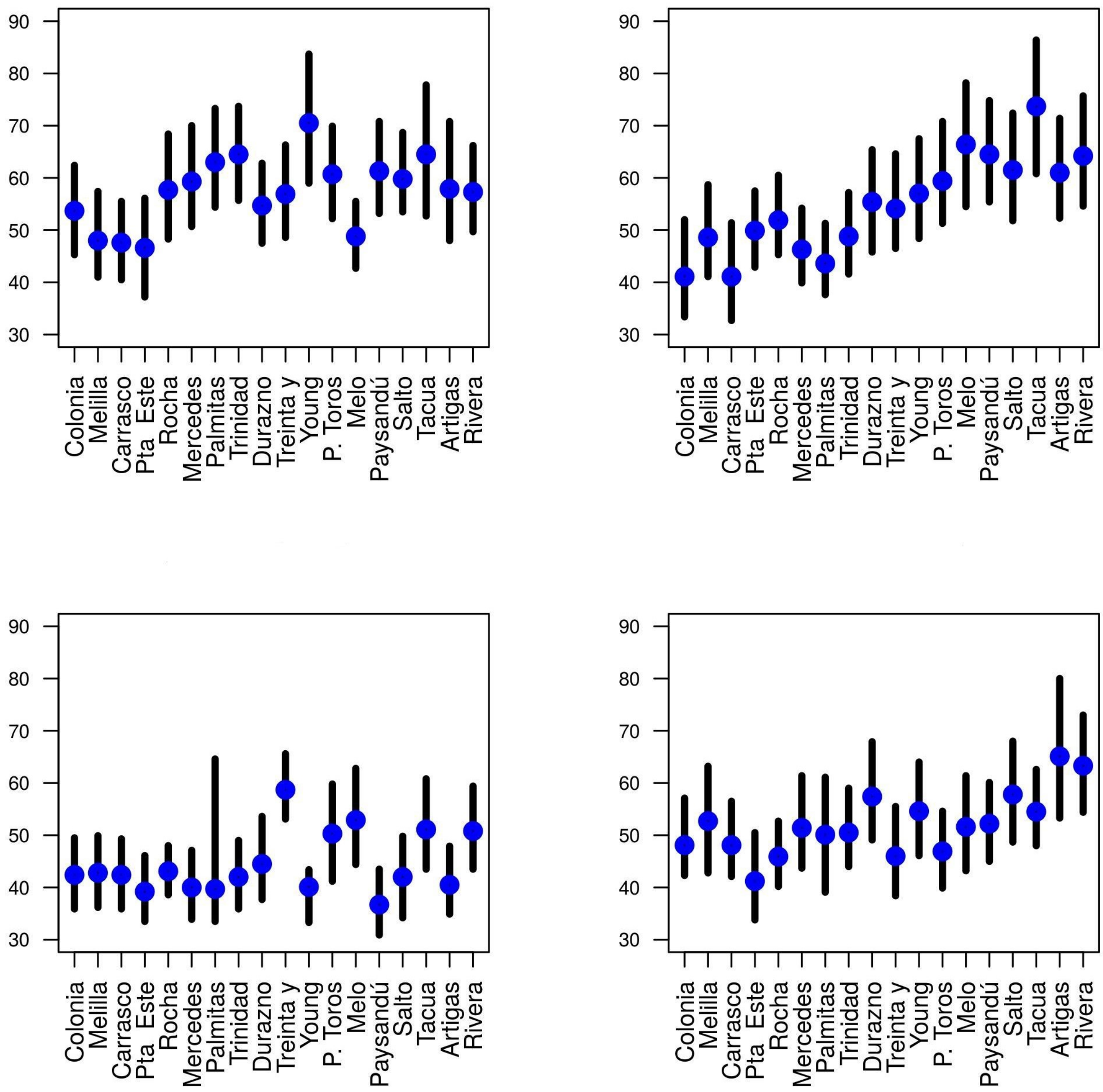

Figure 2: Estimation of the localization parameter (μ) in blue and confidence intervals at 95 % for each one of the stations. Quarter 1 (top left), quarter 2 (top right), quarter 3 (bottom left) and quarter 4 (bottom right).

figures 2 to 4 show that the extreme rainfalls are greater in quarters 2 and 4 than in quarters 1 and 3.

3.2. Model diagnosis and goodness of fit

Both quarterly and semi-annually, the adjustment obtained in the 18 stations through the diagnostic graphs was good, so it can be deduced that the applicability of the Fisher-Tippett theorem even for moderate values such as those of the data set worked ($n = 90$) continues to lead to good results. As an example, Figure 6 shows the four diagnostic charts for the Colonia station in the second quarter. As can be seen from Figure 4 and Figure 6, it is reasonable to test the Gumbel distribution hypothesis for each of the stations. In most cases, the TCVM and LR goodness-of-fit tests led to the same conclusion about the distribution of the different stations. When both tests led to different conclusions, in

