OUTLIERS AND TREND DETECTION TESTS IN RAINFALL EXTREMES.

PANAGIOTA GALIATSATOU[1], PANAYOTIS PRINOS[2]

[1]Ph.D Student, Hydraulics Lab., Dept. of Civil Eng. Aristotle University of Thessaloniki, Greece
(Tel: +30-2310-995856, e-mail: pgaliats@civil.auth.gr)
[2] Professor, Hydraulics Lab., Dept. of Civil Eng. Aristotle University of Thessaloniki, Greece
(Tel: +30-2310-995689, Fax: +30-2310-995672, e-mail: prinosp@civil.auth.gr)

ABSTRACT

In the present work daily rainfall time series is analyzed using Extreme Value methodologies. Four different tests are used for detection of a possible outlier: the estimation of x_{ult}, the conditional probability of the event estimated through a Bayesian formula, a Gumbel plot and a q-q plot for an appropriately selected model for extreme events. Return level plots are created using a point process model for extreme observations, both by including and excluding the possible outlier and discrepancies are discussed. Simple parametric models for extremal trends are then incorporated in the parameters of the point process model. The location and scale parameter of the model are assumed to vary as polynomial functions of time or as sinusoidal terms, treated as entirely distinct or closely inter-related. Alternatively, they are expressed in terms of two other parameters, the rate and the severity of extreme events. The deviance statistic is used to identify the significance of such trends. Daily rainfall from the city of Thessaloniki, Greece is used in the analysis.

Keywords: possible outlier, ultimate event, Gumbel plot, q-q plot, posterior predictive distribution, return level, non-stationarity, point process, polynomial temporal trends, deviance statistic

1. INTRODUCTION

Extreme Value Theory is concerned with probabilistic and statistical subjects related to very high or very low values in sequences of random variables and in stochastic processes. Environment, finance, Internet traffic, reliability, athletics, statistical inference/tests and asymptotic theory for sums are some of the main areas of application of Extreme Value Theory, which is a really rapidly developing subject with a rich mathematical background.

The entire range of possible limit distributions for extreme realizations of a random variable is given by the Extremal Types Theorem. The families of Extreme Value and Generalized Pareto distributions are more widely used to perform Extreme Value analyses. Another particularly appropriate formulation of characterizing the behavior of extremes of a process is derived from the theory of point processes. Coles (2001) concentrates on statistical inference for extremes and uses all the previously mentioned methodologies to analyze different data sets and to produce statistical inference for extremes.

The problem of outliers is of major concern when dealing with extreme events. In statistics, an outlier is a single observation "far away" from the rest of the data that can lead to unrealistic conclusions, especially when considering extrapolation to high enough quantiles of the variables analyzed. The problem of outliers was considered among others by Smith (2003), who examined environmental datasets from this point of view, as well as a remarkable series of track performances achieved by a Chinese athlete including a new world record, which brought up immediate suspicions. Robinson and Tawn (1995) also examined the previously mentioned series to assess how much of an outlier this performance was.

In the context of environmental processes, non-stationarity is often apparent because of seasonal effects, perhaps due to different climate patterns in different months, or in the form of trends, possibly due to long-term trends (Coles, 2001). The purpose of a trend analysis/ test is to determine if the values of a series have a general increase or decrease with increasing time. Trend is often incorporated in the analyses using covariate information. Covariate modeling in the context of the threshold excess model was discussed by Davison and Smith (1990). Coles and Tawn (1990) exploit covariate information consisting of location, inter-site distances and an approximation to the tidal sequence at each site to examine risk assessment for UK coastlines and the issue of sensitivity to climate change. Dixon and Tawn (1999) assess the effect of temporal non-stationarity due to a combination of long- term trend in the mean sea level, the deterministic tidal component, surge seasonality and interactions between tide and surge. Smith (2003) uses different formulations for the parameters of extreme models. The covariates here are taken to be polynomial functions of time and sinusoidal terms. Gaetan and Grigoletto (2004) use a semiparametric approach for smoothing sample extremes, based on nonlinear dynamic modeling of the Generalised Extreme Value distribution. Huerta and Sanso (2005) assume that the ozone observations in Mexico city follow a GEV distribution for which the location, scale or shape parameters define the space-time structure, where the temporal component is determined through a Dynamic Linear Model.

