

Conference Proceedings Paper

An overview of statistical methods for studying the extreme rainfalls in Mediterranean.

Georgia Lazoglou ^{1,*} and Christina Anagnostopoulou ¹

Published: 17 July 2017

¹ Department of Meteorology-Climatology, School of Geology, Aristotle University of Thessaloniki, Thessaloniki, Greece, 54124

* Correspondence: glazoglou@geo.auth.gr; Tel.: +30 2310 998414

Abstract: Extreme rainfall is one of the most devastating natural events. The frequency and intensity of these events has increased. This trend will likely continue as the effects of climate change become more pronounced. As a consequence, it is necessary to evaluate the different statistical methods that assess the occurrence of the extreme rainfalls. This research evaluates some of the most important statistical methods that are used for the analysis of the extreme precipitation events. Extreme Value Theory is applied on ten station data located in the Mediterranean region. Furthermore, its two main fundamental approaches (Block-Maxima and POT) and three commonly used methods for the calculation of the extreme distributions parameters (Maximum Likelihood, L-Moments, and Bayesian) are analyzed and compared. The results showed that the Generalized Pareto Distribution provides better theoretical justification to predict extreme precipitation compared to Generalized Extreme Value distribution while in the majority of stations the most accurate parameters for the highest precipitation levels are estimated with the Bayesian method. Extreme precipitation for return period of 50, 150 and 300 years were finally obtained which indicated that Generalized Extreme Value Distribution with Bayesian estimator presents the highest return levels for western stations, while for the eastern Mediterranean stations the Generalized Pareto Distribution with Bayesian estimator presents the highest ones.

Keywords: GEV; Pareto; Mediterranean; Bayesian

1. Introduction

The frequency and the intensity of extreme rainfall episodes has increased in many land areas around the world [1]. Additionally, in many European countries such as the Czech Republic [2], northern Italy [3] and Germany [4], the extreme precipitation episodes are observed with a high frequency. Due to their severe impacts in many fields (agriculture, economy etc), these episodes have been examined commonly, using either precipitation indices or real data. Although it is necessary to investigate the future behavior of extreme events especially for regions such as the Mediterranean, especially the central Mediterranean, where the flood and drought risks are high [5]. For this purpose, it is useful to evaluate the statistical tools that are used to study extreme climatic episodes which is the main purpose of the present study.

Extreme Value Theory (EVT) is one of the most usual statistical methodologies which is used for the description of rare events (climatic or not) and specifically its two fundamental approaches the Generalized Extreme Value distribution (GEV) and the Generalized Pareto distribution (GPD). EVT analyzes the tail of the studied parameter distribution, which describes the extreme values, commonly produced by the Block Maxima or the Peak Over Threshold (POT) technique. For the application of the Block maxima, the data set is divided into same-sized non overlapping periods and

from each one the maximum (or minimum) value is chosen. This new data set includes the extreme values. According to researches [6-8] the GEV distribution, which combines three different statistical families (Gumbel, Fréchet, and Weibull) can fit the extreme data set with a high accuracy. Another way of defining extreme values is to choose all the values that exceed a specific limit-threshold. This technique is POT and GPD distribution fit this data set satisfactorily [7]. Both Block Maxima and POT methods have some difficulties on their application. Firstly, choosing the most appropriate period-size (for Block Maxima) is not an easy procedure, according to Coles [9] too large sub-periods produce only a few values, while too small sub-periods could lead to biases. Secondly, in POT procedure the choice of the most appropriate threshold includes an uncertainty [10] as it is a scientist's choice and this can affect the final results.

For the analysis and the description of the extremes distribution, three parameters were used: shape, scale and location. The values of these parameters can be calculated, using several statistical methods such as the Maximum likelihood (MLE), the L-moments and the Bayesian method. The MLE is an often-used method due to its reliable results and its simplicity. Mainly it is used for large data sets, as its results could be doubtful for samples minimum then 50 values [11, 12]. The L-moment method is based on the linear combinations of probability weighed moments (PWMs). Finally the Bayesian is an accurate but complex method, which uses the initial data and offers supplementary information about it from alternative sources, through the prior distribution.

