# EVA 2011 - Abstracts of Contributed Talks

(updated May 15)

## Local Asymptotic Normality in $\delta$-Neighborhoods of Standard Generalized Pareto Processes

Stefan Aulbach – *University of Würzburg, Germany*
Email: stefan.aulbach@uni-wuerzburg.de
Co-author: Michael Falk

### Abstract

De Haan and Pereira [1] provided models for spatial extremes in the case of stationarity, which depend on just one parameter $\beta > 0$ measuring tail dependence, and they proposed different estimators for this parameter. This framework was supplemented in [2] by establishing local asymptotic normality (LAN) of a corresponding point process of exceedances above a high multivariate threshold, yielding in particular asymptotic efficient estimators.

The estimators investigated in these papers are based on a finite set of points $t_1, \ldots, t_d$, at which observations are taken. We generalize this approach in the context of functional extreme value theory (EVT). Contrary to multivariate EVT, functional EVT does not investigate finite dimensional random variables but stochastic processes. Actually it turns out that, in a proper setup, multivariate results can easily be deduced from the functional case.

The more general framework of functional EVT allows estimation over some spatial parameter space, i. e., the finite set of points $t_1, \ldots, t_d$ is replaced by $t \in [a, b]$. In particular, we derive efficient estimators of $\beta$ based on those processes in a sample of iid processes in $C[0, 1]$ which exceed a given threshold function.

### References

[1] L. de Haan and T. Pereira, *Spatial extremes: models for the stationary case*, Ann. Statist. **34**, 146-168, 2006.

[2] M. Falk, *Local asymptotic normality in a stationary model for spatial extremes*, J. Multivariate Anal. **102**, 48-60, 2011.

## Quantile estimation for a bivariate sample

Marie-Francoise Barme-Delcroix – *Lab. Painlevé, University Lille 1, France*
Email: marie-francoise.barme@math.univ-lille1.fr
Co-author: Margarida Brito

### Abstract

There is no natural ordering of a multidimensional space and the extension of the definition of univariate quantiles to multivariate distributions is, therefore, not straightforward. We use a definition based on the ordering of a multivariate sample according to an increasing family of curves, we have called isobars. For a given level u, an u-level isobar is defined as a level curve of the conditional distribution function of the radius given the angle (that is, as a conditional quantile ), the sample points being defined by their polar coodinates. In this way, the maximum value of the sample is defined as the point which belongs to the upper level isobar and it is really a sample point. Moreover, the so-defined order statistics may be characterized by an unidimensional approach ([1], [2]). For the applications the distribution function is, in general, unknown and we have to estimate the isobars.

For an unknown distribution and for a given level, here we investigate the behaviour of an histogram-type estimator of this conditional quantile based on a partition of the unit circle. Under additional regularity and mild Lipschitz conditions on the underlying distribution, we establish the uniform asymptotic normality and the uniform almost sure consistency of this estimator.

### References

[1] M. F. Barme-Delcroix and M. Brito, Multivariate stability and strong limiting behaviour of intermediate statistics, *J. Multivariate Anal.* 79, 157–170, 2001.

[2] M. F. Barme-Delcroix and U. Gather, Limit laws for multidimensional extremes, *Statistics and Probability Letters* 77, 1750–1755, 2007.

# Prediction of Catastrophes in Space over Time

Anastassia Baxevani – *University of Gothenburg, chalmers University of Technology, Sweden*
Email: anastass@chalmers.se
Co-authors: Richard Wilson and Manolo Scotto

## Abstract

An optimal alarm policy for detecting catastrophes in spatio-temporal settings plays a central role in the analysis of the occurrence and impact of future rare events. The key research objective is to develop optimal alarm systems based on excursions of an alarm field to predict whether another field will enter a catastrophic situation at a later time.

In this work an optimal alarm system in space over time is introduced and studied in detail. To this extend, we generalize the results obtained by [1] and ([2], [3]) for stationary stochastic processes evolving in continuous time to the present setting. Specifically, we say that a catastrophe is considered to commence at some time if the random field exceeds the location specific level in some spatial neighborhood of the location whatever the situation at the other locations of interest, i.e., catastrophes that start outside the region and work into it are ignored. In this framework we will consider the case that the locations of possible catastrophes are discrete, lie along a smooth bounded one dimensional curve through a two-dimensional space (think of locations along the coastline) and the case the locations lie in an bounded connected region in a two dimensional space.

This latter, is qualitatively more difficult than the setting covered previously as it involves conditioning on a fixed level upcrossing in time for the first time in a spatial region; i.e. it involves conditioning on the occurrence of a point at which the random field has a level upcrossing in time and a local maximum in space. We can do this relatively easy for stationary (both in space and time) Gaussian random fields but it is moderately straightforward for Gaussian random fields stationary in time but non-stationary in space. Obviously, the choice of the alarm would reflect information which gives a higher probability of a catastrophe and should be chosen according to some optimality criteria. The celebrated Rice formula will be used to obtain Palm distributions and the probabilities of interest will be calculated by obtaining Slepian model representations of the alarm and catastrophe random fields, both by conditioning on the event that a catastrophe commences at a specific time or that an alarm commences at a specific time.

A natural extension of the above setting is for the case the underlying field is no longer Gaussian but is substituted by a Laplace field. In this case, we need to develop Palm distributions and Slepian model representations for Laplace fields.

## References

[1] De Maré, J., (1980). Optimal prediction of catastrophes with applications to Gaussian processes. *Ann. Probab.*, **8**, pp. 841-850.

[2] Lindgren, G., (1980). Model processes in nonlinear prediction with application to detection and alarm. *Ann. Probab.*, **8**, pp. 775-792.

[3] Lindgren, G., (1985). Optimal Prediction of Level Crossings in Gaussian Processes and Sequences *Ann. Probab.*, **13**, Number 3, pp. 804-824.

# Light-Tail Estimation: a Negative Moment EVI-estimator and an Application to Environmental Data

Frederico Caeiro – *Universidade Nova de Lisboa, Portugal*
Email: fac@fct.unl.pt
Co-authors: M. Ivette Gomes and Lígia Henriques-Rodrigues

## Abstract

Whenever interested in extreme large events, the most common models in *statistics of extremes* are semi-parametric or even non-parametric in nature, with the imposition of a few "regularity conditions" in the right-tail, $\overline{F}(x) := 1-F(x)$, as $x \to +\infty$, of an unknown model $F$ underlying the available data. The primordial parameter is the *extreme value index* (EVI). For large values, the EVI is the shape parameter $\gamma$ in the distribution function (d.f.) $EV_\gamma(x) = \exp\left(-(1+\gamma x)^{-1/\gamma}\right)$, $1+\gamma x > 0$, the (unified) *extreme value* (EV) distribution. The EVI needs to be estimated in a precise way, because such an estimation is one of the basis for the estimation of other parameters of extreme and large events, like a *high quantile* of probability $1-p$, with $p$ small, the *right endpoint* of the model $F$ underlying the data, $x^F := \sup\{x : F(x) < 1\}$, whenever finite, and the *return period* of a high level, among others.