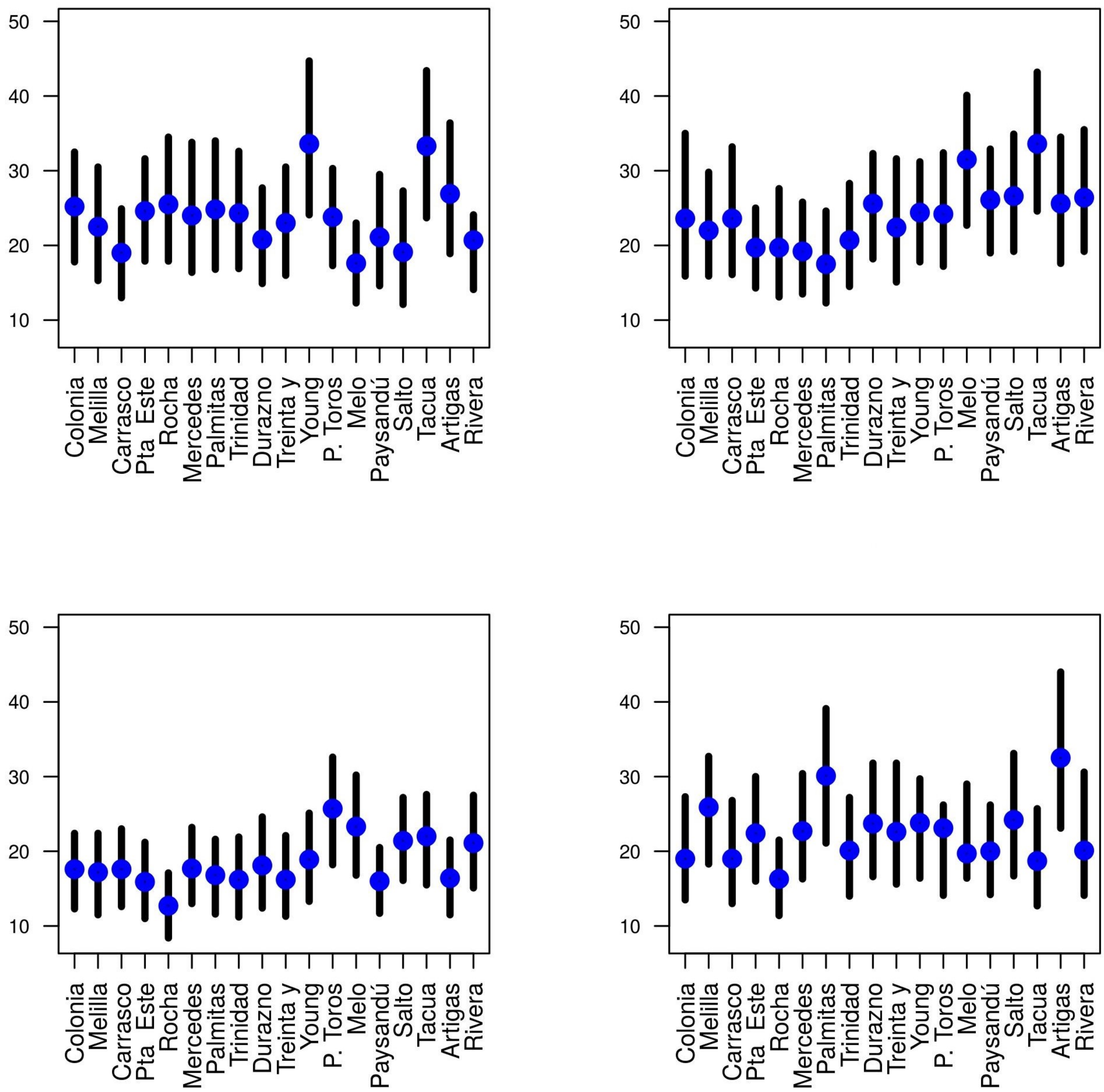

Figure 3: Point and interval estimation at 95 % for the scale parameter (σ) in blue for each station. Quarter 1 (top left), quarter 2 (top right), quarter 3 (bottom left) and quarter 4 (bottom right).

general TCVM seem to performed better, at least in the sense that your results looks more suitable with the results showed in Figure 4 and Figure 6 than the results obtained by the LR test. In particular at the Young and Melo, the estimated value of the shape parameter is far from zero, so it is to be expected that the Gumbel distribution hypothesis test will be rejected. This fact was detected by TCVM test but not by LR as shown in Table 1. Similarly, it can be seen that TCVM seem to perform better than LR at least in the following cases: Colonia (second quarter), Rocha (first semester) and Salto (third quarter). The only case of difference between the TCVM and LR test decision where LR apparently better detects behavior is at the Trinidad station in the third quarter. Table 1 includes for each quarter and semester the distribution of each of the stations according to the joint application of the TCVM test for both Gumbel and Fréchet and Weibull.