In the present study daily rainfall from the city of Thessaloniki, Greece for a time period of 43 years (1958-2000) is analyzed using Extreme Value methodologies. Four different tests are used for detection of a possible outlier: a) the estimation of x_{ult}, b) the conditional probability of the event estimated through a Bayesian formula, c) a Gumbel plot and d) a q-q plot for an appropriately selected model for extreme events of the rainfall process. Return level plots are presented using a point process model both by including and excluding the largest observation considered above and discrepancies are discussed. Simple parametric models for extremal trends are then incorporated in the point process model. These non-stationarities are defined through deterministic functions in the parameters of the model. The location and scale parameter of the model are assumed to vary as polynomial functions of time or as sinusoidal terms, treated as entirely distinct or closely inter-related. Alternatively, they are expressed in terms of two other parameters, the rate and the severity of extreme events. A statistical parameter, the deviance statistic D, is used to identify the significance of such trends.

2. POSSIBLE OUTLIERS

2.1. DETECTION TESTS OF A POSSIBLE OUTLIER

Daily rainfall in the city of Thessaloniki is analyzed for a time period of 43 years (1958-2000). The data are presented in Figure 1 together with the index of day and date, the largest event occurs.

The event of 23/11/1985, measured at 98mm seems to be significantly larger compared to the rest of the time series and also to the adjoining events of 22/11/1985 and 24/11/1985. This is the reason of the following analysis, which is supposed to give some evidence of whether the event should be included in the analysis or considered to be an outlier and omitted from the sample.

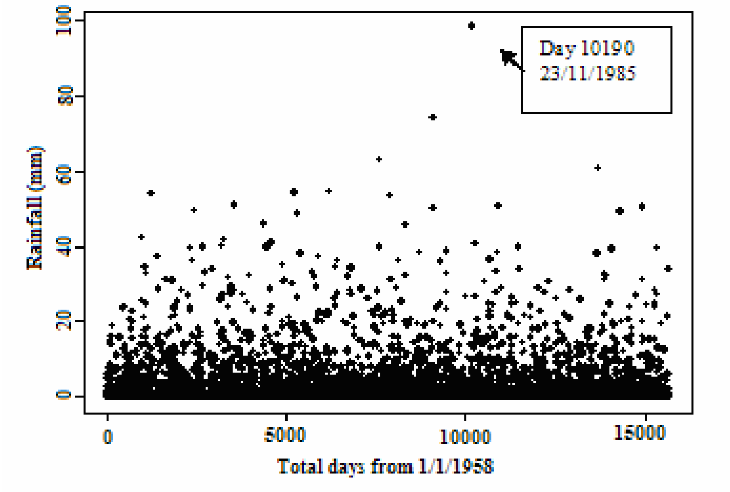

Figure 1. The rainfall time series of Thessaloniki (1958-2000) and its largest events

An outlier is an observation that lies an abnormal distance from other values in a random sample from a population. Their inclusion in the analysis can produce deceivable results that can lead to uneconomical and inappropriate solutions. In the following Sections four different methodologies are used to produce evidence of whether the event of 23/11/1985 should be considered an outlier or not: a) the estimation of the nominal "ultimate rainfall event", x_{ult} b) the calculation of the conditional probability of the event occurring, through a Bayesian formula, c) a Gumbel plot and d) a q-q plot for the GPD (Generalized Pareto Distribution) for the excesses of an appropriately defined threshold of the rainfall process.

2.1.1. "THE ULTIMATE RAINFALL EVEN"

The data from 1958 up to 1981 are analyzed separately, using a Generalized Pareto Distribution (GPD) to produce an estimate of x_{ult} (median and 95% confidence intervals). This period is chosen because of the fact that there are no observations that are completely different from the majority in it. If $Y_i=X_i-u$ are the excesses of the rainfall variable X over a high enough threshold u, in some asymptotic sense, the conditional distribution of excesses follows the GPD:

$$G(y)=1-(1+\frac{\xi y}{\sigma})^{-1/\xi} \qquad (1)$$

where σ, ξ denote the scale and shape parameters, respectively (Pandey et al., 2001). Using the mean residual life plot of the data and the diagram of estimates of parameters σ and ξ with different thresholds u, an appropriate threshold u_o=16mm is chosen for the short rainfall time series. By Maximum Likelihood Estimation (MLE) the parameters and their standard errors are: $\hat{\sigma}$=11.931 (se=1.3436) and $\hat{\xi}$=-0.1331 (se=0.08016). The shape parameter ξ is dominant in determining the qualitative behavior of the GPD. If $\xi<0$ the distribution of excesses has an upper bound of $u-\sigma/\xi$=105.65mm (Coles, 2001). A profile likelihood for x_{ult} may be constructed, as shown in Figure 2, to reveal a 95% confidence interval for the ultimate return level.