For the EVI-estimation, we first refer one estimator, valid for all $\gamma \in \mathbb{R}$, the *moment* (M) estimator (Dekkers *et al.*, 1989), with the functional form $\hat{\gamma}^M_{k,n} := M^{(1)}_{k,n} + \frac{1}{2}\left\{1 - \left(M^{(2)}_{k,n}/[M^{(1)}_{k,n}]^2 - 1\right)^{-1}\right\}$, with $M^{(j)}_{k,n} := \sum_{i=1}^k \left\{\ln X_{n-i+1:n} - \ln X_{n-k:n}\right\}^j / k$, $j \geq 1$. Note that the statistic $M^{(1)}_{k,n}$, is the well-known *Hill* estimator (Hill, 1975), often denoted $\hat{\gamma}^H_{k,n} \equiv M^{(1)}_{k,n} := \sum_{i=1}^k \left\{\ln X_{n-i+1:n} - \ln X_{n-k:n}\right\}/k$, the average of the log-excesses and valid only for $\gamma \geq 0$. We can thus write the *moment* estimator as

$$\hat{\gamma}^M_{k,n} = \hat{\gamma}^H_{k,n} + \hat{\gamma}^{NM}_{k,n}, \qquad \hat{\gamma}^{NM}_{k,n} := \frac{1}{2}\left\{1 - \left(M^{(2)}_{k,n}/[M^{(1)}_{k,n}]^2 - 1\right)^{-1}\right\}, \tag{1}$$

with $NM$ standing for *negative moment* estimator. Indeed, whereas the H-estimator, $\hat{\gamma}^{H}_{k,n}$, is consistent for the estimation of $\gamma_+ := \max(0,\gamma)$, the NM-estimator, in (1), is consistent for the estimation of $\gamma_- := \min(0,\gamma)$, provided that $k = k_n$ is an intermediate sequence, i.e., a sequence of integers such that $k = k_n \to \infty$ and $k_n = o(n)$, as $n \to \infty$. Under these same conditions, the M-estimator is consistent for the estimation of any real $\gamma$, which can be written as $\gamma = \gamma_+ + \gamma_-$. Most of the classical EVI-estimators have usually a high variance for small $k$ and a high bias when $k$ is large. This problem affects both the *moment* and the *Hill* estimator, and leads to a difficult choice of the "optimal" $k$, in the sense of the value $k$ that minimizes the asymptotic mean squared error (MSE).

On the basis of a comment in Fraga Alves (1998), where it is noticed that when $\gamma < 0$, $\hat{\gamma}^{NM}_{k,n}$ and $\hat{\gamma}^{M}_{k,n}$ have the same asymptotic variance, Caeiro and Gomes (2010) were led to the introduction of a semi-parametric class of consistent estimators for $\gamma < 0$, which generalizes both the M and the NM-estimators. Such a class, denoted NM($\theta$), is given by

$$\hat{\gamma}^{NM(\theta)}_{k,n} := \hat{\gamma}^{NM}_{k,n} + \theta\, M^{(1)}_{k,n}, \quad \theta \in \mathbb{R}. \tag{2}$$

Apart from the usual integer parameter $k$, related with the number of top order statistics involved in the estimation, these estimators depend on an extra real parameter $\theta$, which makes them highly flexible and possibly second-order unbiased for a large variety of models in $\mathcal{D}_{\mathcal{M}}(EV_\gamma)_{\gamma<0}$. In this paper, we are interested not only on the adaptive choice of the *tuning* parameters $k$ and $\theta$, but also on an application of these semi-parametric estimators to the analysis of sets of environmental and simulated data. Note that we get the *negative moment* estimator in (1) for $\theta = 0$ and the *moment estimator*, for $\theta = 1$. With the appropriate choice of $\theta$, $\hat{\gamma}^{NM(\theta)}_{k,n}(k)$ enables us to have access to an estimator of a negative EVI with a smaller asymptotic bias and the same asymptotic variance as the M-estimator. We shall further provide an *Algorithm* for the adaptive choice of the *tuning* parameters under play in the semi-parametric estimation of the EVI through such an estimator. Finally, we apply the *Algorithm* to the analysis of a set of environmental data, the daily average wind speeds in knots (one nautical mile per hour), collected in Dublin airport, in the period 1961-1978, as well as to a set of simulated data. Due to the seasonality of wind data, we restrict ourselves to the Spring and Summer months, and analyze the whole data-set, merely as a curiosity.

## References


[1] Caeiro, F., Gomes, M.I. (2010): An asymptotically unbiased moment estimator of a negative extreme value index. *Discussiones Mathematica: Probability and Statistics* **30**:1, 5–19.

[2] Dekkers, A.L.M., Einmahl, J.H.J., Haan, L. de, (1989). A moment estimator for the index of an extreme-value distribution. The Annals of Statistics **17**(4), 1833–1855.

[3] Fraga Alves, M.I. (2001). Weiss-Hill estimator. *Test* **10**, 203–224.

[4] Hill, B.M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *Ann. Statist.* **3**(5), 1163-1174.


# Bias Correction in Extreme Value Statistics with Extreme Value Index around Zero


Juan-Juan Cai – *Tilburg University, The Netherlands*
Email: j.cai@uvt.nl
Co-authors: Laurens de Haan and Chen Zhou


## Abstract


The major difficulty in applying extreme value statistics is to produce accurate estimates of the extreme value index. Most of existing estimators of the extreme value index exhibit convenient statistical properties such as the asymptotic normality under a speed of convergence $k^{-1/2}$, with $k$ denoting the number of high order statistics used in estimation. Although a higher choice of $k$ is desired for better estimation accuracy, it may leads to a higher asymptotic bias. Thus, employing a bias correction procedure allows a higher choice of $k$. Bias correction for estimators of the extreme value index has been studied extensively in literature, however, either for positive (see, e.g. [4]) or negative (see, e.g [5]) extreme value index. In other words, prior knowledge on the sign of the extreme value index is required to perform existing bias correction procedures.

In meteorology and environmental science, empirical literature has documented that most random variables exhibit extreme value indices around zero (see, e.g. [2] ). The probability weighted moment (PWM) estimator proposed by [3] is popular in such a situation (see, e.g. [1]). In this paper, we provide a bias correction procedure for the PWM estimator by estimating the asymptotic bias and subsequently subtracting the bias from the original estimator. We also produce a bias corrected quantile estimator connected to the PWM estimator, when the extreme value index is around zero. Lastly, we provide a bias corrected estimator of the endpoint of a distribution when the extreme value index is negative.

The main advantage of the bias corrected estimators explored in this paper is that one gets more flexibility in choosing the number of high order statistics used in estimation. Compared to the original estimators without bias correction, with choosing a higher choice of $k$ in the bias corrected estimators, the mean squared errors are subsequently of a lower order. This advantage becomes apparent in simulations and an environmental application.

## References

[1] T. A. Buishand, Extreme rainfall estimation by combining data from several sites. *Hydrological Sciences Journal*, 36(4):345–365, 1991.

[2] S. G. Coles and J. A. Tawn, Modelling extremes of the areal rainfall process. *Journal of the Royal Statistical Society. Series B(Methodological)*, 58(2):377–392, 1996.

[3] J. R. M. Hosking and J. R. Wallis , Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29(3):339–349, 1987.

[4] M. I. Gomes and D. Pestana, A sturdy reduced-bias extreme quantile (VaR) estimator. *Journal of the American Statistical Association*, 102(477):280-292, 2007.

[5] D. Li and L. Peng, Does bias reduction with external estimator of second order parameter work for endpoint? *Journal of Statistical Planning and Inference*, 139(6):1937–1952, 2009.


# Nonparametric estimation for multivariate extremes


Miguel de Carvalho – *Ecole Polytechnique Fédérale de Lausanne, Switzerland*
Email: miguel.carvalho@epfl.ch
Co-author: Anthony Davison



## Abstract

The Ramos–Ledford spectral model [2] has received considerable attention as an elegant tool for modelling multivariate extremes. A major advantage over existing models is its ability to unify in a single framework the cases of asymptotic dependence and asymptotic independence. Despite its attractive features, for purposes of estimation only a parametric asymmetric logistic model is available, thus restricting the range of its applications. In this paper we propose nonparametric estimation and inference procedures for the Ramos–Ledford spectral model by empirical likelihood techniques [1].



## References

[1] A. Owen, *Empirical Likelihood*, Boca Raton: Chapman and Hall, 2001.

[2] A. Ramos and A. Ledford, A new class of models for bivariate joint tails, *J. R. Statist. Soc. B*, 71, 219–241, 2009.


# Group theoretic dimension of stationary symmetric $\alpha$-stable random fields


Arijit Chakrabarty – *Indian Statistical Institute, New Delhi, India*
Email: arijit@isid.ac.in
Co-author: Parthanil Roy



## Abstract

The growth rate of the partial maximum of a stationary stable process was first studied in the works of [4, 5], where it was established, based on the seminal works of [1, 2], that the growth rate is connected to the ergodic theoretic properties of the flow that generates the process. The results were generalized to the case of stable random fields indexed by $\mathbb{Z}^d$ in [3], where properties of the group of nonsingular transformations generating the stable process were studied as an attempt to understand the growth rate of the partial maximum process. This work generalizes this connection between stable random fields and group theory to the continuous parameter case, that is, to the fields indexed by $\mathbb{R}^d$.