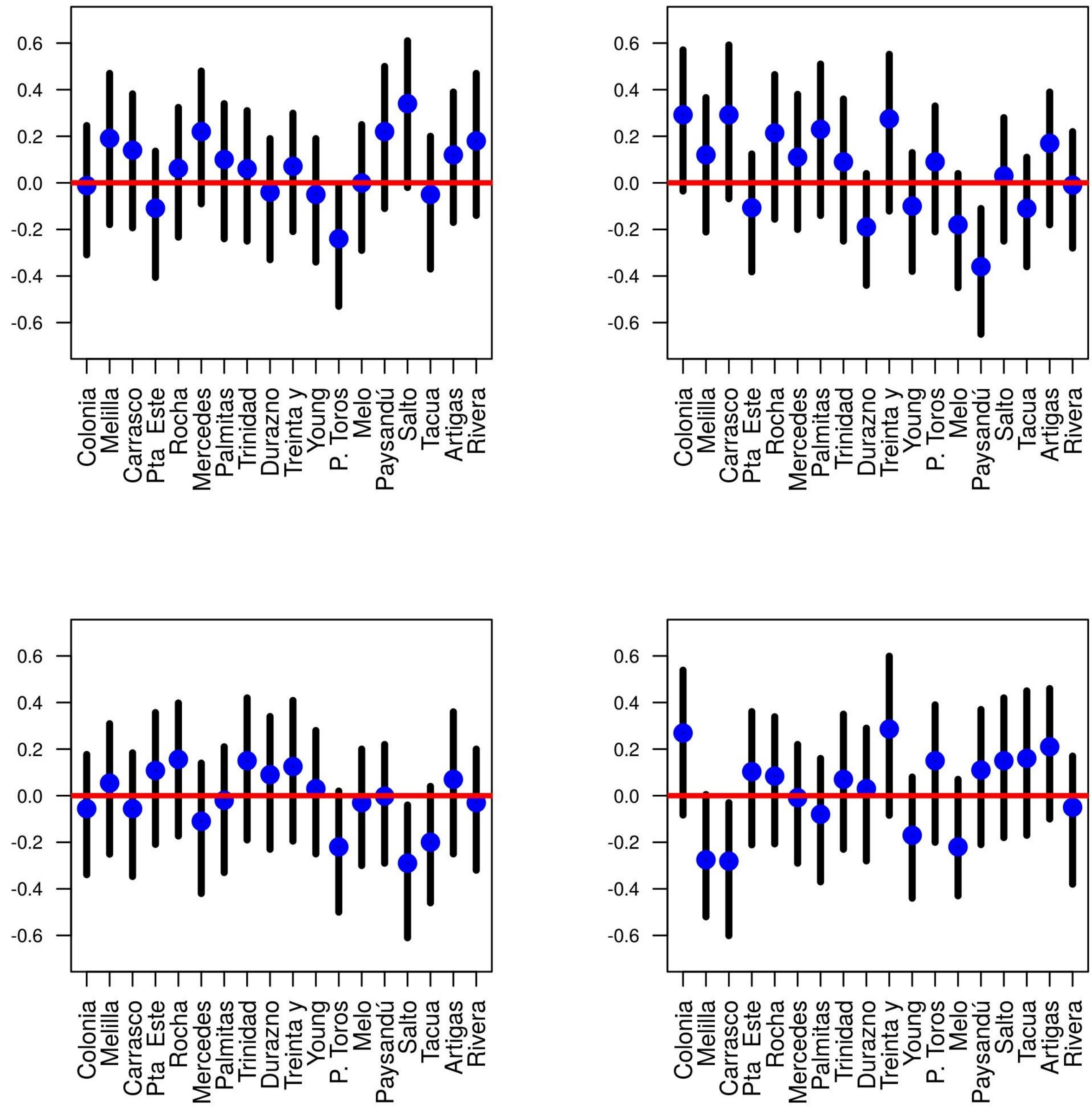

Figure 4: Point and interval estimation at 95 % for ξ. Quarter 1 (top left), quarter 2 (top right), quarter 3 (bottom left) and quarter 4 (bottom right). The red line helps to see the position between the estimation of ξ with respect to zero (Gumbel distribution).

It appears from Table 1 that in the vast majority of cases, there was a good fit to the Gumbel distribution with a few specific cases of Fréchet or Weibull distributions. It is noteworthy that Paysandú is the only station where the three types of distributions (Fréchet, Gumbel and Weibull) were correctly adjusted.

3.3. Clustering of estimated parameters

According to Kaufman (1990), when the Silhouette coefficient takes values between 0.25 and 0.50, it is interpreted as the weak group structure. For both semester data and quarterly, the Silhouette coefficient showed very little heterogeneity in the data. Except in the fourth quarter, the coefficient obtained its maximum for k = 2 groups. In quarter 2, we observed that

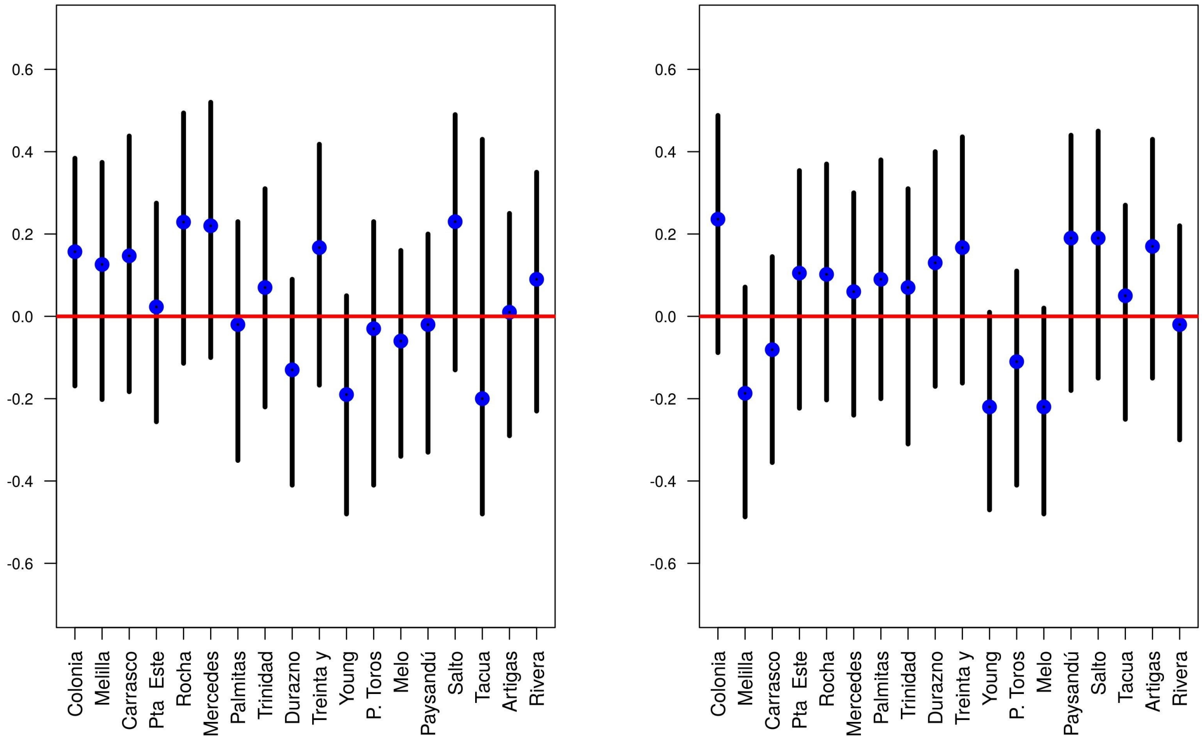

Figure 5: Point and interval estimation at 95 % for ξ for each semester. The red line helps to see the position between the estimation of ξ with respect to zero (Gumbel distribution). Semester 1 (left), semester 2 (right).

the values for k = 7 and k = 8 are slightly higher than the k = 2 case. Anyway for 18 stations and values of the Silhouette coefficient less than 0.5 it is more reasonable to work with k = 2 groups. Figure 7 shows the graph of the Silhouette coefficient for different values of k varying between 2 and 8 groups and for each of the quarters. Table 2 shows the values obtained separating k = 2 groups. Separated into two groups by k-means in quarters 1, 2 and 3 and three groups in quarter 4, below we give the conformation of each of the groups according to quarter or semester.

Quarter 1.

Group 1: Colonia, Melilla, Carrasco, Punta del Este, Durazno, Melo, Paso de los Toros.

