

# Analysis of the tropopause extreme values with radiosounding data

Paul Aizpuru Velasco and R. García-Herrera

Department of Physics of the Earth, Astronomy and Astrophysics II (University Complutense of Madrid)

## Introduction

During the last decades, there is an increasing interest in the study of the connection between the tropopause height and climate change. The tropopause is a thin layer that separates the stratosphere from the troposphere. It is often marked by an abrupt change in the temperature lapse rate from the turbulently mixed troposphere where the temperature decreases with height to the stably stratified stratosphere with constant temperatures or increasing with height. Studies of observations also suggest that the tropopause often marks the location of an abrupt transition in the values of atmospheric properties, (e. g. potential vorticity) and in the concentration of chemical species such as ozone, sulfur dioxide, and various nitrogen oxides [5]. The tropopause can be defined in three different ways:

- Thermal Tropopause (Temperature Lapse Rate) [3]
- Chemical Tropopause (Chemical Compositions) [1]
- Dynamical Tropopause (Potential Vorticity) [5]

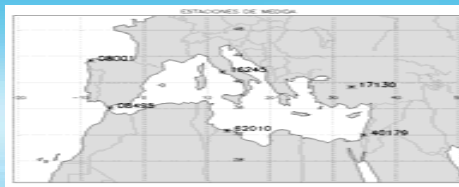


Fig. 1. Station location.

## Objectives and data

**Objective:** Characterize the behaviour of extreme values of the tropopause in the Mediterranean basin using the definition of thermal tropopause, because it is free of spatial limitations.

**Data:** Radiosounding data from IGRA. Daily data over the period of 1981 – 2005 (08495, 16245, 17130 and 40179) and over the period of 1965 – 2005 (08001 and 62010). Using the percentiles 5 and 95 we obtain the extreme values, minimum and maximum respectively [1].

## GEV Analysis of Extremes Values

The analysis of extreme values was made using the GEV analysis. The GEV analysis includes the distribution of Gumbel, Fréchet and Weibull, known like a distribution function of type I, II and III, respectively. To get a better analysis, the three distributions function can be combined into a single one:

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

Defined on the set  $1 + \xi(z - \mu / \sigma) > 0$ , where the parameters satisfy  $-\infty < \mu < \infty$ ,  $\sigma > 0$  and  $-\infty < \xi < \infty$ . This is the generalized extreme value (GEV) family of distributions. The model has three parameters: a location parameter,  $\mu$ ; a scale parameter,  $\sigma$ ; and a shape parameter,  $\xi$ . The type II and type III classes of extreme value distribution correspond respectively to the cases  $\xi > 0$  and  $\xi < 0$  in this parameterization. The subset of the GEV family with  $\xi = 0$  is interpreted as the limit of distribution function as  $\xi \rightarrow 0$ , leading to the Gumbel family [2].

This analysis was made using R project program. This program uses maximum likelihood methods to obtain the parameter values. A potential difficulty with the use of likelihood methods for the GEV concerns the regularity conditions that are required for the usual asymptotic properties associated with the maximum likelihood estimator to be valid. Such conditions are not satisfied by the GEV model because the end-points of the GEV distribution are functions of the parameter values:  $[(\mu - \sigma) / \xi]$  is an upper end-point of the distribution when  $\xi < 0$ , and lower end-point when  $\xi > 0$ . This violation of the usual regularity conditions means that the standard asymptotic likelihood results are not automatically applicable. The problems are solved for the following cases:  $-1 < \xi < -0.5$ ,  $\xi < -1$  and  $\xi \leq -0.5$  [4].

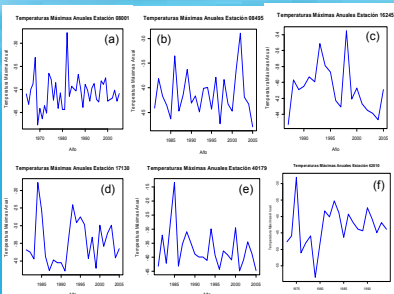


Fig. 2. Annual maximum temperatures variation of stations 08001 (a), 08495 (b), 16245 (c), 17130 (d), 40179 (e) and 62010 (f).

