MODELLING OF WIND SPEED WITH EOT TECHNIQUES

D. Lahoz and M. San Miguel

Abstract. The Excesses Over Threshold (EOT)(cf. [1]) techniques have been applied in another researches over environmental data set with good results. They have been used for the creation of model of periods of drought (cf. [2]), hurricanes (cf. [3]), air pollution(cf. [4]),... . In this paper the series of wind speed are fitness with EOT techniques. Another magnitudes associated are also studied, the duration of wind squall or the maximum wind in the squall, for example. The techniques EOT are applied to wind speed data in the Ebro Valley.

Keywords: Excesses Over Threshold, Wind speed, SYNOP, Wind Farm

AMS classification: 60G70,60K40

§1. Introduction

This paper is included in the implementation of the Physical- Mathematical Model ([7]) for the study of the wind and their proprieties in wind farms of the Ebro Valley. This article studies the fitness of the wind speed to know probability distributions with techniques *Excesses Over Thresholds* (EOT).

In the first part of the paper the EOT techniques are described, the notation, the probability distributions used for the fitness and the definitions necessaries for to work. The test for the selection of the thresholds and the techniques for the validation of the fitness for the fixed thresholds are showed. These one are divided in graphical tools and numerical test.

The second part shows the descriptive studio of the wind speed series. The EOT techniques have appliqued a three different series. The first collection are extracted of SYNOP data of Airport of Zaragoza collected for the INM (Instituto Nacional de Meteorología). The another series are a data collection of two wind farms sited in the Ebro Valley.

The application of EOT techniques to the studied series are described in the following section. This shows the test and graphical results for the calculation of threshold of fitness for the three studied series. Besides, the series are fitness a Compound Poisson Process for the fixed thresholds.

The last part says the conclusions of the article and the possible actions for to improve the results of this work.

§2. Excesses Over Thresholds (EOT)

The Excesses Over Thresholds statistical techniques (EOT) are a particular method of the techniques of Extremes Values ([1], [4], [6]) , and they are used for the study probabilistic of one data, and the fitness of this to probabilistic distribution. The EOT techniques proportioned tools for the calculation of probabilistic distribution of fitness and tests for the verification of goodness of this.
The initial hypothesis in the techniques EOT is that the processes of occurrences of the extreme event behaves like a know probability distribution. Usually the selected probability distribution for the number of event in a interval is a Poisson distribution. Then the recurrent times follows a Exponential.
The purpose of the EOT techniques is to verify if the times between two consecutive increasing crossing over a prefixed thresholds are independent and equally distributed exponential accomplishments ([1, 2]).
The figure 1 shows the excesses over thresholds 5 m/s in one part of the wind speed in the Airport of Zaragoza. There are two variables associated to an excesses over thresholds processes, the number of event in a time interval and the times between two consecutive crosses over the umbral (T_i in the figure 1).

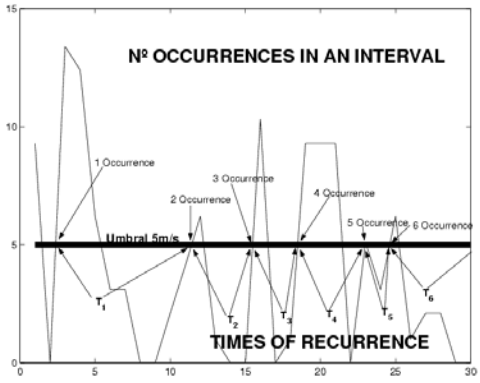

Figure 1: Excesses over Thresholds, Occurrences and Times of Recurrence in a EOT processes in the wind speed series

In the figure 1 the Times of Recurrence are measure between one excess over threshold and the next excess, this is, between the initial point and the next initial point. There are others forms to take the references points as show the figure 2. Others references are between maximum point of the rank and maximum point of the next rank and between medium point of the squall and medium point of the next cross.