Group 2: Rocha, Palmitas, Trinidad, Young, Tacuarembó, Artigas, Mercedes, Treinta y tres, Paysandú, Salto, Rivera.

Quarter 2.

Group 1: Colonia, Melilla, Carrasco, Punta del Este, Rocha, Mercedes, Trinidad, Palmitas, Treinta y tres.

Group 2: Durazno, Melo, Paso de los Toros, Young, Paysandú, Salto, Tacuarembó, Artigas, Rivera.

Quarter 3.

Group 1: Colonia, Melilla, Carrasco, Punta del Este, Rocha, Mercedes, Palmitas, Trinidad, Durazno, Paysandú, Treinta y Tres, Young, Artigas.

Group 2: Paso de los Toros, Melo, Salto, Tacuarembó, Rivera.

Quarter 4.

Group 1: Melilla, Carrasco, Mercedes, Palmitas, Young, Melo.

Group 2: Durazno, Salto, Artigas, Rivera.

Group 3: Colonia, Punta del Este, Rocha, Trinidad, Treinta y Tres, Paso de los Toros, Paysandú, Tacuarembó.

Semester 1.

Group 1: Colonia, Melilla, Carrasco, Punta del Este, Rocha, Durazno, Melo, Paso de los Toros, Palmitas, Trinidad, Mercedes, Treinta y tres,

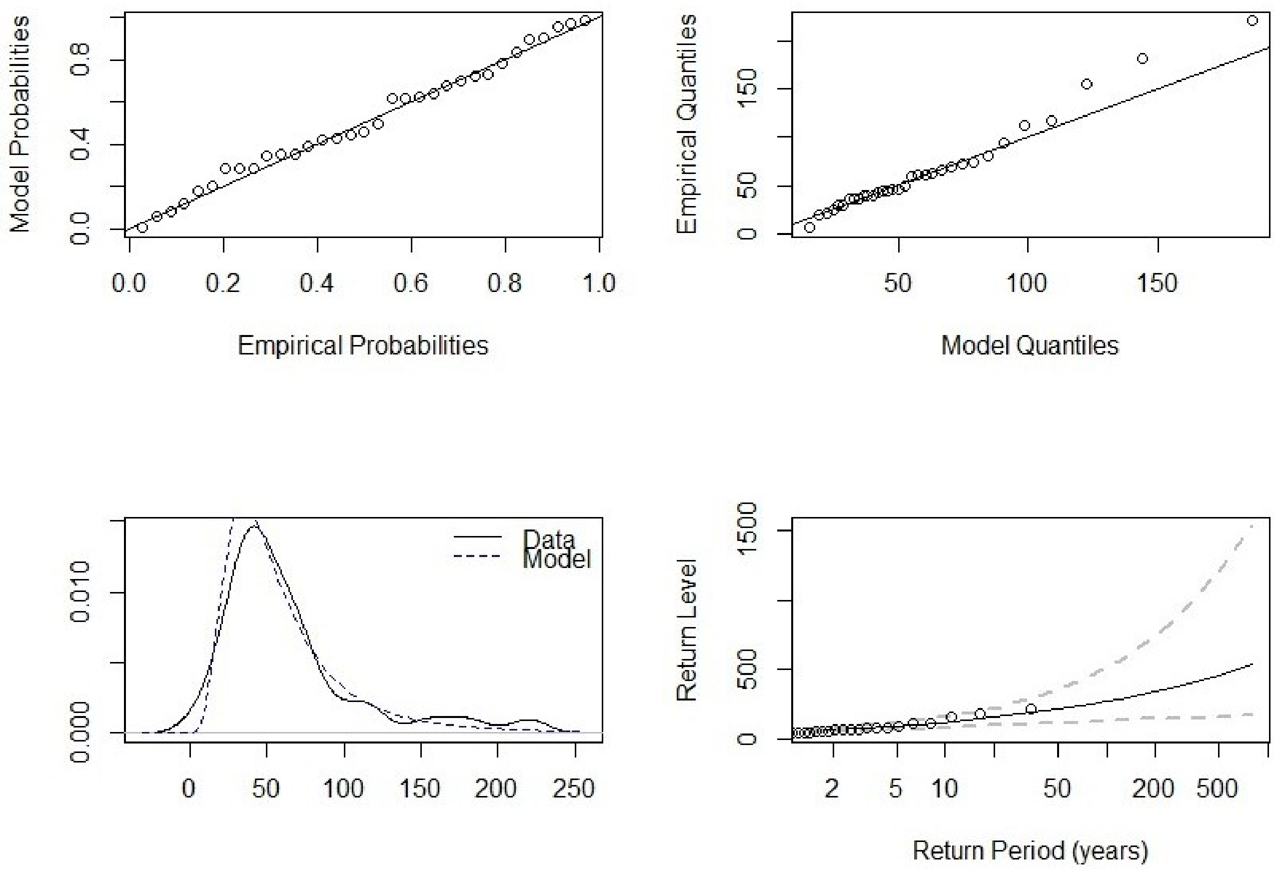

Figure 6: Diagnosis plots for Colonia station in the second quarter. pp-plot (top left), qq-plot (top right), empirical and model densities (bottom left) and return level plot (bottom right).

Paysandú, Salto, Rivera.
Group 2: Young, Tacuarembó, Artigas.
Semester 2.
Group 1: Colonia, Punta del Este, Rocha, Palmitas, Paysandú, Salto, Mercedes.
Group 2: Melilla, Carrasco, Trinidad, Durazno, Treinta y Tres, Young, Paso de los Toros, Melo, Tacuarembó, Artigas, Rivera.

It is observed that the southernmost stations of Uruguay (Colonia, Melilla, Carrasco, Punta del Este and Rocha) are in the same group in quarters 1, 2 and 3 (except for Rocha in quarter 3). In Figure 8 it is shown that separating in k = 2 groups for quarters 1 to 3 and k = 3 groups for quarter 4, k-means works well. On the other hand, if we consider the easternmost stations in Uruguay (Punta del Este, Rocha, Melo and Treinta y Tres) and the westernmost stations (Colonia, Mercedes, Palmitas, Young, Paysandú and Salto) it is observed that they are mixed in different groups in each quarter.

