

## Modelling hydrometeorological extremes associated to the moisture transport driven by the Great Plains Low-Level Jet

Luis Gimeno-Sotelo · P. de Zea Bermudez · Iago Algarra · Luis Gimeno

Received: date / Accepted: date

**Abstract** The Great Plains Low-Level Jet system consists of very strong winds in the lower troposphere that transport a huge amount of moisture from the Gulf of Mexico to the American Great Plains. This paper aims to study the extremes of the Transported Moisture (TM) from the GPLLJ source region to the jet domain; and, for low and high TM, to analyze the extremal dependence between the upper tail of the precipitation in the GPLLJ sink region and the lower tail of the tropospheric stability in that region ( $\omega$ ).

The declustered extremes of TM were analyzed using *Peaks Over Threshold* (POT). A non-stationary Exponential model was fitted to the cluster maxima. Estimated return levels show that the extremes of TM are expected to decrease in the future. This is meteorologically congruent with the known displacement of the western edge of the North Atlantic Subtropical High, which controls atmospheric circulation in the North Atlantic, and to a higher scale with the change of phase from negative to positive of the Atlantic Multidecadal Oscillation.

Bilogistic and Logistic models were fitted to the extremes of ( $-\omega$ , precipitation) for low and high TM, respectively. The extremal dependence between " $-\omega$ " and precipitation proves to be stronger in the case of high TM. This confirms that dynamical instability represented by " $-\omega$ " is the most important parameter for achieving high values of pre-

---

Luis Gimeno-Sotelo

Centro de Investigación Mariña, Universidade de Vigo, Environmental Physics Laboratory (EPhysLab), Ourense, Spain ; Departamento de Estatística e Investigación Operacional, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal.

P. de Zea Bermudez

Departamento de Estatística e Investigación Operacional, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal ; CEAUL - Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal.

(Corresponding author) E-mail: pbermudez@fc.ul.pt

Iago Algarra

Centro de Investigación Mariña, Universidade de Vigo, Environmental Physics Laboratory (EPhysLab), Ourense, Spain.

Luis Gimeno

Centro de Investigación Mariña, Universidade de Vigo, Environmental Physics Laboratory (EPhysLab), Ourense, Spain.

precipitation once there is a mechanism that allows the continuous supply of large amounts of moisture, such as the derived from a low-level jet system.

**Keywords** Moisture transport · Great Plains Low-Level Jet · Extreme Value Analysis · *Peaks Over Threshold* methodology · Bivariate Threshold Excess Models

## 1 Introduction

The World Climate Research Programme (WCRP), which is the United Nations programme defining climate research priorities, identifies Weather and Climate extremes as one of the big challenges: it has been included as an independent chapter in all Intergovernmental Panel on Climate Change (IPCC) reports (e.g. Field et al. (2012); Qin et al. (2014); Masson-Delmotte et al. (2018) ). Within the study of these extremes, the analysis of combined events, defined as "the combination of multiple climate drivers that contributes to societal or environmental risk", has gained great importance, being multiple the publications devoted to them in high-impact journals due to their enormous socioeconomic importance (e.g. Raymond et al. (2020); Ridder et al. (2020); Zscheischler et al. (2020)). Initially focused on the analysis of the simultaneous or consecutive occurrence of local phenomena, such as droughts and heat waves, the studies involving precipitation as one of the variables have been abundant. However, studies trying to link precipitation extremes to large-scale atmospheric circulation patterns have been much less frequent and, to the best of our knowledge, the role of the large-scale moisture transport has never been considered from this perspective.

Moisture transport from oceans to continents is the primary component of the atmospheric branch of the water cycle and forms the link between evaporation from the ocean and precipitation over the continents (Gimeno et al., 2012). There has been an important number of studies on the role of anomalies in the transport of moisture during natural hydrometeorological hazards, extreme drought (e.g., Drumond et al. (2019)) or intense precipitation (e.g. Stohl and James (2004)). The close relation between moisture transport and extreme precipitation events is maximized when this is studied in the areas of influence of the two major global mechanisms of atmospheric moisture transport, namely Low-Level Jet (LLJ) systems and Atmospheric Rivers (ARs), two large-scale dynamical/meteorological structures, the former being key in tropical and subtropical regions and the latter in extratropical regions (Gimeno et al., 2016).

A LLJ is a system of very strong winds in the lower troposphere, typically in the first 1000 meters height (Stensrud, 1996). As water vapour is mainly confined in the lower troposphere, LLJs are major mechanisms of moisture transport at planetary scale. When LLJs are active, they transport a huge amount of moisture favoring high precipitation in the downwind regions. In contrast, in periods when LLJs are absent, downwind regions can suffer from drought events (Gimeno et al., 2016). Within these systems, the Great Plains Low-Level Jet (GPLLJ) is the most studied one because of its socioeconomic effects. It transports a huge amount of moisture from the Gulf of Mexico to the American Great Plains and it is mainly active in the summer (Burrows et al., 2019). Broadly speaking, the GPLLJ carries one-third of all water vapour entering continental United States (Helfand and Schubert, 1995), and it is associated with 10 % – 45 % of the summer precipitation of the American Great Plains region (Hodges and Pu, 2019). In Fig. 1 it is possible to see the climatology of the Great Plains Low-Level Jet System for the months of June, July and August.

The economic importance of the GPLLJ is enormous in the sense that it determines the average and extreme precipitation of a large agricultural region, whose production depends

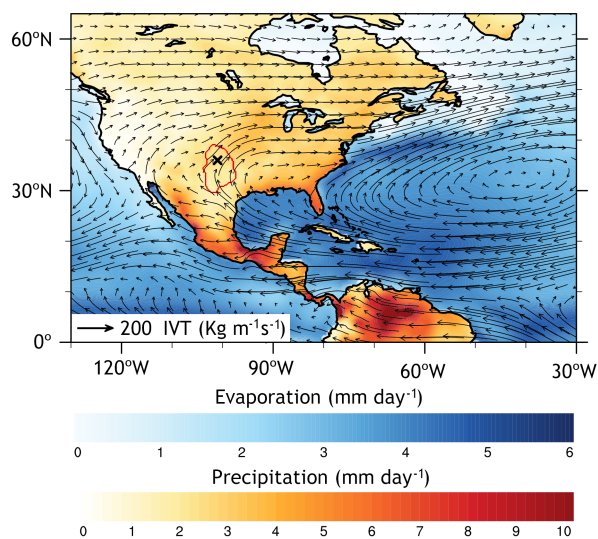


Fig. 1: Climatology of the Great Plains Low-Level Jet System for June, July and August. The region with the highest occurrence of LLJs is inside the red curve, with the cross indicating the point at which the proportion of days on which the LLJ occurs is the highest one. Bluish colors represent the evaporation (mm/day; data from OAFLEX), reddish colors indicate the precipitation (mm/day, data from CPC) and the arrows symbolize the flux of moisture at each point of the grid under consideration (IVT: Integrated Vapor Transport;  $\text{Kg m}^{-1} \text{s}^{-1}$ , data from ERA5). The figure was created following the methodology in Algarra et al. (2019).

on precipitation, occurring large losses from floods and droughts (Basara et al., 2013). It is also important in the determination of the wind resource and especially in the damage generated by severe weather, as GPLLJ is closely related to the development of mesoscale convective systems (Chen et al., 1998) and they are associated with heavy precipitation, supercellular storms and tornado development (Weaver et al., 2012).

The GPLLJ affects precipitation by increasing its frequency, modifying its spatial distribution and increasing its intensity (Pitchford and London, 1962; Mo et al., 1995; Walters and Winkler, 2001; Schumacher and Johnson, 2009; Squitieri and Gallus, 2016; Squitieri and Gallus Jr, 2016). The underlying mechanism to the relationship between the GPLLJ and the precipitation is a strong moisture and heat transport at low levels from the Gulf of Mexico. Moreover, wind convergence at low levels implies atmospheric instability in the output area of the GPLLJ, favoring upward movement. Therefore, it is evident that transported moisture and atmospheric instability are two factors that play an important role in precipitation.

This paper has two main purposes motivated by the dynamics of the complex climatological system extensively described above. The first one is the analysis of TM from an extremal point of view. The second one is to assess the relation between the maximum values of precipitation and the minimum observations of "omega" (tropospheric stability in the GPLLJ sink area), when TM is low (25% of the lowest values) and for values of high TM (25% of the highest values). The Extreme Value Theory (EVT) possesses the proper tools to do so. For extensively details both theoretical and from a practical perspective see e.g.

Gumbel (1958), Kotz and Nadarajah (2000), Coles (2001), Beirlant et al. (2004), de Haan and Ferreira (2006) and Embrechts et al. (2013).

The structure of the paper is as follows. The data is carefully introduced in Section 2. Section 3 presents one of most widely used methods for modelling extreme observations, the *Peaks Over Threshold* (POT) and the associated Generalized Pareto Distribution (GPD). Some fundamental aspects of bivariate extreme value modelling are given in Section 4. Sections 5 and 6 contain the univariate extremal study of TM and the extreme bivariate analysis of precipitation and "-omega" (the negative sign enables the transformation of a sample of minima into a sample of maxima). Finally, the results are discussed in Section 7. Future work is also presented in this section.

## 2 Data

In a recent paper (Algarra et al., 2019), a state-of-the-art Lagrangian approach is used in order to identify the main moisture sources and sinks associated to the GPLLJ (Fig. 2).

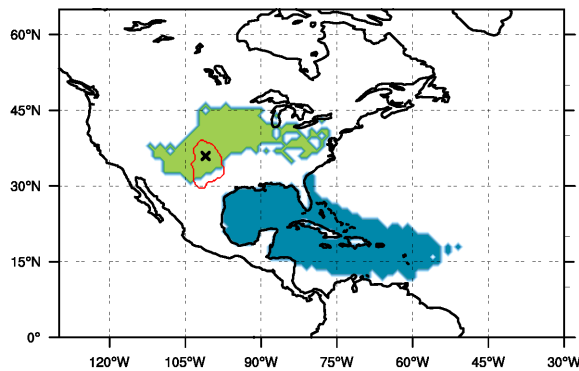


Fig. 2: Key regions associated to the GPLLJ: Region with the highest occurrence of LLJs (inside the red curve, with the cross indicating the point at which the proportion of days on which the LLJ occurs is the highest one); the GPLLJ major oceanic moisture source region (in blue) and its major moisture sink region (in green). The figure was created following the methodology in Algarra et al. (2019).

The area inside the red curve is the jet domain, that is, it is the region with the highest occurrence of LLJs during the period May-October, being the cross the geographical point at which they occur most frequently (36°N, 101°W, 500m height); the area in blue identifies the major oceanic source region for the moisture reaching the jet domain; and the area in green corresponds to the main sink of that moisture, once it has been transported by the jet. So, there are two regions of interest in our analysis: the moisture source and sink regions, connected by the GPLLJ structure in a temporal domain of several days from the evaporation in the source to the precipitation in the sink.