The main purpose of this study is to evaluate some of the most commonly used statistical methods and techniques that are used for the extreme analysis. That is achieved using the extreme precipitations around the Mediterranean region. Additionally it offers an overview of three statistical techniques, for the estimation of the extreme distribution parameters. Finally based on the above methods and techniques this study presents a measurement of the rainfall risk for the Mediterranean stations through the return levels.

2. Experiments

2.1 Data

For this research daily values of precipitation from 10 Mediterranean stations (Malaga, Barcelona, Nice, Bastia, Cagliari, Verona-Villa-Franca, Bologna, Gospic, Split-Marjan, Athens, and Thessaloniki) were used. Studying extreme climatic events requires an accurate data set in order to achieve reliable results. As a consequence the 10 studied stations, which cover the whole Mediterranean region, were chosen as their percentage of missing values does not exceed 2%. For the majority of stations, the data came from the "European Climate Assessment & Dataset" (ECAD) [13], while for Athens and Thessaloniki, the data was provided by the National Observatory of Athens and by the Department of Meteorology and Climatology of the Aristotle University of Thessaloniki respectively. Furthermore, the data covers a period of 60 years (1951-2010) and the location of each station can be observed in Figure 1.

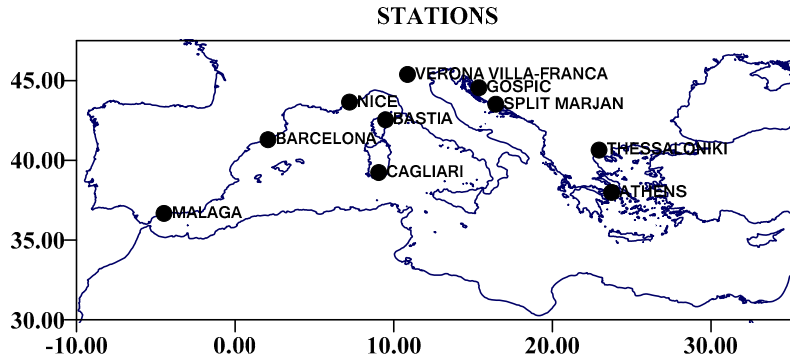


Figure 1. The geographical distribution of the 10 studied stations.

2.2 Methodology

In the present study the two fundamental approaches of EVT were applied on the extreme precipitation events of the Mediterranean stations. Firstly the extreme values of precipitation were chosen for each station, using both the Block Maxima and the POT technique. Due to the fact that the data set is daily values of precipitation, the Block Maxima applied on annual sub-periods. As a result, the maximum precipitation episode of each year was chosen as extreme. According to the second technique, POT, the rainfall episodes with amounts higher than 99% were chosen as extremes. The percentage of 99% is in agreement with Anagnostopoulou and Tolika [14] who proved that this is the most appropriate threshold for extreme rainfalls in Europe. The two new produced data sets were fitted on both the Generalized Extreme Value Distribution (GEV) and the Generalized Pareto Distribution (GPD) respectively. Moreover, to check if the chosen distributions can describe the extreme data sets satisfactorily, both the Kolmogorov Smirnov and the Anderson Darling tests were applied. Finally, three statistical methods that are used for finding the distributions parameters were evaluated. These methods are : the Maximum Likelihood estimation method (MLE), the L-moments method and the Bayesian method.

3. Results

3.1 Climatology of the 10 Mediterranean Stations

In order to study the climatic extreme events of a region, it is helpful to know its climatology. Table 1 presents the mean daily precipitation and the absolute maximum rainfall amount that was recorded on each station.