3.4. Comparison between distributions

The application of the Kolmogorov-Smirnov test for equality of distributions (applied in pairs at two stations) in most cases did not reject the hypothesis of equality of distributions. As an example, Table 3 shows the results corresponding to the fourth quarter that among the stations further south with respect to the stations further north. For example in row 1 we show the p-value to the test between Colonia station and each of the other and in the final column we show the p-value to the test between Artigas station and each of the other. In most cases rejects the equality of distributions at

	Quarter 1			Quarter 2			Quarter 3			Quarter 4			Semester 1			Semester 2		
	TCVM	LR	D	TCVM	LR	D	TCVM	LR	D	TCVM	LR	D	TCVM	LR	D	TCVM	LR	D
Colonia	**0.499**	**0.875**	G	0.037	**0.058**	F	**0.342**	**0.708**	G	**0.125**	**0.063**	G	**0.519**	**0.237**	G	**0.273**	0.028	G
Melilla	**0.304**	0.047	G	**0.278**	**0.734**	G	**0.990**	**0.907**	G	**0.069**	**0.064**	G	**0.731**	**0.250**	G	**0.361**	**0.210**	G
Carrasco	**0.616**	**0.210**	G	**0.443**	**0.889**	G	**0.412**	**0.470**	G	**0.135**	**0.070**	G	**0.725**	**0.267**	G	**0.606**	**0.696**	G
Punta	**0.618**	**0.406**	G	**0.785**	**0.150**	G	**0.499**	**0.288**	G	**0.424**	**0.965**	G	**0.303**	**0.549**	G	**0.730**	**0.801**	G
Rocha	**0.121**	**0.376**	G	**0.271**	**0.098**	G	**0.657**	**0.181**	G	**0.750**	**0.562**	G	0.023	**0.278**	F	**0.713**	**0.515**	G
Mercedes	0.016	0.005	F	**0.135**	**1.000**	G	**0.830**	**1.000**	G	**0.493**	**1.000**	G	0.013	0.042	F	**0.598**	**0.681**	G
Trinidad	**0.933**	**0.549**	G	**0.250**	**0.695**	G	0.042	**0.951**	F	**0.250**	**0.604**	G	**0.474**	**0.495**	G	**0.131**	**0.851**	G
Young	**0.683**	**1.000**	G	**0.393**	**0.309**	G	**0.891**	**0.727**	G	**0.116**	**0.389**	G	**0.184**	**0.291**	G	0.034	**0.340**	F
Palmitas	**0.717**	**0.295**	G	**0.077**	**0.104**	G	**0.347**	**0.845**	G	**0.632**	**0.409**	G	**0.613**	**0.985**	G	**0.476**	**0.952**	G
Durazno	**0.846**	**0.693**	G	**0.217**	**0.167**	G	**0.712**	**0.540**	G	**0.639**	**0.956**	G	**0.269**	**0.376**	G	**0.509**	**0.282**	G
Treinta	**0.602**	**0.765**	G	**0.056**	0.025	G	**0.559**	**0.425**	G	0.035	0.012	F	**0.279**	**0.245**	G	**0.442**	**0.215**	G
P. Toros	**0.196**	**0.103**	G	**0.501**	**0.689**	G	**0.061**	**0.178**	G	**0.701**	**0.150**	G	**0.931**	**0.580**	G	**0.291**	**0.566**	W
Melo	**0.947**	**0.547**	G	**0.196**	**0.254**	G	**0.836**	**0.953**	G	0.042	**0.197**	G	**0.650**	**0.868**	G	0.034	**0.331**	W
Paysandú	0.036	0.004	F	0.041	0.007	w	**0.505**	**0.765**	G	**0.345**	**0.374**	G	**0.341**	**0.819**	G	0.049	**0.141**	F
Salto	0.007	0.014	F	**0.480**	**0.867**	G	0.035	**0.065**	w	**0.487**	**0.195**	G	**0.151**	**0.029**	G	**0.233**	**0.148**	G
Tacuarembó	**0.194**	**0.876**	G	**0.699**	**0.484**	G	**0.126**	**0.199**	G	**0.287**	**0.275**	G	**0.440**	**0.082**	G	**0.534**	**0.940**	G
Rivera	**0.442**	**0.199**	G	**0.721**	**0.907**	G	**0.755**	**0.933**	G	**0.624**	**0.735**	G	**0.440**	**0.844**	G	**0.726**	**0.913**	G
Artigas	**0.845**	**0.229**	G	**0.111**	**0.240**	G	**0.853**	**0.451**	G	**0.209**	**0.077**	G	**0.297**	**0.562**	G	**0.340**	**0.152**	G

Table I: p-value for the TCVM and LR tests. Column "D" means adjusted distribution according to TCVM test at 5 %: G (Gumbel), F (Fréchet), W (Weibull). In bold the p-values greater than 0.05.

Quarter 1	Quarter 2	Quarter 3	Quarter 4	Semester 1	Semester 2
0.3117	0.3138	0.3520	0.3014	0.3072	0.3019

Table II: Mean value of the Silhouette coefficient for each one of the different semesters and quarters separating in k = 2 groups.

10 %. Similar results were obtained in the other quarters. In turn, taking two stations from the south or two stations from the north, the null hypothesis of equality of distributions is not rejected.

The results obtained through this test are consistent with what was informally expressed in subsection 3.1 from the visual inspection of figures 2 to 4, where small differences are seen in the estimates of the different stations, but this test gives us a tool more precise with respect to the equality or not of the distribution of the different stations. On the other hand, the results reported in Table 3 are in line with the estimates of μ shown in Figure 2.

