REVSTAT – Statistical Journal
Volume 18, Number 2, April 2020, 189–202

TESTING FOR TRENDS IN EXCESSES OVER A THRESHOLD USING THE GENERALIZED PARETO DISTRIBUTION

Authors: Sergio Juárez
– Universidad Veracruzana, Veracruz, México

Javier Rojo
– Korvis Professor of Statistics, Oregon State University, Oregon, USA
javier.rojo@oregonstate.edu

Received: February 2019 Revised: September 2019 Accepted: September 2019

Abstract:

- The Generalized Pareto Distribution (GPD) is used for modeling exceedances over thresholds. The general form of the GPD depends on three parameters: the location parameter μ; the scale parameter ($\beta > 0$); and the shape parameter ($-\infty < \xi < \infty$). This work restricts attention to the case where $\mu = 0$ and shows that, as ξ decreases while β is kept fixed, the family of GPD(ξ, β) distributions increases in the usual stochastic order. This property is used for testing the significance of trends in the size of the exceedances over high thresholds in a time series consisting of ozone measurements.

Key-Words:

- *Stochastic Order; Peaks Over a Threshold; Ozone Concentrations; Likelihood Ratio Tests.*

1. INTRODUCTION

Let X be a random variable with continuous distribution function F and corresponding survival function $\bar{F} = 1 - F$. Let x^* be the right endpoint of the support of F defined by $x^* = \sup\{x \in \mathbb{R} \colon F(x) < 1\}$. Given a real number $u < x^*$, referred to as the threshold, an *exceedance* over the threshold u occurs when $X > u$. The residual life function of F at time u, the probability that $X > u + x$ given that $X > u$, is

$$(1.1) \qquad \bar{F}_u(x) = P\big(X - u > x \mid X > u\big) = \frac{\bar{F}(x+u)}{\bar{F}(u)}\,, \qquad 0 < x < x^* - u\,.$$

The random variable $X - u$ is called the *excess* over the threshold u and $\bar{F}_u$ is the *excess* survival function of X over u. When F belongs to the domain of attraction of one of the extreme value distributions, it follows that, for sufficiently large u, the distribution function of $X - u$ can be approximated by the Generalized Pareto Distribution (GPD). The distribution function of a GPD(ξ, β) is

$$(1.2) \qquad F(x; \xi, \beta) = \begin{cases} 1 - (1 - \xi x/\beta)^{1/\xi}\,, & \xi \neq 0\,,\ \beta > 0\,, \\ 1 - \exp(-x/\beta)\,, & \xi = 0\,,\ \beta > 0\,, \end{cases}$$

where ξ and β are the shape and scale parameters, respectively. When $\xi < 0$ the support of $F(x; \xi, \beta)$ consists of the positive reals. When $\xi > 0$, the support is the interval $(0, \beta/\xi)$. The case $\xi = 0$ corresponds to the exponential distribution with mean β. When $\xi = 1$, the GPD distribution corresponds to the uniform distribution on $[0, \beta]$.

More precisely, let $X_1, ..., X_n$ be a sequence of independent and identically distributed random variables with continuos distribution H. Let $M_n = \max\{X_1, ..., X_n\}$. Suppose that there are sequences $a_n > 0$ and b_n of real numbers such that

$$(1.3) \qquad P\big\{a_n(M_n - b_n) \leq z\big\} \to G(z)\,, \qquad \text{as} \quad n \to \infty\,.$$

Then $G(z)$ is a member of the generalized extreme value distribution family defined by

$$G(z) = \exp\left\{-\left\{1 - \xi\left(\frac{z-\mu}{\sigma}\right)\right\}^{1/\xi}\right\}.$$

The precise technical justification for modeling excesses using the GPD — expression (1.2) — was provided by Smith [32] and is based on the fact that

$$\lim_{u \to x^*} \sup_{0 < x < x^* - u} \big|F_u(x) - F\big(x; \xi, \beta(u)\big)\big| = 0\,,$$

for fixed ξ and some positive function $\beta(u)$, if and only if F is in the domain of attraction of some extreme value distribution. This result is from the parallel work done by Balkema and de Haan [1] and Pickands [23]. Since most of the common continuous distributions belong to the domain of attraction of one of the three extreme value distributions, this result makes the GPD the natural model for the excess distribution of the random variable X when the threshold is high.

Starting with the early works by Smith [31] and Davison [6], the GPD has been used by many authors to model excesses over high thresholds in several fields such as river floods, air pollution, wind velocity, sea waves, insurance claims, etc. For the details of these applications see Hosking and Wallis [12], Smith [33], Dargahi-Noubary [5], Grimshaw [10], Rootzen and Tajvidi [29], Castillo and Hady [4], and Parisi and Lund [22]. Embrechts *et al.* [8], Falk *et al.* [9], and Reiss and Thomas [24] present detailed and elegant accounts of the theoretical underpinnings and the practical aspects of the modeling of extremes including discussions on the modeling of exceedances and excesses.