## References

[1] J. Rosiński (1995): On the structure of stationary stable processes. *The Annals of Probability* 23:1163–1187.

[2] J. Rosiński (2000): Decomposition of stationary $\alpha$–stable random fields. *Annals of Probability* 28:1797–1813.

[3] P. Roy and G. Samorodnitsky (2008): Stationary symmetric $\alpha$-stable discrete parameter random fields. *Journal of Theoretical Probability* 21:212–233.

[4] G. Samorodnitsky (2004a): Extreme value theory, ergodic theory, and the boundary between short memory and long memory for stationary stable processes. *Annals of Probability* 32:1438–1468.

[5] G. Samorodnitsky (2004b): Maxima of continuous time stationary stable processes. *Advances in Applied Probability* 36:805–823.


# Alarm systems and catastrophes from a different point of view


Pasquale Cirillo – *University of Bern, Switzerland*
Email: pasquale.cirillo@stat.unibe.ch
Co-author: Jürg Hüsler

## Abstract

Using a chain of urns, we build a Bayesian nonparametric alarm system to predict catastrophic events.
In observing a given phenomenon, it is often important to predict if the variable or quantity of interest will or will not reach a predetermined critical level sometime in the near future. This type of analysis becomes fundamental when the studied phenomenon may have a strong impact on human life, as in the case of floods, tides, epidemics or large power outages. In these rare and extreme situations, the critical level frequently corresponds to some sort of catastrophe, and the possibility of giving an alarm in due time is invaluable.
An alarm system is hence a tool that predicts the occurrence of a catastrophic event at a specified time in the future on the basis of the available information. To be more exact, an alarm systems is meant to activate when the probability of a catastrophe in a given time horizon overcomes a certain threshold of tolerance.
In the literature (e.g. [1]), an alarm system is said optimal when, for a set of available data, it possesses the highest probability of correct alarm. This means that when the alarm is given, the probability that the catastrophic event actually manifest itself is the highest among all possible systems.
A naive alarm system is represented by the predictor $\hat{X}_{t+h} = E[X_{t+h}|X_s, -\infty < s \leq t, h > 0]$, where an alarm is cast every time the predictor exceeds some risk level. Naturally this system is not optimal at all, because it does not show good performances in detecting exceedances, in correctly locating them in time and in reducing the number of false alarms.
Differently from other alarm systems in the literature, we propose a model that is constantly updated on the basis of the available information, according to the Bayesian paradigm. Thus one of our main assumptions is that catastrophic events over time may be assumed to be exchangeable in the sense of de Finetti, i.e. their joint distribution is immune to permutations.
Our idea is to use a special urn process that can only assume a finite number of values over time but that, thanks to Polya-like reinforcement mechanism, is able to learn from the past. Every value must be seen as a level of risk, on a scale that goes from 0, no risk, to L, catastrophe. We will assume time to be discrete, in order to have an intuitive description of the model. Nevertheless, we also show that it is possible to extend our alarm system to continuous time.
The use of a finite-valued process may seem reductive in many cases where the observed phenomenon assumes a continuous range of values. In reality, if $\{X_n\}$ is our process, it can be seen as the "simplified" (or discretized) version of another underlying process $\{Y_n\}$, characterized by a (much) larger space of states.
The construction we propose is a generalization of the model presented in [2], which belongs to the wider class of reinforced urn processes (RUP). RUP have been introduced in [3] as reinforced random walks on a state space of urns.
In the papers we discuss some interesting probabilistic properties of our model, showing how they can be used to predict catastrophic events. The construction we propose is rather flexible and it can be applied to very different problems.
In the last part of the work, we test our alarm system on a couple of interesting time series related to catastrophic events.

## References

[1] M. Antunes, M.A. Amaral-Turkman and F.K. Turkman, *A Bayesian approach to event prediction*, Journal of Time Series Analysis 24, 631-646, 2003.

[2] P. Cirillo and J. Hüsler, *Extreme shock models: an alternative perspective*, Statistics and Probability Letters 81, 25-30, 2010.

[3] P. Muliere, P. Secchi and S. Walker, *Urn schemes and reinforced random walks*, Stochastic Processes and their Applications 88, 59-78, 2000.

# Extremes values for continuous diffusions and its skeleton: the bounded case and FARCH approximations.

Didier Dacunha-Castelle – *Université Paris Sud, France*
Email: Didier.Dacunha-Castelle@math.u-psud.fr
Co-author: Thi Tu Huong Hoang

## Abstract

We address extreme theory for bounded ergodic diffusion with inaccessible boundary and non zero drift at the boundary.

We show that in this case the square volatility is linear at the boundary and it depends only of the shape coefficient of extremes if it exists a limit GEV theorem. We then give a detailed study of the copula of the skeleton imbedded Markov chain observes at equispaced times k$\Delta$ . This kind of copula has very specific properties. The basic tool is a representation of the transition density given by [1].This is a way to study different problems for the dependence on $\Delta$ of the extremes and on the link continuous /discrete time for this theory. The method can be applied to every ergodic stationary diffusion on R with of course more complicated behaviors of extremes.

We apply these results to temperature data. These results are used also to get important improvement for models of simulation of temperatures. Here the main tool are FARCH (functional autoregressive process with functional variance coefficient) approximations of the previous skeleton, FARCH processes being studied as misspecified ones. These processes are also geometrically ergodic, we give a proof using the approximated renewal atom technic.

## References

[1] D. Dacunha-Castelle, D. Florens-Zmirou. Estimation of the coefficients of a diffusion for discrete observations. Stochastics 19, 1986.

[2] R.A.Davis. Minimum and maximum of one dimensional diffusions. Stochastic processes and their applications 13, 1982.

[3] Robinson & J.A. Tawn. Extremal analysis of processes sampled at different frequencies. J. Roy. Statist. Soc. Ser B 62, 2000.


# A methodology for hyperbolic graph efficiency assessment via extremes


Abdelaati Daouia – *Toulouse School of Economics, France*
Email: daouia@cict.fr



## Abstract

We consider a sample of independent businesses in a sector of production activity where a vector of inputs $X$ is used to produce multiple outputs $Y$. An important question is : what are the most efficient businesses that might be useful to emulate? Using ideas from extreme value theory in conjunction with the concept of hyperbolic graph efficiency measure, we have been able to come up with a reasonable answer. From a statistical viewpoint, the set of all possible producers may be viewed as the joint support of the population of businesses $(X, Y)$. The identification of the top sample observations lying near its optimal boundary is clearly a problem belonging to statistical modeling of extremes.

The hyperbolic distance function measures the potential input decrease and output increase along a hyperbolic path to achieve the efficient support boundary. Its free disposal hull estimator and two alternative extreme-value based estimators are discussed. Practical guidelines to effect the necessary computations of these estimators are described and generated data sets illustrate how they work out in practice. The Dekkers-Einmahl-de Haan [1] type of estimator appears to give frank improvement in fit to data and performs better than the free disposal hull estimator [2] in terms of both bias and mean-squared error even in presence of outliers. A case study is provided on frontier and efficiency analysis of the delivery activity of French postal services.



## References

[1] A.L.M. Dekkers, J.H.J. Einmahl and L. de Haan, *A moment estimator for the index of an extreme-value distribution*, *The Annals of Statistics*, 17(4), 1833–1855, 1989.

[2] D.C. Wheelock and P.W. Wilson, *Non-parametric, Unconditional Quantile Estimation for Efficiency Analysis with an Application to Federal Reserve Check Processing Operations*, *Journal of Econometrics*, 145 (1-2), 209–225, 2008.


