A NEW UNSUPERVISED THRESHOLD DETERMINATION FOR HYBRID MODELS

Nehla DEBBABI [1,2] *and Marie KRATZ* [3]

[1] CReSTIC, Faculté des Sciences Exactes et Naturelles, Université de Reims Champagne Ardenne
[2] Research Lab. COSIM, Engineering College of Communications of Tunis, University of Carthage
[3,*] ESSEC Business School, CREAR Risk research center
nehla.debbabi@supcom.tn, kratz@essec.edu

ABSTRACT

A Gauss-GPD hybrid model that links a Gaussian distribution to a Generalized Pareto Distribution (GPD) is considered for asymmetric heavy tailed data. The paper proposes a new unsupervised iterative algorithm to find successively the junction point between the two distributions and to estimate the hybrid model parameters. Simulation results show that this method provides a reliable position for the junction point, as well as an accurate estimation of the GPD parameters, which improves results when compared with other methods. Another advantage of this approach is that it can be adapted to any hybrid model.

Index Terms— Heavy-tailed data modelling, Hybrid density estimation, Extreme Value Theory (EVT), Generalized Pareto distribution (GPD), Unsupervised algorithm.

1. INTRODUCTION

Modelling non-homogeneous and multi-component data is a problem that challenges scientific researchers in several fields [1, 2, 3, 4, 5, 6]. In general, it is not possible to find a simple and closed form probabilistic model to describe such data. Therefore, it seems natural to consider non-parametric approaches, such as *e.g.* kernel density estimation ones [7] or non-parametric Bayesian methods [8, 9] to name a few. However, when the multiple components are separable, parametric modelling becomes again tractable. Several hybrid models have been proposed in such context, combining two or more densities [1, 2, 3, 4, 5]. In this work, we are interested in the hybrid model that links a Gaussian distribution to a Generalized Pareto Distribution (GPD), noted G-GPD, to model asymmetric heavy tailed data, as e.g. in [3]. Why such a choice of modelling? Since we are interested in fitting the whole distribution underlying asymmetric heavy tailed data, the idea is to consider both the mean and tail behaviors, and to use limit theorems for each one, as suggested and developed analytically in [10]. On one hand, the Central Limit Theorem (CLT) justifies to introduce a Gaussian distribution for the mean behavior, on the other hand Pickand's theorem [11] tells us that the tail of the distribution may be evaluated through a GPD above a high threshold, using the Peak Over Threshold (POT) method of the Extreme Value Theory (EVT). Hence, if the G-GPD modelling may appear restrictive at first glance, it is in fact quite general since it could apply to any underlying distribution.

This modelling might be useful in various fields, under the presence of asymmetric heavy tailed data, as for instance in (re)insurance when dealing with asymmetric heavy-tailed claims, or in finance [3, 12]. Other examples can be given in signal processing, when considering the spike detection in neural signals in biomedicine [13], or the energy detection for unknown signals over a fading channels in telecommunication [14], or bearing defect early detection in vibratory signals in machine diagnostics [15].

The main issue with this modelling is the determination of the junction point (or threshold) between the distributions, and the estimation of the different parameters. Much literature in EVT has been dedicated to the determination and estimation of the threshold above which the observations can be modeled by a GPD, via standard POT methods (mean excess plot method, Hill estimator, QQ-estimator; see e.g. [13, 16, 17, 18, 19, 20]). The difficulty faced when applying these methods is that they are graphical ad hoc approaches. The offline solution of those methods represents an important disadvantage especially when complexity burden and/or delay processing are critical. To overcome this difficulty, especially in practice, unsupervised methods have been developed (e.g. [2, 3]), in particular in [3] for the G-GPD model.

In this paper we propose a new unsupervised approach for the selection of the junction point, as well as for the estimation of the G-GPD parameters. It starts by enforcing the continuity and the differentiability of the G-GPD probability density function (pdf) at the junction point, as in [3], but then proceeds in an iterative way to determine successive thresholds and parameters of the two distributions that are involved. This approach does not depend on the choice of initial parameters; the algorithm starts with a selection of an arbitrary value of the threshold and stops, when it converges, to a fixed threshold with respect to some given error. We are going to see that it provides good results, improving the estimation of the

[*] M. Kratz is also member of MAP5, UMR8145, Univ. Paris Descartes

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

model when compared with other methods, in particular with [3]. Note that another great benefit of our method is that it can be adapted to any other hybrid model, hence can resolve some non-homogenous data modeling problems.
The remainder of this paper is organized as follows. In section 2 we describe our iterative algorithm for the unsupervised threshold determination. Simulation results are presented in section 3. A comparison of our algorithm with the one presented in [3], as well as with standard methods of GPD parameters estimation, is discussed in the same section. Conclusions follow in the last section.