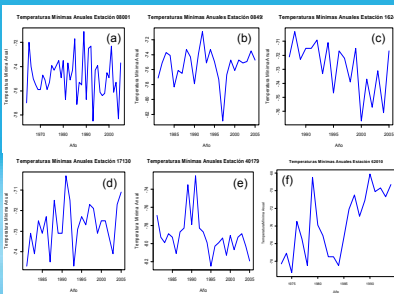


Fig. 5. Annual minimum temperatures variation of stations 08001 (a), 08495 (b), 16245 (c), 17130 (d), 40179 (e) and 62010 (f).

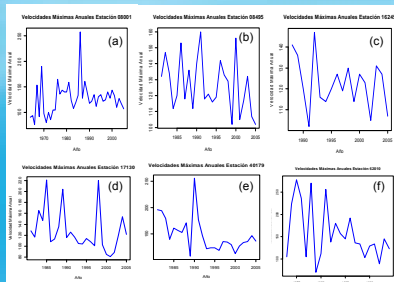


Fig. 8. Annual wind speed variation of stations 08001 (a), 08495 (b), 16245 (c), 17130 (d), 40179 (e) and 62010 (f).

Parameter Stations	$\mu$	$\sigma$	$\xi$
08001 (La Coruña)	-42.18 (0.50)	2.95 (0.34)	-0.07 (0.08)
08495 (Gibraltar)	-43.33 (0.76)	3.36 (0.56)	0.02 (0.16)
16245 (Pratica di Mare)	-42.11 (0.64)	2.42 (0.46)	-0.03 (0.19)
17130 (Ankara)	-37.82 (1.14)	4.63 (0.91)	0.12 (0.24)
40179 (Bet Dagan)	-40.51 (0.97)	4.09 (0.79)	0.23 (0.20)
62010 (Tripoli Airport)	-45.38 (1.41)	6.14 (0.93)	-0.26 (0.10)

Table 1. Maximum Likelihood Estimates for stationary GEV for annual maximum temperatures (the value in bracket is the standard error).

Parameter Stations	$\mu$	$\sigma$	$\xi$	$\mu_1$	$\sigma_1$
08001 (La Coruña)	-75.42 (0.28)	1.64 (0.20)	-0.27 (0.10)	—	—
08495 (Gibraltar)	-76.03 (0.53)	2.45 (0.35)	-0.43 (0.09)	—	—
16245 (Pratica di Mare)	-73.12 (0.49)	1.99 (0.42)	-0.68 (0.17)	—	—
17130 (Ankara)	-73.12 (0.25)	1.11 (0.17)	-0.29 (0.13)	—	—
40179 (Bet Dagan)	-80.03 (0.36)	1.61 (0.25)	0.02 (0.12)	—	—
62010 (Tripoli Airport)	-74.90 (0.64)	0.81 (0.16)	-0.21 (0.14)	5.41 (0.96)	-0.73 (0.35)

Table 2. Maximum Likelihood Estimates for stationary and non-stationary (62010) GEV for annual minimum temperatures (the value in bracket is the standard error).

Parameter Stations	$\mu$	$\sigma$	$\xi$
08001 (La Coruña)	113.47 (4.11)	24.22 (2.90)	-0.01 (0.08)
08495 (Gibraltar)	119.62 (3.45)	14.50 (2.59)	-0.12 (0.21)
16245 (Pratica di Mare)	118.19 (3.02)	11.35 (2.16)	-0.24 (0.18)
17130 (Ankara)	109.81 (5.21)	23.03 (4.10)	0.19 (0.16)
40179 (Bet Dagan)	133.27 (4.58)	19.99 (3.62)	0.18 (0.17)
62010 (Tripoli Airport)	128.45 (10.61)	43.13 (8.08)	0.09 (0.19)

Table 3. Maximum Likelihood Estimates for stationary GEV for annual maximum wind speed (the value in bracket is the standard error).