# Exploratory Plots in Extreme Value Analysis


Bikramjit Das – *ETH Zürich, Switzerland*
Email: bikram@math.ethz.ch
Co-author: Souvik Ghosh



## Abstract

Exploratory plotting tools are widely used to diagnose the goodness-of-fit of data sets to a hypothesized distribution. Some of them have found extensive use in diverse areas of finance, telecommunication, environmental science, etc. in order to detect sub-exponential or heavy-tailed behavior in observed data. Here we look at two such methodologies: the Quantile-Quantile plots for heavy-tails and the Mean Excess plots. Under the assumption of heavy-tailed behavior of the underlying sample the convergence in probability of these plots to a fixed set in a suitable topology of closed sets of $\mathbb{R}^2$ has been studied in [1] and [2]. These results give theoretical justifications for using the plots to test the null hypothesis that the underlying distribution is heavy-tailed by checking if the observed plot is "close" to the limit under the null hypothesis. In practice though one set of observations would lead to only one plot of each kind. This is often not good enough to verify proximity of the plot to the fixed set of interest under the null hypothesis of heavy-tails. Here we provide weak limits for these two plots when the underlying distribution of the sample is heavy-tailed and as an application we are able to construct confidence bounds around the plots which enables us to check whether the underlying distribution is heavy-tailed or not.



## References

[1] B. Das and S.I. Resnick. QQ plots, random sets and data from a heavy tailed distribution. *Stochastic Models*, 24(1):103–132, 2008.

[2] S. Ghosh and S.I. Resnick. A discussion on mean excess plots. *Stochastic Processes and their Applications*, (120):1492–1517, 2010.

# A new semi-parametric family of estimators for the second order parameter

El-hadji Deme – *Université Gaston Berger de Saint-Louis, Sénégal*
Email: eltinoyaw@yahoo.fr
Co-authors: Laurent Gardes and Stéphane Girard

## Abstract

An important parameter in extreme value theory is the extreme value index $\gamma$. It controls the first order behavior of the distribution tail. In the literature, numerous estimators of this parameter have been proposed especially in the case of heavy tailed distributions (which is the situation considered here). The most known estimator was proposed by [4]. It depends on the $k$ largest observations of the underlying sample. The bias of the tail index estimator is controlled by the second order parameter $\rho$. In order to reduce the bias of $\gamma$'s estimators or to select the best number $k$ of observations to use, the knowledge of $\rho$ is essential. Some estimators of $\rho$ can be found in the literature, see for example [1, 2, 3]. We propose a semiparametric family of estimators for $\rho$ that encompasses the three previously mentioned estimators. The asymptotic normality of these estimators is then proved in an unified way. New estimators of $\rho$ are also introduced.

## References

[1] G. Ciuperca and C. Mercadier. Semi-parametric estimation for heavy tailed distributions. *Extremes*, **13**, 55–87, 2010.

[2] M.I. Fraga Alves, M.I. Gomes, and L. de Haan. A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica*, **60**(2), 193–213, 2003.

[3] Y. Goegebeur, J. Beirlant, and T. de Wet. Kernel estimators for the second order parameter in extreme value statistics. *Journal of Statistical Planning and Inference*, **140**, 2632–2652, 2010.

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

# Robust estimators of extreme value index and linear models

Jan Dienstbier – *Technical University of Liberec, Czech Republic*
Email: jan.dienstbier@tul.cz

## Abstract

In the contribution we deal with estimation of extremal properties of heavy-tailed linear models. We use quantile sensitive linear regression estimators such as regression quantiles. The idea is to plug their residuals or intercepts into the robust tail index estimators proposed in [1]. Derived estimators of extreme value index are consistent and asymptotically normal which can be proven using the theory of smooth functionals of tail quantile function [2] and an improved version of the law of iterated logarithm for regression quantiles similar to the one derived in [3]. We show that described location and scale invariant estimators work for simulated datasets with a reasonable reliability. A practical adaptation of the method is also demonstrated on Condroz dataset of calcium levels in Belgium Condroz region.

## References

[1] J. Beirlant, J.L. Teugels, and P. Vynckier, *Practical Analysis of Extreme Values*, Leuven University Press, Leuven, 1996.

[2] H. Drees, On Smooth Statistical Tail Functionals, *Scandinavian journal of Statistics* 25, 1998, 187–210.

[3] J. Jurečková, Regression rank scores testes against heavy-tailed alternatives, *Bernoulli* 5, 1999, 659-676.

# Statistical Analysis of Extremal Dependence: Bootstrapping the Extremogram

Holger Drees – *University of Hamburg, Germany*
Email: holger.drees@uni-hamburg.de

## Abstract

The complexity of the serial dependence between extreme values of a time series cannot be captured by a single number like the extremal index. Recently Davis and Mikosch [1] introduced the extremogram which is given by the limiting conditional probability that an observation lagged by a given number of time points belongs to some extreme set $tB$ given the present observation belongs to another set $tA$ as $t \to \infty$. They examined the asymptotics of an empirical version of the extremogram for fixed sets $A$ and $B$, but the *joint behavior* of these estimators over an infinite family of sets is needed to get a full picture of the extremal serial dependence.

In the talk we will present the asymptotics of a process of empirical extremograms indexed by sets. As the limiting Gaussian process usually has a quite complex covariance structure, we indicate how to use a multiplier bootstrap to construct uniform confidence bands. The main technical tools are limit theorems for a general class of empirical cluster functionals established by Drees and Rootzén [2] and a recent result which proves consistency of bootstrap versions of these processes.

### References


[1] Richard A. Davis and Thomas Mikosch, *The extremogram: A correlogram for extreme events*, Bernoulli 15 (4), 977–1009, 2009.

[2] H. Drees and H. Rootzén, *Limit Theorems for Empirical Processes of Cluster Functionals*, Annals of Statistics 38 (4), 2145–2186, 2010.


## Evaluating extreme snow avalanches in long term forecasting


Nicolas Eckert – *Cemagref UR ETNA, France*
Email: nicolas.eckert@cemagref.fr
Co-author: Ophelie Guin


### Abstract


Mountain communities are exposed to snow avalanche risk in winter. Hence, evaluating extreme avalanches is a crucial question for hazard zoning and the design of defense structures, i.e. long term forecasting. Different damageable quantities (traveled distance, impact pressure, flow depth, deposit volumes) have to be considered, but multivariate extreme value theory is for now not used is this field. Even the use of univariate EVT remains difficult because the most critical variable, the traveled distance, strongly depends on topography [1]. The recently developed alternative is the combination of a mechanical model for flow propagation with a stochastic model describing the variability of the different inputs/outputs, leading to a multivariate Peak Over Threshold (POT) model where the correlation between the different magnitude variables is constrained by physical rules [2]. Crucial technical problems are model identifiability and finding a reasonable compromise between precision of the description of the flow and computation times. These points can be addressed with a depth-averaged set of equation describing the propagation of snow avalanches within a hierarchical Bayesian framework [3]. First, the joint posterior distribution of model unknowns is estimated using a sequential Metropolis-Hastings algorithm. Second, the point estimates are used to predict the joint distribution of different variables of interest for hazard mapping. With regards to a purely statistical approach, this has the advantage of introducing reliable physics into the modeling. On the other hand, consistency with EVT in terms of attraction domain and asymptotic dependence of the different variables is not granted and remains an open space for further research. The different steps of the method are illustrated and discussed with real case studies from the French Alps.


### References


[1] D. McClung and K. Lied, *Statistical and geometrical definition of snow-avalanche runout*, Cold Regions Science and Technology, 13:107-119, 1987.

[2] M. Barbolini and C. J. Keylock, *A new method for avalanche hazard mapping using a combination of statistical and deterministic models*, Natural Hazards and Earth System Sciences, 2:239-245, 2002.

[3] N. Eckert, M. Naaim and E. Parent, *Long-term avalanche hazard assessment with a Bayesian depth-averaged propagation model*, Journal of Glaciology, 56:563-586, 2010.


## A Novel Approach to Brown-Resnick Processes and Mixed Moving Maxima Processes


Sebastian Engelke – *Georg-August-Universitaet Goettingen, Germany*
Email: sengelk@uni-goettingen.de
Co-authors: Martin Schlather and Zakhar Kabluchko


### Abstract


Brown and Resnick [1] introduced a max-stable process that is obtained as a limit of maxima of independent Ornstein-Uhlenbeck processes. This so-called Brown-Resnick process and its generalizations are in fact essentially the only limits which arise as maxima of independent Gaussian processes [5]. Another representation of the Brown-Resnick process is given by the maximum of Brownian motions $(B_t)_{t\in\mathbb{R}}$ drifting to negative infinity for $t \to \pm\infty$. Stoev [4] showed that substituting the Brownian motions in this construction for suitable Lévy processes still yields a stationary, max-stable process. This results, however, only holds for one-sided processes on the positive axis.
To obtain a two-sided stationary process we cannot simply reflect the Lévy process at the $y$-axis as in the Brownian case. In fact, we show that a certain transformation of the Lévy process is necessary to ensure stationarity on the whole real line.