Therefore, the series to analyze, based on the sources and sink areas of moisture linked to the GPLLJ, are:

1. **Transported Moisture (TM)** from the GPLLJ source region to the jet domain (mm/day), as calculated in Algarra et al. (2019). In this study, a Lagrangian approach was used to track air parcels reaching the jet domain from the source region. The TM is then computed by adding the moisture gains of the parcels in the source region before arriving at the jet domain.
2. **Precipitation** in the GPLLJ sink region (mm/day): daily series of precipitation integrated in the whole moisture sink region of the GPLLJ taken from the Climate Prediction Center (CPC) dataset (Xie et al., 2010), which is a state-of-the-art precipitation dataset (see Sun et al. (2018) for a review on gridded precipitation data).
3. Tropospheric Stability in the GPLLJ sink region (**omega**, measured in Pa/s): daily series of vertical velocity computed as the mean of omega at 850 hPa in the sink region, taken from the reanalysis ERA-5 (Hersbach et al., 2020). Omega is defined as the vertical component of velocity in pressure coordinates (these three-dimensional coordinates are defined by replacing the usual z-coordinate by atmospheric pressure ( $p$ )). This is,  $\omega := \frac{dp}{dt}$ , so negative values of  $\omega$  represent ascending movements and positive values correspond to descending movements. The level of 850 hPa (about 1500 m height) is considered for  $\omega$  as it represents the vertical movement at the lower troposphere, where the GPLLJ occurs and most of the moisture is confined.

The series consist of 6992 observations daily recorded from 1 May 1980 to 31 October 2017. This period comprises the extended summer periods since the inclusion of satellite data in the reanalysis, which occurred in 1979<sup>1</sup>. However, in the statistical analysis to be done in this article, we will only use the summer months, that is, the June-July-August periods, because they are the most interesting ones meteorologically speaking (in these months, the GPLLJ is more active, with occurrence close to 70% of the days). Therefore, we will initiate our study with series that have 3496 observations, although in fact we have 874 observations in each of the groups of TM.

### 3 The Generalized Pareto Distribution and the POT approach

The POT approach consists of fitting an asymptotic model to the excesses above a high (enough) threshold  $u$ . Let  $X_1, X_2, \dots$  be a sequence of i.i.d. random variables, each having distribution function  $F$ , then the random variable  $Y = X - u | X > u$  represents the excesses of  $X$  above  $u$ . The GPD is used as an approximation of  $F_u$  as long as  $u$  is high enough (see Pickands (1975); Balkema and De Haan (1974)). The distribution function of the GPD is:

$$H_{\xi}(y|\sigma_u) = \begin{cases} 1 - \exp\left(-\frac{y}{\sigma_u}\right), & \xi = 0, \\ 1 - \left(1 + \frac{\xi y}{\sigma_u}\right)^{-\frac{1}{\xi}}, & \xi \neq 0; \end{cases} \quad (1)$$

If  $\xi \geq 0$ ,  $y \in (0, \infty)$  and if  $\xi < 0$ ,  $y \in (0, -\sigma_u/\xi)$ . The scale and shape parameters satisfy, respectively,  $\sigma_u > 0$  and  $-\infty < \xi < \infty$ .  $\sigma_u$  is used to indicate that the scale parameter depends on the threshold  $u$ .

See e.g. de Zea Bermudez and Kotz (2010a), de Zea Bermudez and Kotz (2010b) and Mackay et al. (2011) for a review of GPD parameter estimation methods. It is possible to estimate interesting quantities using the estimates of the parameters of the GPD, such as tail probabilities and extremal quantiles.

<sup>1</sup> *Extended summer* refers to the period of May to October

### 3.1 Threshold Selection and Model Assessment

A very important question in the POT approach is how to choose the threshold  $u$ . The problem is in selecting a value that allows a trade-off between the large variance of the estimators that occurs for too high values of  $u$  and the large bias that occurs for too small values of this threshold. Methods such as the Mean Excess Function, MEF (see Davison and Smith (1990)) and the stability of the parameter estimates are commonly used to determine an adequate threshold. Once the threshold is selected, attention focuses on assessing the fit of the model. For doing so, two goodness-of-fit tests for the GPD are commonly used: the Cramér-von Mises (CvM) and Anderson-Darling (AD) tests (see Choulakian and Stephens (2001)). The Likelihood Ratio Test (LRT) is also used to assess if the GPD model can be reduced to the Exponential distribution (see Coles (2001)). In this situation, the fit to the Exponential distribution is usually assessed by the Lilliefors-corrected Kolmogorov-Smirnov (LcKS) test (see Lilliefors (1969)).

### 3.2 Dealing with Dependent Sequences

So far, we have assumed that the excesses above a threshold  $u$  are i.i.d. However, this assumption is often unrealistic when working with time series, as there is usually, at least, short-term temporal dependence that may affect our analysis. The most popular way of addressing this issue is to carry out a **declustering** process. The "runs-declustering", which is explained in Coles (2001), consists on fitting a GPD model to the sample of the maxima of each cluster of excesses, where clusters are defined as follows: exceedances (observations above  $u$ ) separated by less than  $r$  non-exceedances are included in the same cluster.

In applications, it is most important to estimate, for a high number  $m$ , the  $m$ -observation return level ( $x_m$ ), which satisfies  $P(X > x_m) = p$ , where  $p = \frac{1}{m}$ . That is,  $x_m$  is exceeded once in every  $m$  observations. It can be estimated as follows:

$$\widehat{x}_m = \begin{cases} u + \frac{\widehat{\sigma}}{\widehat{\xi}} \left[ \left( m^{\frac{N_u}{n}} \widehat{\theta} \right)^{\widehat{\xi}} - 1 \right], & , \widehat{\xi} \neq 0, \\ u + \widehat{\sigma} \log \left( m^{\frac{N_u}{n}} \widehat{\theta} \right), & , \widehat{\xi} = 0, \end{cases} \quad (2)$$

where  $\widehat{\sigma}$  and  $\widehat{\xi}$  are the estimates of the parameters of the GPD model fitted to the cluster maxima,  $N_u$  is the number of excesses above the threshold  $u$ ,  $n$  is the total number of observations of the series and  $\widehat{\theta} = \frac{N_c}{N_u}$  is the estimate of the extremal index, with  $N_c$  being the number of clusters of excesses.

## 4 Bivariate Threshold Excess Models

### *Parametric models*

Let  $F$  be the joint distribution function of  $(X_1, X_2)$ . It may be approximated by a parametric model  $G$  within the region  $x_1 > u_1$ ,  $x_2 > u_2$ . Some parametric bivariate models are the Logistic model (Gumbel, 1960), the Asymmetric Logistic Model (Tawn, 1988), the Husler-Reiss Model (Hüsler and Reiss, 1989), the Negative Logistic Model (Joe, 1990), the Asymmetric Negative Logistic Model (Joe, 1990), the Bilogistic Model (Smith, 1990), the Negative Bilogistic Model (Coles and Tawn, 1994) and the Coles-Tawn Model (Coles and Tawn, 1991). Let  $\tilde{y}_j = 1/\tilde{x}_j$ , with  $\tilde{x}_1$  and  $\tilde{x}_2$  being standard Fréchet-transformed values of

$x_1$  and  $x_2$ , respectively. Then, for  $x_1 > u_1$  and  $x_2 > u_2$ ,  $F(x_1, x_2)$  may be approximated, for example, by:

1. If  $G$  is the **Logistic** model:  $G(x_1, x_2) = \exp \left[ - \left( \tilde{y}_1^{1/\alpha} + \tilde{y}_2^{1/\alpha} \right)^\alpha \right]$ , where  $0 < \alpha \leq 1$ .
2. If  $G$  is the **Bilogistic** model:  $G(x_1, x_2) = \exp \left\{ -\tilde{y}_1 q^{1-\alpha} - \tilde{y}_2 (1-q)^{1-\beta} \right\}$ , where  $q = q(\tilde{y}_1, \tilde{y}_2; \alpha, \beta)$  is the root of the equation  $(1-\alpha)\tilde{y}_1(1-q)^\beta - (1-\beta)\tilde{y}_2 q^\alpha = 0$ . The two parameters ( $\alpha$  and  $\beta$ ) lie in  $(0, 1)$ .

(see, e.g., Beirlant et al. (2004) and Coles (2001) for further details about bivariate models).

#### *Pickands dependence function and extremal coefficients*

It is possible to express a parametric model  $G$  as follows:

$$G(x_1, x_2) = \exp \left\{ - \left( \frac{1}{\tilde{x}_1} + \frac{1}{\tilde{x}_2} \right) A \left( \frac{\tilde{x}_1}{\tilde{x}_1 + \tilde{x}_2} \right) \right\}, \quad (3)$$

where the function  $A(\cdot)$  is called **Pickands dependence function**.  $A(\cdot)$  is a convex function defined on  $[0, 1]$  with  $\max(t, 1-t) \leq A(t) \leq 1$  for all  $0 \leq t \leq 1$ . If  $X_1$  and  $X_2$  are perfectly dependent,  $A(t) = \max(t, 1-t)$ ,  $\forall t \in [0, 1]$ ; if they are independent,  $A(t) = 1$ ,  $\forall t \in [0, 1]$ .

There are some **extremal coefficients** that can be calculated using  $A(\cdot)$ , such as the hereinafter referred to as *Dependence coefficient*, defined as  $2(1 - A(1/2))$ . Independence corresponds to *Dependence* = 0 and perfect dependence to *Dependence* = 1: the strength of dependence increases as *Dependence* increases.

(see, e.g., Beirlant et al. (2004) for extensive information about this topic).

#### *Censored-likelihood method of inference*

Let  $(x_{1,1}, x_{1,2}), \dots, (x_{n,1}, x_{n,2})$  be independent realizations of  $(X_1, X_2)$ . The plane is divided into four regions:

$$\begin{aligned} D &= (-\infty, u_1) \times (-\infty, u_2) & A &= [u_1, \infty) \times (-\infty, u_2) \\ B &= (-\infty, u_1) \times [u_2, \infty) & C &= [u_1, \infty) \times [u_2, \infty) \end{aligned}$$

The **censored-likelihood function** is defined as

$$L(\theta; (x_{1,1}, x_{1,2}), \dots, (x_{n,1}, x_{n,2})) = \prod_{i=1}^n \psi(\theta; (x_{i,1}, x_{i,2})), \quad (4)$$

where  $\theta$  is the parameter vector of the model and

$$\psi(\theta; (x_1, x_2)) = \begin{cases} \frac{\partial^2 F}{\partial x_1 \partial x_2} \Big|_{(x_1, x_2)}, & (x_1, x_2) \in C, \\ \frac{\partial F}{\partial x_1} \Big|_{(x_1, u_2)}, & (x_1, x_2) \in A, \\ \frac{\partial F}{\partial x_2} \Big|_{(u_1, x_2)}, & (x_1, x_2) \in B, \\ F(u_1, u_2), & (x_1, x_2) \in D. \end{cases}$$

Using the likelihood function in (4), it is possible to obtain maximum likelihood estimates and asymptotic confidence intervals for the parameters of the bivariate threshold excess models.