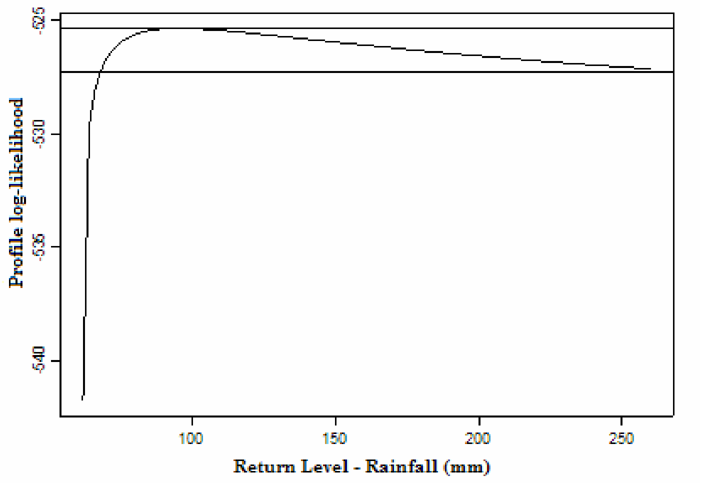

Figure 2. Profile log-likelihood for x_{ult}

Figure 2 shows a large 95% confidence interval in the range of (60, 260)mm for x_{ult} =105.65mm. The record of 98mm of 23/11/1985 lies within the confidence interval, so on the basis of the analysis, there is no real evidence that the record is unrealistic. The alternative Bayesian analysis considers the problem, not as one about estimating the ultimate limit parameter, but estimates a prediction interval, which will give more precise information about what is likely to happen in a given year, considering the previously recorded observations (Smith, 2003).

2.1.2. THE CONDITIONAL PROBABILITY OF THE EVENT TO OCCUR

A primary objective of an extreme value analysis is often prediction. Let z denote a future observation with density function $f(z|\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta$. The posterior predictive density of z, given observed data **x** is:

$$f(z|\mathbf{x}) = \int_{\boldsymbol{\Theta}} f(z|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|x)d\boldsymbol{\theta} \tag{2}$$

Predictive distributions reflect the uncertainty in the model and the uncertainty due to the variability of future observations (Stephenson and Tawn, 2004). Using expression (2), the posterior predictive distribution of a future observation z is given by:

$$\Pr(Z \le z|\mathbf{x}) = \int_{\boldsymbol{\Theta}} \Pr(Z \le z|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{x})d\theta \tag{3}$$

From equation (3) it is obvious that the conditional probability can be expressed as an analytic function of the three parameters $\boldsymbol{\theta}$=(μ, σ, ξ) of a point process model, say φ(μ,σ,ξ), and hence estimated through the Bayesian formula (Smith, 2003):

$$\iiint \varphi(\mu,\sigma,\xi)\pi(\mu,\sigma,\xi|\mathbf{x})d\mu d\sigma d\xi \tag{4}$$

where $\pi(\ldots|\mathbf{x})$ denotes the posterior density given past data X. The prior distribution for the parameters used to construct this posterior density is a trivariate normal distribution. This distribution enables the specification of independent parameters. Setting φ=logσ, we might choose a prior density function $f(\mu,\sigma,\xi)=f_\mu(\mu)\cdot f_\varphi(\varphi)\cdot f_\xi(\xi)$, where $f_\mu(\cdot)$, $f_\varphi(\cdot)$, $f_\xi(\cdot)$ are normal density functions with mean zero and variances υ_μ, υ_φ, υ_ξ, respectively, corresponding to a specification of prior independence in the parameters μ, φ and ξ. The absence of genuine prior information leads to the use of priors that have very high variance, or equivalently, are nearly flat. The choice of normal densities is arbitrary and variances are chosen as $\upsilon_\mu=\upsilon_\varphi=10^4$ and $\upsilon_\xi=10^2$ (Coles, 2001).

If Z_L is the maximum daily rainfall over a period of L years, then allowing for the uncertainty in the estimation of the parameter components, the predictive distribution of Z_L is defined as (Coles and Tawn, 1996):

$$\Pr(Z_L < z|\mathbf{x}) = \int_{\boldsymbol{\Theta}} \Pr(Z < z|\boldsymbol{\theta})^L \pi(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta} \tag{5}$$

where Z denotes the annual maximum. If L=1, equation (5) reduces to (2). The result in this case for $P(Z > 98|\mathbf{x})$ is equal to 0.001111111 for a future period of 1 year. For a future period of four years $P(Z > 98|\mathbf{x})$=0.004545455. Both probabilities are not low enough to provide strong evidence that the event of 23/11/1985 is an outlier.