2. DESCRIPTION OF THE ITERATIVE APPROACH FOR UNSUPERVISED THRESHOLD SELECTION

Let us introduce the G-GPD hybrid model that we will consider, linking a Gaussian distribution and a GPD at a junction point denoted by u. It is defined by the pdf:

$$h(x;u,\mu,\sigma,\xi,\beta) = \begin{cases} \gamma f(x;\mu,\sigma), & \text{if } x \leq u \\ \gamma g(x-u;\xi,\beta), & \text{if } x > u \end{cases}$$

where $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$ are the mean and the standard deviation, respectively, of the Gaussian pdf f expressed as

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2}), \quad \forall x \in \mathbb{R}$$

while ξ and β represent the tail index and the scale parameter respectively of the GPD pdf g defined by

$$g(x;\xi,\beta) = \begin{cases} \frac{1}{\beta}(1+\frac{\xi}{\beta}x)^{-1-\frac{1}{\xi}}, & \text{if } \xi \neq 0 \\ \frac{1}{\beta}\exp(-\frac{x}{\beta}), & \text{if } \xi = 0 \end{cases} \quad \forall x \in \mathcal{D}(\xi,\beta)$$

where

$$\mathcal{D}(\xi,\beta) = \begin{cases} [0,\infty), & \text{if } \xi \geq 0 \\ [0,-\frac{\beta}{\xi}], & \text{if } \xi < 0 \end{cases}$$

The regulator factor γ ensures that the hybrid h is a pdf, *i.e.*

$$\int_{\mathbb{R}} h(x;u,\mu,\sigma,\xi,\beta)dx = 1.$$

It is expressed as a function of the Gaussian cumulative distribution function (cdf) F, with parameters μ and σ, evaluated at u as

$$\gamma = \frac{1}{F(u;\mu,\sigma)+1}.$$

The corresponding G-GPD cdf H is then expressed as

$$H(x;u,\mu,\sigma,\xi,\beta) = \begin{cases} \gamma F(x;\mu,\sigma), & \text{if } x \leq u \\ \gamma(F(u;\mu,\sigma) + G(x-u;\xi,\beta)), & \text{if } x > u \end{cases}$$

where G denotes the GPD cdf with parameters ξ and β.
In order to obtain a smooth pdf, the continuity and the differentiability of the G-GPD pdf at u are enforced, as in e.g. [3] *i.e.* $f(u;\mu,\sigma) = g(0;\xi,\beta)$ and $f'(u;\mu,\sigma) = g'(0;\xi,\beta)$. From these relations a deterministic ξ and β are deduced as functions of the three remaining parameters u, μ and σ, namely as

$$\beta_u = f(u;\mu,\sigma)^{-1}. \quad (1)$$

$$\xi_u = \beta\frac{(u-\mu)}{\sigma^2} - 1. \quad (2)$$

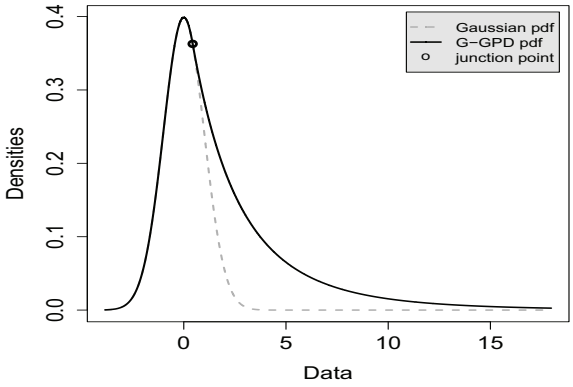

Fig. 1. Gaussian pdf (dotted curve) with parameters $\mu = 0$ and $\sigma = 1$ and *G-GPD* pdf (continuous curve) with parameters $\mu = 0$, $\sigma = 1$, $u = 0.4354$, $\xi = 0.2$ and $\beta = 2.7558$.