## Annual Maximum Temperatures

- All the stations have a stationary behaviour.
- The  $\mu$  values are quite similar except for the 17130 station, which has the hottest T. This could indicate that the distribution of all the stations are centered on a similar point except for the corresponding to the 17130 one.
- The values of  $\sigma$  are similar for all the station except for the 17130, 40179 and 62010 ones, whose values are quite greater. This is the reason why all the stations have a similar dispersion, except for the stations with the greatest values. In those cases, we obtain a greatest dispersion getting a greater variability. These variations in the variability make changes in the return level, especially in the 100 years return level.
- The  $\xi$  values can be separated in two groups: the lowest ones corresponding to the three first stations and the highest ones for the rest. In all the stations, except for the 62010 one, the standard error has a greater value than the parameter value, that is to say, it is not significantly different from 0 and therefore the parameter  $\xi \rightarrow 0$ , what in turn implies that the distributions are of Gumbel type. This is not the case of the 62010 station, where  $\xi < 0$  and the standard error is not higher than the parameter value, thus it is a Weibull distribution.

## Annual Minimum Temperatures

- All the stations have a stationary behaviour, except for the 62010 station (with a linear trend in  $\mu$  and an exponential one in  $\sigma$ ).
- The values of  $\mu$  are quite similar, (mean value of annual minimum T is analogous in all the stations) except for the 40179 station, which has a lower one.
- The values of  $\sigma$  are also quite similar, except for the 62010 station, which it has a lower value. Thus, all the stations have a similar dispersion, except for the 62010 station with less variability for  $t = 0$ .
- The  $\xi$  parameter is similar except for the 40179 station, where we find a lower absolute value, than the others, and for the 08495 and 16245 stations, with higher absolute values.
- The standard errors of  $\xi$  are lower than the parameter values. Hence, negative values of  $\xi$  are significant and thus, the data distributions belong to Weibull type. However, for the 40179 and 62010 stations,  $\xi$  show a value not significantly different from 0 (Data distribution of Gumbel type).

## Annual Maximum Wind Speed at the Tropopause Level

- All the stations have a stationary behaviour.
- Like in the other cases the  $\mu$  values are quite similar, except for the 40179 and 62010 stations, where we get higher values. This could indicate that the distribution of all the stations are centered on a similar point, except for the 40179 and 62010 ones.
- Although the value of  $\sigma$  is similar in all the stations, we can find two different groups: the stations with high values of  $\sigma$  and then greater variability (08001, 17130, 40179 and 62010); the stations with low values of  $\sigma$  (08495 and 16245). This variation in  $\sigma$  is also appreciated in the return levels plot, especially in the 100-years return level, obtaining a greater interval.
- $\xi$  has a greater standard error than the parameter value as it happens with the annual maximum temperatures. Hence,  $\xi$  is not significantly different from 0 (Gumbel type distributions, for all cases).

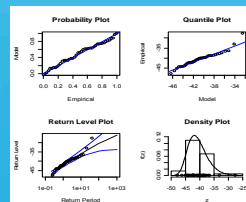


Fig. 3. Diagnostic plots for GEV fit for station 08001.

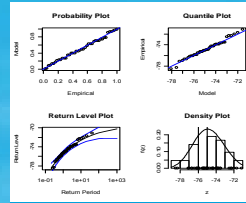


Fig. 6. Diagnostic plots for GEV analysis for station 08001.

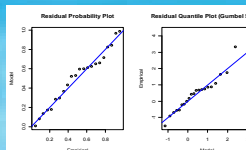


Fig. 7. Diagnostic plots for the non-stationary GEV analysis for station 62010.