One of the main objectives of modeling excesses over high thresholds with the GPD is the estimation of tails of probability distributions — Smith [32]. But the GPD has also been used to detect and test for trends in the excesses. The papers by Smith [33], Davison and Smith [7], Smith and Huang [35] and Rootzen and Tajvidi [29] are some examples of such applications. Our interest in this article is also in testing for the existence of a long term trend in the excesses of a time series. The main difference with other works is our use of the concept of stochastic orderings of distribution functions. In Section 2 it is shown that given k GPD distributions $F(\cdot;\xi_j,\beta)$, $(j=1,...,k)$, if $\xi_1<\xi_2<\cdots<\xi_k$, then $F(x;\xi_1,\beta)>F(x;\xi_2,\beta)>\cdots>F(x;\xi_k,\beta)$ for all x. That is, we give a sufficient condition for the GPD family to be stochastically ordered. This condition is used in Section 3 to develop a simple procedure based on a likelihood ratio statistic for testing $H_0\colon \xi_1=\xi_2=\cdots=\xi_k$ *vs.* the isotonic alternative $H_a\colon \xi_1\leq\xi_2\leq\cdots\leq\xi_k$. Our procedure is desirable when it is believed a priori that the GPDs satisfy the stochastic order restriction and, hence, it is desirable to have a test that is more powerful than an omnibus test.

The test being proposed here belongs to the field of restricted inference. There is a vast literature in this area. The literature consists of roughly two large subareas: shaped-restricted inference, and order-restricted inference. Barlow *et al.* [2] is a classic pioneering work based on isotonic regression ideas and the Pool-Adjacent-Violators-Algorithm. Robertson *et al.* [25] and the many references therein, summarize and extend the work of Barlow *et al.* and adopt the Nonparametric Maximum Likelihood paradigm proposed by Kiefer and Wolfowitz [14]. Kiefer and Wolfowitz [15] seem to have pioneered the area of shape-restricted inference. Wang [36, 37, 38], extended ideas of Kiefer and Wolfowitz to the estimation of distribution functions under the restriction of being star-shaped or being Increasing Failure Rate on Average. Lo [19], Rojo [26, 27], and Rojo and Ma [28], provide nonparametric estimators for distribution functions that are stochastically ordered. One recent monograph that examines shape-restricted inference is Groeneboom and Jongbloed [11]. Marshall and Olkin [20] and Shaked and Shanthikumar [30] provide excellent treatises on the topic of partial orders of distribution functions.

Finally, in Section 4, we apply our procedure to test for the existence of a monotonic trend in the size of the excesses of a time series of ozone measurements.

2. STOCHASTIC ORDERING OF THE GPD

The concept of stochastic order permeates the theory and applications of statistics. The concept was introduced in the seminal paper by Lehmann [17] and was used to study the power properties of certain tests.

Definition 2.1. Let X and Y be random variables such that

$$P(X > x) \leq P(Y > x), \qquad -\infty < x < \infty .$$

Then X is said to be *stochastically smaller* than Y. This is denoted by $X <^{\text{st}} Y$.

We can also state that Y is stochastically larger than X and write $Y >^{\text{st}} X$. If F and G represent the cumulative distribution functions (*cdfs*) of X and Y respectively, then $X <^{\text{st}} Y$ if and only of $F(x) \geq G(x)$ for all $x \in \mathbb{R}$, and then we write $F <^{\text{st}} G$. As discussed by Lehmann [17], a convenient situation arises when the stochastic order is induced by the parameter as it varies monotonically in the parameter space. That is, a parametric family of cdfs $\{F(x;\theta) : \theta \in \Theta \subset \mathbb{R}\}$ is stochastically increasing in θ if $\theta_1 < \theta_2$ implies that $F(\cdot;\theta_1) <^{\text{st}} F(\cdot;\theta_2)$. Similarly, $\{F(x;\theta) : \theta \in \Theta \subset \mathbb{R}\}$ is stochastically decreasing in θ if $\theta_1 < \theta_2$ implies that $F(\cdot;\theta_2) <^{\text{st}} F(\cdot;\theta_1)$. Lehmann and Rojo [18] provided simple characterizations of this and other related orders.

Sufficient conditions are provided here for the family of GPD distribution functions $\mathcal{F} = \{F(x;\xi,\beta) : -\infty < \xi < \infty,\ \ \beta > 0\}$, to be stochastically ordered. Since β is a scale parameter it is clear that the family $\mathcal{F}$ is stochastically ordered in β for fixed ξ. The following Proposition states that the family $\mathcal{F}$ is stochastically decreasing in ξ for fixed β.

Proposition 2.1. *Let $F_1, F_2 \in \mathcal{F}$ with shape parameters ξ_1 and ξ_2, respectively and equal scale β. If $\xi_1 < \xi_2$ then $F_2 <^{\text{st}} F_1$.*

Proof: The proof of Proposition 2.1 uses the following result.

Proposition 2.2 (Mitrinovic [21], pp. 266, inequality 3.6.1)**.** *If $a > 0$ and $x > 0$, then*

$$e^{-x} \leq \left(\frac{a}{ex}\right)^{a}. \tag{2.1}$$

Setting $x = 1/u$ and $a = 1$ in (2.1) we obtain

$$u \geq e^{1-1/u}, \qquad u > 0 . \tag{2.2}$$

Now we prove Proposition 2.1. Let $F(\cdot;\xi,\beta) \in \mathcal{F}$ for β fixed. From the definition of the usual stochastic order, it is enough to show that $F(\cdot;\xi,\beta)$ is an increasing function of the parameter $\xi \in \mathbb{R}$. This is true if and only if

$$h(\xi) \;=\; \log\bigl[1 - F(x;\xi,\beta)\bigr] \;=\; (1/\xi)\log(1 - \xi x/\beta)$$

is a decreasing function. First we analyze the case $\xi \neq 0$, for which the problem reduces to showing that

$$h'(\xi) \;=\; -(1/\xi^2)\log(1 - \xi x/\beta) - \frac{x}{\xi\beta(1 - \xi x/\beta)} \;<\; 0 . \tag{2.3}$$