Table 1. An overview of rainfalls in the 10 studied stations

STATIONS		PRECIPITATION	
Name	Country	Mean daily (mm)	Absolute maximum (mm)
Malaga	Spain	1.55	313.00
Barcelona	Spain	1.72	194.80
Nice	France	2.20	191.40
Bastia	France	2.12	232.40
Cagliari	Italy	1.14	109.60
Verona Villa Franca	Italy	2.17	198.00
Gospic	Croatia	3.79	141.00
Split Marjan	Croatia	2.23	131.60
Athens	Greece	1.09	116.00
Thessaloniki	Greece	1.24	98.00

According to Table 1 Gospic is the station with the highest mean daily precipitation, while Malaga has the highest absolute maximum amount. As it concerns the lowest values, both the minimum mean daily precipitation and the lowest absolute maximum rainfall amount observed in Greek stations (Athens and Thessaloniki respectively). In general, it is obvious from Table 1 that the mean daily precipitation is not related with the absolute maximum amount. For instance, Malaga had the most extreme rainfall episode, although its mean daily precipitation is low.

3.2 Goodness of fit test with Shape Diagram

After the application of two goodness of fit tests (Kolmogorov-Smirnov, Anderson Darling), it is proved that GEV and GPD distributions can characterize the extreme rainfall behavior with high accuracy. Following that, the parameters of each distribution were estimated using three different methods (MLE, L-Moments and the Bayesian methods). Shape parameter is a helpful tool for checking whether the chosen distribution is appropriate or not. Particularly as rainfalls haven't got an upper threshold the appropriate distribution should not be bounded above. For example the Weibull distribution (negative shape parameter), is inappropriate for precipitation. On the contrary, the Frechet distribution (unbounded above distribution) which has positive shape value, could be appropriate for precipitations. Based on that, Figure 9, presents an overview of the shape diagrams of each station. It is shown that GEV-L is not appropriate for the description of extreme rainfalls in the majority of stations. Moreover, GPD-M and GEV-M are inappropriate for Gospic, while GPD-L for Verona-Villa-Franca.

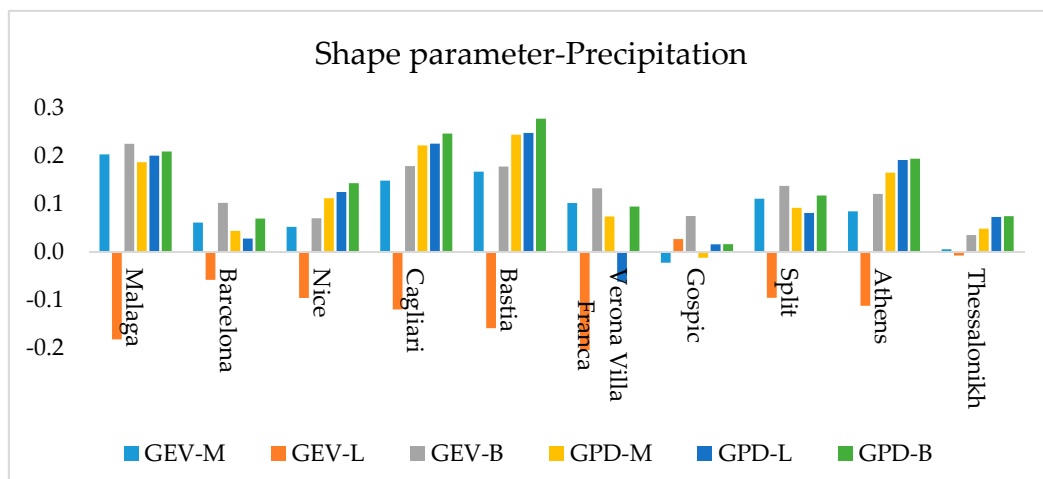


Figure 2. Shape parameter of the 10 studied stations, estimated by six different statistical methods (GEV-M, GEV-L, GEV-B, GPD-M, GPD-L, GPD-B)

3.2 QQ-plots

Quantile-quantile plot (QQ plot) is a useful tool used to check if the assumed distribution fits satisfactorily the studied data set. In Figures 3A-3J GEV and GDP distributions using the MLE, L-moments and Bayesian techniques are compared to each station. According to Figure 3 both GEV and GPD distributions characterize satisfactorily the extreme rainfalls in the majority of the stations. Generally the QQ plots reveal that the GPD-M, GPD-L and GPD-B methods are more appropriate for small and medium rainfall values, compared with GEV-M, GEV-L and GEV-B which deviate a little from the reference line. As it concerns the greatest values of the extreme rainfalls, they are accurately fitted by both GEV-B and GPD-B distributions in almost all stations.