2.1.3. THE GUMBEL PLOT

Suppose that the annual maxima over the 43 years are $Y_1, Y_2, \ldots, Y_{43}$ ordered as $Y_{1:43} \leq \ldots \leq Y_{43:43}$, then $Y_{i:43}$ for $1 \leq i \leq 43$, is plotted against the reduced value $x_{i:43}$, where $x_{i:43} = -\log(-\log p_{i:43})$, with $p_{i:43}=(i-0.5)/43$ being the ith plotting position (Figure 3).

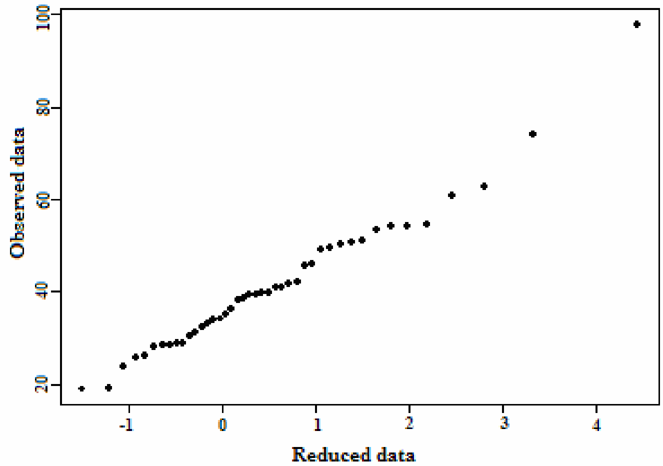

Figure 3. Gumbel plot for rainfall data

A straight line indicates a good fit to the Gumbel distribution. There is no systematic evidence of curvature upwards or downwards in the annual maxima of the rainfall process, nor do there seem to be evidence of outliers.

2.1.4. Q-Q PLOTS FOR EXCESSES

Let $\widehat{F}(x)$ be an estimate of F(x) based on $x_1, x_2, \ldots, x_n$.The scatter plot of the points $\widehat{F}^{-1}(p_{i:N})$ versus $x_{i:N}$, i=1,…,N is called a q-q plot. Thus, the q-q plot shows the estimated versus the observed quantiles. If the model fits the data well, the pattern of points on the q-q plot will exhibit a 45-degree straight line (Castillo et al, 2005). For the GPD model and assuming $\xi \neq 0$, the quantile plot consists of the pairs: $\{(\widehat{H}^{-1}(i/(k+1)), y(i)\}$ for i=1,.., k (Coles, 2001), where k is the number of excesses of an appropriately defined threshold. For two different thresholds of 18mm and 9mm, the q-q plots of points exceeding both thresholds are shown in Figure 4(a,b).