Making the change of variable $u = 1 - \xi x/\beta$ we get $h'(\xi) = h'\bigl(\beta(1-u)/x\bigr) = g(u)$, where

$$g(u) \;=\; -\bigl[x/\beta(1-u)\bigr]^2 \bigl(\log u + (1/u) - 1\bigr) ,$$

for $0 < u < 1$ when $\xi > 0$, and $1 < u < \infty$ when $\xi < 0$. Then, $g(u) < 0$ if and only if $\log u + (1/u) - 1 > 0$, if and only if $u > e^{1-1/u}$, $u > 0$. But this is the strict inequality in (2.2). Hence (2.3) holds and therefore $F(x;\xi,\beta)$ is increasing in ξ for $\xi \in \mathbb{R}\backslash\{0\}$. Now

$$\lim_{\xi\to 0}\left\{1-(1-\xi x/\beta)^{1/\xi}\right\} = 11 - e^{-x/\beta}.$$

This means that $F(x;\xi,\beta)\uparrow F(x;\xi{=}0,\beta)$ as $\xi\uparrow 0$, and $F(x;\xi,\beta)\downarrow F(x;\xi{=}0,\beta)$ as $\xi\downarrow 0$. Then, from the proved monotonicity of $F(x;\xi,\beta)$ in $\xi\in\mathbb{R}\backslash\{0\}$, the proposition follows. □

Thus, the following result is obtained.

Corollary 2.1. *Let $F(x;\xi;\beta)$ denote the GPD distribution with scale parameter β and shape parameter ξ as defined by (1.2). Then,*

If $\xi^* = \xi$ and $\beta < \beta^*$, $F(\cdot;\xi^*,\beta^*) \geq^{\mathrm{st}} F(\cdot;\xi,\beta)$.

If $\xi > \xi^*$ and $\beta = \beta^*$, $F(\cdot;\xi^*,\beta^*) \geq^{\mathrm{st}} F(\cdot;\xi,\beta)$.

When $\xi > -1$, the expected value μ of a GPD(ξ,β) is $\mu = \beta(1+\xi)^{-1}$. Then $\xi = \xi(\mu) = (\beta/\mu) - 1$. Thus the shape parameter ξ is a decreasing function of the mean μ. So, if $X_1 \sim$ GPD(ξ_1,β) and $X_2 \sim$ GPD(ξ_2,β), with $\xi_1,\xi_2 > -1$, and we assume that the means $\mu_j = EX_j$ $(j=1,2)$ are such that $\mu_2 \leq \mu_1$, then $\xi_1 < \xi_2$. Thus if $\mu_2 \leq \mu_1$ then $X_2 <^{\mathrm{st}} X_1$. The converse is also true. To see this, let F_j be the cdf of X_j and assume $X_2 <^{\mathrm{st}} X_1$, then we have $1-F_2(x) \leq 1-F_1(x)$ for all x, and since the GPD only takes positive values, it follows that

$$\mu_2 = \int_0^\infty \bigl[1-F_2(x)\bigr]\,dx \leq \int_0^\infty \bigl[1-F_1(x)\bigr]\,dx = \mu_1\,.$$

We can put together all these results in the following corollary.

Corollary 2.2. *Let $X_j \sim GPD(\xi_j,\beta)$, (or if $X_j \sim GPD(\xi,\beta_j)$), $(j=1,...,k)$. Suppose that $E(X_j) = \mu_j$ exists for all j. Then the following propositions are equivalent.*

a) $X_1 \overset{\mathrm{st}}{>} X_2 \overset{\mathrm{st}}{>} \cdots \overset{\mathrm{st}}{>} X_k$.

b) $\xi_1 < \xi_2 < \cdots < \xi_k$, $(\beta_1 > \beta_2 > \cdots > \beta_k)$.

c) $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_k$.

3. TESTING FOR A LINEAR TREND IN THE EXCESSES

Let $X_j \sim$ GPD(ξ_j,β), $(j=1,...,k)$, and denote equality in distribution by $\overset{\mathcal{D}}{=}$. Suppose we want to test the null hypothesis

$$H_0\colon\ X_1 \overset{\mathcal{D}}{=} X_2 \overset{\mathcal{D}}{=} \cdots \overset{\mathcal{D}}{=} X_k$$

vs. the alternative

$$H_a\colon\ X_1 >^{\mathrm{st}} X_2 >^{\mathrm{st}} \cdots >^{\mathrm{st}} X_k\,.$$

From Corollary 2.2, we see that this would be equivalent to testing the null hypothesis

$$H_0\colon\ \xi_1 = \xi_2 = \cdots = \xi_k$$

vs. the alternative hypothesis

$$H_a\colon\ \xi_1 < \xi_2 < \cdots < \xi_k\,.$$

Similarly, the hypothesis $H_a\colon\ X_1 <^{\rm st} X_2 <^{\rm st} \cdots <^{\rm st} X_k$ can be tested by using $H_a\colon\ \xi_1 > \xi_2 > \cdots > \xi_k$. From Corollary 2.2, a test for the stochastic order could also be based on the means of the GPD's. However the means do not always exist. Therefore we test the hypothesis of stochastic order on the basis of the shape parameter. Assume that for each X_j we have a random sample of size n_j, $x_j = (x_{1j}, ..., x_{n_j j})'$ and let $x = (x_1, x_2, ..., x_k)$ be the full data vector. Furthermore, assume that we observe the X_j's sequentially along time, and let t_j be the epoch at which the random sample x_j was observed. To detect a linear time trend, we introduce a third parameter θ by writing $\xi_j = \xi + \theta t_j$, $(j = 1, ..., k)$. When the t_j's are equally spaced, t_j can be set as $t_j = j$. Thus, we can test the hypothesis of order restriction by testing