3.5. Independence test based on recurrence rates

The application of the independence test confirmed the expected dependence between values corresponding to geographically close stations. For example, at the level of 10 %, the independence is rejected between Melilla (X) and Carrasco (Y) in quarter 1 (p-value =

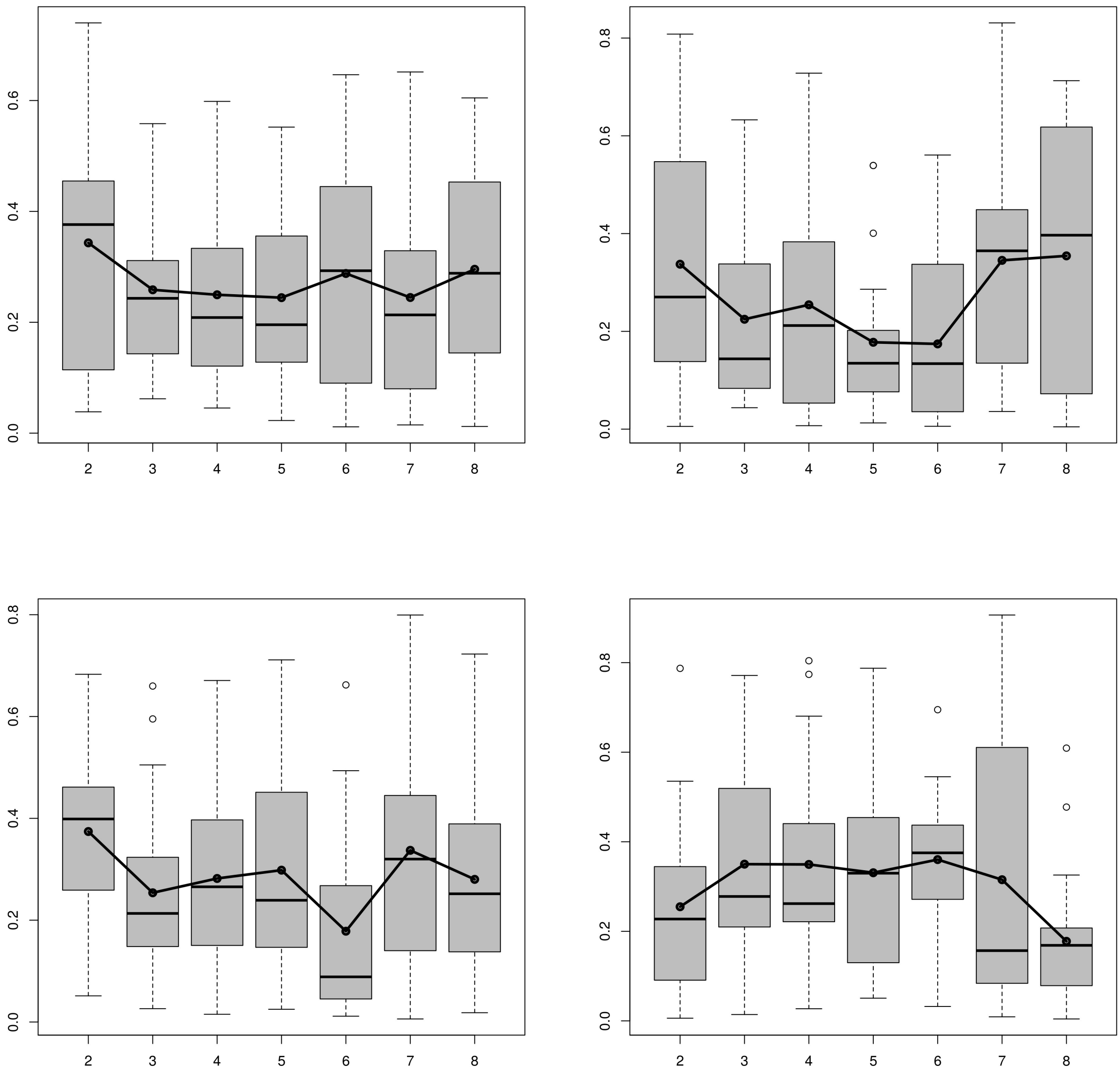

Figure 7: Silhouette coefficient from k = 2 groups to k = 8 groups. Left to right and up to down quarter 1, quarter 2, quarter 3 and quarter 4.

0) or between Rivera (X) and Artigas (Y) in quarter 1 (p-value = 0.029). In general terms and in agreement with what was observed in the clustering section, it was observed that the maximum values observed in the 5 southernmost stations were independent of the maximums observed in the 4 northernmost stations. Table 4 shows the decisions made by the independence test between the vectors X = (Colonia, Melilla, Carrasco, Punta del Este, Rocha) and Y = (Salto, Tacuarembó, Rivera, Artigas) in each of the quarters and semesters.

It is known that in Uruguay it rains more in quarters 1 to 3 in the north than in the south, see the annual accumulate rainfall in Uruguay given in Figure 1, this fact is reflected in terms of extreme rainfall too, according to the results shown in Table 4.

Finally, Table 5 shows the decision resulting from the application of the independence test