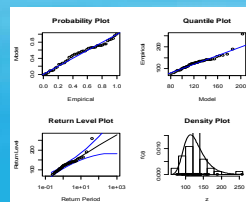


Fig. 9. Diagnostic plots for GEV analysis for station 08001.

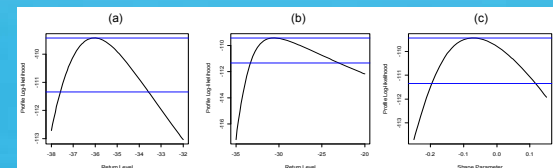


Fig. 4. Profile likelihood for (a) 10 and 100 years (b) return level, and profile likelihood for  $\xi$  (c), for station 08001.

## Stationary behaviour

- Four different diagnostic plots are obtained. With the probability and quantile plots (upper panels of Fig. 3, 6 and 9), a comparison between the different empirical properties and the distribution ones obtained with the GEV is made: data over the diagonal indicate the goodness of fit with this distribution.
- The return level and density plot give more information about the GEV analysis. The return level plot represents the different return levels for different return period with a 95% confidence level. The density plot gives information about the data distributions.
- We can get the profile log-likelihood for the 10-and 100-years return levels with a 95% confidence interval. We can also get the profile log-likelihood for  $\xi$  with a 95% confidence interval. This helps us to determine the type of distribution. (Fig 4 (a), (b) and (c)).

## Non-Stationary behaviour

- Apart from the values of the parameters of location ( $\mu$ ), scale ( $\sigma$ ) and shape ( $\xi$ ) of the distribution, other values are obtained ( $\mu_1$ ,  $\sigma_1$  and  $\xi_1$ ), which represent the temporal dependencies of the parameters. The  $i$  subindex values can be 0,1,2,..., where 0 refers to the values of the parameters in the instant  $t=0$  and the others to the changing rate of each parameter.
- In this case, we obtain only two residual diagnostic plots: probability and quantile plots (Fig. 7). Just like in the stationary case, the data show the validity of the GEV model.

$$\mu(t) = \mu_0 + \mu_1 t \quad \sigma(t) = \exp(\sigma_0 + \sigma_1 t) \quad \xi = cte$$

## Conclusions and ongoing work

- All the stations have a stationary behaviour for all the variables, except for the annual minimum temperature on station 62010, where we find a non-stationary behaviour with a linear trend in  $\mu$ , and an exponential trend in  $\sigma$ .
- The parameters show homogeneous values, except for the stations: 17130 in the annual maximum temperatures, 40179 in the annual minimum temperatures and 40179 and 62010 in the annual maximum wind speed.
- The parameters have a different behaviour for different variables, as a result the variability and then the return level (especially the 100 year one) are different. We also find a different behaviour in the distributions: although the great majority of the stations have a Gumbel distribution, there are some cases with a Weibull one, (e.g. annual minimum temperature stations 08001, 08495, 16245 and 17130, and annual maximum temperature station 62010).
- Ongoing work: Identification of the mechanism that explains those heterogeneity; the roles of the local and synoptic scale factors and the role of the teleconnections as well.

## References

[1] Ariel Cabanellas, J. A., 2007. Análisis Climático de la tropopausa mediante datos de radiosondeo. Dpto. de Física Aplicada. Grupo de Física de la Atmósfera y del Océano, Universidad de Vigo.

[2] Coles, S., 2001. *An Introduction to Statistical Modeling of Extreme Values*, Springer Series in Statistics.

[3] Hoinka, K. P., 1998. *Statistics of the Global Tropopause Pressure*, American Meteorological Society, Vol. 126, 3303-3325.

[4] Smith, R. L., 1995. *Maximum likelihood estimation in a class of non-regular cases*, Biometrika, 72, 67-90.

[5] WMO, 1998. *Atmospheric ozone 1985: Global ozone research and monitoring report*, Report 16, World Meteorological Organization, Geneva, Switzerland, 392 pp.