$$H_0\colon\ \theta = 0$$

vs. the alternative hypothesis

$$H_a\colon\ \theta \neq 0\,. \tag{3.1}$$

Although other forms of monotonic trends could occur, e.g. $\xi_j = \xi\exp(\theta t_j)$, a test without assuming a particular form of the monotone trend would require a semiparametric model that would provide protection against misspecification of the functional form of the trend but would not perform as well as the current test for the specific alternative of a monotonic linear trend.

Modeling the parameters of the GPD in order to assess a trend is similar to the approach described in other works such as those by Smith [34], Smith and Huang [35] and Rootzen and Tajvidi [29]. For instance, Rootzen and Tajvidi model the scale parameter as $\beta = \exp(\alpha_0 + \alpha_1 t)$ where t is time in years, and keep the shape parameter ξ constant. In this work we reverse this procedure.

Let $\underset{\sim}{X}$ represent the data vector $X_1, X_2, ..., X_n$. For testing the hypothesis (3.1), we use the Likelihood Ratio Test (LRT) based on $\lambda(\underset{\sim}{X}) = L(\hat{\xi}, \hat{\beta})/L(\hat{\xi}, \hat{\theta}, \hat{\beta})$, where L denotes the likelihood function and the estimators are maximum likelihood estimators (MLE). Then $-2\log\lambda(\underset{\sim}{X})$ follows asymptotically a chi-square distribution with one degree of freedom. The detailed expression for $-2\log\lambda(\underset{\sim}{X})$ is given in the Appendix.

4. AN APPLICATION TO OZONE DATA

The data we analyze was collected in Yosemite National Park Wanona Valley and consists of hourly measurements of ozone (ppm) taken from April 1, 1987 to October 31, 1996. The time series contains 84,011 observations with 9412 missing values. The main concern is the detection of a long term trend in the extremal behavior of the time series.

More precisely, the problem is to detect either a decreasing or increasing trend in the size of the excesses over a certain high threshold, if in fact a trend exists. Table 1 displays the monthly number of exceedances over 0.08 ppm. The observations have a strong seasonal component with two periods: the exceedances period which extends from the month of April trough the month of October and the no-exceedances period in the remaining months. The frequency of exceedances increases in the summer months and then decreases in the fall months. Moreover, exploring the data we found that the ozone levels also tend to increase in the summer months and decrease in the fall months. Since the interest lies on the extremal behavior of the data, the analysis was based on the months from April to October.

Table 1: Monthly exceedances over 8 ppm.

Year	Apr	May	Jun	Jul	Aug	Sep	Oct	Total (N_u)	n
1987	4	14	70	55	75	50	23	291	4742
1988	9	2	11	83	71	92	21	289	4856
1989	0	6	9	32	29	7	0	83	4913
1990	1	8	34	91	65	63	3	265	4630
1991	0	0	2	19	1	38	17	77	4463
1992	0	2	14	27	49	21	11	124	4736
1993	0	0	3	20	11	21	0	55	3860
1994	6	6	3	14	3	0	0	32	4720
1995	0	0	0	6	50	27	0	83	4804
1996	0	0	4	39	29	22	2	96	4636
Total	20	38	150	386	383	341	77	1395	46360

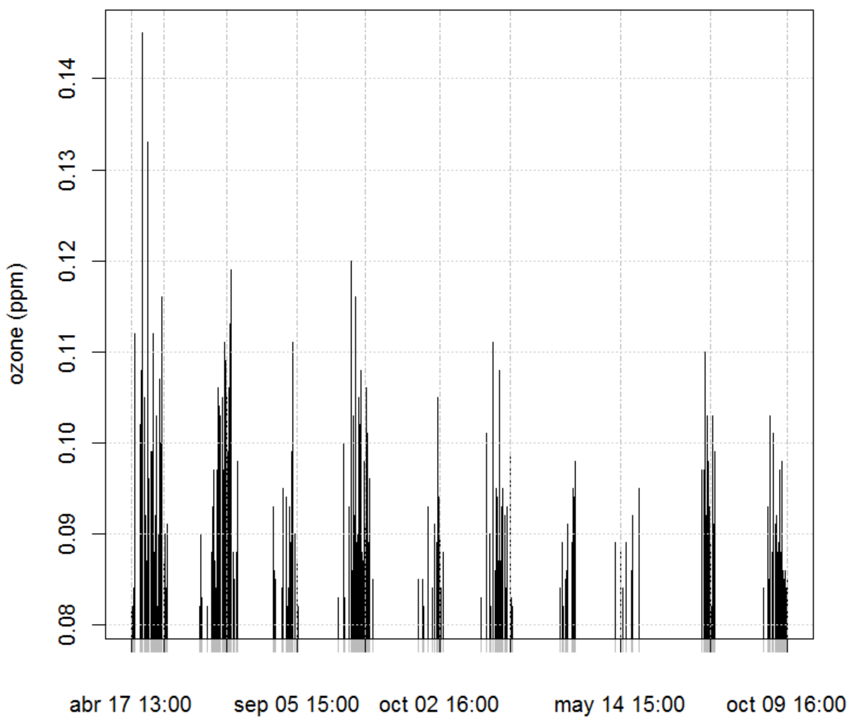

Figure 1: Excesses over 0.08 ppm.