(see e.g. Coles (2001) for further details about this method of inference).

#### *Asymptotic Independence and Asymptotic Dependence*

The bivariate threshold excess models presented in this paper rely on the assumption of *asymptotically dependent* variables. Considering  $F_1$  and  $F_2$  as the marginal distributions of  $X_1$  and  $X_2$  respectively, the following coefficient is defined:

$$\chi := \lim_{u \rightarrow 1} P(F_2(X_2) > u \mid F_1(X_1) > u) \quad (5)$$

$\chi$  takes values between 0 and 1 : when  $X_1$  and  $X_2$  are asymptotically independent,  $\chi = 0$ ; and when they are asymptotically dependent,  $0 < \chi \leq 1$ . Regarding asymptotically dependent variables, the extremal dependence is stronger as  $\chi$  increases. The coefficient  $\chi$  defined in (5) may also be obtained as follows:

$$\chi = \lim_{u \rightarrow 1} \chi(u) = \lim_{u \rightarrow 1} \left[ 2 - \frac{\log P(F_1(X_1) \leq u, F_2(X_2) \leq u)}{\log P(F_1(X_1) \leq u)} \right], \quad 0 < u < 1. \quad (6)$$

(see e.g. Coles (2001) for further details about this issue).

## 5 Univariate analysis of Transported Moisture

In this section we will address the univariate analysis of the series of Transported Moisture (TM) from the GPLLJ source region to the jet domain. We will firstly present a brief exploratory analysis of the series and, afterwards, the POT analysis with declustering that was carried out.

### 5.1 A brief exploratory analysis

The series of TM for the summer periods (months of June, July and August) is expressed in **mm/day** and has 3496 observations (38 summers). The data was daily recorded from 1980 to 2017. The plot of the series as well as the histogram with the kernel density estimate are presented in Fig. 3.

The plot of the series suggests that it is reasonably stationary except for the largest values, for which a declining trend seems to exist. In the left-hand plot of Fig. 4 we compare the values of TM which were observed in the first 19 years with the ones recorded during the latest 19 years; in the right-hand plot of that figure we analyze the TM's yearly evolution. The most relevant aspect which can be observed in these figures is that the magnitude of the large values clearly decreases as time goes by.

### 5.2 Threshold Models Approach

We model the data by the POT methodology due to its advantages when compared to the traditional block maxima method. We used a declustering scheme for the exceedances over the chosen threshold in order to deal with the short-term temporal dependence existing between them. First, we will present the threshold selection procedure and, afterwards, the POT analysis with four different *run lengths*.



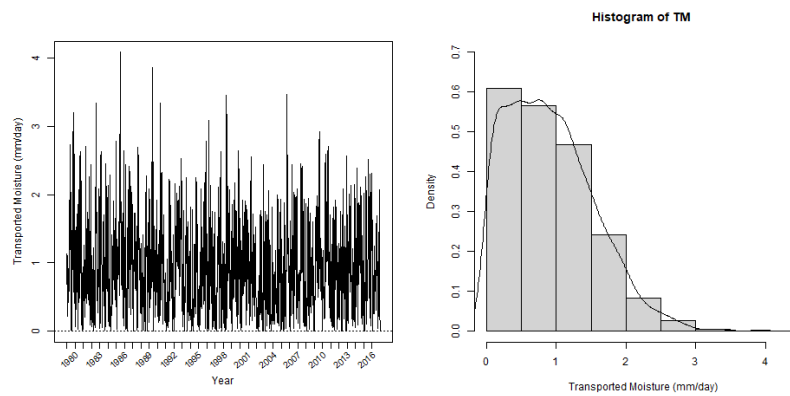


Fig. 3: Plot of the series of TM from the GPLLJ source region to the jet domain (left) and histogram and kernel density estimate for the TM series (right)

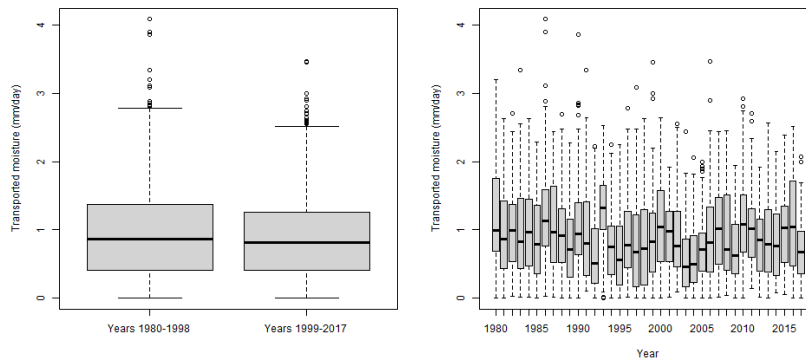


Fig. 4: Boxplots for the observations of the first and second half of the TM series (Years 1980-1998 and Years 1999-2017), on the left; and boxplots for the observations of each summer from 1980 to 2017 for the TM series, on the right.

### Threshold selection

The two methods presented in Subsection 3.1 were applied to the series under study. The estimated MEF presented in Fig. 5 leads to think that a value around 2 might be an appropriate threshold, as a linearity pattern is clearly visible to the right of that value. The Maximum Likelihood (ML) estimates for the shape parameter, plotted in Fig. 5, are approximately constant above  $u = 2$ , which supports that  $u = 2$  might be a reasonable choice. For that choice of  $u$ , the number of exceedances is  $N_u = 201$  ( 5.75% largest observations of TM).

### POT analysis with declustering

For correct application of the POT approach, it is necessary to verify if there is some evidence of excess clustering. The exceedances above the threshold  $u = 2$  are plotted in

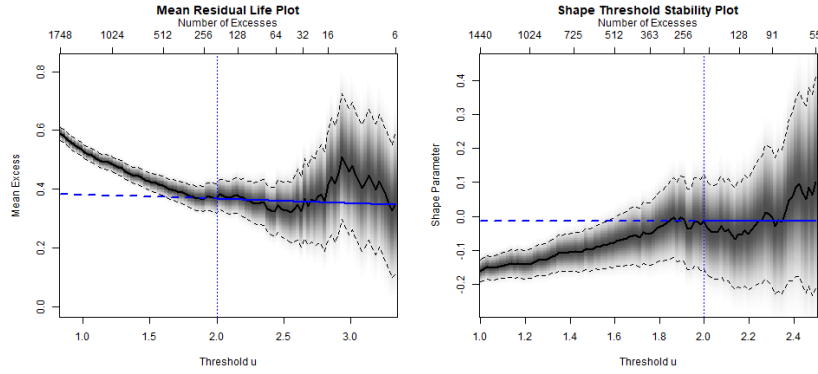


Fig. 5: Estimated MEF of TM (solid black line), with 95% confidence intervals as black dashed lines and fitted solid blue line beyond  $u = 2$  (left). ML estimates for the shape parameters of the GPD models fitted to the TM series, as a function of  $u$  (right).

Fig. 6 (left). There is some visual evidence that there might be some temporal dependence between the exceedances. Thus, the *R* package *evd* is used to perform "runs-declustering" with *run length* ( $r$ ) equal to 1, 2, 3 and 4. After a thorough analysis of the results obtained for these values of  $r$ , we came to the conclusion that the latter choice seems to be the best. Fig. 6 (right) shows the cluster maxima for  $r = 4$ . The corresponding dates can be found in Appendix A.

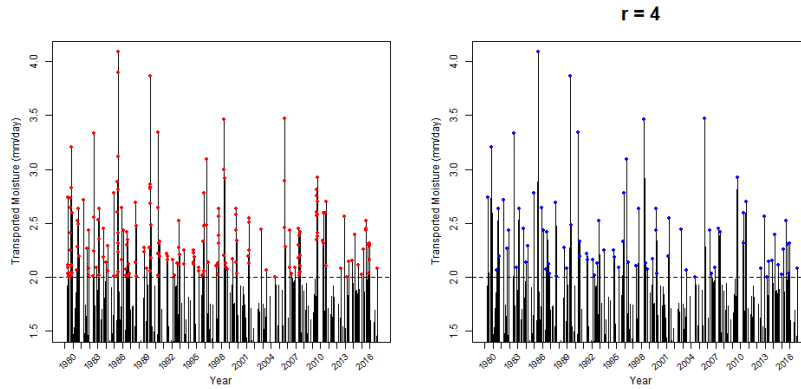


Fig. 6: Exceedances of the TM series above the threshold  $u = 2$  (left). Cluster maxima of excesses of the TM series above  $u = 2$ , performing declustering with *run length* ( $r$ ) equal to 4 (right).

Table 1 contains the results with regard to the number of clusters obtained ( $N_c$ ), the estimate of the extremal index ( $\hat{\theta} = N_c/N_u$ ) and the ML estimates for the parameters of the GPD ( $\hat{\xi}$ ,  $\widehat{\sigma}_{GPD}$ ) and the Exponential model ( $\widehat{\sigma}_{EXP}$ ), with their corresponding standard errors.

In the overall, the results are very much alike and clearly support that the Exponential model seems to be a better alternative than the GPD.

Table 1: Results of the POT analysis for the TM series considering the threshold  $u = 2$  and  $r = 1, 2, 3$  and 4.

	$r = 1$	$r = 2$	$r = 3$	$r = 4$
$N_c$	102	96	90	83
$\hat{\theta}$	0.507	0.478	0.448	0.413
$\hat{\xi}$ (Std.Err)	0.052 (0.114)	0.052 (0.118)	0.022 (0.118)	0.022 (0.123)
$\hat{\sigma}_{GPD}$ (Std.Err)	0.367 (0.055)	0.373 (0.058)	0.401 (0.064)	0.408 (0.067)
$\hat{\sigma}_{EXP}$ (Std.Err)	0.387 (0.038)	0.393 (0.040)	0.410 (0.043)	0.418 (0.046)

Now the question is if we consider the GPD model or the Exponential for modelling the cluster maxima. The *profile log-likelihood* 95% confidence interval for  $\xi$  is  $(-0.178, 0.321)$  and so it is consistent with the hypothesis that  $\xi = 0$  (Exponential model). The Exponential and the GPD QQ-Plots are presented in Fig. 7. The figure clearly shows that both the Exponential and the GPD models are both appropriate for modelling the cluster maxima.

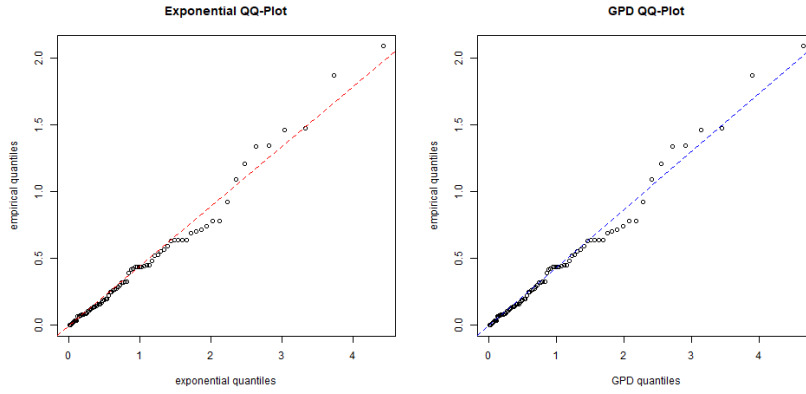


Fig. 7: Exponential and GPD QQ-Plots for the cluster maxima of excesses of the TM series, for  $u = 2$  and  $r = 4$ .