Moreover, we analyze properties of these processes and show that they have a mixed moving maxima representation under certain conditions. For the special case when the Lévy process is a Brownian motion, i.e. the original Brown-Resnick process, we explicitly derive the distribution of the processes used in its mixed moving maxima representation. It turns out that it equals the distribution of the Brownian motion conditioned to stay negative. This fact naturally entails a fast and efficient simulation method of the Brown-Resnick process [2].
Furthermore, we replace the Lévy-process by a d-dimensional Lévy-Mori field [3] and obtain a new class of max-stable, stationary random fields in $d$ dimensions.

## References


[1] B. M. Brown and S. I. Resnick, Extreme values of independent stochastic processes, *J. Appl. Probability*, 14:139-156, 1977.

[2] S. Engelke, Z. Kabluchko and M. Schlather, An equivalent representation of the Brown-Resnick process, *Stat. Probab. Lett.*, In Press, 2011.

[3] T. Mori, Representation of linearly additive random fields, *Probab. Theory Related Fields*, 92(1):91–115, 1992.

[4] S. A. Stoev, On the ergodicity and mixing of max-stable processes, *Stochastic Process. Appl.*, 118(9):1679–1705, 2008.

[5] Z. Kabluchko, M. Schlather and L. de Haan, Stationary max-stable fields associated to negative definite functions, *Ann. Probab.*, 37:2042–2065, 2009.


# Sojourn Times and the Fragility Index


Michael Falk – *University of Würzburg, Germany*
Email: falk@mathematik.uni-wuerzburg.de
Co-author: Martin Hofmann


## Abstract


We investigate the sojourn time above a high threshold of a continuous stochastic process $Y = (Y_t)_{t\in[0,1]}$ on $[0,1]$ with continuous identical margins. It turns out that the limit, as the threshold increases, of the expected sojourn time given that it is positive, exists if the copula process corresponding to $Y$ is in the functional domain of attraction of an extreme value process $\eta$ as defined in [1]. This limit coincides with the limit of the fragility index ([2]) corresponding to $(Y_{i/n})_{1\le i\le n}$ as $n$ and the threshold increase. It turns out that this limit is the reciprocal of the mass of the (finite) spectral measure corresponding to $\eta$ ([3]) or, equivalently, of the functional $D$-norm of the constant function one ([1]).

If the process is in a certain neighborhood of a generalized Pareto process, then we can replace the constant threshold by a general threshold function and we can compute the asymptotic sojourn time distribution. An extreme value process is a prominent example.


## References


[1] S. Aulbach, M. Falk and M. Hofmann, *On extreme value processes and the functional D-norm.* Technical report, University of Würzburg, 2011.

[2] J. L. Geluk, L. de Haan and C. G. de Vries, *Weak and strong financial fragility.* Tinbergen Institute Discussion Paper, TI 2007-023/2, 2007

[3] L. de Haan and A. Ferreira, *Extreme Value Theory. (An Introduction).* Springer Series in Operations Research and Financial Engineering, Springer, New York, 2006.


# Poisson process approximation for point processes of exceedances using Stein's method


Anne Feidt – *University of Zurich, Switzerland*
Email: anne.feidt@math.uzh.ch


## Abstract


A fundamental result of extreme value theory is that a point process of exceedances can be shown to asymptotically behave like a Poisson process. Stein's method for point process approximation, developed by Barbour and Brown in [1], additionally provides bounds on the errors involved in approximating a point process by a Poisson process with the same mean measure. These bounds depend on the sample size $n$ and thus give a more precise idea on the accuracy of the approximation for finite sample sizes.

With a generalization to higher dimensions in mind, we study the behaviour of the point process of occurrences of iid bivariate random variables in sets in which they can be considered extreme. We aim to use Stein's method to approximate by a Poisson process with an intensity measure that reflects the copula structure of the distribution of the random variables.

We show this on the example of a point process of simultaneous exceedances over some thresholds of both components of iid bivariate geometric random pairs whose survival copula belongs to the Marshall-Olkin family of copulas.

More precisely, we first determine the accuracy in total variation of the approximation by a Poisson process having the same intensity as the original point process, defined on the lattice $\mathbb{Z}_+^2$. For more flexibility, we would prefer a Poisson process with a continuous intensity defined on $\mathbb{R}_+^2$. To approximate further by such a process, the total variation metric proves too strong and we need to use a weaker Wasserstein metric to obtain sharp error bounds as well as rely on the distributional parameters to vary with the sample size at an appropriate rate. The latter condition however implies that the continuous intensity of the approximating Poisson process also depends on $n$. By imposing conditions on the parameters we are able to show that the approximation by a further Poisson process with a continuous intensity no longer varying with $n$ will not add errors of a bigger order to the final error bound.

## References


[1] A. D. Barbour and T. C. Brown, Stein's method and point process approximation, *Stochastic Processes and their Applications*, 43(1): 9-31, 1992.


# The Generalized Pareto Process and its Domain of Attraction


Ana Ferreira – *ISA, Universidade Técnica de Lisboa and CEAUL, Portugal*
Email: anafh@isa.utl.pt
Co-author: Laurens de Haan


## Abstract


The Generalized Pareto distribution has played a fundamental role in Extreme Value Theory ([1], [2] and e.g. [3] for possible extension to multivariate theory). We present a generalization of this concept for extremes in function space. We call this generalization the 'Generalized Pareto Process', a concept to formalize in agreement with some desirable properties. We present ways to construct it and characterize its domain of attraction.


## References


[1] A.A. Balkema and L. de Haan: Residual life time at great age. Ann. Probab. 2, 792–804 (1974)

[2] J. Pickands III: Statistical inference using extreme order statistics. Ann. Statist. 3, 119–131 (1975)

[3] H. Rootzén and N. Tajvidi: Multivariate generalized Pareto distributions. Bernoulli 12, 917–930 (2006)


# Investigating the interplay of excess kurtosis and tail events in financial series


Anna M. Fiori – *University of Milano-Bicocca, Italy*
Email: anna.fiori@unimib.it


## Abstract


The well established *fat tail issue* of financial returns is routinely supported by high values of the conventional kurtosis statistic [2]. Based on fourth powers of the observations, the measure is however unable to discriminate between heavy tailed distributions as d.g.p. or simply the presence of a few extremes in the sample ([3]).

For this reason, we suggest alternative measures of right and left kurtosis which are less sensitive to outlying observations and are consistent with common risk perceptions of investors and risk managers. Based on a recent interpretation of kurtosis as inequality at either side of the median ([4]), the new measures *R(D)* and *R(S)* presume a partial ordering of distributions induced by nested *kurtosis curves* and are easily computed by *Gini indices* in both symmetric and asymmetric contexts ([1]).

Using the theory of L-statistics, we construct consistent estimators of the new measures in the i.i.d. case and prove their asymptotic normality under minimal assumptions on the underlying distribution. While the sampling variance of the conventional kurtosis coefficient is related to the population moment of order eight, the new measures can be appropriately estimated under the milder requirement that second moments are finite. In addition to an explicit formula for the asymptotic variance, we estimate the latter by nonparametric and bootstrap techniques. Extensive Monte Carlo simulations are designed to compare the traditional kurtosis measure with its right and left counterparts in finite samples.

For purposes of financial applications, the interplay of kurtosis and clustering in the volatility dynamics needs to be considered. Consequently, we first fit ARMA-GARCH models to several series of daily stock market returns by QML techniques. The kurtosis curves and related measures *R(D)* and *R(S)* are then estimated on standardized residuals. While the stylized facts about excess kurtosis in financial series may have been accepted too readily, the new approach provides a deeper insight into the interplay of peakedness, tail events and right/left excess kurtosis.

## References

[1] A. M. Fiori, Measuring kurtosis by right and left inequality orders, *Communications in Statistics - Theory and Methods*, 37(17), 2665-2680, 2008.