Figure 1 shows the empirical marked point processes of exceedances over 0.08 ppm. A clear decreasing trend in the size of the excesses appears. We assess the significance of this trend using the LRT from Section 3.

The LRT requires the excesses to be independent of one another. There is, however, a strong dependence between the exceedances because they tend to occur in clusters. That is, an exceedance tends to attract other exceedances. Several procedures to deal with dependent data have been proposed. One such procedure is to identify clusters of exceedances for which it can be assumed that the excesses within any cluster are independent of the excesses within any other cluster, and then select the maximum excesses within each cluster.

The practical problem with this approach is the identification of independent clusters. Two methods have been used. One is to select a time length b (called block length) and then partition all the observations into consecutive blocks of length b. Then consider all the exceedances within a block as a cluster of exceedances. These are called block-clusters. See Leadbetter [16] for the formal justification of this approach as well as for some applications.

The second approach is to select a positive integer r (called the run length) and then decide that any run of at least r consecutive observations below the threshold separates two clusters of exceedances and then assume that such clusters are independent. These are called run-clusters. See Smith [33] for an application of this approach. In this work we use the run-cluster approach with 72 hours (three days) separation. This window of 72 hours is the common practice when analyzing ozone data. Once we have identified the run clusters, we take the maximum excess within each cluster. To distinguish from the *Exceedances over a Threshold* we call these values the *Peaks over a Threshold*, (POT's). Table 2 shows the POT's that we analyze in this work.

Table 2: POT, run-clusters, 72 hours.

Year	Peaks									
1987	0.002	0.004	0.032	0.022	0.065	0.025	0.053	0.032	0.027	0.036
	0.010	0.011								
1988	0.002	0.010	0.002	0.013	0.017	0.007	0.017	0.026	0.023	0.025
	0.031	0.039	0.008	0.018						
1989	0.013	0.005	0.015	0.014	0.004	0.013	0.009	0.019	0.031	0.010
	0.007	0.002								
1990	0.003	0.020	0.003	0.013	0.040	0.036	0.007	0.018	0.026	0.016
	0.005									
1991	0.005	0.005	0.002	0.013	0.004	0.011	0.007	0.025	0.010	0.008
1992	0.003	0.021	0.010	0.031	0.006	0.015	0.007	0.028	0.012	0.013
	0.019	0.003	0.002							
1993	0.004	0.009	0.011	0.009	0.015	0.018				
1994	0.009	0.008	0.004	0.009	0.006	0.012	0.015			
1995	0.017	0.017	0.030	0.023	0.018	0.009	0.023	0.011	0.019	
1996	0.004	0.013	0.023	0.021	0.012	0.009	0.017	0.018	0.006	0.006
	0.004	0.005								

Figure 2 shows the POT's for all the years of the observation period. The decreasing trend in the POT's is evident. Under H_0 the estimates of the parameters are $\hat{\xi} = 0.2121$ and

$\hat{\beta} = 0.0179$. Under H_a we have $\hat{\xi} = 0.164$, $\hat{\theta} = 0.0575$, and $\hat{\beta} = 0.0209$. The positive value of the estimate of ξ is consistent with the observed decrease in the excesses of the ozone levels. The observed value of the LRT is $-2 \log \lambda(x) = 17.24$ which has a p-value of 0.000033. Thus we conclude that the observed decrease in the size of the excesses from 1987 to 1996 is statistically significant.

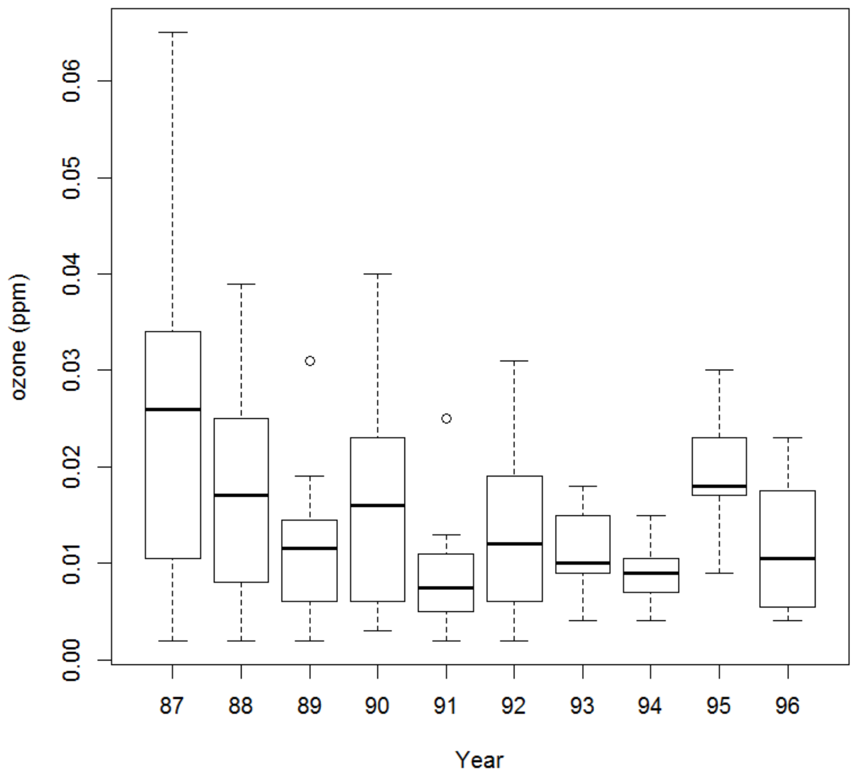

Figure 2: Maximum excesses within run-clusters grouped by years.