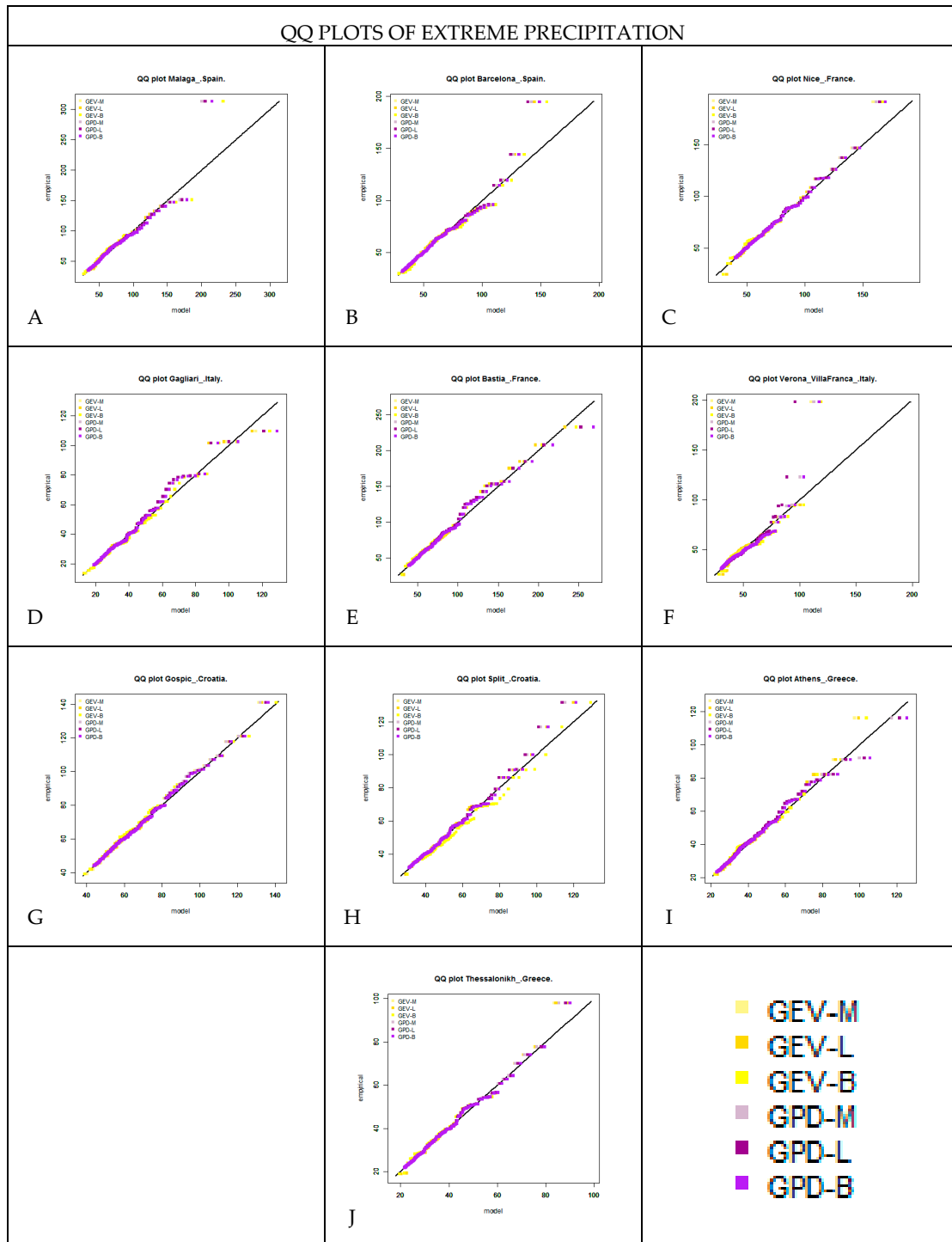


Figure 3. QQ plots for the 10 studied stations using six different statistical methods (GEV-M, GEV-L, GEV-B, GPD-M, GPD-L, GPD-B).