For all values of  $r$  considered, the Cramér-von Mises (CvM) and Anderson-Darling (AD) tests did not reject the null hypothesis that the cluster maxima come from a GPD. Therefore, we can conclude that the GPD model fits well to the data.

The fact that the GPD model fits the declustered excess data does not necessarily mean that it is better than the Exponential model (limiting case of the GPD when  $\xi \rightarrow 0$ ). In order to test  $H_0 : \xi = 0$  vs.  $H_1 : \xi \neq 0$ , a Likelihood Ratio Test (LRT) was performed. It is possible to conclude that the Exponential model is more appropriate to model the cluster maxima of the excesses above  $u = 2$ , at all the usual significance levels.

The Lilliefors-corrected Kolmogorov-Smirnov (LcKS) test did not reject the null hypothesis that the cluster maxima of the excesses above  $u = 2$  come from an Exponential distribution. For all values of  $r$  analyzed, the conclusion from this LcKS test is that the Exponential model fits well to the data, for the usual levels of significance.

(see Gimeno-Sotelo (2021) for further details about the results of the statistical tests that were carried out.)

Hence, we will use the Exponential model for the cluster maxima. As previously said, Fig. 4 seems to indicate a declining trend in the largest values of the series, reflecting non-stationarity as time evolves. In this framework, it is reasonable to allow the scale parameter of the Exponential distribution to vary according to time. That corresponds to introduce the year of observation as a covariate. Taking into account that the scale parameter is always positive, the *log* link function is used. Thus, the expression for the scale parameter of the Exponential model that we fitted is the following:

$$\sigma_t = \exp\{\phi_0 + \phi_1 t\}, \quad (7)$$

where  $t = Year - 1979$ . The aim of this location modification is to enable time to vary the 38 summer periods.

The results shown in Table 2 indicate that, as suspected, the ML estimate of the parameter  $\phi_1$  is negative for all the values of  $r$  considered, what means that the estimate of the scale parameter of the Exponential model is lower in more recent years when compared to the initial period. This decrease in the estimate of the scale parameter with time seems to be more important as  $r$  increases.

Table 2: ML estimates for the parameters of the non-stationary Exponential model fitted to the cluster maxima of excesses of the TM series, choosing  $u = 2$  and  $r = 1, 2, 3$  and 4.

	$r=1$	$r=2$	$r=3$	$r=4$
$\hat{\phi}_0$ (Std.Err)	-0.726 (0.180)	-0.705 (0.187)	-0.603 (0.195)	-0.538 (0.211)
$\hat{\phi}_1$ (Std.Err)	-0.014 (0.009)	-0.014 (0.009)	-0.018 (0.009)	-0.020 (0.010)

Now the issue lies in assessing if the non-stationary Exponential model is better when compared to the stationary one. As usual in the case of nested models, a Likelihood Ratio Test (LRT) can be used: the null hypothesis of that test in this case is  $\phi_1 = 0$  (that is, the stationary model is more appropriate) vs. an alternative,  $\phi_1 \neq 0$ . Table 3 shows that the non-stationarity in the large values of TM starts to become evident for  $r = 4$ , at level  $\alpha = 0.05$ .

Table 3: Observed value of the LRT statistic and corresponding  $p$ -value, for  $r = 1, 2, 3$  and 4.

$H_0 : \phi_1 = 0$	$r=1$	$r=2$	$r=3$	$r=4$
LRT statistic	2.375	2.306	3.442	3.926
$p$ -value	0.123	0.129	0.064	0.048

### 5.2.1 Estimating return levels

As previously said, return level estimation is one of the main focus of any extreme value analysis. From now on, we will consider the non-stationary Exponential model. For each value of  $t \in \{1, 2, \dots, 38\}$  the corresponding  $\widehat{x}_m(t)$  is obtained by using  $\widehat{\sigma}_t = \exp\{\widehat{\phi}_0 + \widehat{\phi}_1 t\}$  in expression (2) for  $\xi = 0$ . We will just show the plot corresponding to  $r = 4$  (see Fig. 8). The figure reveals that, in summer 1980 the estimated  $m$ -observation return levels are higher than

the ones observed in summer 2017, for  $m = 92 \times 38$ ,  $m = 92 \times 50$  and  $m = 92 \times 100$  (there are 92 observations per year in the TM series, corresponding to the daily observations of June, July and August). It is also interesting to highlight that the differences between the estimated return levels become smaller over time. We extracted from Fig. 8 the values of the estimated return levels in the first and last summers of that series (see Table 4).

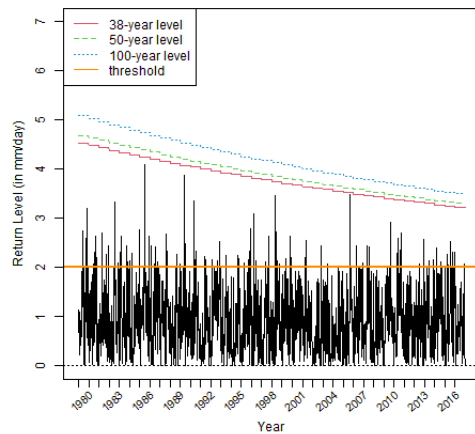


Fig. 8: Estimated return levels for the TM series, considering a non-stationary Exponential model for  $r = 4$ .

Table 4: Estimated 38-year, 50-year and 100-year return levels for the TM series in the first and the last summers of the period considered, using the non-stationary Exponential model for  $r = 4$ .

	38-year return level	50-year return level	100-year return level
Summer 1980	4.531	4.688	5.085
Summer 2017	3.228	3.304	3.497

It is possible to see in Table 4 that the ratio between the estimated 38-year return level for the last summer of the TM series and the first summer of that series is approximately equal to 0.712 (representing a decrease of approximately 28.8% in the estimated 38-year return level from the beginning to the end of the series). In the case of the 50-year return level, the ratio mentioned before is approximately equal to 0.705 (decrease of approximately 29.5% in the estimated 50-year return level); and in the case of the 100-year return level, the ratio equals approximately 0.688 (decrease of approximately 31.2% in the estimated 100-year return level). Thus, as it is obvious, the interpretation of the results of this table is in line with the interpretation of Fig. 8. Moreover, the other comment we made on that figure can also be checked in Table 4, since the difference between the estimated 100-year return level and the 38-year return level is approximately 0.554 for the first summer of the TM series and approximately 0.269 for the last summer of that series. That is, over the period under study,

the difference between those estimated return levels has approximately decreased 51.4% of the value corresponding to the first summer.

From these results, and provided the atmospheric conditions evolve in the current manner, it is possible to say that we expect to observe a persistent decrease in the extreme values of TM as time goes by (see Fig. 8).

The univariate modelling was performed by means of the *R* packages *evd* (Stephenson, 2002) and *extRemes* (Gilleland and Katz, 2016). The code used for the computations is available from the authors upon request.

### 5.3 Meteorological implications

For the meteorological interpretation of the analysis of TM, we consider as *days with extreme TM* those days corresponding to the cluster maxima of excesses of TM for  $r = 4$  (see Appendix A). In order to explain the meaning of our results in meteorological terms, Fig. 9 displays the anomalies of the 500 hPa geopotential, moisture fluxes, and precipitation for all days when TM was extreme (panel a)), for extreme days in the first half of the study period 1980-2017 (panel b)), and for extreme days in the second half (panel c)). A geopotential tripole with positive anomalies on the central and eastern continent and negative ones on both sides of the Atlantic Ocean and the Pacific coasts favours atmospheric circulation from the Caribbean Sea and the Mexican Gulf to the Great Plains, and consequently also implies high values of TM and precipitation in our region of interest. This tripole is intensified for the first half of our study period and weakens in the second half, indicating that very high TM is more difficult in the first two decades of the XXI century than it was in the last two decades of the XX century. This accords with a displacement of the western edge of the North Atlantic Subtropical High (NASH), the semi-permanent structure that controls atmospheric circulation in the North Atlantic, which is in turn controlled by the Atlantic Multidecadal Oscillation (AMO), a dominant mode of climate variability in the North Atlantic Ocean with period of about 60-80 years (Trenberth et al., 2021). When the AMO is positive, the NASH is weak, low-level TM from the Caribbean to central and SE North America is reduced, and precipitation on the Great Plains is also reduced (Seager et al., 2014). This is exactly what happened in our period of interest, in which the AMO changed its phase from negative to positive in the mid-nineties (see key figures in Trenberth et al., 2021), with a continuous increase in the AMO index. This conceptual meteorological scheme is in complete agreement with our results in Fig. 8, which shows a decrease in the estimated 38-year, 50-year and 100-year return levels for the TM series over the study period.

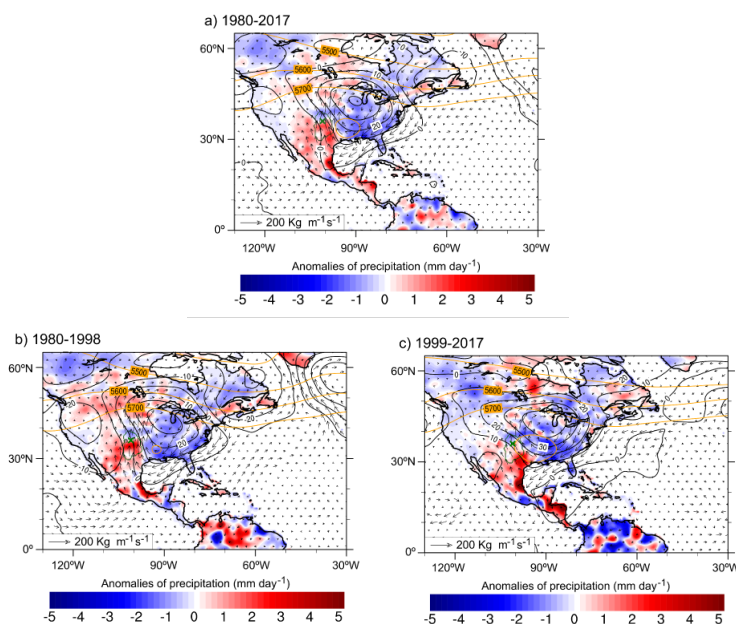


Fig. 9: Anomalies of precipitation (in colors, mm/day). The vectors represent the anomalies of IVT (Integrated Vapor Transport,  $\text{Kg m}^{-1} \text{s}^{-1}$ ); the orange lines show the 500 hPa geopotential field together with their anomalies (black lines). The green cross indicates the point of maximum GPLLJ intensity and green contour refers to the 75th percentile of the GPLLJ index value. Panel a) : *days with extreme TM* in the period 1980-2017; panels b) and c) : *days with extreme TM* corresponding to the periods 1980-1998 and 1999-2017, respectively.