Once we have found statistical evidence for the decreasing trend in the excesses, we estimate the upper tail of the ozone levels as in Davison and Smith [7] or Embrechts *et al.* [8]. From (1.1) one gets

$$1 - F(u + x) = \gamma_u \big[1 - F_u(x)\big],$$

where $\gamma_u = \Pr(X > u) = 1 - F(u)$. Thus, if N_u is the number of exceedances over u and n is the number of observations, then an estimator of γ_u is $\hat{\gamma}_u = N_u/n$, and an estimator of the upper tail of F_X is given by

$$1 - \hat{F}(u + x) = \hat{\gamma}_u \big[1 - \hat{F}_u(x)\big] = \frac{N_u}{n}\left(1 - \hat{\xi}\,\frac{x}{\hat{\beta}}\right)^{1/\hat{\xi}}, \qquad x > 0\,. \tag{4.1}$$

Estimators of the quantiles of F are obtained by solving $\hat{F}(x_p) = p$ for x_p in (4.1), $0 \le p \le 1$. This yields

$$\hat{x}_p = u + \frac{\hat{\beta}}{\hat{\xi}}\left[1 - \left(\frac{n(1-p)}{N_u}\right)^{\hat{\xi}}\,\right]. \tag{4.2}$$

When $\hat{\xi} > 0$ by setting $p = 1$ we obtain the estimator of the right end point $\hat{x}^* = u + \hat{\beta}/\hat{\xi}$.

The ozone levels are not independent. So, to simplify, we assume that within the exceedances period in the year (from April to October) the ozone levels come from a strongly stationary process. Then, from the Ergodic Theorem — see Breiman [3], pp. 118 —, we have that $(1/n)\sum_{i=1}^{n} 1_{\{X_i>u\}} = N_u/n$ converges almost surely to $1-F(u)$, where now F is the marginal distribution of the ozone levels. Thus N_u/n may be used as an estimator of $1-F(u)$, and then we can use (4.2) to estimate the upper tail and high quantiles of the distribution of the ozone levels. Table 3 shows the estimates of the shape parameters and from Table 1 we get the number of observations (ozone measurements) and the number of exceedances per year. With this information we can estimate the extreme quantiles of the ozone levels. For instance, for 1987, we have

$$\hat{x}_p = 0.08 + (0.0209)\Big(1 - \big[4742\,(1-p)/291\big]^{0.22}\Big)/0.22\,, \qquad 0 \le p \le 1\,.$$

Figure 3 shows the estimated 0.99, 0.999 quantiles as well as the right endpoints of the marginal distribution of the ozone levels. The decreasing trend is evident.

Table 3: Estimated shape parameters.

t_j	1	2	3	4	5	6	7	8	9	10
$\hat{\xi}_j$	.22	.278	.336	.394	.452	.51	.568	.626	.684	.742

Figure 3: Estimated 0.99 and 0.999 quantiles, and estimated right endpoints of the distribution of ozone levels for a threshold of 0.08 ppm.

5. CONCLUSIONS

An exploratory data analysis of the extreme values of a time series of ozone levels made clear the existence of a decreasing linear trend in the size of the excesses over the threshold 8 ppm. We fitted the GPD to the POT's of the time series. By modeling the shape parameter of the GPD as a linear function of time in years, we were able to test the significance of a trend in the size of the excesses. More specifically, consider the years s and t with $s,t = 1987,...,1996$. Then we can say that the ozone excesses over 8 ppm for year s were more likely to take larger values then the ozone excesses over 8 ppm for year t, when $s < t$.

A. APPENDIX – Maximum Likelihood Calculations

The density function of a GPD(ξ,β) is

$$f(x;\xi,\beta) = \begin{cases} (1/\beta)\,(1-\xi x/\beta)^{(1/\xi)-1}\,, & \xi \neq 0\,,\ \beta > 0\,, \\ (1/\beta)\exp(-x/\beta)\,, & \xi = 0\,,\ \beta > 0\,. \end{cases}$$

Let $X_j \sim \mathrm{GPD}(\xi_j,\beta)$, and let $x_j = (x_{1j},...,x_{n_j j})'$ be a random sample from X_j, $(j = 1,...,k)$. Write $\xi_j = \xi + \theta t_j$. Then the log-likelihood function under $H_0\colon\ \theta = 0$ is

$$\text{(A.1)}\qquad l(\xi,\beta) \;=\; \sum_{j=1}^{k}\sum_{i=1}^{n_j} \log f(x_{ij};\xi,\beta) \;=\; -n\log\beta + (\xi^{-1}-1)\sum_{j=1}^{k}\sum_{i=1}^{n_j}\log(1-\xi x_{ij}/\beta)\,,$$