The EOT techniques also permits to relate to the Poisson Processes another magnitudes different to the variable studied. For example,in this case, besides of wind speed, the duration and the medium speed of the gust are the two variables related to the processes. This new processes is defined as Compound Poisson Processes:

Definition 1. A Poisson Processes is defined as Compound Poisson Processes when all occurrence of the Poisson Processes have got associated a magnitudes vector $(X_1, ..., X_n)$ indepen-

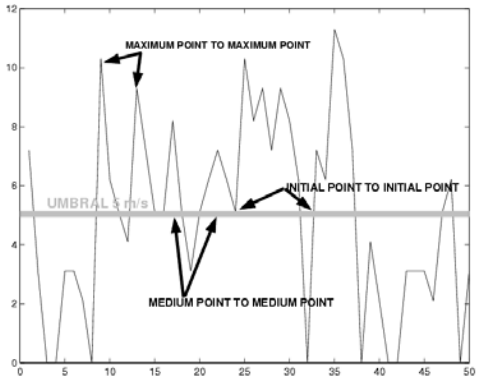

Figure 2: Different references points for the definition of Times of Recurrence

dents and equality distributed with marginal distributions $F_1, ..., F_n$.

In summary, to proof that the wind speed series are a Compound Poisson Processes it has followed the next steps:

- To find the thresholds for which the wind speed data is a Poisson Processes. The Times of Recurrence must have a Exponential character and the Number of Occurrences must have a Poisson character. The selection of the thresholds can have by means of climatical, economic or numerical criterions. This paper only studies the numerical criterions.

- Next, with the fixed thresholds, the collection is fitness to a Poisson Processes. Different tests are applied to proof that the Times of Recurrence is a Exponential Distribution and the Number of Recurrences is a Poisson Distribution.

- For last, if the process is a Compound Poisson Process the magnitudes associates must be independents between its and independents with the Times of Recurrence.

§3. Wind speed series studied

The EOT techniques have been applied to three different wind speed series. The first one is the wind speed in the Airport of Zaragoza measured for the INM, and it is a part of the SYNOP data. The another two series are the wind speed in wind farms sited in the Ebro Valley.
The descriptive study of the series is showed in the next subsections, and a greater study can be found in [7].

3.1. Wind speed in the Airport of Zaragoza

The data collection have 40279 measurements. The INM measures each six hours (at 0, 6, 12 and 18 o'clock) a SYNOP data, including the wind speed. The measure goes from 1972 to 1999. The missing data (the 5% of the all data) is completed for the mean of the four neighbours (two previous and two back data).
The description of wind speed in the Airport of Zaragoza can be seen in the figure 3

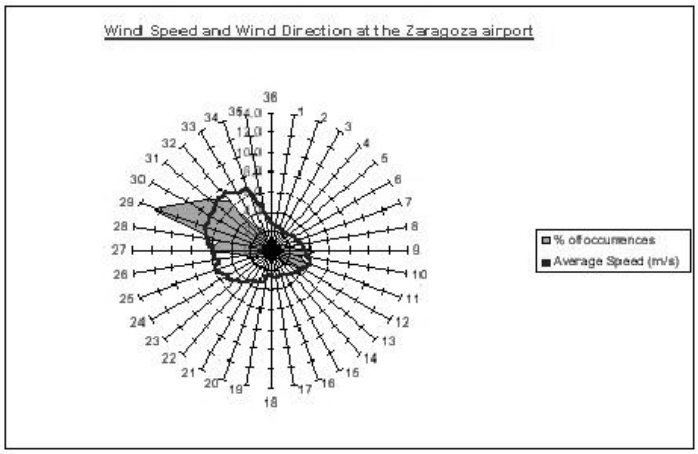

Figure 3: Data distribution of the series of the Airport of Zaragoza

3.2. Wind speed in the Wind Farm1

The second data are measurements realized in a wind farm sited in the Ebro Valley. In this farm some wind generator have got tools to be able to measure different meteorologic variables: wind speed, wind direction, pressure, ... Each ten minutes a new measure. Usually they take measures to three different levels. In this article the data used is the maximum wind speed to 40 metres of floor level.
The series begin in January of 1997 and they finish in March of 2000, with a total of 170747 measures. The collection only have got twenty missing data.
The data distribution of the wind speed is showed in figure 4 (left figure).