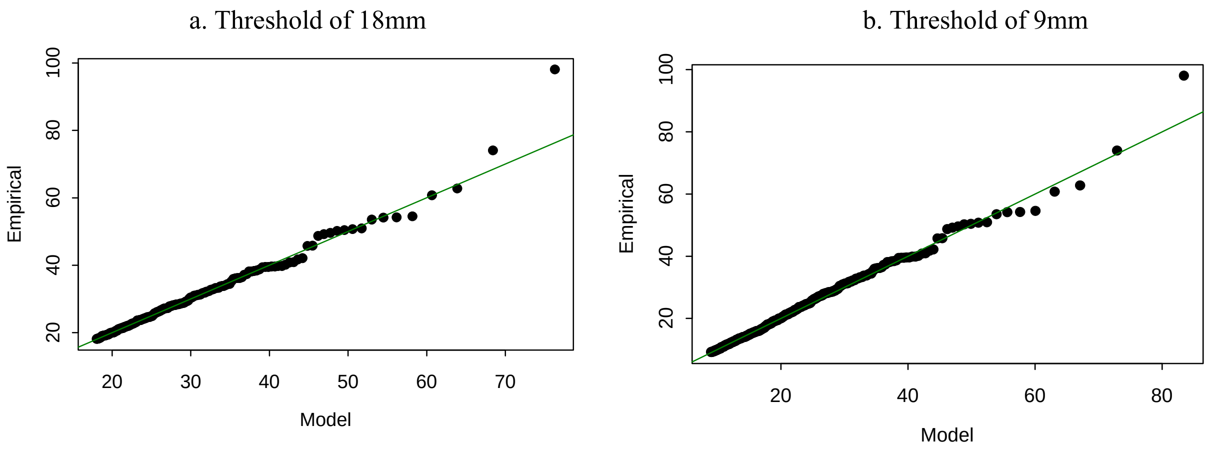

Figure 4. Q-Q plots using two different thresholds

The shape of the plot for both thresholds (especially for the lower threshold) can not support the treatment of the final data point as an outlier. In fact, the largest point is not very far from the straight line, especially when considering the lower threshold. It should be noted here that in the following a threshold of 18mm is used, because it is considered to be more consistent with the methodology of threshold selection mentioned in 2.1 and described in detail by Coles (2001). Smith (2003) emphasizes the fact that observations which first appear to be outliers, may not in fact be inconsistent with the rest of the data if they come from a long-tailed distribution. By maximum likelihood fitting of the GPD model using both previously mentioned thresholds, estimates of the shape parameter ξ are 0.0213 and 0.101 for thresholds 18 and 9mm respectively, confirming a long-tailed distribution.

2.2. RETURN LEVEL ESTIMATION WITH AND WITHOUT THE LARGEST EVENT

The R-year return level is calculated using the formula:

$$x_R = G^{-1}(1-\frac{1}{\lambda R}) + u \tag{6}$$

where $G^{-1}(p)$ denotes the Pareto quantile function. If m is the number of years of observations and k the total number of exceedances, then the mean crossing rate is $\lambda=k/m$. Figure 5(a, b) shows estimates of the median and 95% confidence intervals for the return level, when the largest observation is incorporated or omitted from the sample.

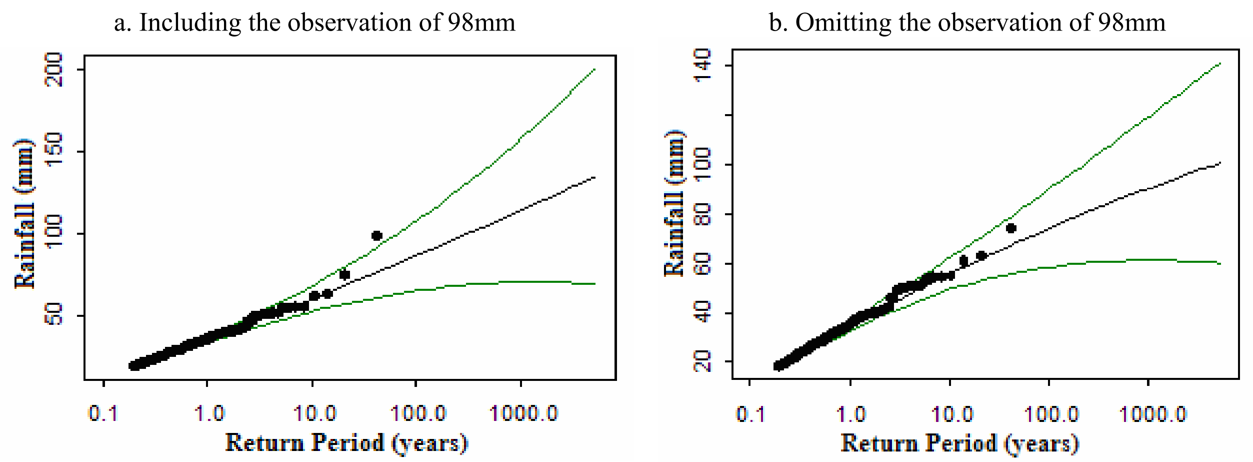

Figure 5. Return level of rainfall (mm) against return period (years)

Estimates of return levels are generally lower, when the largest observation of 98mm is omitted from the sample. From Figure 5b, it is easily noticed that the distribution of excesses of the threshold is now upper- bounded (the shape parameter ξ is negative) in contrast to Figure 5a, where the observation of 98mm is included in the analysis. 95% confidence bands in Figure 5a seem not to include the observation of 98mm. These confidence intervals are estimated assuming a normal distribution for the parameters of the distribution. This is not considered as a cause of concern, because it is observed that, when using for example the Bayesian approach with flat priors for the parameters, the most extreme observation is included in the confidence bands (Figure 6).

Figure 6 shows that the confidence interval constructed using the Bayesian approach incorporates the largest observation of the 43 years of data available. It can be concluded that in this case the results of a Bayesian analysis are closer to real observations and to real conditions, compared to MLE procedure and therefore they lead to a more conservative solution (confidence intervals include higher values of the rainfall process).