An example of the G-GPD pdf is illustrated in Fig. 1, where the difference between this latter and the Gaussian distribution is clearly noticeable.
The main idea of our method, after enforcing the continuity and the differentiability of the G-GPD pdf at the junction point, is to proceed in an iterative way. Each iteration is composed of two main steps. The first one consists in fitting the Gaussian distribution, minimizing the Squared Error (SE) between the G-GPD cdf obtained when considering the threshold determined in the previous iteration, and the empirical G-GPD one, using the Levenberg Marquardt algorithm [21, 22]. Once the Gaussian parameters μ and σ are determined, the second step aims at determining the next threshold by minimizing the SE between the G-GPD cdf, admitting these Gaussian parameters, and the empirical G-GPD one, using again the Levenberg Marquardt algorithm. Note that the GPD parameters are implicitly determined using the equations given for the junction point.

2.1. Construction of our iterative algorithm

First, we consider an n-sample $X = (x_i)_{1 \leq i \leq n}$ with a G-GPD as a parent distribution.

Step 1: We start our algorithm by determining the empirical

G-GPD cdf H_n of the n-sample X defined by

$$H_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_i \leq x\}}, \quad \forall x \in \mathbb{R},$$

where 1 denotes the indicator function.

Step 2: We choose an arbitrary initial threshold equal to a certain quantile of X, say u_0, *e.g.* $u_0 = q_{50\%}$, the 50% quantile. Then, we evaluate the mean μ_0 and the standard deviation σ_0 of the Gaussian distribution of the first iteration by minimizing the SE between the G-GPD cdf H, when considering the threshold u_0, and the empirical one H_n using the Levenberg-Marquardt method [21, 22]. As a matter of fact, we chose to use this method since it is a robust one for nonlinear system resolution based on a minimization in the least squares sense. Hence, the estimates $\widehat{\mu}_0$ and $\widehat{\sigma}_0$ of the parameters of the Gaussian distribution of the first iteration are obtained as a solution of the minimization problem:

$$(\widehat{\mu}_0, \widehat{\sigma}_0) = \underset{(\mu,\sigma)\in\mathbb{R}\times\mathbb{R}_+}{argmin} \|H(X; u_0, \mu, \sigma, \xi_{u_0}, \beta_{u_0}) - H_n(X)\|_2^2,$$

where ξ_{u_0} and β_{u_0} are the GPD parameters obtained using equations (1) and (2) and $\| \, . \, \|_2$ denotes the Euclidean norm.

Step 3: In the previous step we fitted the Gaussian part of the distribution, in this step we will fit the GPD one. To do so, we keep the threshold $\widehat{u}_1$ that minimizes the SE between the G-GPD cdf, using the estimated Gaussian parameters $\widehat{\mu}_0$ and $\widehat{\sigma}_0$, and the empirical one using as well the Levenberg-Marquardt method. In other words, $\widehat{u}_1$ is the solution of the minimization problem:

$$\widehat{u}_1 = \underset{u\geq 0}{argmin} \ \ \|H(X; u, \widehat{\mu}_0, \widehat{\sigma}_0, \xi_u, \beta_u) - H_n(X)\|_2^2,$$

where $\forall u \geq 0$, ξ_u and β_u are the GPD parameters relative to u, obtained using equations (1) and (2).

Step 4: At this stage, we proceed iteratively to fit the normal distribution and the GPD, simultaneously like we did in step 2 and 3. It means, for each iteration, the estimated Gaussian parameters are used to find the junction point while this latter will be used to evaluate the normal parameters of the next iteration, and so on. This can be written mathematically as follows: $\forall k \geq 1$

- $(\widehat{\mu}_k, \widehat{\sigma}_k) \leftarrow \underset{(\mu,\sigma)\in\mathbb{R}\times\mathbb{R}_+}{argmin} \|H(X; \widehat{u}_k, \mu, \sigma, \xi_{\widehat{u}_k}, \beta_{\widehat{u}_k}) - H_n(X)\|_2^2$
- $\widehat{u}_{k+1} = \underset{u\geq 0}{argmin} \ \ \|H(X; u, \widehat{\mu}_k, \widehat{\sigma}_k, \xi_u, \beta_u) - H_n(X)\|_2^2$
- The algorithm stops when the following condition is satisfied
$$\|u_{k+1} - u_k\|_2^2 < \epsilon$$
with $\epsilon > 0$, small enough.

Hereafter, we present the pseudocode associated to this algorithm.