where $n = \sum_{j=1}^{k} n_j$, $(\xi,\beta) \in \Theta_0 = \big\{(\xi,\beta)\colon\ \xi<0,\ \beta>0\big\} \cup \big\{(\xi,\beta)\colon\ \xi>0,\ \beta>0,\ \text{and}\ \beta/\xi > \max_{ij}(x_{ij})\big\}$. Making the reparametrization $(\xi,\beta) \mapsto (\xi,\tau)$, where $\tau = \xi/\beta$, the log-likelihood function becomes

$$l(\xi,\tau) \;=\; -n\log\xi + n\log\tau + (\xi^{-1}-1)\sum_{j=1}^{k}\sum_{i=1}^{n_j}\log(1-\tau x_{ij})\,,$$

where $\big\{\xi<0,\ \tau>0\big\} \cup \big\{0<\xi\leq 1,\ \tau < 1/\max_{ij}(x_{ij})\big\}$. The log-likelihood equations are

$$\text{(A.2)}\qquad \frac{\partial l}{\partial \xi} \;=\; (n/\xi) - (1/\xi^2)\sum_{j=1}^{k}\sum_{i=1}^{n_j}\log(1-\tau x_{ij}) \;=\; 0\,,$$

$$\text{(A.3)}\qquad \frac{\partial l}{\partial \tau} \;=\; (n/\tau) - (\xi^{-1}-1)\sum_{j=1}^{k}\sum_{i=1}^{n_j}\frac{x_{ij}}{1-\tau x_{ij}} \;=\; 0\,.$$

Solving equation (A.2) for ξ we obtain

$$\text{(A.4)}\qquad \xi(\tau) \;=\; -(1/n)\sum_{j=1}^{k}\sum_{i=1}^{n_j}\log(1-\tau x_{ij})\,.$$

Since equation (A.4) gives ξ as an explicit function of τ, we can substitute $\xi(\tau)$ of (A.4) in equation (A.3), and obtain

$$(n/\tau) - \big(\xi(\tau)^{-1} - 1\big) \sum_{j=1}^{k} \sum_{i=1}^{n_j} \frac{x_{ij}}{1 - \tau x_{ij}} \;=\; 0\,,$$

which can be solved numerically for τ. If $\hat{\tau}$ is the solution, then the MLE's of ξ and β are given by $\hat{\xi} = \xi(\hat{\tau})$ and $\hat{\beta} = \hat{\xi}/\hat{\tau}$, respectively. This is the standard procedure to find the MLE's of the parameters of the GPD. For a detailed analysis of this procedure see Grimshaw [10]. Under $H_a\colon\ \theta > 0$ the log-likelihood function is

$$\begin{aligned} l(\xi,\theta,\beta) \;&=\; \sum_{j=1}^{k} \sum_{i=1}^{n_j} \log f(x_{ij};\xi,\theta,\beta) \\ &=\; -n \log\beta + \sum_{j=1}^{k} \big[(\xi + \theta t_j)^{-1} - 1\big] \sum_{i=1}^{n_j} \log\big[1 - (\xi + \theta t_j)\, x_{ij}/\beta\big]\,, \end{aligned}$$

where $(\xi,\theta,\beta) \in \Theta_a = \big\{(\xi,\theta,\beta)\colon \xi+\theta t_j < 0,\ j = 1,...,k,\ \beta > 0,\ \theta > 0\big\} \cup \big\{(\xi,\theta,\beta)\colon \xi+\theta t_j > 0,\ j = 1,...,k,\ \beta > 0,\ \theta > 0 \text{ and } \beta/(\xi+\theta t_j) > \max_i(x_{ij}),\ j = 1,...,k\big\}$. Let $x_{(n_j)j} = \max_i(x_{ij})$, and note that the restriction $\beta/(\xi+\theta t_j) > x_{(n_j)j}$ is equivalent to $\beta - \xi\, x_{(n_j)j} - \theta\, x_{(n_j)j}\, t_j > 0$. So, the parameter space $\Theta_a \subset \mathbb{R}^3$ is given by all the $(\beta,\xi,\theta)'$ that satisfy the linear pointwise restrictions

$$\begin{bmatrix} 1 & -x_{(n_1)1} & -x_{(n_1)1}t_1 \\ 1 & -x_{(n_2)2} & -x_{(n_2)2}t_2 \\ \vdots & \vdots & \vdots \\ 1 & -x_{(n_k)k} & -x_{(n_k)k}t_k \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta \\ \xi \\ \theta \end{bmatrix} > \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}.$$

Finding the MLE's of ξ, θ, and β becomes a problem of maximization with linear constraints. There are several numerical algorithms to deal with this type of problem. In this work we used the Price's controlled random search procedure. See Khuri [13], pp. 334–336, for the details of this algorithm. The calculations were performed with R. The test statistic is given by $-2\log\lambda(x) = 2\big[l(\hat{\xi},\hat{\theta},\hat{\beta}) - l(\hat{\xi},\hat{\beta})\big]$.

ACKNOWLEDGMENTS

This work was supported by NSF Grant DMS 1731082 and NSA grant H98230-19-1-0016.

REFERENCES

[1] Balkema, A.A. and de Haan, L. (1974). Residual life at great age, *Annals of Probability*, **2**, 792–804.

[2] Barlow, R.E.; Bartholomew, D.J.; Bremner, J.M. and Brunk, H.D. (1972). *Statistical Inference Order Restrictions. The Theory and Application of Isotonic Regression*, London–New York–Sidney, Wiley.

[3] Breiman, L. (1968). *Probability*, Reading Massachussetts, Addisson Wesley.

[4] Castillo, E. and Hadi, A.S. (1997). Fitting the generalized Pareto distribution to data, *Journal of the American Statistical Association*, **92**, 1609–1620.