3.3. Wind speed in the Wind Farm2

The Wind Farm2, also, is sited in the Ebro Valley but in different location that the Wind Farm1. The measurements are as in Wind Farm1. Here the series begin in January of 1998 and they finish in February of 2000, with 113566 measures.
The right plot of the figure 4 shows the distribution of the maximum wind speed at 40 metres in the Wind Farm2.

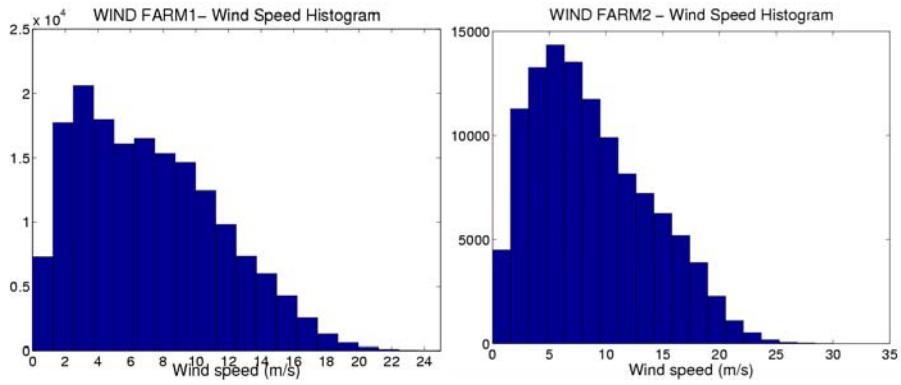

Figure 4: Data distribution of wind speed in the Wind Farm1(left) and in the Wind Farm2(right)

§4. EOT techniques in wind speed series

This section goes to describe the results obtained for the EOT techniques of section 2 in the series shows of section 3.

4.1. Selection the threshold

The first step is to find the umbral for which the series is a Poisson Process. There are two aspects to study. The number the cross over thresholds in a interval fixed is a Poisson distribution and the time of recurrence is a Exponential distribution.
The tests are the same to proof the two aspects, the rate plot and the Castro plot.
The rate plot shows the coefficient variation in the Exponential distribution and the rate between mean and variance in the Poisson distribution, for different thresholds. The rate must to be 1 for Exponential and Poisson distributions. In figure 5 and 6 are showed the Poisson rate plot and the Exponential rate plot for the three series. For the Poisson distribution the series are divided in intervals of same amplitude and in each one of thats the number of occurrence is summed.
The Castro plot ([3],[2]) permits to proof the Poisson character of a variable. If a data collec-

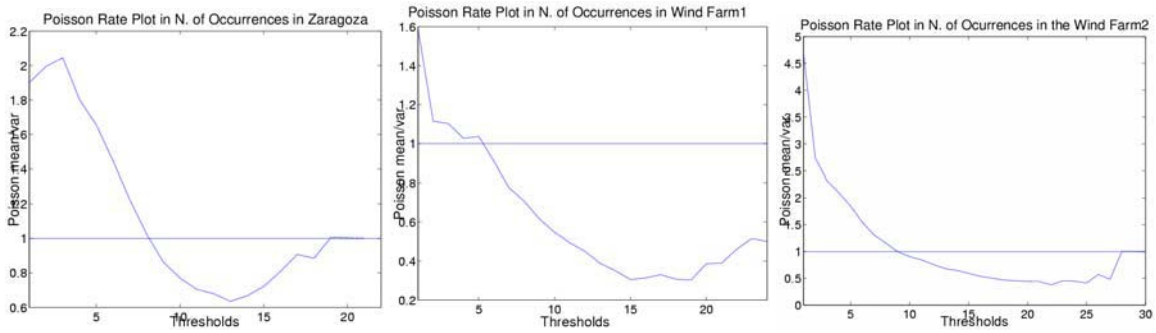

Figure 5: Poisson rate plot for wind speed in the Airport of Zaragoza(left), in the Wind Farm1(center) and in the Wind Farm2(right)