## 6 Bivariate analysis of Precipitation and "-omega"

In Section 1, we introduced the series of **precipitation** (measured in mm/day) in the GPLLJ sink region and the series of tropospheric stability in that region (**omega**, measured in Pa/s). As it was already referred, these series consist of 3496 observations, corresponding to the daily observations of the summer months (June, July and August) of the period 1980-2017. Now, interest focuses on studying the extremal dependence between precipitation and "-omega" (the sign of "omega" is reversed because the meteorological interest lies on studying the joint behaviour of the upper tail of precipitation and the lower tail of "-omega"). In fact, our study consists in analyzing the bivariate extremes of precipitation and "-omega" for two subsamples of the series: for the days when the TM from the GPLLJ source region to the jet domain is high and when it is low. Thus, one subsample consists of the days with the 25% lowest values of TM, whereas the other one includes the 25% highest values of that variable (consequently, each subsample includes 874 observations). It is important to mention that the TM series was lagged 1 day with respect to the series of precipitation and "-omega", that is, for example, for an observed pair of (-omega, precipitation) occurring on 2 June 1980, the corresponding value of TM is the one that occurred on 1 June 1980. The reason for doing so is meteorological: precipitation and "-omega" are observed in the GPLLJ sink region, while the TM is computed on its way from the source region to the jet domain.

Hence, the moisture arrives at the sink region (approximately) 1 day after it is observed, and that is why the adjustment that we carried out was necessary <sup>2</sup>.

### 6.1 Preliminary analysis

The data of interest are the observed pairs of (-omega,precipitation) for the days with low TM and for those corresponding to high TM. As we can see in Fig. 10, the boxplots show that there are higher extreme values of "-omega" when the TM is low than when it is high. In contrast, there are higher extreme values of precipitation when the TM is high than when it is low.

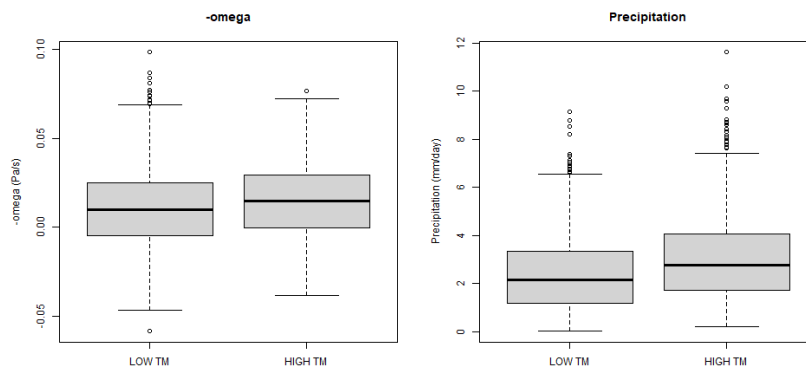


Fig. 10: Boxplots of "-omega" (left) and precipitation (right) for low TM and high TM

<sup>2</sup> For assessing the influence of the time period, other alternatives were considered. As such, we analyzed the scenarios of moisture arrival at the sink region 2, 3 and 4 days after it is observed, and the results were analogous.



## 6.2 Fitting the Bivariate Threshold Excess Models

### *Univariate threshold models for the margins*

The fitting of the bivariate threshold excess models requires previous fit of GPD models to the excesses over appropriate thresholds for each margin.

We came to the conclusion that  $u_1 = 0.03$  is a suitable threshold for "-omega" and  $u_2 = 5.2$  is an adequate threshold for the precipitation, for both low TM and high TM. See Gimeno-Sotelo (2021) for further details about these issues.

The results of the POT analysis can be found in Table 5 . As we can see in that table, the ML estimates for the shape parameter of the GPD are all negative and larger than  $-0.5$ , which guarantees the asymptotic properties of ML estimation (de Zea Bermudez and Kotz, 2010a). Moreover, the GPD model boundary constraints are satisfied ( $x^F > x_{n:n}$ ).

Table 5: Results of the POT analysis for the "-omega" and precipitation series in the cases of low TM and high TM, considering thresholds  $u_1 = 0.03$  for "-omega" and  $u_2 = 5.2$  for precipitation.

	-omega(low TM)	prec. (low TM)	-omega(high TM)	prec. (high TM)
Threshold	0.03	5.2	0.03	5.2
Number of excesses	170	55	211	98
Percentage of excesses	19.5 %	6.3 %	24.1 %	11.2 %
$\hat{\xi}$ (Std.Err)	-0.180 (0.072)	-0.160 (0.128)	-0.311 (0.059)	-0.163 (0.087)
$\hat{\sigma}$ (Std.Err)	0.018 (0.002)	1.185 (0.219)	0.017 (0.001)	1.676 (0.222)
$\widehat{x^F}$	0.132	12.583	0.084	15.499
$x_{n:n}$	0.098	9.134	0.077	11.609

The *profile log-likelihood* 95% confidence intervals of  $\xi$  for "-omega" and precipitation in the cases of low and high TM. The confidence intervals for  $\xi$  for precipitation contain the value 0, which suggests that the limiting GPD model (Exponential) is more appropriate. In contrast, the confidence intervals of  $\xi$  for "-omega" only contain negative values.

In what concerns the CvM and AD tests, for "-omega" and precipitation, it is concluded that the GPD model fits well to the data in both cases of low and high TM, for all usual significance levels. The results of the LRT show that for "-omega", both for low and high TM, at the level of significance 0.05, the GPD model is more appropriate than the Exponential one. For precipitation, it is possible to conclude that the Exponential model is more adequate than the GPD one. See Gimeno-Sotelo (2021) for further details.

By means of the LcKS test, we can conclude that the Exponential distribution fits well to the precipitation excess data. The estimated values of  $\sigma$  and the corresponding standard errors (in brackets) obtained are: 1.020 (0.138) and 1.442 (0.146) for low TM and high TM, respectively.

### *Bivariate analysis*

The marginal models for the excesses of (-omega,precipitation) are GPD. In the case of the precipitation, the limiting form of the GPD (Exponential model) was considered for simplicity purposes.

In Fig. 11 the observed points of (-omega,precipitation) are represented, for high and low TM, along with two lines representing the thresholds ( $u_1 = 0.03$ ,  $u_2 = 5.2$ ). This enables the definition of three "extremal" quadrants as follows:

- A - large values only in "-omega".
- B - large values only in precipitation.
- C - large values in both variables.

In the plots we also indicate the number of points belonging to each of the quadrants and the corresponding percentage in terms of the total sample size. It should be mentioned that the largest difference in the percentages is observed in quadrant C (which reflects the situation of extremes in both variables).

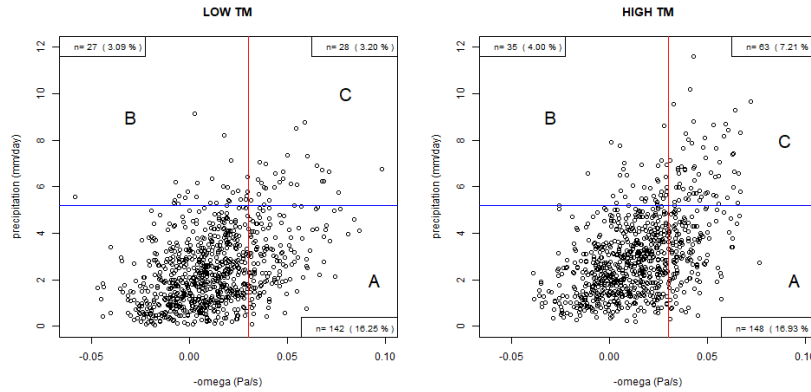


Fig. 11: Scatterplots for  $(-\text{omega}, \text{precipitation})$  in the cases of low TM (left) and high TM (right). The red line refers to the threshold for "-omega" (0.03 in both cases of low and high TM) and the blue line corresponds to the one for precipitation (5.2 in both cases of low and high TM).

Fig. 12 presents the *chi plots* for  $(-\text{omega}, \text{precipitation})$  in the cases of low TM and high TM. *Chi plots* are plots of  $u \in (0, 1)$  against empirical estimates of  $\chi(u)$ . The dashed lines refer to the approximate 95% confidence intervals. It can be seen that the empirical estimates of  $\chi(u)$  are larger than 0 in both cases for the values of  $u$  close to 1, so it is consistent with  $\chi > 0$ , and consequently with the fact that "-omega" and precipitation are **asymptotically dependent** in both situations. Therefore, we can assume that the models presented here are appropriate for these variables.

The joint distribution function of  $(X_1, X_2) = (-\text{omega}, \text{precipitation})$  in the cases of low TM and high TM can be approximated by one bivariate parametric models within the region C. In order to estimate the parameters of the models, a one-step censored-likelihood method was used. That is, the maximization of the censored-likelihood function, given in (4), enables to obtain simultaneously the ML estimates for the marginal and dependence parameters.

In Table 6 it is possible to see the results that we obtained when fitting several usual parametric bivariate models to the pair  $(-\text{omega}, \text{precipitation})$  in the cases of low and high TM.

In Table 6, in each case, the model with the lowest AIC appears in bold: for low TM, the best model is the Bilogistic one (AIC=270.826) and for high TM, the Logistic one (AIC=311.341). It is important to point out that, within each case, the values of AIC are quite similar.

With respect to the coefficient *Dependence*, defined as  $2(1 - A(1/2))$ , where  $A(\cdot)$  is the corresponding Pickands dependence function. It can be seen in Table 6 that the value of the

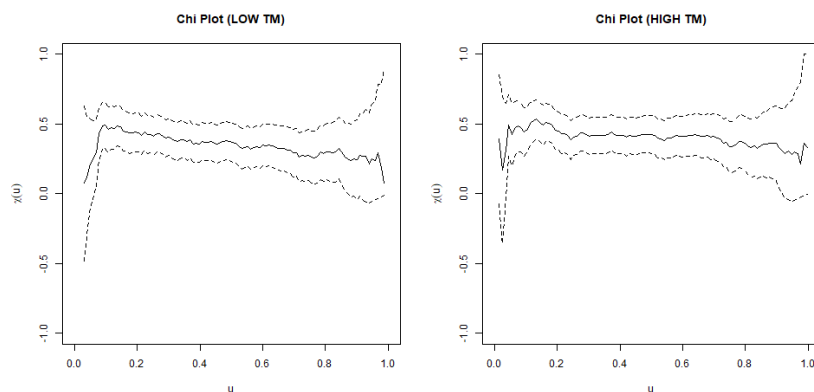


Fig. 12: *Chi plots* for  $(-\omega, \text{precipitation})$  in the cases of low TM and high TM.

*Dependence* coefficient is larger for high TM than for low TM for all the parametric models considered, which means that the extremal dependence between  $-\omega$  and precipitation is stronger in the case of high TM than when TM is low.

Table 6: AIC and *Dependence* coefficient for the parametric models that we fitted to model the joint distribution function of  $(-\omega, \text{precipitation})$  in the cases of low and high TM within the region C in both cases.