[2] A. M. Fiori and M. Zenga, Karl Pearson and the origin of kurtosis, *International Statistical Review*, 77(1), 40-50, 2009.

[3] T. H. Kim and H. White, On more robust estimation of skewness and kurtosis, *Finance Research Letters*, 1(1), 56-73, 2004.

[4] M. Zenga, Kurtosis, in: Kotz, S., Read, C.B., Balakrishnan, N. and Vidakovic, B. (Eds.), *Encyclopedia of Statistical Sciences*, 2nd ed., New York: Wiley, 2006.


# On double exceedances of stationary Gaussian processes


Jasmin Furrer – *University of Bern, Switzerland*
Email: jasmin.furrer@stat.unibe.ch
Co-author: Jürg Hüsler



## Abstract

We investigate the clustering of double exceedances of stationary Gaussian processes $X(t),\ t \in [0,T]$. Exceedances over identical levels have been analyzed thoroughly in [1]. Based on these results, we consider the case with two different thresholds. Applications of such a setup will be particularly interesting in climate sciences, e.g. earth quakes which are of different strength. Exceedances over different thresholds which are separated by at least a time lag $\varepsilon$ are considered. We assume that the correlation function has the following properties:

$$r(t) = 1 - |t|^\alpha + o(|t|^\alpha) \quad \text{as} \quad t \to 0,$$
$$r(t) < 1 \quad \forall\, t > 0,$$

with $\alpha \in (0,2)$. Additionally, we assume that in the interval $[\varepsilon, T]$ there exists only one point of maximal correlation $t_m$, being an interior point of the interval, and that $r$ is twice continuously differentiable in a neighborhood of $t_m$. Furthermore, the relationship between the two thresholds $u$ and $v$ is the following: $v = c \cdot u$ with $c \in (r(t_m), 1)$. With these assumptions, the following probability of a cluster of $\varepsilon$-separated exceedances can be derived exactly where the levels to be exceeded by the extreme events tend to $\infty$:

$$P(u, c; \varepsilon, T) = P(\exists (s,t) : s,t \in [0,T], |t-s| \geq \varepsilon, X(s) > v, X(t) > u).$$

The shape of the excursion sets depends on the behavior of the conditional mean $E(X(t) \,|\, X(0) = u,\ X(t_m) = v)$, a function defined by $u, c$ and the correlation function $r$. We will discuss some examples to indicate the pattern of exceedances depending on the given correlation function.



## References

[1] Hüsler, J., Ladneva, A. and Piterbarg, V. I. (2010) On clusters of high extremes of Gaussian stationary processes with $\varepsilon$-separation. *Electron. J. Probab.* **15** (59), 1825-1862.


# On the estimation of spatial max-stable models using threshold exceedances


Carlo Gaetan – *Dipartimento di Scienze Ambientali, Informatica e Statistica, Università Ca' Foscari, Venezia, Italy*
Email: gaetan@unive.it
Co-author: Jean-Noel Bacro



## Abstract

In the last decade spatial extreme problems have received more and more attention [4, 2, 1, 7]. Max-stable processes [3] are processes which naturally extend the multivariate extreme value properties to spatial data [10, 9, 5]. A natural and effective framework for the parametric inference for these models could be the likelihood, but the joint distribution of the observation is generally unavailable. Recently, composite likelihood methods have been proposed in order to circumvent this problem [7]. In that paper the spatial max-stable model is estimated using a composition of bivariate distributions of block maxima over a given period of time.

However it is well known that modelling only block maxima wastes a lot of information if other data on extremes are available. In this regard a statistical inference based on exceedances over a high threshold could be more effective.

In the talk we introduce composite likelihoods (CLs) based on two distinct models of the bivariate threshold exceedances. The first model [6] is an approximation of the bivariate distribution over the region where both exceedances are positive. The second model [8] uses the conditional distribution of the exceedances given that at least one is positive.

We compare the relative merits of both CLs through a simulation study where strong, moderate and weak extremal spatial dependencies are taken into accounts.

## References

[1] L. Bel, J.N. Bacro and C. Lantuéjoul, Assessing extremal dependence of environmental spatial fields, *Environmetrics*, 19, 163-182, 2008.

[2] T. A. Buishand, L. de Haan and C. Zhou, On spatial extremes: with application to a rainfall problem, *The Annals of Applied Statistics*, 2, 624-642, 2008.

[3] L. de Haan, A spectral representation for max-stable processes, *The Annals of Probability*, 12, 1194-1204, 1984.

[4] L. de Haan and T. Pereira, Spatial extremes: models for the stationary case, *The Annals of Statistics*, 34, 146–168, 2006

[5] Z. Kabluchko, M. Schlather and L. de Haan, Stationary max-stable fields associated to negative definite functions, *The Annals of Probability*, 37, 2042-2065, 2009.

[6] A.W. Ledford and J. Tawn, Statistics for near independence in multivariate extreme values, *Biometrika*, 83, 169-187, 1996.

[7] S. A. Padoan, M. Ribatet and S.A. Sisson, Likelihood-based inference for max-stable processes, *Journal of the American Statistical Association*, 105, 263-277, 2010.

[8] H. Rootzén and N. Tajvidi, Multivariate generalized Pareto distributions, *Bernoulli*, 12, 917-930, 2006.

[9] M. Schlather, Models for stationary max-stable random fields, *Extremes*, 5, 33-44, 2002.

[10] R. L. Smith, Max-stable processes and spatial extremes, *Technical Report*, University of North Carolina, 1990.


# Asymptotic comparison at optimal levels of reduced-bias EVI-estimators


M. Ivette Gomes – *University of Lisbon, Portugal*
Email: ivette.gomes@fc.ul.pt
Co-author: Frederico Caeiro



## Abstract

We are interested in the asymptotic comparison, at optimal levels, of a set of semi-parametric reduced-bias extreme value index (EVI) estimators, valid for a wide class of heavy-tailed models, underlying the available data. For heavy-tailed models in the domain attraction for maxima of an extreme value distribution $EV_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma}),\ x > -1/\gamma,\ \gamma > 0$, with $\gamma$ the EVI, the classical EVI-estimators are the Hill estimators (Hill, 1975), which are the average of the scaled log-spacings as well as of the log-excesses, given by $U_i := i\,\{\ln X_{n-i+1:n} - \ln X_{n-i:n}\}$ and $V_{ik} := \ln X_{n-i+1:n} - \ln X_{n-k:n}$, $1 \le i \le k < n$, respectively. But the Hill EVI-estimators have often a strong asymptotic bias for moderate up to large values of $k$. Consequently, the adequate accommodation of the bias of Hill's estimators has been extensively addressed in recent years by several authors. We mention the pioneering papers by Beirlant *et al.* (1999) and Feuerverger and Hall (1999), among others. In all these papers, authors are led to second-order reduced-bias EVI-estimators, with asymptotic variances larger than or equal to $(\gamma\ (1-\rho)/\rho)^2$, the minimal asymptotic variance of an "asymptotically unbiased" estimator in Drees' class of functionals, where $\rho < 0$ is a "shape" second-order parameter ruling the rate of convergence of the normalized sequence of maximum values towards the limiting *extreme value* random variable (r.v.). Recently, Caeiro *et al.* (2005) and Gomes *et al.* (2008), among others, considered, in different ways, the problem of corrected-bias EVI-estimation, being able to reduce the bias without increasing the asymptotic variance, which was shown to be kept at the value $\gamma^2$, the asymptotic variance of Hill's estimator, the maximum likelihood estimator of $\gamma$ for an underlying Pareto distribution. Those estimators, called *minimum-variance reduced-bias* (MVRB) EVI-estimators, are all based on an adequate external estimation of a pair of second-order parameters, $(\beta, \rho) \in (\mathbb{R}, \mathbb{R}^-)$, done through estimators denoted $(\hat\beta, \hat\rho)$, and outperform the classical estimators for all $k$. We shall here consider a set of MVRB statistics, denoted generally $UH_{\hat\beta,\hat\rho}(k)$, with $UH$ standing for *unbiased Hill*. We shall also consider, similarly to what has been done in Beirlant *et al.* (1999), among others, the statistics