According to Fisher, (1922) the rejection of observations is too crude to be defended and unless there are other reasons for rejection than the mere divergences from the majority, it would be more philosophical to accept these extremes, not as gross errors but as the indications that the distribution of errors is not normal (Reiss and Thomas, 2001). In the rest of the analysis the observation will not be omitted from the sample, because it is not considered as an outlier.

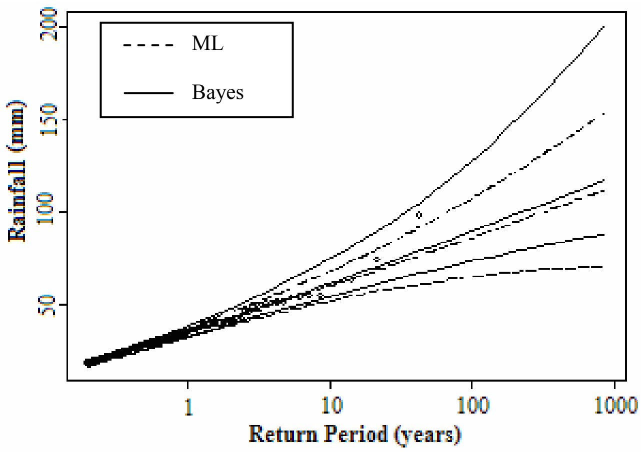

Figure 6. Return Level estimation (median and 95% confidence intervals) using ML and Bayes approach

3. DETECTION TESTS FOR TRENDS

Estimates of return levels based on the models analyzed in Section 2 assume that the behavior of the process will remain unchanged in future years. Many hydrological time series exhibit trending behavior or non-stationarity. In fact, the trending behavior is a type of non-stationarity. The purpose of a trend analysis is to determine if the values of a series have a general increase or decrease with time.

If $X_1, X_2,.., X_n$ is a series of independent random variables with a common distribution function F, for any large fixed threshold u the sequence $\{X_1, X_2,.., X_n\}$, viewed on the interval $[u, \infty)$, is approximately a non-homogeneous Poisson process with intensity function: $\lambda_{\boldsymbol{\theta}}(x) = \frac{1}{\sigma}\{1 + \xi(\frac{x-\mu}{\sigma})\}_+^{-(\xi+1)/\xi}$, where $\sigma>0$ and $\boldsymbol{\theta}=(\mu, \sigma, \xi)$ is the vector of parameters determined by the tail behavior of F. The approximation outlined leads to a likelihood for $\boldsymbol{\theta}$ based on an observed set of exceedances $\mathbf{x}=(x_1,x_2,..,x_{nu})$ of a high threshold u, given by:

$$L(\boldsymbol{\theta};\mathbf{x}) = \exp(-\frac{n}{n_y}\Lambda_{\boldsymbol{\theta}}[u,\infty))\prod_{i=1}^{n_u}\lambda_{\boldsymbol{\theta}}(x_i) \tag{7}$$

where n_y is the number of observations per year and $\Lambda_{\boldsymbol{\theta}}[u,\infty) = \int_u^{\infty}\lambda_{\boldsymbol{\theta}}(x)dx = \{1+\xi(\frac{u-\mu}{\sigma})\}_+^{-1/\xi}$ (Stephenson and Tawn, 2004). Simple polynomial models are implemented to describe the structure of $\boldsymbol{\theta}(t)=(\mu(t), \sigma(t), \xi(t))$, in order to allow the parameters of the point process model to vary with time. We are interested to draw inference about the latent trend $\boldsymbol{\theta}(t)$, where $t \in [t_1,..,t_n]$ which corresponds to years 1 to n. Three different approaches of modeling trends are examined here: a) the modeling of μ and σ as polynomial functions of time when these parameters are treated as entirely distinct, b) the modeling of μ and σ as polynomial functions of time when they are treated as closely inter-related and c) the modeling of parameters λ and φ, which represent the rate at which extreme events occur and their severity, respectively (Butler, 2005).

3.1. TRENDS IN RAINFALL EXTREMES

3.1.1 PARAMETERS AS DISTINCT POLYNOMIAL FUNCTIONS OF TIME

Model parameters μ and σ are treated here as distinct polynomial functions of time of order p_μ and p_σ respectively, while ξ is supposed to be constant over time (Butler, 2005):

$$\mu(t) = \sum_{k=0}^{p_\mu}(t-t_1)^k\mu_k \ , \quad \sigma(t) = \sum_{k=0}^{p_\sigma}(t-t_1)^k\sigma_k \quad \text{and} \quad \xi(t)=\xi \tag{8}$$

This model contains $p_\mu+p_\sigma+3$ parameters. In the analyses of this paper values of k=1,2 are implemented to the parameters.