Parametric Model	Low TM		High TM	
	AIC	<i>Dependence</i> $2(1 - A(1/2))$	AIC	<i>Dependence</i> $2(1 - A(1/2))$
Logistic	272.498	0.232	<b>311.341</b>	<b>0.359</b>
Asymmetric Logistic	273.721	0.200	331.227	0.323
Husler-Reiss	271.328	0.227	311.952	0.352
Negative Logistic	271.351	0.229	311.366	0.357
Asymmetric Negative Logistic	273.722	0.205	315.197	0.348
Bilogistic	<b>270.826</b>	<b>0.207</b>	319.310	0.418
Negative Bilogistic	271.813	0.216	313.031	0.352
Coles-Tawn	271.244	0.203	312.850	0.353

Numerical results slightly vary from Gimeno-Sotelo (2021) due to the optimization method chosen to maximize the censored-likelihood for each model.

The Bilogistic model proves to be the most adequate to model the joint distribution function of  $(-\omega, \text{precipitation})$  within the region C in the case of low TM, whereas the Logistic model is chosen in the case of high TM.

The ML estimates for the marginal and dependence parameters of those models, as well as the corresponding standard errors, can be found in Table 7.

The Pickands dependence functions corresponding to the models presented in Table 7 can be found in Fig. 13. In that figure it is possible to see that the Pickands dependence function corresponding to the Logistic model for high TM is closer to  $A(t) = \max(t, 1 - t)$ ,  $t \in [0, 1]$  (the perfect dependence case), which means that the extremal dependence between  $-\omega$  and precipitation is stronger when there is high TM than when the TM is low, as we had concluded before.

Table 7: ML estimates (standard errors in brackets) for the marginal and dependence parameters of the bilogistic model for (-omega,precipitation) in the case of low TM; and the logistic model for (-omega,precipitation) in the case of high TM.

Bilogistic Model (Low TM)				
$\widehat{\sigma}_1$	$\widehat{\xi}_1$	$\widehat{\sigma}_2$	$\widehat{\alpha}$	$\widehat{\beta}$
0.018 (0.002)	-0.154 (0.077)	1.025 (0.132)	0.911 (0.033)	0.662 (0.114)
Logistic Model (High TM)				
$\widehat{\sigma}_1$	$\widehat{\xi}_1$	$\widehat{\sigma}_2$	$\widehat{\alpha}$	
0.016 (0.001)	-0.264 (0.067)	1.498 (0.145)	0.715 (0.034)	

$\widehat{\sigma}_1$  and  $\widehat{\xi}_1$  - estimates of the scale and shape parameter of the GPD for the excess data ("-omega")  
 $\widehat{\sigma}_2$  - estimate of the scale parameter of the Exponential (precipitation)  
 $\widehat{\alpha}$  and  $\widehat{\beta}$  - estimates of the dependence parameters of the Bilogistic model  
 $\widehat{\alpha}$  - estimate of the dependence parameter of the Logistic model.

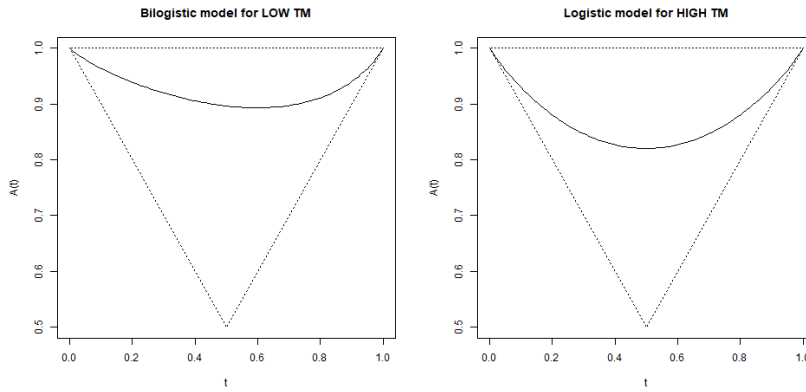


Fig. 13: Pickands dependence functions corresponding to the fitted Bilogistic model for (-omega,precipitation) in the case of low TM (left) ; and to the Logistic model for (-omega,precipitation) in the case of high TM (right) . The dashed black lines refer to the independence ( $A(t) = 1, t \in [0, 1]$ ) and perfect dependence case ( $A(t) = \max(t, 1 - t), t \in [0, 1]$ ).

The quantile curve of a joint distribution function  $F$  at lower tail probability  $p$  is denoted as  $Q(F, p)$ , that is,  $Q(F, p) := \{(x_1, x_2) : F(x_1, x_2) = p\}$ . In Fig. 14 the estimates of the quantile curves  $Q(F_j, 0.95)$ ,  $Q(F_j, 0.975)$  and  $Q(F_j, 0.99)$  are plotted for each  $j \in \{1, 2\}$ , where  $F_1$  denotes the joint distribution function of (-omega,precipitation) in the case of low TM and  $F_2$  in that of high TM. In order to construct those estimated curves, the models presented in Table 7 were used. The dates of the 10 days that exceed the estimate of  $Q(F_1, 0.95)$  and of the 13 days that are beyond the estimate of  $Q(F_2, 0.95)$  are presented in Table 8. This reflects a stronger extremal dependence when TM is high when compared to low TM. The meteorological analysis of these concurrent extreme days will be presented in the next Sub-section.

The R package *evd* was used (Stephenson, 2002) to perform the bivariate analysis of the data. The code used for the computations is available from the authors upon request.

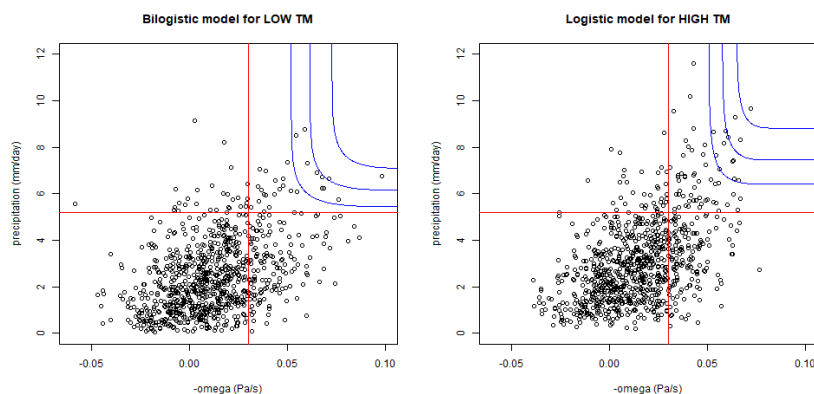


Fig. 14: Estimated quantile curves at lower tail probabilities  $p = 0.95, 0.975, 0.99$  for the joint distribution function of  $(-\omega, \text{precipitation})$  in the cases of low TM (left) and high TM (right), using the fitted models. In these plots, the red vertical line refers to the threshold for  $-\omega$  and the horizontal one corresponds to the threshold for precipitation.

Table 8: Concurrent extreme days of  $(-\omega, \text{precipitation})$  for low TM (first column) and high TM (second column). They corresponds to the dates beyond the estimate of  $Q(F_1, 0.95)$  and the dates beyond the estimate of  $Q(F_2, 0.95)$ , respectively.

Low TM	High TM
1983-6-28	1980-8-15
1994-6-8	1980-8-16
1994-6-23	1981-7-28
1995-6-9	1987-8-26
1998-6-8	1990-7-12
1998-6-9	1990-7-21
2004-7-23	1993-6-18
2008-6-5	1993-6-30
2014-6-15	1993-7-14
2015-8-18	2004-6-10
	2005-7-26
	2007-8-19
	2007-8-24

Year-Month-Day

### 6.3 Meteorological implications

According to a simple approximation, values of high precipitation (HP) scale with precipitable water (IWV, the amount of vapour in a vertical sense, which for simplicity we sometimes refer to as moisture) and with a metric of vertical velocity (Emori and Brown, 2005)<sup>3</sup>. In our analysis, we use a measure of vertical velocity in the quantity  $-\omega$  at 850Pa obtained from the ERA5 reanalysis, which represents the vertical velocity due to large-scale motions (synoptic scale, ranging from few hundred to several thousand kilometres), derived primarily from dynamically driven instability such as that associated with extratropical fronts and cyclones. At a smaller horizontal scale, some other vertical motions

<sup>3</sup> In this meteorological context, HP refers to precipitation higher than a local percentile, usually 95th.

also occur as a result of thermodynamic instability either at the mesoscale (from a few to several hundred kilometres, such as Mesoscale Convective Systems (MCS)), and at the microscale (up to a few kilometres only, such as those seen in isolated, clustered, or embedded MCS storm cells). The quantity " $\omega$ " is not that effective in reproducing vertical motion linked to these phenomena, and performs poorly in the case of tropical cyclones (structures between the mesoscale and synoptic scales).

In their analysis of extreme daily precipitation events, Kunkel et al. (2020) corroborated the results of previous authors who found that although HP events are strongly related to extremes of IWV, they are not related to extremes of " $\omega$ " as much. For higher values of HP, the limiting factor is IWV, not " $\omega$ ". This is because for very high IWV, thermodynamic effects play an important role (mesoscale convection, MCS, supercell storms), which is not the case with low IWV, where only large-scale dynamic instability factors (fronts, extra-tropical cyclones... ) are of importance. This implies that: *(a) higher values of HP may be achieved with high than with low IWV, b) with low IWV, much higher and more persistent values of dynamic instability (" $\omega$ ") are required than for high IWV, when HP can occur with lower values of " $\omega$ ".*

Kunkel et al. (2012) showed that for our region of interest (Great Plains) in the summer months, large-scale synoptic systems such as fronts and extratropical cyclones were responsible for the largest number of HP events, accounting for about 80% of events, with 8% related to tropical cyclones and 12% to MCS and other smaller-scale convective phenomena. This implies that *dynamical instability (" $\omega$ ") is the most important parameter for achieving high values of precipitation once there is a mechanism that allows the continuous supply of large amounts of moisture (a moisture transport mechanism).*

A large amount of TM from the Caribbean source to the region of interest driven by the GPLLJ (large and sustained transported moisture from the source to the sink of the GPLLJ system) results in large values of IWV in the region of interest, which do not occur for low TM. This implies that: *a) precipitation has higher values when TM is high than when it is low, due to the higher values of IWV. b) high precipitation events can occur for moderate-to-high values of " $\omega$ " when TM is high, but only for high to very high values of " $\omega$ " when TM is low.*

This meteorological rationale is coherent with the results presented in this paper. The analysis of the dependence of the extremes of precipitation and " $\omega$ " for days with TM above the 75th percentile compared with those below the 25th percentile confirms the stronger dependence (a greater probability that an extreme of one variable will be accompanied by an extreme of the other) for high than for low TM. For low TM, an extreme of " $\omega$ " does not guarantee a precipitation extreme because it may be accompanied by low IWV. For low values of IWV, only very extreme values of " $\omega$ " may lead to extreme precipitation.