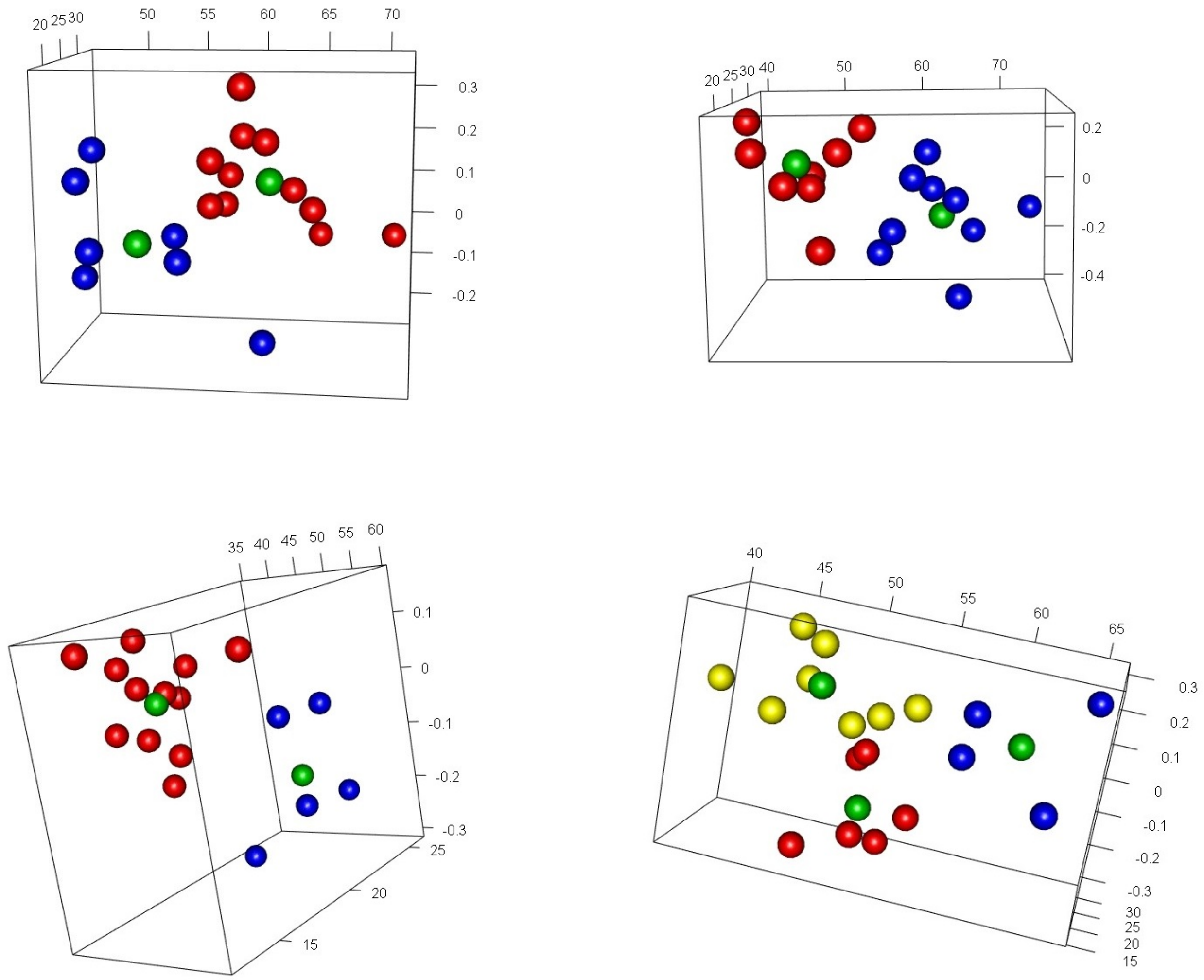

Figure 8: Graph of the 18 triples (μ, σ, ξ) in each quarter separated into 2 groups (quarters 1, 2 and 3) and three groups in quarter 4. In red those belonging to group 1, in blue those belonging to group 2, in yellow group 3) and in green the centroid of each cluster. Quarter 1 (top left), Quarter 2 (top right), Quarter 3 (bottom left), and Quarter 4 (bottom right).

between both groups separated through k-means for each of the quarters and semesters.

As seen in Table 5, except for quarter 4 and semester 2 in the other cases, the hypothesis of independence between the groups is not rejected. The explanation in the case of quarter 4 (where the groups give dependents), is due to the fact that Carrasco is in group 1 while the very close Melilla station is in group 2, with Carrasco and Melilla being two stations very close between them. The nearby stations are highly dependent. In semester 2, something similar occurs between the Salto station (which belongs to group 1) and Tacuarembó station (which belongs to group 2).

Summarizing, by combining these three statistical tools, and concerning to maximum rainfall in each quarter, small difference were found between south and north but not between east and west. This result can be interesting because it is well-known that in winter the accumulated rainfall distribution gradient is

	Melilla	Carrasco	Pta Este	Rocha	Salto	Tacua	Rivera	Artigas
Colonia	NR(0.843)	NR(0.640)	NR(0.448)	NR(0.843)	R(0.096)	NR(0.172)	R(0.025)	R(0.096)
Melilla		NR(0.843)	NR(0.286)	NR(0.286)	NR(0.843)	NR(0.645)	NR(0.448)	NR(0.172)
Carrasco			NR(0.287)	NR(0.287)	NR(0.480)	NR(0.843)	NR(0.172)	R(0.025)
Pta Este				NR(0.843)	R(0.051)	R(0.096)	R(0.012)	R(0.005)
Rocha					R(0.025)	R(0.051)	R(0.005)	R(0.005)
Salto						NR(0.645)	NR0.843)	NR(0.453)
Tacua							NR(0.172)	R(0.025)
Rivera								NR(0.646)

Table III: Application of the Kolmogorov-Smirnov test to pairs of stations for data from quarter 4, at the significance level of 10 %. ”NR”means that the null hypothesis of equality of distributions is not rejected, while R”means that we reject the null hypothesis. In parentheses the p-value of the test.

Quarter 1	Quarter 2	Quarter 3	Quarter 4	Semester 1	Semester 2
NR (0.537)	NR (0.651)	R (0.041)	NR (0.519)	NR (0.102)	NR (0.573)

Table IV: Decision at 10 % based on the independence test between the southern and northern areas: X = (Colonia, Melilla, Carrasco, Punta del Este, Rocha) and Y = (Salto, Tacuarembó, Rivera, Artigas). ”NR”means that the null hypothesis of independence between X and Y is not rejected, while R”means that we reject the null hypothesis. The p-value of each test is included in parentheses.

Quarter 1	Quarter 2	Quarter 3	Quarter 4	Semester 1	Semester 2
NR (0.287)	NR (0.393)	NR (0.268)	R (0.000)	NR (0.640)	R (0.008)

Table V: Decision at 10 % from the independence test between group 1 (X) and group 2 (Y). ”NR”means that the null hypothesis of independence between X and Y is not rejected, while R”means that we reject the null hypothesis. The p-value of each test is included in parentheses.

west- east and south-north in the rest of the seasons. This is not reflected (according to the results we have obtained) when we work with maximum rainfall.