3.1.2. PARAMETERS AS CLOSELY INTER-RELATED POLYNOMIAL FUNCTIONS OF TIME

This model assumes that $\mu(t)$ is a polynomial of order p_μ and the ratio $\zeta(t)= \mu(t)/\sigma(t)$ is constant over time. It follows that (Butler, 2005):

$$\mu(t) = \sum_{k=0}^{p_\mu}(t-t_1)^k\mu_k \ , \quad \sigma(t) = (1/\zeta)\sum_{k=0}^{p_\sigma}(t-t_1)^k\mu_k \quad \text{and} \quad \xi(t)=\xi \tag{9}$$

where ζ is an unknown parameter. This model contains less parameters, namely $p_\mu+3$. The ζ parameter is called the coefficient of extremal variation and in the following analysis it is assumed to be constant over time for k=1.

3.1.3. TRENDS IN THE NUMBER OF EXTREME STORMS AND THE SEVERITY OF STORMS

A widely used model (Coles and Tawn, 1990) treats the rate at which extreme events occur $\lambda(t;u)$ and the severity (magnitude) of extreme events $\varphi(t;u)$ as polynomials in time of order p_λ and p_φ, respectively while keeping ξ constant. The parameter u is defined as the 90% quantile of the raw data. This model contains $p_\lambda+p_\varphi+3$ parameters and implies that:

$$\mu(t;u) = u - \xi^{-1}(1-\xi)[1-\{\sum_{k=0}^{p_\lambda}(t-t_1)^k\lambda_k\}^{\xi}]\{\sum_{k=0}^{p_\varphi}(t-t_1)^k\varphi_k\} \tag{10}$$

$$\sigma(t;u) = \{\sum_{k=0}^{p_\lambda}(t-t_1)^k\lambda_k\}^{\xi}\{\sum_{k=0}^{p_\varphi}(t-t_1)^k\varphi_k\}(1-\xi) \qquad \text{and} \qquad \xi(t)=\xi$$

3.2. TREND DETECTION

Using the models for the parameters analyzed in Section 3.1. estimates for the parameters μ, σ, ξ are produced using MLE. A simple measure, the deviance statistic D, is used to produce evidence of trends of the previously mentioned categories in the data. The deviance statistic is defined as: $D=2\{l_1(M_1)-l_0(M_0)\}$, where $l_1(M_1)$, $l_0(M_0)$ are log-likelihoods for models M_1 and M_0 respectively (M_0 is the likelihood of the model without any trends and M_1 the likelihood assuming different trends). If D is lower to χ_f^2, where f are the degrees of freedom of the model with trends incorporated, there is no evidence of the trends examined in the data. Table 1 shows the results of all the models used. It should be noted that t corresponds to time (days), while t^* to a normalized vector of time considering that on 1/1/1980 all different parameters of Table 1 are stable.

Table 1. Different polynomial temporal trend models for rainfall extremes

Model	Log-likelihood	$\lvert D \rvert$
Time constant	594.816	0.000
Linear trend in μ ($\mu=\mu_o+\mu_1 t^*$)	594.816	0.000
Linear trend in μ ($\mu=\mu_o+\mu_1 t^*+\mu_2 t^{*2}$)	594.816	0.000
Linear trend in σ ($\sigma=\sigma_o+\sigma_1 t^*$)	594.807	0.017
Linear trend in μ and σ ($\mu=\mu_o+\mu_1 t^*$ and $\sigma=\sigma_o+\sigma_1 t^*$)	594.782	0.068
Trends in μ ($\mu=\mu_o+\mu_1 t^*+\mu_2\sin(2\pi t/365)+\mu_3\cos(2\pi t/365)$)	594.815	0.002
Trends in σ ($\sigma=\sigma_o+\sigma_1 t^*+\sigma_2\sin(2\pi t/365)+\sigma_3\cos(2\pi t/365)$)	593.834	1.964
Linear trends in μ ($\mu=\mu_o+\mu_1 t^*$) and σ ($\sigma=(1/\zeta)(\sigma_o+\sigma_1 t^*)$)	594.809	0.014
Trend in number of storms ($\lambda=\lambda_o+\lambda_1 t^*$)	594.966	0.300
Trend in severity of storms ($\varphi=\varphi_o+\varphi_1 t^*$)	594.934	0.237
Trends in $\lambda=\lambda_o+\lambda_1 t^*$ and $\varphi=\varphi_o+\varphi_1 t^*$	594.931	0.229

Estimates of the log-likelihood and the deviance statistic, shown in Table 1, give no evidence of a polynomial temporal trend of the form described in Section 3.1. in the location or the scale parameters of the point process model. The largest divergence is observed for the case where $\sigma=\sigma_o+\sigma_1 t^*+\sigma_2\sin(2\pi t/365)+\sigma_3\cos(2\pi t/365)$), but still the deviance statistic is not considerable. It should be noted that the deviance statistic gives only some evidence of temporal trend existence and to provide powerful results, different tests have to be used.