These results may be better visualized by plotting anomalies of geopotential at 500 hPa, moisture fluxes and precipitation for the days with TM above the 75th percentile (high TM) versus the days with TM below the 25th percentile (low TM). In panel b) of Fig. 15 the days with high TM are used, and there it can be seen that moisture transport is enhanced when a positive geopotential anomaly is seen over Central and Southeastern North America and a negative one occurs to the West and to the East, which then favours atmospheric circulation from the Caribbean and the Mexican Gulf towards the Great Plains. On the other hand, a positive geopotential anomaly indicates an inhibited vertical movement in this region. The effect of the enhanced moisture transport is greater than the effect of inhibited vertical movement because positive anomalies of precipitation occur over the Great Plains. The opposite pattern is seen when the TM is low (panel a). This clearly accords with the results discussed

above wherein high precipitation is much more sensitive to IWV and vertical movement (" $-\omega$ ").

These patterns are slightly different for the concurrent extreme days of " $-\omega$ " and precipitation (composites of the same meteorological variables but only for the days included in Table 8 ). In panel d) of Fig. 15 the concurrent extreme days of ( $-\omega$ ,precipitation) corresponding to high TM are used. In that panel it can be seen that the geopotential anomaly tripole is displaced slightly to the West, but this still permits strong atmospheric circulation from the Caribbean (strong moisture transport) whilst favouring vertical movement over the Great Plains (negative anomalies of geopotential). This combination of strong moisture transport and moderate vertical movement (but extreme when compared with all the days with high TM) results in stronger precipitation over the Great Plains than on all the days when there is high TM. In contrast, the concurrent extreme days of ( $-\omega$ ,precipitation) with low TM are used in panel c). There it is possible to visualize that the negative geopotential anomalies in Central and Southeastern North America are intensified, suggesting a stronger vertical movement displaced slightly to the West, permitting greater moisture transport from the Caribbean, although this is outside (to the East) of the GPLLJ box used in this study (TM as defined in our study thus continues to be low). This combination of greater vertical movement with some moisture transport also results in heavier precipitation over the Great Plains than for all the days when TM is low.

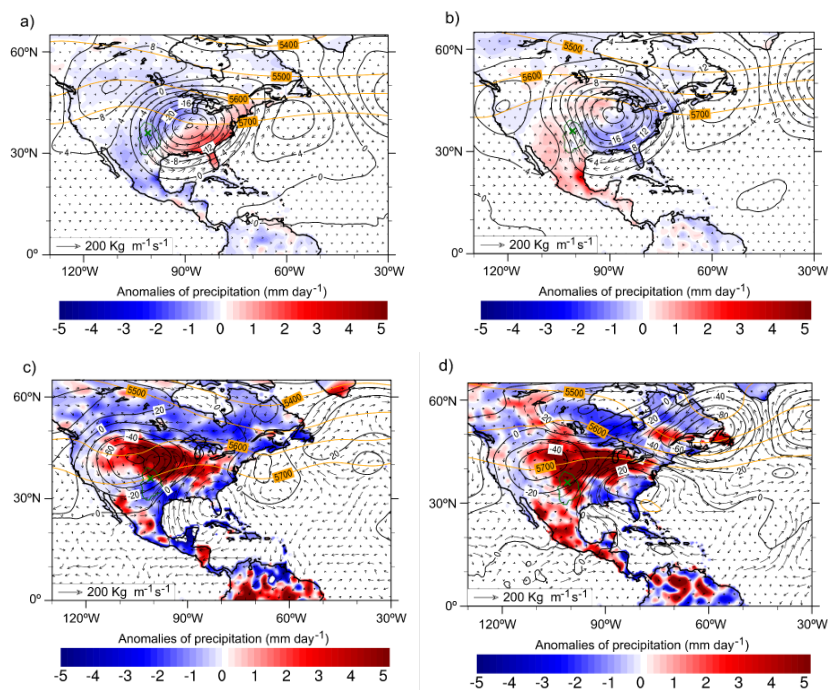


Fig. 15: Anomalies of precipitation (in colors, mm/day). The vectors represent the anomalies of IVT (Integrated Vapor Transport,  $\text{Kg m}^{-1} \text{s}^{-1}$ ); the orange lines show the 500 hPa geopotential field together with their anomalies (black lines). The green cross indicates the point of maximum GPLLJ intensity and green contour refers to the 75th percentile of the GPLLJ index value. The days used for the composites are: (a) those with TM below the 25th percentile (low TM); (b) those with TM above the 75th percentile (high TM); (c) the concurrent extreme days of (-omega,precipitation) for low TM ; and (d) the concurrent extreme days of (-omega,precipitation) for high TM.

## 7 Conclusions and Future Work

The scenario of this paper is set on the Great Plains Low-Level Jet (GPLLJ) system, which is a system of very strong winds in the lower troposphere that transports a huge amount of moisture from the Gulf of Mexico to the American Great Plains and is mainly active during the summer months.

This work aimed to analyze the extremal behaviour of the TM from the GPLLJ source region to the jet domain; and, in the cases of low and high TM, to study the extremal dependence between the upper tail of the precipitation in the GPLLJ sink region and the lower tail of the tropospheric stability in the GPLLJ sink region ( $\omega$ ). For this purpose, we used the series of daily observations of TM, precipitation and " $\omega$ " of all June-July-August periods from 1980 to 2017, which amounts to 3496 observations.

In terms of the univariate analysis of TM, we used the POT methodology. The "runs-declustering" was used in order to reduce the temporal dependence between the exceedances. The declustering process was performed for several values of *run length* ( $r$ ), namely 1,2,3 and 4. We came to the conclusion that the Exponential model was more appropriate than the GPD to model the cluster maxima of excesses over the chosen threshold. Moreover, we



concluded that in the case of  $r = 4$ , the non-stationary model was more adequate than the stationary one. The non-stationarity over time starts to be evident for  $r$  equal to 4, reflected by a declining scale parameter. The estimated 38-year, 50-year and 100-year return levels for the TM series also decreased over time and, additionally, the difference between them became smaller. It is important to recall that a "t-year" return level corresponds to a value which will be exceeded once in "t-summer" periods to come. Therefore, it is possible to say that we expect to observe lower extreme values of TM in the future.

This result is in meteorological agreement with some of the observed changes in the North Atlantic atmospheric circulation and in the climate variability mode that controls the decadal scale, the Atlantic Multidecadal Oscillation (AMO). The change of phase of AMO from negative to positive in the mid-nineties and the continuous increase in the AMO index up to the end of the studied period result in a displacement of the western edge of the North Atlantic Subtropical High and reduced low-level moisture transport from the Caribbean to central and SE North America.

Moreover, we analyzed the bivariate extremes of (-omega,precipitation) in the cases of low and high TM. The sign of "omega" was changed because, meteorologically speaking, the interest lied on the study of the joint behaviour of the upper tail of precipitation and the lower tail of "omega". The series of precipitation and "-omega" were lagged 1 day with respect to the TM series due to the temporal nature of the GPLLJ system. The fit of bivariate threshold excess models requires previous fit of univariate threshold models to the margins. Two GPD models were fitted to the margins. In the case of precipitation, the shape parameter of the fitted GPD was very close to zero. Considering that the Exponential distribution results as the limiting case of the GPD when  $\xi$  tends to zero, the GPD was reduced to its limiting form. The censored-likelihood method was used for fitting several different parametric models. The most parsimonious models were the Bilogistic and the Logistic, for low and high TM, respectively. The extremal dependence between "-omega" and precipitation proved to be stronger in the case of high TM than when TM is low. This conclusion can be visualized by means of the estimated Pickands dependence functions. Additionally, the *chi plots* allowed us to conclude that it is reasonable to assume that the variables are asymptotically dependent, in both the cases of low and high TM.

The results of the bivariate analysis have two important meteorological implications. Firstly, they confirm that dynamical instability, as quantified by "-omega", is the most important parameter for achieving high values of precipitation once there is a moisture transport mechanism such as a low-level jet system, which continuously supply large amounts of moisture. Secondly, they confirm that moderate-to-high values of dynamical instability are necessary to generate high precipitation when TM is high, but high to very high values are required when TM is low.

**Future work** should involve extending worldwide the univariate analysis of the moisture transport driven by the different LLJ systems (Algarra et al., 2019). It would also be interesting to perform an analogous analysis for the TM by the other major moisture transport systems, the atmospheric rivers; see Algarra et al. (2020). In that work, an increasing trend of moisture taken by atmospheric rivers in their main sources was identified. However, only the mean values were studied and the analysis of the extremes of the moisture transport remain to be analyzed in order to assess if they exhibit the same behavior. Regarding the bivariate analysis involving precipitation and vertical movement, it would also be convenient to extend the analysis at a global level for the other LLJ systems. Additionally, it would be interesting to discriminate according to the type of precipitation, separating large-scale precipitation from another more localized, such as that derived from MCS or isolated storms. Moreover, in that case of localized precipitation, adding other variables that

---

reproduce vertical movements on a smaller scale would also be of interest, for example the Convective Available Potential Energy (CAPE), which better reproduces movements derived from thermodynamic instability. Finally, due to the extensive data that we possess regarding the climate processes addressed in this paper, we intend to carry out in the near future extremal analyses which will take into account the spatial characteristics of the meteorological phenomena (see Davison et al., 2012 and Huser and Wadsworth, 2020).

## Appendix A Dates corresponding to the cluster maxima of excesses for $r=4$

See Table 9, which contains the dates corresponding to the cluster maxima of excesses for  $r = 4$ .

Table 9: Dates corresponding to the cluster maxima of excesses for  $r = 4$ .

Year	Month-Day
1980	7-8, 8-13
1981	7-10, 7-29, 8-14
1982	7-1, 8-5, 8-26
1983	7-20, 8-26
1984	6-15, 8-11
1985	6-5, 6-25
1986	6-3, 7-18, 8-25
1987	6-10, 7-10, 7-25, 8-10, 8-25
1988	7-31, 8-7
1989	7-27, 8-27
1990	7-8, 7-19
1991	6-30, 7-17, 7-23
1992	7-11, 7-21
1993	6-16, 7-3, 7-29, 8-19
1994	7-17
1995	8-2, 8-9
1996	6-28, 8-9, 8-25
1997	6-23, 7-10
1998	7-5, 8-1
1999	7-5, 7-23, 8-4
2000	7-8, 8-12, 8-17, 8-24
2002	6-22, 7-5
2003	8-16
2004	7-13
2005	7-22
2006	8-5
2007	6-28, 7-21, 8-22
2008	7-6, 7-25
2010	7-29
2011	7-3, 7-12, 8-1
2013	7-17, 8-28
2014	6-22, 7-12, 8-23
2015	6-17, 7-30
2016	6-3, 6-26, 7-19, 8-8, 8-18, 8-28
2017	8-21

**Acknowledgements** Luis Gimeno-Sotelo, Iago Algarra and Luis Gimeno acknowledge the financial support received from the Spanish Government within the LAGRIMA project (Grant No. RTI2018-095772-B-I00) and the partial support obtained from the Xunta de Galicia under Project ED431C 2017/64-GRC Programa de Consolidación e Estructuración de Unidades de Investigación Competitivas (Grupos de Referencia Competitiva) and Consellería de Educación e Ordenación Universitaria, both co-funded by the ERDF, in the framework of the Operational Program Galicia 2014–2020. P. de Zea Bermudez was supported by FCT (Fundação para a Ciência e a Tecnologia, Portugal), through the projects PTDC/MAT-STA/28649/2017 and UID/MAT/00006/2020.