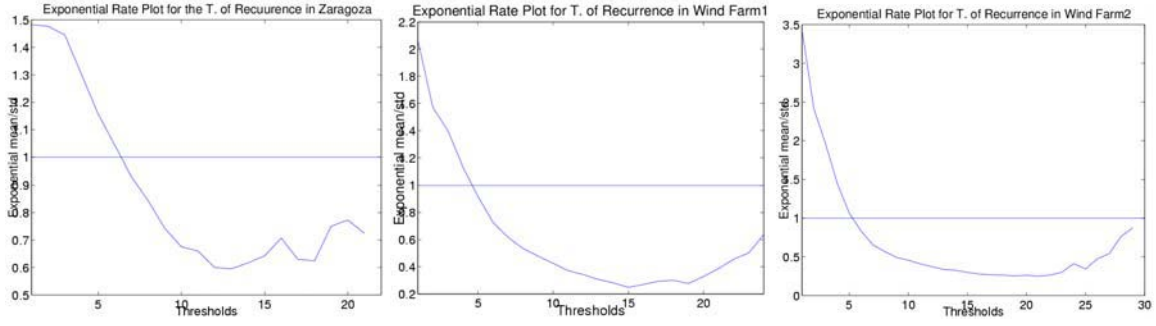

Figure 6: Exponential rate plot for wind speed in the Airport of Zaragoza(left), in the Wind Farm1 (center) and in the Wind Farm2(right)

tion is a Poisson distribution then the series have a lineal character in $\widehat{Y}_c(g)$ (the logarithmic of the empiric probability generator functional) respect to c in different functions g. In the figure

7 can to see the Castro plot in the three series with the thresholds selected to work in this paper for each one. The functions g_i of the figure 7 have got different comportment, the first one is lineal increased, the second lineal decreased, other lineal to steps and the last convex.

The umbral selected for to continue the study for the three series are showed in the figure 7.

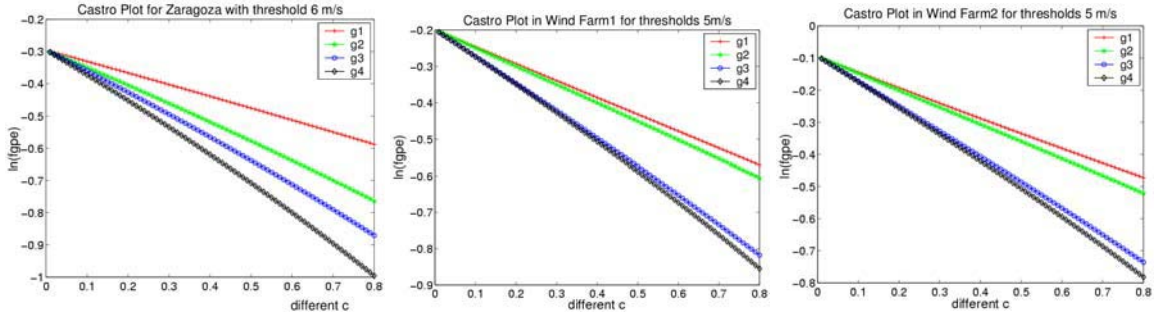

Figure 7: Castro plot for the wind speed in the Airport of Zaragoza and thresholds 6 m/s(left), in the Wind Farm1 and thresholds 5 m/s(center) and ln the Wind Farm2 with thresholds 5 m/s(right)

Those have been elected because numerical or graphics criterions.
Others different criterions are possible, economics (usually with plus of 4 or 5 metres/second the wind generator generate power) or climatics (depend of the localization, different for each series).

4.2. Fitness to Poisson Process

The Q-Q plot, the rate and the χ^2 and Kolmogorov- Smirnov tests are used to probe that the series of Number of Occurrences and Times of Recurrence forme a Poisson Process for the selected thresholds in the section before. The first one is fitted to a Poisson Distribution and the second one to a Exponential Distribution. Also it is analogous to prove that the Times of Waiting $W_1, ..., W_n$ have the same distribution as the order statistic corresponding to n independent random variables uniformly distributed on the interval $(0, t)$.([5]). The Times of Waiting are the instants of time of the occurrences in the studied interval.
In the figures 8 and 9 can to see the Q-Q plot for the three series with the Poisson distribution and the Number of Occurrences (figure 8) and with the Exponential Distribution and the Times of Recurrence (figure 9) for the fixed thresholds.