4. CONCLUSIONS

In this investigation, the distribution of the maximum rainfall in each quarter was obtained for each one of the 18 meteorological stations distributed throughout the entire Uruguayan territory. The vast majority had a good fit to the Gumbel distribution and in a few cases Fréchet or Weibull. Taking advantage of the geographical location of the different stations, this information was used to draw conclusions at the spatial level. From the adjusted distributions, combining three statistical techniques, clustering applying k-means, test of independence and the test of equality of distributions, it was obtained as a fundamental conclusion that the behavior of the maximum rainfall at the quarterly level is homogeneous throughout the entire Uruguayan territory with slightly differences between the southern and northern stations, which suggests a separation (although not clearly marked) between two regions, one corresponding to the southern region and the other to the northern region. Also, differences between the east and west are not observed. Another important conclusion of the work is from the statistical point of view, is that in general TCVM seem to performed better than the results obtained by the LR test. Given that the TCVM applied is an intuitive adaptation of the one proposed for the normal distribution in Kalemkerian (2019), as future work the theoretical development of this tool applied to the Gumbel distribution would be of interest, as well as the comparison with other tests related to the Gumbel for other data sets.

REFERENCES

Anderson, T., Corso, C., Mendez, L. A., Gois Barboza, A. and Schneider, V. E., 2020: Determinação de equações de chuvas intensas para a encosta superior do nordeste do Rio Grande do Sul. Ciência e Natura; 42.

Bettolli, M.L., Solman, S.A., Da Rocha, R.P., Llopart, M., Gutiérrez, J.M., Fernández, J., Olmo, M.E., Lavin-Gullon, A., Chou, S.C., Rodrigues, D. Carneiro, Coppola, E., Balmaceda Huarte, R.; Barreiro, M.; Blázquez, J.; Doyle, M., Feijoó, M., Huth, R., Machado, L., Cuadra, S. Vianna. The CORDEX Flagship Pilot Study in southeastern South America: a comparative study of statistical and dynamical downscaling models in simulating daily extreme precipitation events. Climate Dynamics.

Carril, A.F., Cavalcanti, IFA., Menendez, C.G., Sörensson, A., López-Franca, N., Rivera, J.A., Robledo, F., Zaninelli, P.G., Ambrizzi, T., Penalba, O.C., da Rocha, R.P., Sánchez, E., Bettolli, M.L., Pessacg, N., Renom, M., Ruscica, R., Solman, S., Tencer, B., Grimm, A.M., Rusticucci, M., Cherchi, A., Tedeschi R., Zamboni, L. (2016). Extreme events in the La Plata basin: a retrospective analysis of what we have learned during CLARIS-LPB project. Clim Res 68(2–3):95–116.

Cavalcanti, IFA. (2012). Large scale and synoptic features associated with extreme precipitation over South America: a review and case studies for the frst decade of the 21st century. Atmos Res 118:27–40. https://doi.org/10.1016/j.atmosres.2012.06.012

Cavalcanti, IFA, Carril, A.F., Penalba, O.C., Grimm, A.M., Menéndez, C.G., Sanchez, E., Cherchi, A., Sörensson, A., Robledo, F., Rivera, J., Pántano, V., Bettolli, M.L., Zaninelli, P., Zamboni, L, Tedeschi, RG, Dominguez, M, Ruscica, R, Flach, R (2015) Precipitation extremes over La Plata Basin—review and new results from observations and climate simulations. J Hydrol. 23:211–230. https://doi.org/10.1016/j.jhydrol.2015.01.028.

Coles, S., Bawa, J.; Trenner, L. and Dorazio, P., 2001: An Introduction to statistical modeling of extreme values, Springer, vol 208.

Fisher, R. A. and Tippett, L. H. C., 1928: Limiting forms of the frequency distribution of the largest or smallest member of a sample, Mathematical Proceedings of the Cambridge Philosophical Society 180-190.

Gnedenko, B., 1943: Sur la distribution limite du terme maximum d'une serie statistical techniques, Springer (2008). aleatoire, Annals of mathematics, 423-453.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C. and Wallis, J. R., 1979: Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form, Water resources research 15-5, 1049-1054.

Hernández, A., Guenni, and Sansó, B., 2011: Características de la precipitación extrema en algunas localidades de Venezuela. Interciencia (3) 185–191.

Jenkinson, A. F., 1955: The frequency distribution of the annual máximum (or mínimum) values of meteorological elements. Quarterly Journal of the Royal Meteorological Society, 348 (81) 158-171.

Kalemkerian, J., 2019: A truncated Cramér–von Mises test of normality, Communications in Statistics-Theory and Methods, 48:16, 3956-3975.

Kalemkerian, J. and Fernández, D., 2020: An Independence Test Based on Recurrence Rates. Journal of multivariate analysis, v 178 (C), 104624.

Kalemkerian, J. and Fernández, D., 2022: An Independence test based on recurrence rates. An empirical study and applications to real data. Communications in Statistics-Simulation and Computation: 1-21. https://doi.org/10.1080/03610918.2022.2037637.

Kaufman, P. R. L., 1990: Finding groups in Data: An Introduction to Cluster Analysis. Wiley.

Medeiros, E. S., Alvez, M. A. and Souza, S. A., 2019: Estimação de nível de retorno da precipitação máxima diária no município de Jataí, Goiás. Ciência e Natura; 41:1-08.

Package "extremes". https://cran.r-project.org/web/packages/extRemes/extRemes.pdf.

Rousseeuw, P., 1986: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, J. Comp App Math.

Santiñaque, F., 2020: Análisis estadístico de precipitaciones extremas en Uruguay, Facultad de Ingeniería – UdelaR. https://hdl.handle.net/20.500.12008/24577.

Santiñaque, F., Kalemkerian, J. and Renom, M., 2021: https://arxiv.org/pdf/2112.05179.pdf

Silva, L. V., Casaroli, D., Pego, A.W., Alves, J. and Battisti, R., 2019: Rainfall Intensity-Duration-Frequency Relationship for Risk Analysis in the Region of Matopiba, Brazil. Revista Brasileira de Meteorologia, 34(2): 247-254.

Vich, A. I. J., Norte, F. A. and Lauro, C., 2014: Análisis regional de frecuencias de caudales de ríos pertenecientes a cuencas nacientes en la cordillera de los Andes. Meteorologica, 39: 3-26.

Von Mises, R., 1936: La distribution de le plus grande de n valeurs. Rev. math. Union interbalcanique, (1) 141-160.