3.3 Return Levels

The estimation of the precipitation return levels, offers a common way to estimate the climatic risk, usually based on historical data. In this study, the return levels of extreme rainfalls were calculated for both GEV and GPD distribution using MLE, L-moments and Bayesian techniques (Table 2). According to Table 2, Bastia presents the highest return levels of extreme precipitation while

Thessaloniki the lowest. That is observed in the three return periods, regardless of the method and the distribution which is used. In Table 2 it appears that GPD-B and GEV-B give the highest return levels for the three return periods and for all stations except of Verona-Villa-Franca. For the lowest return levels the results are more complex, as in almost five stations the GEV distribution (mainly GEV-M) assesses the lowest return levels whereas GPD distribution (GPD-M and GPD-L) presents the lowest of the rest of the stations. What is also observed in Table 2, is that the Bayesian methods do not give any lowest return level neither with GEV nor with the GPD distribution. Finally, in the western Mediterranean and in stations with the highest altitude, the highest precipitation return levels are recorded with the GEV-B method while for the eastern stations with the GPD-B. Moreover, the GPD distribution either with MLE or with the L- moments method present the lowest return values for the western stations, while the GEV-M for the eastern ones.

Table 2. The precipitation return levels for the 10 stations and for three return periods (50, 150 and 300 years). The red cells shows the highest values of each station while the blue shows the lowest ones.

	Malaga	Barcelona	Nice	Gagliari	Bastia	Verona Villa- Franca	Gospic	Split	Athens	Thessaloni khi
50										
GEV-M	194.9	140.1	153.6	110.8	222.2	106.4	128.3	115.9	93.9	81.2
GEV-L	192.9	139.7	158.7	109	221.8	113.1	129.7	115.7	95.8	81.9
GEV-B	216.4	150.4	162.6	117.3	231.8	111.9	134.5	122.9	99.3	85.5
GPD-M	191.6	138	156	115.1	240.1	109.3	129	111.9	112.4	83
GPD-L	195.7	135.3	158.6	115.1	239.3	94.1	131.9	110.4	116.2	85.6
GPD-B	203.2	143.8	163.9	120.7	255.7	113.1	132.4	116.3	117.9	86.7
150										
GEV-M	258	167.9	183.4	142.4	287.1	128.4	145.3	140.9	113	94
GEV-L	252.4	167.3	194	137.9	285.6	146	146.9	139.8	117.1	95
GEV-B	297.7	185.2	195.6	150.4	309.3	138.8	155.5	154.2	122.3	102.8
GPD-M	248.9	163.6	190.3	153.7	325.7	130.2	147	134.4	143.2	98.1
GPD-L	256.8	159	194.9	154.1	325.4	105.1	151.3	131.8	150.7	102.4
GPD-B	270	172.8	203.6	163.9	354.8	136.3	152	141.6	152.7	104.1
300										
GEV-M	305.5	186.4	203.1	165	334.4	143.6	155.7	158.2	126	102.1
GEV-L	296.4	185.5	218.2	158.1	331.8	171.3	157.5	156.4	131.9	103.3
GEV-B	361.6	209.2	217.5	174	368	158	168.9	176.8	138.4	114.3
GPD-M	291.6	180.3	214.1	183.4	392.8	144.2	158.3	149.8	165.6	108
GPD-L	302.9	174.4	220.5	184.1	393.1	111.6	163.8	146.4	176.4	113.8
GPD-B	321	192.3	232	197.7	434.3	152.1	164.5	159.3	178.6	115.8