[5] Dargahi-Noubary, G.R. (1989). On tail estimation: an improved method, *Mathematical Geology*, **21**, 829–842.

[6] Davison, A.C. (1984). *Modeling excesses over high thresholds, with application.* In "Statistical Extremes and Applications" (J. Tiago de de Oliveira, Ed.), Dordrecht, Reidel Publishing Company.

[7] Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high threshold (with discussion), *J. R. Statistical Society B*, **52**, 393–442.

[8] Embrechts, P.; Kluppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events*, Berlin, Springer Verlag.

[9] Falk, M.; Husler, J. and Reiss, R.D. (1994). *Laws of Small Numbers: Extreme and Rare Events*, Basel, Birkhäuser Verlag.

[10] Grimshaw, S.D. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution, *Technometrics*, **35**, 185–191.

[11] Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric Estimation under Shape Constraints*, New York, Cambridge University Press.

[12] Hosking, J.R.M. and Wallis, J.R. (1987). Parameter and quantile estimation for the Generalized Pareto distribution, *Technometrics*, **29**, 339–349.

[13] Khuri, A.I. (1993). *Advanced Calculus with Applications in Statistics*, New York, Wiley.

[14] Kiefer, J. and Wolfowitz, J. (1956). Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters, *The Annals of Mathematical Statistics*, **27**, 887–906.

[15] Kiefer, J. and Wolfowitz, J. (1976). Asymptotically minimax estimation of concave and convex distribution function, *Z. Wahrsch. verw. Gebiete.*, **34**, 73–85.

[16] Leadbetter, M.R. (1995). On high level exceedance modeling and tail inference, *Journal of Statistical Planning and Inference*, **45**, 247–260.

[17] Lehmann, E.L. (1955). Ordered Families of Distributions, *Annals of Mathematical Statistics*, **26**, 399–419.

[18] Lehmann, E.L. and Rojo, J. (1992). Invariant Directional Orderings, *Annals of Statistics*, **20**, 2100–2110.

[19] Lo, S.H. (1987). Estimation of Distribution Functions Under Order Restrictions, *Statistics and Decisions*, **5**, 251–262.

[20] Marshall, A.W. and Olkin, O. (2007). *Life Dsitributions — Structure of Nonparametric, Semiparametric, and Parametric Families*, Springer.

[21] Mitrinovic, D.S. (1970). *Analytic Inequalities*, New York, Springer Verlag.

[22] Parisi, F. and Lund, R. (2000). Seasonality and Return Periods of Landfalling Atlantic Basin Hurricanes, *Australian New Zeland Journal of Statistics*, **42**(3), 271–282.

[23] Pickands, J. (1975). Statistical inference using extreme order statistics, *Annals of Statistics*, **3**, 119–131.

[24] Reiss, R.D. and Thomas, M. (1997). *Statistical Analysis of Extreme Values*, Basel, Birkhäuser Verlag.

[25] Robertson, T.; Wright, F.T. and Dykstra, R.L. (1988). *Order Restricted Statistical Inference*, Chichister, Wiley.

[26] Rojo, J. (1995). On the weak convergence of certain estimators of stochastically ordered survival functions, *Journal of Nonparametric Statistics*, **4**(4), 349–363.

[27] Rojo, J. (2004). On the estimation of survival functions under stochastic order constraint, *Institute of Mathematical Statistics: Lecture Notes – Monograph Series*, **44**, 37–61, "The First Erich L. Lehmann Symposium – Optimality" (J. Rojo and V. Pérez-Abreu, Eds.).

[28] Rojo, J. and Ma, Z. (1988). On the Estimation of Stochastically Ordered Survival Functions, *Journal of Statistical Computation and Simulation*, **55**, 1–21.

[29] Rootzen, H. and Tajvidi, N. (1997). Extreme value statistics and wind storm losses: A case study, *Scandinavian Actuarial Journal*, **1**, 70–94.

[30] Shaked, M. and Shanthikumar, J.G. (1994). *Stochastic Orders and Their Applications*, Boston, Academic Press.

[31] Smith, R.L. (1984). *Threshold methods for sample extremes.* In "Statistical Extremes and Applications" (J. Tiago de Oliveira, Ed.), Dordrecht, Reidel Publishing Company.

[32] Smith, R.L. (1987). Estimating tails of probability distributions *Annals of Statistics*, **15**, 1174–1207.

[33] Smith, R.L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground level ozone, *Statistical Science*, **4**, 367–393.

[34] Smith, R.L. (1990). *Extreme value theory*, Chapter 14 of "Handbook of Applicable Mathematics Supplement" (W. Ledermann, E. Lloyd, S. Vajda and C. Alexander, Eds.), John Wiley, Chichester, 437–472.

[35] Smith, R.L. and Huang, L.-S. (1993). *Modeling high threshold exceedances of urban ozone*, Technical Report # 6, National Institute of Statistical Sciences, Research Triangle Park.

[36] Wang, J.L. (1987a). Estimating IFRA and NBU survival curves based on censored data, *Scandinavian Journal of Statistics*, **14**, 199–210.

[37] Wang, J.L. (1987b). Estimators of a distribution function with increasing failure rate average, *Journal of Statistical Planning and Inference*, **16**, 415–427.

[38] Wang, J.L. (1988). Optimal estimation of a star shaped distribution function, *Statisticcs and Decisions*, **6**, 21–32.