2.2. Pseudocode of the proposed algorithm

Algorithm 1 Iterative algorithm for unsupervised threshold determination and G-GPD parameters estimation

Initialization of u_0, ϵ

1: Determination of the empirical G-GPD cdf H_n of the n-sample $X = (x_i)_{1\leq i\leq n}$

2: Estimation of $\widehat{\mu}_0$ and $\widehat{\sigma}_0$ when considering u_0;

$$(\widehat{\mu}_0, \widehat{\sigma}_0) \leftarrow \underset{(\mu,\sigma)\in\mathbb{R}\times\mathbb{R}_+}{argmin} \|H(X; u_0, \mu, \sigma, \xi_{u_0}, \beta_{u_0}) - H_n(X)\|_2^2$$

(β_{u_0}, ξ_{u_0}) are satisfying (1) and (2).

3: Determination of the first threshold $\widehat{u}_1$, using the fixed Gaussian parameters $(\widehat{\mu}_0, \widehat{\sigma}_0)$;

$$\widehat{u}_1 \leftarrow \underset{u\geq 0}{argmin} \ \ \|H(X; u, \widehat{\mu}_0, \widehat{\sigma}_0, \xi_u, \beta_u) - H_n(X)\|_2^2$$

4: Iterative process;
$k \leftarrow 1$
while $\|\widehat{u}_k - \widehat{u}_{k-1}\|_2^2 \geq \epsilon$

$$(\widehat{\mu}_k, \widehat{\sigma}_k) \leftarrow \underset{(\mu,\sigma)\in\mathbb{R}\times\mathbb{R}_+}{argmin} \|H(X; \widehat{u}_k, \mu, \sigma, \xi_{\widehat{u}_k}, \beta_{\widehat{u}_k}) - H_n(X)\|_2^2$$

$$\widehat{u}_{k+1} \leftarrow \underset{u\geq 0}{argmin} \ \ \|H(X; u, \widehat{\mu}_k, \widehat{\sigma}_k, \xi_u, \beta_u) - H_n(X)\|_2^2$$

$k \leftarrow k + 1$
end while
return $(u_{k+1}, \ \mu_k, \ \sigma_k, \ \xi_{u_{k+1}}, \ \beta_{u_{k+1}})$

3. SIMULATION RESULTS AND DISCUSSION

To assess the performance of our algorithm in terms of accurate estimation of the *G-GPD* model, we build on Monte Carlo simulations. In order to do this, we simulate 100 data sets of length $n = 10^5$, denoted by $X^k = (x_i^k)_{1\leq i\leq n}$, $\forall 1 \leq k \leq 100$, following the G-GPD distribution for a given fixed parameters $\theta = (u, \mu, \sigma, \xi_u, \beta_u)$. For each data set, we determine the estimate $\widehat{\theta}^k = (\widehat{u}^k, \widehat{\mu}^k, \widehat{\sigma}^k, \xi_{\widehat{u}^k}, \beta_{\widehat{u}^k})$ of θ using our algorithm. Then, we calculate the Kullback-Leibler (K-L) divergence, measuring the difference between the G-GPD pdf with parameter θ and the corresponding estimated one. It is defined as follows:

$$D_{KL}(h_k \parallel \widehat{h}_k) = \frac{1}{n}\sum_{i=1}^{n} h_k(x_i^k) \log\left(\frac{h_k(x_i^k)}{\widehat{h}_k(x_i^k))}\right),$$

where $h_k(X^k) = h(X^k; u, \mu, \sigma, \xi_u, \beta_u)$
and $\widehat{h}_k(X^k) = h(X^k; \widehat{u}^k, \widehat{\mu}^k, \widehat{\sigma}^k, \xi_{\widehat{u}^k}, \beta_{\widehat{u}^k})$, $\forall 1 \leq k \leq 100$.

The closest to 0 the K-L divergence is, the most accurate evaluation of the G-GPD model we obtain. To test the reliability of the G-GPD model parameters estimation, we choose the average of $D_{KL}(h_k \parallel \widehat{h}_k)$, $1 \leq k \leq 100$, as evaluation criterion. Table 1 gives the average K-L divergence obtained for three different examples of simulated data sets. The first G-GPD model is simulated with $\xi = 0.2$, $\mu = 0$ and $\sigma = 1$, the second one with $\xi = 1$, $\mu = 2$ and $\sigma = 1$ while the third with $\xi = 0.5$, $\mu = 1$ and $\sigma = 1$. We mention that for each case the exact values of β and u are determined using the obtained relations between the model parameters at the junction point. Table 1 exhibits the performance of our algorithm, the average K-L divergence being less than 10^{-3}. Comparing the results obtained with our method and Method [3], we see that the G-GPD model evaluation is enhanced using our algorithm.