4. Discussion and Conclusions

The main aim of the present study is to evaluate different statistical methods that are used for the description of extreme rainfall events. For this purpose, daily values of precipitation are used, coming from 10 stations located around the Mediterranean region. The stations expand from Spain to Greece and the time period starts from 1951 to 2010. To analyse the extreme rainfalls, the data was organized into new data sets, the Block Maxima data set which was fitted with the GEV distribution and POT data set fitted with GPD distribution. Finally, their parameters were estimated with MLE, L-moments and Bayesian methods.

The choice of the distribution that can characterize the extreme rainfall's behavior became using goodness of fit tests. In accordance with many researches [14-17] GEV and GPD distribution seem to be the most appropriate. The importance of choosing an appropriate distribution is great, as it can affect the final results [18]. Consequently, an extra test for choosing an appropriate estimation method was achieved through the QQ plots. The results showed that GPD method with all of the three techniques could characterize sufficient the extreme rainfall episodes especially the minimum and medium extreme values. Moreover, it should be mentioned that both GEV-B and GPD-B are appropriate for the maximum values of extreme precipitation. Taking into account both the QQ plots and the shape diagram it is observed that the less suitable method for fitting the extreme episodes is GEV-L, despite the fact that is commonly used [15, 19].

Furthermore, in this research the return levels of extreme rainfalls were calculated for three return periods (50, 150 and 300 years). The results showed that in the stations of Bastia and Malaga the highest return levels were recorded while in the eastern the lowest ones. That is explained from the fact that many of these stations such as Bastia are located in the western Mediterranean close to cyclogenesis center (Gulf of Genoa) [20,21], while other stations such as Malaga located along the Atlantic depression trajectories which are moving into the Mediterranean [20]. It is also found that GEV-B method, also used by Dyrddal et al [22], assess the highest return levels of precipitation for the western Mediterranean stations, while the GPD-B, also used by Cooley et al. [23] for the eastern ones.

Author Contributions: All authors conceived and initiated the study. Georgia Lazoglou analyzed the data. All authors contributed to the discussion and interpretation of the results and the writing of the manuscript.

Conflicts of Interest: "The authors declare no conflict of interest."

Abbreviations

The following abbreviations are used in this manuscript:

GEV-M: Generalized Extreme Value Distribution with the Maximum Likelihood estimation method

GEV-L: Generalized Extreme Value Distribution with the L-moments estimation method

GEV-B: Generalized Extreme Value Distribution with the Bayesian estimation method

GPD-M: Generalized Pareto Distribution with the Maximum Likelihood estimation method

GPD-L: Generalized Pareto Distribution with the L-moments estimation method

GPD-B: Generalized Pareto Distribution with the Bayesian estimation method

GEV-L: Generalized Extreme Value Distribution with the L-moments estimation method

MLE: Maximum Likelihood Estimation method

EVT: Extreme Value Theory

GEV: Generalized Extreme Value Distribution

GPD: Generalized Pareto Distribution

POT: Peaks Over Threshold

References

1. Pachauri, R. K.; Allen, M. R.; Barros, V. R., Broome, J.; Cramer, W.; Christ, R.; ... & Dubash, N. K. Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change, **2014**.
2. Kyselý, J. Trends in heavy precipitation in the Czech Republic over 1961–2005. *International Journal of Climatology* **2009** 29(12), 1745-1758.
3. Brunetti, M.; Maugeri, M.; Nanni, T. Changes in total precipitation, rainy days and extreme events in northeastern Italy. *International Journal of Climatology* **2001** 21, 861–871.
4. Hundecha, Y.; Bardossy, A. Trends in daily precipitation and temperature extremes across western Germany in the second half of the 20th century. *International Journal of Climatology* **2005**, 25, 1189–1202.
5. Pal, J.; Giorgi, F.; Bi, X. Consistency of recent European summer precipitation trends and extremes with future regional climate projections. *Geophysical Research Letters* **2004** 31(13), 4, DOI: 10.1029/2004GL019836.