The results of the test of hypothesis, p-value of χ^2 and Kolmogorov- Smirnov test, for the three variables in the three series are showed in the table 1.

4.3. Magnitudes relates

In section 2 it has been defined the Compound Poisson Processes (definition 1) which permit to study the magnitudes relates to the Poisson Processes as marginal distributions of process. The Poisson Process can be considerate Compound if the magnitudes are independents between them and independents with the occurrence process.
The magnitudes associates with the Poisson Process here studied are the duration of the excesses and the medium speed over the thresholds. To proof the independent the series associates have used the Spearman and the Kendall test.
The study of the associates magnitudes with that tests results negative in the two magnitudes in the three series. The Spearman test and the Kendall test refused the independent of the magnitudes series and the independent of those with the times of recurrence and the fixed thresholds.

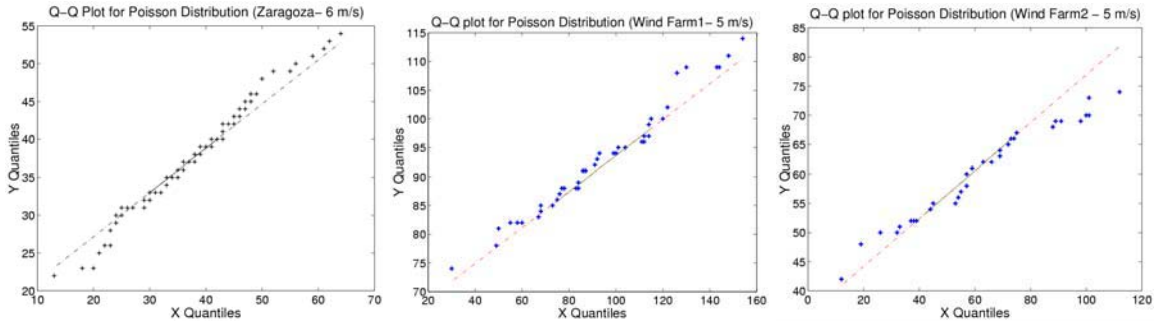

Figure 8: Q-Q plot for the number of occurrences for the wind speed in the Airport of Zaragoza and thresholds 6 m/s(left), in the Wind Farm1 and thresholds 5 m/s(center) and in the Wind Farm2 with thresholds 5 m/s(right)

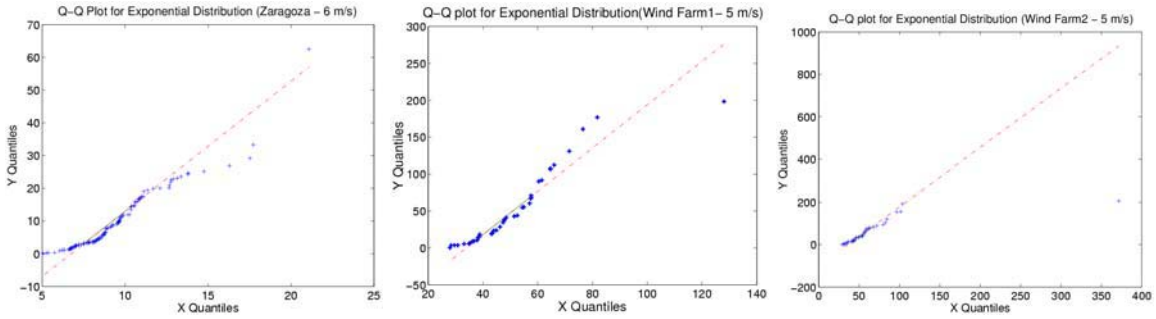

Figure 9: Q-Q plot for the times of recurrence for the wind speed in the Airport of Zaragoza and thresholds 6 m/s(left), in the Wind Farm1 and thresholds 5 m/s(center) and in the Wind Farm2 with thresholds 5 m/s(right)

§5. Conclusions

In this work it has been applied the EOT techniques for the fitness of wind speed. It has been showed the methodology of EOT and the results obtained at three series, all sited in the Ebro Valley. The conclusions of the study is that EOT techniques are a good tool for the probabilistic fitness for the wind speed series. Also, in cases applied, the results obtained in a initial study with the wind speed series and EOT techniques are the next:

- for the series of the wind speed in the Airport of Zaragoza the adjustment to Poisson Process can be acceptable to near to 6 m/s. The test refused the hypothesis of the Compound Poisson Process. The magnitudes associates are not independent.