Note that, based on simulations, our algorithm converges regardless the choice of the initial value of threshold. It is obvious that the convergence will be slower if the chosen value of initial threshold is far from the real junction point. For the simulations, we chose $u_0 = H_n^{-1}(0.5)$ and $\epsilon = 10^{-12}$ as initial parameters where H_n^{-1} denotes the inverse empirical cdf.

Besides the reliable determination of the threshold, our algorithm shows a good performance in terms of the GPD parameters estimation. To highlight this point, we compare the obtained results using our method with standard statistical methods [23, 24]. Several methods in literature have been developed for the estimation of the GPD parameters, among which the MOments Method (MOM) based on the use of the first and the second moment of the GPD. Knowing that the GPD distribution admits a finite moment of order k if and only if the tail index $\xi < \frac{1}{k}$, the performance of the MOM decreases when $\xi \geq \frac{1}{2}$. The Probability Weighted Moments (PWM) method, has then been proposed in [23] extending the MOM when $\xi < 1$, but the estimation of the GPD parameters is still unreliable for $\xi \geq 1$. Another statistical method, well used in literature, is the Maximum Likelihood Method (MLM) [23]. The limitation of this method is that it can returns a local maximum instead of the global one. To overcome the encountered problem, Zhang has proposed a specific method in [24] for any value of ξ based on a Bayesian analysis. Nevertheless, it remains an ad-hoc method.

We show in Table 2 that our method for G-GPD parameters estimation goes over those encountered problems. Indeed, when revisiting Examples 2 and 3, we can compare in Table 2 the mean value of the estimated GPD tail parameter $\xi_{\widehat{u}}^k$, $\forall 1 \leq k \leq 100$, as well as the mean value of the GPD shape parameter $\beta_{\widehat{u}}^k$, $\forall 1 \leq k \leq 100$, when using our algorithm and existing methods. We can easily infer the reliable estimation of the GPD parameters obtained via our algorithm, when computing the mean squared error (mse) between the exact and the estimated parameters via the different methods.

	Example 1	**Example** 2	**Example** 3
Method [3]	$1.6664\ 10^{-4}$	$8.9004\ 10^{-5}$	$5.3537\ 10^{-5}$
Our meth.	$1.6591\ 10^{-4}$	$4.3661\ 10^{-5}$	$3.2163\ 10^{-5}$

Table 1. Comparison of the average K-L divergence obtained using our method and the one presented in [3] for three different examples: Example 1 for $\xi = 0.2$, $\mu = 0$ and $\sigma = 1$, Example 2 for $\xi = 1$, $\mu = 2$ and $\sigma = 1$ while the third with $\xi = 0.5$, $\mu = 1$ and $\sigma = 1$.

	Our meth.	**MOM**	**PWM**	**MLM**	**Zhang**
Generated data sets with $\xi = 1$, $\mu = 2$, **and** $\sigma = 1$					
$\xi = 1$	0.9992	0.4992	0.835	0.5007	0.5007
mse(ξ) (10^{-3})	0.2347	250.74	30.2165	249.327	249.313
$\beta = 3.0905$	3.0917	25.0006	6.6813	8.2346	8.2347
mse(β)	$0.24\ 10^{-3}$	877.28	12.991	26.464	26.465
Generated data sets with $\xi = 0.5$, $\mu = 1$, **and** $\sigma = 1$					
$\xi = 0.5$	0.4991	0.4191	0.105	0.1876	0.1877
mse(ξ) (10^{-3})	0.1346	7.5503	156.55	97.609	97.532
$\beta = 2.8727$	2.8735	4.2269	5.5153	5.6406	5.636
mse(β)	$0.16\ 10^{-3}$	1.8692	13.275	7.6647	7.6367

Table 2. Comparison of the mean estimates of the GPD parameters using our method and existing ones, obtained from 100 generated data sets for two different cases, $(\xi, \mu, \sigma) = (1, 2, 1)$ and when $(\xi, \mu, \sigma) = (0.5, 1, 1)$, respectively.

4. CONCLUSION

This paper proposes a new numerical method, with an iterative algorithm for unsupervised threshold detection for a Gauss-GPD hybrid model that links a Gaussian distribution to a GPD. Simulation results show that our method enhances the evaluation of the junction point as well as the estimation of the two distributions parameters, in particular compared with the method proposed in [3]. Besides the accurate determination of the threshold (or junction point), we obtain a better estimation of the GPD parameters than with standard methods. We also avoid the issue of hand supervising. As a follow up, it would be interesting to explore a new hybrid model with two junction points introducing an intermediate behavior between the Gaussian distribution and the GPD.