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

- Algarra I, Eiras-Barca J, Nieto R, Gimeno L (2019) Global climatology of nocturnal low-level jets and associated moisture sources and sinks. *Atmospheric Research* 229:39–59
- Algarra I, Nieto R, Ramos AM, Eiras-Barca J, Trigo RM, Gimeno L (2020) Significant increase of global anomalous moisture uptake feeding landfalling atmospheric rivers. *Nature communications* 11(1):1–7
- Balkema AA, De Haan L (1974) Residual life time at great age. *The Annals of Probability* pp 792–804
- Basara JB, Maybourn JN, Peirano CM, Tate JE, Brown PJ, Hoey JD, Smith BR (2013) Drought and associated impacts in the Great Plains of the United States—a review. *International Journal of Geosciences* 4(6B):72–81
- Beirlant J, Goegebeur Y, Segers J, Teugels JL (2004) *Statistics of extremes: theory and applications*. John Wiley & Sons
- Burrows DA, Ferguson CR, Campbell MA, Xia G, Bosart LF (2019) An objective classification and analysis of upper-level coupling to the Great Plains low-level jet over the twentieth century. *Journal of Climate* 32(21):7127–7152
- Chen C, Tao WK, Lin PL, Lai GS, Tseng S, Wang TCC (1998) The intensification of the low-level jet during the development of mesoscale convective systems on a mei-yu front. *Monthly Weather Review* 126(2):349–371
- Choulakian V, Stephens MA (2001) Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics* 43(4):478–484
- Coles S (2001) *An introduction to statistical modeling of extreme values*, vol 208. Springer
- Coles SG, Tawn JA (1991) Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B (Methodological)* 53(2):377–392
- Coles SG, Tawn JA (1994) Statistical methods for multivariate extremes: an application to structural design. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 43(1):1–31
- Davison AC, Smith RL (1990) Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)* 52(3):393–425
- Davison AC, Padoan SA, Ribatet M, et al. (2012) Statistical modeling of spatial extremes. *Statistical science* 27(2):161–186
- de Haan L, Ferreira A (2006) *Extreme value theory: an introduction*, vol 21. Springer
- de Zea Bermudez P, Kotz S (2010a) Parameter estimation of the generalized Pareto distribution—Part I. *Journal of Statistical Planning and Inference* 140(6):1353–1373
- de Zea Bermudez P, Kotz S (2010b) Parameter estimation of the generalized Pareto distribution—Part II. *Journal of Statistical Planning and Inference* 140(6):1374–1388
- Drumond A, Stojanovic M, Nieto R, Vicente-Serrano SM, Gimeno L (2019) Linking anomalous moisture transport and drought episodes in the IPCC reference regions. *Bulletin of the American Meteorological Society* 100(8):1481–1498
- Embrechts P, Klüppelberg C, Mikosch T (2013) *Modelling extremal events: for insurance and finance*, vol 33. Springer Science & Business Media
- Emori S, Brown S (2005) Dynamic and thermodynamic changes in mean and extreme precipitation under changed climate. *Geophysical Research Letters* 32(17)

- Field CB, Barros V, Stocker TF, Dahe Q (2012) Managing the risks of extreme events and disasters to advance Climate Change adaptation: special report of the Intergovernmental Panel on Climate Change. Cambridge University Press
- Gilleland E, Katz RW (2016) extRemes 2.0: An extreme value analysis package in R. *Journal of Statistical Software* 72(8):1–39, DOI 10.18637/jss.v072.i08
- Gimeno L, Stohl A, Trigo RM, Dominguez F, Yoshimura K, Yu L, Drumond A, Durán-Quesada AM, Nieto R (2012) Oceanic and terrestrial sources of continental precipitation. *Reviews of Geophysics* 50(4)
- Gimeno L, Dominguez F, Nieto R, Trigo R, Drumond A, Reason CJ, Taschetto AS, Ramos AM, Kumar R, Marengo J (2016) Major mechanisms of atmospheric moisture transport and their role in extreme precipitation events. *Annual Review of Environment and Resources* 41:117–141
- Gimeno-Sotelo L (2021) Univariate and bivariate extremes in meteorology: an application to the Great Plains Low-Level Jet System. Master's thesis, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal, <http://hdl.handle.net/10451/49021>
- Gumbel EJ (1958) *Statistics of Extremes*. Columbia University Press, New York.
- Gumbel EJ (1960) Distributions des valeurs extremes en plusieurs dimensions. *Publ Inst Statist Univ Paris* 9:171–173
- Helfand HM, Schubert SD (1995) Climatology of the simulated Great Plains low-level jet and its contribution to the continental moisture budget of the United States. *Journal of Climate* 8(4):784–806
- Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* 146(730):1999–2049
- Hodges D, Pu Z (2019) Characteristics and variations of low-level jets in the contrasting warm season precipitation extremes of 2006 and 2007 over the Southern Great Plains. *Theoretical and Applied Climatology* 136(1):753–771
- Huser R, Wadsworth JL (2020) Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews: Computational Statistics* p e1537
- Hüsler J, Reiss RD (1989) Maxima of normal random vectors: between independence and complete dependence. *Statistics & Probability Letters* 7(4):283–286
- Joe H (1990) Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistics & Probability letters* 9(1):75–81
- Kotz S, Nadarajah S (2000) *Extreme Value Distributions: Theory and Applications*. Imperial College Press, London.
- Kunkel KE, Easterling DR, Kristovich DA, Gleason B, Stoecker L, Smith R (2012) Meteorological causes of the secular variations in observed extreme precipitation events for the conterminous United States. *Journal of Hydrometeorology* 13(3):1131–1141
- Kunkel KE, Stevens SE, Stevens LE, Karl TR (2020) Observed climatological relationships of extreme daily precipitation events with precipitable water and vertical velocity in the contiguous United States. *Geophysical Research Letters* 47(12):e2019GL086721
- Lilliefors HW (1969) On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. *Journal of the American Statistical Association* 64(325):387–389
- Mackay EB, Challenor PG, Bahaj AS (2011) A comparison of estimators for the generalised Pareto distribution. *Ocean Engineering* 38(11-12):1338–1346
- Masson-Delmotte V, Zhai P, Pörtner HO, Roberts D, Skea J, Shukla PR, Pirani A, Moufouma-Okia W, Péan C, Pidcock R, et al. (2018) Global Warming of 1.5 C. An IPCC Special Report on the impacts of Global Warming of 1.5 C 1:1–9

- Mo KC, Nogues-Paegle J, Paegle J (1995) Physical mechanisms of the 1993 summer floods. *Journal of Atmospheric Sciences* 52(7):879–895
- Pickands J (1975) Statistical inference using extreme order statistics. *Annals of Statistics* 3(1):119–131
- Pitchford KL, London J (1962) The low-level jet as related to nocturnal thunderstorms over Midwest United States. *Journal of Applied Meteorology and Climatology* 1(1):43–47
- Qin D, Plattner G, Tignor M, Allen S, Boschung J, Nauels A, Xia Y, Bex V, Midgley P, et al. (2014) *Climate Change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds TF Stocker et al) pp 5–14
- Raymond C, Horton RM, Zscheischler J, Martius O, AghaKouchak A, Balch J, Bowen SG, Camargo SJ, Hess J, Kornhuber K, et al. (2020) Understanding and managing connected extreme events. *Nature Climate Change* 10(7):611–621
- Ridder NN, Pitman AJ, Westra S, Ukkola A, Do Hong X, Bador M, Hirsch AL, Evans JP, Di Luca A, Zscheischler J (2020) Global hotspots for the occurrence of compound events. *Nature Communications* 11(1):1–10
- Schumacher RS, Johnson RH (2009) Quasi-stationary, extreme-rain-producing convective systems associated with midlevel cyclonic circulations. *Weather and Forecasting* 24(2):555–574
- Seager R, Neelin D, Simpson I, Liu H, Henderson N, Shaw T, Kushnir Y, Ting M, Cook B (2014) Dynamical and thermodynamical causes of large-scale changes in the hydrological cycle over North America in response to global warming. *Journal of Climate* 27(20):7921–7948
- Smith RL (1990) Extreme Value Theory. *Handbook of Applicable Mathematics* 7:437–471
- Squitieri BJ, Gallus WA (2016) WRF forecasts of Great Plains nocturnal low-level jet-driven MCSs. Part I: Correlation between low-level jet forecast accuracy and MCS precipitation forecast skill. *Weather and Forecasting* 31(4):1301–1323
- Squitieri BJ, Gallus Jr WA (2016) WRF forecasts of Great Plains nocturnal low-level jet-driven MCSs. Part II: Differences between strongly and weakly forced low-level jet environments. *Weather and Forecasting* 31(5):1491–1510
- Stensrud DJ (1996) Importance of low-level jets to climate: A review. *Journal of Climate* 9(8):1698–1711
- Stephenson AG (2002) evd: Extreme value distributions. *R News* 2(2):0, URL <https://CRAN.R-project.org/doc/Rnews/>
- Stohl A, James P (2004) A Lagrangian analysis of the atmospheric branch of the global water cycle. Part I: Method description, validation, and demonstration for the August 2002 flooding in Central Europe. *Journal of Hydrometeorology* 5(4):656–678
- Sun Q, Miao C, Duan Q, Ashouri H, Sorooshian S, Hsu KL (2018) A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics* 56(1):79–107
- Tawn JA (1988) Bivariate Extreme Value Theory: models and estimation. *Biometrika* 75(3):397–415
- Trenberth K, Zhang R, et al. (2021) *The Climate Data Guide: Atlantic Multi-decadal Oscillation (AMO)*. Last modified 05 Jun 2021, retrieved from <https://climatedataguide.ucar.edu/climate-data/atlantic-multi-decadal-oscillation-amo>.
- Walters CK, Winkler JA (2001) Airflow configurations of warm season southerly low-level wind maxima in the Great Plains. Part I: Spatial and temporal characteristics and relationship to convection. *Weather and Forecasting* 16(5):513–530

- Weaver SJ, Baxter S, Kumar A (2012) Climatic role of North American low-level jets on US regional tornado activity. *Journal of Climate* 25(19):6666–6683
- Xie P, Chen M, Shi W (2010) CPC unified gauge-based analysis of global daily precipitation. In: Preprints, 24th Conf. on Hydrology, Atlanta, GA, Amer. Meteor. Soc, vol 2
- Zscheischler J, Martius O, Westra S, Bevacqua E, Raymond C, Horton RM, van den Hurk B, AghaKouchak A, Jézéquel A, Mahecha MD, et al. (2020) A typology of compound weather and climate events. *Nature Reviews Earth & Environment* 1(7):333–347