$$UH^*_{\hat\rho}(k) := UH_{\hat\beta(k;\hat\rho),\hat\rho}(k), \tag{3}$$

with an asymptotic variance that is no longer $\gamma^2$ but $\gamma^2(1-\rho)^2/\rho^2 > \gamma^2$ for every $\rho$. As mentioned before, in the $UH_{\hat\beta,\hat\rho}(k)$ and $UH^*_{\hat\rho}(k)$ classes, $(\hat\beta, \hat\rho)$ or $\hat\rho$, respectively, need to be adequate consistent estimators of the second-order parameters, if we want to keep the same properties of $UH_{\beta,\rho}(k)$ or $UH^*_\rho(k)$, the associated r.v.'s. A third possibility is to estimate both the scale and shape second-order parameters at the same leke $k$, used for the estimation of $\gamma$, like in Feuerverger and Hall (1999), also among others. We could then work with statistics of the type,

$$UH^{**}(k) := UH_{\hat\beta(k;\hat\rho(k)),\hat\rho(k)}(k), \tag{4}$$

with an asymptotic variance surely larger than $\gamma^2(1-\rho)^2/\rho^2$, for every $\rho$. We shall here consider the estimator in Feuerverger and Hall (1999), denoted $FH^{**}(k)$, as an illustration of estimators of the type of the ones in (4). We shall compare asymptotically, at optimal levels, the above mentioned set of MVRB statistics, denoted generically $UH(k)$, the reduced-bias statistics $UH^*(k)$, in (3) (assuming thus that $\beta$ and $\rho$ are known or adequately estimated) and the $FH^{**}(k)$-statistics, as illustrative of reduced-bias EVI-estimators associated with an "internal" estimation of second-order parameters, like the ones in (4). Again, as in the classical case, there is not any estimator that can always dominate the alternatives, but interesting clear-cut patterns are found. Consequently, and in practice, a suitable choice of a set of EVI-estimators will jointly enable us to better estimate the EVI, the primary parameter of extreme events.

## References

[1] Beirlant, J., Dierckx, G., Goegebeur, Y. and Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes* **2**:2, 177-200.

[2] Caeiro, F. Gomes, M.I. and Pestana, D.D. (2005). Direct reduction of bias of the classical Hill estimator. *Revstat* **3**:2, 111-136.

[3] Feuerverger, A. and Hall, P. (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Annals Statistics* **27**, 760-781.

[4] Gomes, M.I., de Haan, L. and Henriques-Rodrigues, L. (2008). Tail Index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *J. Royal Statistical Society* **B70**:1, 31-52.

[5] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Annals Statistics* **3**, 1163-1174.


# Bias-reduced estimators for bivariate tail modelling


Armelle Guillou – *Strasbourg University and CNRS, France*
Email: armelle.guillou@math.unistra.fr
Co-authors: Jan Beirlant and Goedele Dierckx



## Abstract

Ledford and Tawn [3] introduced a flexible bivariate tail model based on the coefficient of tail dependence and on the dependence of the extreme values of the random variables. In this paper, we extend the concept by specifying the slowly varying part of the model as done by Hall [2] with the univariate case. Based on Beirlant et al. [1], we propose a bias-reduced estimator for the coefficient of tail dependence and for the estimation of small tail probabilities. We discuss the properties of these estimators via simulations and a real-life example. Furthermore, we discuss some theoretical asymptotic aspects of this approach.



## References

[1] Beirlant, J., Joossens, E. and Segers, J., 2009. Second-order refined peaks-over-threshold modelling for heavy-tailed distribution. *Journal of Statistical Planning and Inference* 139, 2800-2815.

[2] Hall, P., 1982. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society: Series B* 44, 37-42.

[3] Ledford, A.W. and Tawn, J.A., 1997. Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B* 59, 475-499.


# Lifting wind storms


Laurens de Haan – *Erasmus University, Netherlands*
Email: ldhaan@ese.eur.nl



## Abstract

In view of establishing a novel method for determining hydraulic boundary conditions Deltares is planning to use wind data not only from a single source but over a wide area as input. These (artificial) wind field data should be used to gain insight on how a severe storm behaves (over time and space), one that is so severe that it is not expected to show up in the data. The idea is that extreme value theory can help to perform this task.

The above is a simplification of the problem as stated in Deltares report 1202120-001-HYE-0002 but it may capture the essence of the problem.

A method is proposed to produce these extreme storms not by estimating the structure of the limiting max-stable process but by "lifting" less extreme storms.


# Products in conditional extreme value model


Rajat S. Hazra – *Indian Statistical Institute, Kolkata, India*
Email:rajatmaths@gmail.com
Co-author: Krishanu Maulik

## Abstract

Classical multivariate extreme value theory tries to capture the extremal dependence between the components under a multivariate domain of attraction condition and it requires each of the components to be in domain of attraction of a univariate extreme value distribution as well. Multivariate extreme value (MEV) model has a rich theory but has some limitations as it fails to capture the dependence structure in presence of asymptotic independence. A different approach to MEV was given by [1], where they examined MEV distributions by conditioning on one of the components to be extreme. Here we assume one of the components to be in Fréchet or Weibull domain of attraction and study the behavior of the product of the components under this conditional extreme value model.

## References

[1] Janet E. Heffernan and Jonathan A. Tawn. A conditional approach for multivariate extreme values. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66(3):497–546, 2004. ISSN 1369-7412. With discussions and reply by the authors.

# $D$-Norm of Extreme Value Processes and Functional Domain of Attraction

Martin Hofmann – *University of Würzburg, Germany*
Email: hofmann.martin@mathematik.uni-wuerzburg.de
Co-authors: Stefan Aulbach and Michael Falk

## Abstract

We introduce the distribution function $G$ of a continuous max-stable process on $[0,1]$ (cf. [2]) and it is shown that $G$ can be represented via a norm on functional space, called $D$-norm. This is in complete accordance with the multivariate case and leads to a functional domain of attraction approach for stochastic processes, which is more general than the usual one based on weak convergence, cf. [3], [4].

This framework leads to a natural definition of functional generalized Pareto distributions (GPD) $W$ in $C[0,1]$, which satisfy $W = 1 + \log(G)$ in their upper tails. Moreover, we characterize the functional domain of attraction condition for copula processes in terms of tail equivalence with a functional GPD and introduce $\delta$-neighborhoods of a functional GPD. It is shown that these are characterized by a polynomial rate of convergence of functional extremes, which all is in great analogy with the well-known finite-dimensional case (cf. [1]).

## References

[1] M. Falk, J. Hüsler and R.-D. Reiss, *Laws of Small Numbers: Extremes and Rare Events.* 3rd ed., Birkhäuser, Basel, 2010.

[2] E. Giné, M. Hahn, and P. Vatan, *Max-infinitely divisible and max-stable sample continuous processes.* Probab. Theory Related Fields 87, 139-165, 1990.

[3] L. de Haan and A. Ferreira, *Extreme Value Theory. (An Introduction).* Springer Series in Operations Research and Financial Engineering, Springer, New York, 2006.

[4] L. de Haan and T. Lin, *On convergence toward an extreme value distribution in $C[0,1]$,* Ann. Probab. **29**, 467-483, 2001

# An Efficient Estimation Method for Risk Models

Mei Ling Huang – *Dept. of Mathematics, Brock University, St. Catharines. Ontario, Canada L2S 3A1*
Email: mhuang@brocku.ca
Co-author: Percy Brill

## Abstract

The extreme value distributions are Pareto type distributions which have many applications in risk analysis, survival analysis and queueing networks ([1]). There is difficulty in the inference of the Pareto distribution which has infinite moments in the heavy tailed situation. This paper studies the truncated Pareto distribution to avoid this difficulty. We also propose an optimal weighted estimation method for the truncated Pareto distribution. For the new estimator, an exact efficiency function relative to the classical estimators, and its properties, are derived. L1-optimal weights and L2 -optimal weights are also proposed. Monte Carlo simulation results confirm the theoretical conclusions. The new estimation method has been applied to analyze hurricane and forest fire loss data. These studies show that the new estimation method is more efficient relative to existing methods, and fits data well in risk analysis.

## References

[1] P. Embrechts and C. Kluppelberg and T. Mikosch, *Modelling Extremal Events, for Insurance and Finances*, Springer, New York, 2003.