4. CONCLUSIONS

In the present study daily rainfall from the city of Thessaloniki, Greece was analyzed using Extreme Value methodologies to a) detect possible outliers present in the series and b) investigate the existence of polynomial temporal trends of specified forms in the parameters of the Extreme Value model used.

Four different methodologies were used to decide upon the inclusion of the largest observation (the possible outlier) in the data: a) the estimation of x_{ult}, b) the conditional probability of the event estimated through a Bayesian formula, c) a Gumbel plot and d) a q-q plot for an appropriately selected model for extreme events of the rainfall process.

Finally, simple polynomial temporal trends for the location and the scale parameters were incorporated in the point process model and a simple test, the deviance statistic D, was used to identify the significance of such trends.

The main conclusions of the analysis are:

a. Estimates of return levels are significantly lower (up to 20.8% for 1000-year return level), when the largest observation of 23/11/1985 is omitted from the sample. The distribution of

excesses of the threshold is upper- bounded (the shape parameter ξ is negative) in contrast to the case including the observation in the analysis.
b. The largest event of 23/11/1985 doesn't seem to be an outlier. All four previously mentioned approaches (especially b and c) have proven that there is no evidence that this event should be omitted from the sample.
c. There is no evidence of polynomial temporal trends of the forms described in Section 3.1.

5. ACKNOWLEDGMENTS

The authors want to thank the Dutch National institute RIKZ for providing data of the Dutch coast. The work described in this publication was supported by the European Community's Sixth Framework Programme through the grant to the budget of the Integrated Project FLOODsite, Contract GOCE-CT-2004-505420. This paper reflects the authors' views and not those of the European Community. Neither the European Community nor any member of the FLOODsite Consortium is liable for any use of the information in this paper.

6. REFERENCES

Butler, A., 2005, Statistical Modelling of Synthetic Oceanographic Extremes. Ph.D Thesis, (2005), Lancaster University.

Castillo, E., Hadi, A. S., Balakrishnan, N. and Sarabia J. M., 2005, Extreme Value and Related Models with Applications in Engineering and Science, Wiley Series in Probability and Statistics.

Coles, S. G. and Tawn, J. A., 1990, "Statistics of coastal flood prevention", Philos. Trans. R. Soc. Lond., A., **332**, pp. 457-476.

Coles, S. G and Tawn, J. A., 1996, "A Bayesian analysis of extreme rainfall data", Applied Statistics, **45**, No. 4, pp 463-478.

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

Davison, A. C. and Smith, R. L., 1990, "Models for exceedances over high thresholds (with discussion), J. R. Stat. Soc., Ser. B, **52**, pp. 393-442.

Dixon, M., J. and Tawn, J. A. (1999), "The effect of non-stationarity on sea-level estimation", Applied Statistics, **48**, pp. 135-151.

Gaetan, C. and Grigoletto, M., 2004, "Smoothing sample extremes with Dynamic Models", Extremes **7**, pp. 221-236

Huerta, G. and Sanso, B., 2005, "Time-varying models for extreme values", Department of Mathematics and Statistics, University of New Mexico, www.ams.ucsc.edu/reports/ams2005-04.pdf

Pandey, M. D., Van Gelder, P. H. A. J. M. and Vrijling, J. K, 2001, "The estimation of extreme quantiles of wind velocity using L-moments in the peaks-over-threshold approach", Journal of Structural Safety, **23**(2), pp. 179-192.

Reiss, R.-D. and Thomas, M., 2001. Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields, Second Edition, Birkhaüser.

Robinson, M. E. and Tawn, J. A., 1995, "Statistics for exceptional athletic records", Applied Statistics, **44**, pp. 499-511.

Smith, R. L., 2003, "Statistics of extremes with applications in Environment, Insurance and Finance", University of North Carolina, http:/ www.stat.unc.edu/postscript/rs/semstatrls.ps.

Stephenson, A. and Tawn, J., 2004, "Bayesian inference for extremes: accounting for the three Extremal Types", Extremes, **7**, No. 4, pp. 291-307.