6. Kharin, V.V.; Zwiers, F.W. Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM. *Journal of Climate* **2000**, *13*, 3760–3788.
7. Katz, R.W.; Parlange, M.B.; Naveau, P. Statistics of extremes in hydrology. *Advances in water resources* **2002**, *25*, 1287–1304.
8. García, J.A.; Serrano, A.; Gallego, M.C. A spectral analysis of Iberian Peninsula monthly rainfall. *Theoretical Applied Climatology* **2002**, *71*, 77–95.
9. Coles, S. *An Introduction to Statistical Modeling of Extreme Values*. Publisher: Springer Series in Statistics, London, 2001.
10. Beguería, S. Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value. *Journal of Hydrology* **2005**, *303*, 215–2309.
11. Hosking, J.R.M.; Wallis, J.R. *Regional Frequency Analysis: An Approach Based on LMoments*. Publisher: Cambridge University Press, New York, 1997.
12. Coles, S.; Dixon, M.J. Likelihood-based inference for extreme value models. *Extremes* **1999**, *2*(1), 5–23
13. Klein Tank, A.M.G. et al. (38 co-authors) Daily dataset of 20th century surface air temperature and precipitation series for the European Climate Assessment. *International Journal of Climatology* **2002**, *22*, 1441–1453.
14. Anagnostopoulou, C., Tolika, K. Extreme precipitation in Europe: Statistical threshold selection based on climatological criteria. *Theoretical and Applied Climatology* **2012** *107*(3–4), 479–489, DOI:10.1007/s00704-011-0487-8.
15. Kysely, J. Coverage probability of bootstrap confidence intervals in heavy-tailed frequency models, with application to precipitation data. *Theoretical and Applied Climatology* **2010**, *101*(3), 345–361, DOI:10.1007/s00704-009-0190-1.
16. Dyrddal, A.V.; Lenkoski, A.; Thorarinsdottir, T.L.; Stordal, F. Bayesian hierarchical modeling of extreme hourly precipitation in Norway. *Environmetrics* **2015**, *26*(2), 89–106, DOI: 10.1002/env.2301.
17. Roth, M.; Buishand, T.A.; Jongbloed, G.; Klein Tank A.M.G.; van Zanten, J.H. Projections of precipitation extremes based on a regional, non-stationary peaks-over-threshold approach: A case study for the Netherlands and north-western Germany. *Weather and Climate Extremes* **2014**, *4*, 1–10, DOI:10.1016/j.wace.2014.01.001.
18. El Adlouni, S.; Ouarda, T. B. M. J.; Zhang, X.; Roy, R.; Bobée, B. Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research* **2007**, *43*(3).
19. Lee, S.H.; Maeng, S.J. Frequency analysis of extreme rainfall using L-moment. *Irrigation and Drainage* **2003**, *52*(3), 219–230, DOI:10.1002/ird.090.
20. Trigo, I.; Davies, T.; Bigg, G. Objective climatology of cyclones in the Mediterranean region. *Journal of Climate* **1999**, *12*, 1685–1696.
21. Maheras, P.; Flocas, A.H.; Patrikas, I.; Anagnostopoulou, C. A 40 year objective climatology of surface cyclones in the Mediterranean region: spatial and temporal distribution. *International Journal of Climatology* **2001**, *21*, 109–130.
22. Dyrddal, A.V.; Lenkoski, A.; Thorarinsdottir, T.L.; Stordal, F. Bayesian hierarchical modeling of extreme hourly precipitation in Norway. *Environmetrics* **2015**, *26*(2), 89–106, DOI:10.1002/env.2301.
23. Cooley, D.; Nychka, D.; Naveau, P. Bayesian Spatial Modeling of Extreme Precipitation Return Levels. *Journal of the American Statistical Association* **2007**, *102*(479), 824–840, DOI: 10.1198/016214506000000780.



© 2017 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).