REFERENCES

[1] A. K. Mandava, L. Shahram, and E. E. Regentova, "Reliability assessment of microarray data using fuzzy classification methods: A comparative study," in *Advances in Computing and Communications*. 2011, vol. 190 of *Communications in Computer and Information Science*, pp. 351–360, Springer Berlin Heidelberg.

[2] M. Knecht and S. Küttel, "The czeledin distribution function," in *XXXIV ASTIN Colloquium, Berlin*, August 2003.

[3] J. Carreau and Y. Bengio, "A hybrid pareto model for asymmetric fat-tailed data: the univariate case," *Extremes*, vol. 1, pp. 53–76, 2009.

[4] W. Gehin, "Modélisation des queues de distribution des rendements des actif financiers," M.S. thesis, EURIA, Brest, 2011.

[5] V. P. Singh C. Li and A. K. Mishra, "Simulation of the entire range of daily precipitation using a hybrid probability distribution," *WATER RESOURCES RESEARCH*, vol. 48, pp. 1–17, 2012.

[6] M. E. Kuhl and P.S. Bhairgond, "Nonparametric estimation of nonhomogeneous poisson processes using wavelets," in *Simulation Conference, 2000. Proceedings. Winter*, 2000, vol. 1, pp. 562–571 vol.1.

[7] K. JooSeuk and D.S Clayton, "Robust kernel density estimation," *Journal of Machine Learning Research*, vol. 13, pp. 2529–2565, 2012.

[8] S. G. Walker, P. Damien, P. W. Laud, and A. F. M. Smith, "Bayesian nonparametric inference for random distributions and related functions," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 61, no. 3, pp. 485–527, 1999.

[9] P. Orbanz and Y.W. Teh, "Bayesian nonparametric models," in *Encyclopedia of Machine Learning*. Springer, 2010.

[10] M. Kratz, "There is a var beyond usual approximations," 2013, preprint arXiv:1311.0270.

[11] J. Pickands, "Statistical inference using extreme order statistics," *Annals of Statistics*, vol. 3, pp. 119–131, 1975.

[12] J. Carreau and Y. Bengio, "A hybrid pareto mixture for conditional asymmetric fat-tailed distributions," *Neural Networks, IEEE Transactions on*, vol. 20, no. 7, pp. 1087–1101, 2009.

[13] N. Debbabi, M. Kratz, M. Mboup, and S. El Asmi, "Combining algebraic approach with extreme value theory for spike detection," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, 2012, pp. 1836–1840.

[14] F. F. Digham, M. S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," *Communications, IEEE Transactions on*, vol. 55, no. 1, pp. 21–24, 2007.

[15] Y. Yang, D. Yu, and J. Cheng, "A roller bearing fault diagnosis method based on {EMD} energy entropy and {ANN}," *Journal of Sound and Vibration*, vol. 294, no. 12, pp. 269 – 277, 2006.

[16] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag, 1997.

[17] S. Tanaka, K. Takara, A. Snorrason, H. Finnsdottir, and E. M. Moss, "A study on threshold selection in pot analysis of extreme floods," *IAHS-AISH publication*, , no. 271, pp. 299–304, 2002, eng.

[18] B. M. Hill, "A simple general approach to inference about the tail of a distribution," *The Annals of Statistics*, vol. 3, pp. 1163–1174, 1975.

[19] M. Kratz and S. I. Resnick, "The qq-estimator and heavy tails," *Communications in Statistics. Stochastic Models*, vol. 12, no. 4, pp. 699–724, 1996.

[20] Y. Demichel, A. Estrade, M. Kratz, and G. Samorodnitsky, "How fast can the chord length distribution decay?," *Advances in Applied Probability*, vol. 43, pp. 504–523, 2011.

[21] K. Levenberg, "A method for the solution of certain nonlonear problems in least squares quart," *Applied Math*, vol. 2, pp. 164–168, 1944.

[22] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM Journal on Applied Mathematics*, vol. 11, pp. 431–441, 1963.

[23] J. R. M. Hosking and J. R. Wallis, "Parameter and quantile estimation for the generalized pareto distribution," *Technometrics*, vol. 29, no. 3, pp. 339–349, 1987.

[24] J. Zhang and M. A. Stephens, "A new and efficient estimation method for the generalized pareto distribution," *Technometrics*, vol. 51, no. 3, pp. 316–325, 2009.