- for the series in the wind farm1 and the wind farm2, for the threshold 5 m/s is accepted the Poisson character in some tests and plots but not the PPC proprieties.

As it can see in the article not all tests accept the previous results, a better fitness are necessaries and the end of the article some ideas are presented. If one series is fitting to a Poisson Process for a fixed umbral, then for superior thresholds it also have a Poisson comport. If in the wind farm1 for the threshold 5 m/s the series is a Poisson Process, for 6, 7 o more m/s too have a Poisson character. This properties aren't reliable in none of the three series.
For the study of the Compound Poisson Process is necessary to take other times intervals,

months, years, weather stations,... Another possible change is to study the series only in little intervals of time where the comport of the wind speed is more homologous. For example, to study the fitness only for months, weather stations,...
At last, it can be probe the study the wind speed only for some wind directions. As it can be seen in figure 3 the wind speed depend of the wind direction. In the Ebro Valley there are two direction dominated with different comport, this are denominated "Cierzo" and "Bochorno". The EOT techniques can be applied only to this directions.

References

[1] A. C. Davison and R. L. Smith, *Model for exceedances over high thresholds*. J. R. Stat. Soc. B 52(num. 3), pages. 393-442(1990).

[2] A. C. Cebrián. *Análisis, modelización y predicción de episodios de sequia*. Memoria doctoral. Zaragoza University(1999).

[3] J. Castro and V. Pérez- Abreu. *Some statistical analysis of the cluster point processes of hurricanes on the coast of Mexico* .Statistics for the enviorenment 2: Water related issues. John Wiley and Sons.

[4] R. L. Smith. *Thresholds methods for sample extremes*. Statistical extremes and application. NATO Advanced Study Institute, Reidel, pages. 621-638(1990)

[5] E. Parzen. *Stochastic Processes*. Holden-Day Inc.. pages 132- 142 (1962).

[6] P. F. Rasmussen et al. *The POT method for flood estimation*, K. Hipel ed. Extreme values: floods and droughts, Vol 1.Klumer Academic Publishers, pp. 15-26 (1994)

[7] J. A. Escudero Lázaro and D. Lahoz Arnedo. *Physical- Mathematical wind prediction model*. VII Jornadas Zaragoza- Pau de Matemática Aplicada y Estadística.2001

[8] R. L. Smith. *Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone*. Statistical Science 4(4). Pag. 376-393.(1989)

Lahoz D., San Miguel M.
Department: Métodos Estadísticos. Facultad de Ciencias
University: Universidad de Zaragoza
davidla@unizar.es and msm@unizar.es

	Number of Occurrences			Times of Recurrence			Times of Waiting	
Airport of Zaragoza(6m/s)	mean/variance	χ^2	K-S	V.C.	χ^2	K-S	χ^2	K-S
	1.6	0.50	0.0004	1.15	0.92	0.00000	0.10	0
Wind Farm1(5m/s)	mean/variance	χ^2	K-S	V.C.	χ^2	K-S	χ^2	K-S
	1.23	0.51	0.0055	0.92	0.45	0.00004	0.47	0
Wind Farm2(4m/s)	mean/variance	χ^2	K-S	V.C.	χ^2	K-S	χ^2	K-S
	1.8	0.76	0.0220	1.06	0.77	0.00320	0.60	1

Table 1: p-Value for χ^2 and Kolmogorov -Smirnov tests in the three series and fixed thresholds .The table shows the rate mean/std deviation and the rate mean/varianze for the Times of Recurrence (Exponential distribution) and the Number of Occurrence (Poisson distribution), respectively.