[2] I. B. Aban and M. M. Meerschaert and A. K. Panorska, Parameter Estimation for Truncated Pareto Distribution, *Journal of the American Statistical Association*, 2006, Vo. 101, No. 473, pp. 270-277.

# On the shape of excursions of Gaussian processes

Jürg Hüsler – *University of Bern, Switzerland*
Email: juerg.huesler@stat.unibe.ch
Co-author: Vladimir Piterbarg

## Abstract

We investigate the shape of excursions above a high level $u$ by a stationary Gaussian process $X(t)$. Typically the excursions of the process $X(t)$ above a large value $u$ are rather short and tend to 0 as $u$ tends to $\infty$. We call such a case a cluster of excursions since the path may cross the level $u$ many times. The probability of such excursion events can be accurately approximated for large $u$, as is well-known.

However, the length of excursions can be much larger if one conditions on two clusters of excursions which are separated in time by a positive time lag, not depending on $u$. Under certain assumptions on the correlation function of the process, the asymptotic behavior of the probability of such a pattern of clusters of exceedances is derived exactly where the level $u$ tends to $\infty$ (see [1]). The shape of such excursions is quite interesting and does depend on the conditioned mean and covariances of the underlying process.

We focus on the shape of paths in-between the two clusters or after one such cluster. The paths vary slightly around a deterministic trend. In addition, the probability of such events can be determined asymptotically exact for $u \to \infty$.

## References

[1] J. Hüsler, A. Ladneva and V. Piterbarg, *On clusters of high extremes of Gaussian stationary processes with $\varepsilon$-separation.* Elect. J. of Proabability **15**, 1825-1862, 2010.

# Assessing the Index of Regular Variation for Random Difference Equations

Anja Janssen – *University of Hamburg, Germany*
Email: janssen@math.uni-goettingen.de
Co-author: Ulf-Rainer Fiebig

## Abstract

We analyze multivariate random difference equations (RDEs), which are of the form

$$\mathbf{Y}_t = \mathbf{A}_t \mathbf{Y}_{t-1} + \mathbf{B}_t, \quad t \in \mathbb{N}, \tag{5}$$

for $\mathbf{Y}_t \in \mathbb{R}^d$, where $\mathbf{A}_t$ are random $\mathbb{R}^{d\times d}$-matrices and $\mathbf{B}_t$ are random $\mathbb{R}^d$-vectors. We assume that $(\mathbf{A}_t, \mathbf{B}_t)_{t\in\mathbb{N}}$ are i.i.d. and independent of $\mathbf{Y}_0$.

In a seminal paper [3], Kesten derived conditions for $(\mathbf{A}_1, \mathbf{B}_1)$ (mainly for the case of non-negative entries of $\mathbf{A}_1$ and $\mathbf{B}_1$) which guarantee the existence of a stationary solution $(\mathbf{Y}_t)_{t\in\mathbb{N}_0}$ to (5). Furthermore, he derived a property of the stationary distribution which could later be shown to imply multivariate regular variation, cf. [1]. Now, an important characteristic for the extremal behavior of the process $(\mathbf{Y}_t)_{t\in\mathbb{N}_0}$ is given by the index of regular variation $\kappa > 0$, which satisfies

$$\lim_{y\to\infty} \frac{P(\|\mathbf{Y}_t\| > xy)}{P(\|\mathbf{Y}_t\| > y)} = x^{-\kappa}$$

for all $x > 0$ and an arbitrary norm $\|\cdot\|$ on $\mathbb{R}^d$. However, for most cases and especially for higher-dimensional matrices, the formula for the derivation of $\kappa$ which is stated in [3], namely that $\kappa > 0$ is the unique solution to

$$0 = \lim_{n\to\infty} \frac{1}{n} \ln E\|\mathbf{A}_n \cdot \ldots \cdot \mathbf{A}_1\|_{\mathrm{op}}^{\kappa},$$

where $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm, is not analytically solvable. This instance is especially problematic since $\kappa$ is not only of interest in itself but needed for many further derivations and simulations of the extremal behavior of a stationary solution to (5). We will make use of a characteristic of $\kappa$ that is introduced quite implicitly in [3] and which does not involve a limit but a certain spectral measure $\nu_\kappa$ on $\mathbb{S}_+^{d-1} = \{\mathbf{x} \in \mathbb{R}_+^d : \|\mathbf{x}\| = 1\}$ which satisfies

$$\int_{\mathbb{S}_+^{d-1}} E\left[\|\mathbf{A}_1\mathbf{x}\|^\kappa f\left(\frac{\mathbf{A}_1\mathbf{x}}{\|\mathbf{A}_1\mathbf{x}\|}\right)\right] \nu_\kappa(\mathrm{d}\mathbf{x}) = \int_{\mathbb{S}_+^{d-1}} f(\mathbf{x})\, \nu_\kappa(\mathrm{d}\mathbf{x})$$

for all continuous functions $f$ on $\mathbb{S}_+^{d-1}$. We propose a numerical algorithm which allows us to approximate $\nu_\kappa$ and thereby evaluate $\kappa$.

The talk is based on [2].

## References

[1] J. Boman and F. Lindskog, Support theorems for the Radon transform and Cramér-Wold theorems, J. Theoret. Probab., **22**(3) 683710: 2009

[2] A. Janßen, *On Some Connections between Light Tails, Regular Variation and Extremes*, Ph.D. thesis, University of Göttingen: 2010

[3] H. Kesten, *Random difference equations and renewal theory for products of random matrices*, Acta Math., **131** 207248: 1973


# Asymptotic behaviour of a test of no breaks versus fixed number of breaks


Daniela Jaruskova – *Czech Technical University, Prague, Czech Republic*
Email: jarus@mat.fsv.cvut.cz



## Abstract

A sequence of independent random variables $X_i, \ldots, X_n$ is observed. We test the null hypothesis $H_0$ claiming that all observations are identically distributed with the same mean value $\mu$:

$$H_0 : X_i = \mu + e_i, \quad i = 1, \ldots, n,$$

against the alternative that there exists at least one and at most $d$ breaks in the mean (at unknown positions):

$$A_d : \quad \text{there exist} \quad 1 < n_1 < n_2 < \cdots < n_d < n \quad \text{satisfying}\, n_1 \geq \epsilon\, n, n_2 - n_1 \geq \epsilon\, n, \ldots, n - n_d \geq \epsilon\, n$$
$$X_i = \mu_1 + e_i,\ i = 1, \ldots, n_1,\ X_i = \mu_2 + e_i,\ i = n_1 + 1, \ldots, n_2,\ \ldots,\ X_i = \mu_{d+1} + e_i,\ i = n_d + 1, \ldots, n,$$

such that $\mu_{j_0} \neq \mu_{j_0+1}$ for some $1 \leq j_0 \leq d$. The $\{e_i\}$ are i.i.d. satisfying $E\, e_i = 0$, $E\, e_i^2 = 1$ and $E\, |e_i|^{2+\Delta} < \infty$.

Under $H_0$ the test statistic that is based on the standardized differences of least squares esimates of $\{\mu_j\}$ converges in distribution to a maximum of a $\chi^2-$ process (see [1]):

$$\max_{\substack{0<t_1<t_2<\cdots<t_d<t_{d+1}=1 \\ t_1 \geq \epsilon, t_2 - t_1 \geq \epsilon, \ldots, 1 - t_d \geq \epsilon}} X_1^2(\boldsymbol{t}) + \ldots X_d^2(\boldsymbol{t}),$$

where

$$X_i(\boldsymbol{t}) = \frac{1}{\sqrt{t_{i+1} t_i (t_{i+1} - t_i)}} \Big( t_i W(t_{i+1}) - t_{i+1} W(t_i) \Big).$$

The approximate critical values may be obtained from the approximation of the exceedance probability (see [2] and [3]):

$$P\Big( \max_{\substack{0<t_1<t_2<\cdots<t_d<t_{d+1}=1 \\ t_1 \geq \epsilon, t_2 - t_1 \geq \epsilon, \ldots, 1 - t_d \geq \epsilon}} X_1^2(\boldsymbol{t}) + \ldots X_d^2(\boldsymbol{t}) > u^2 \Big) \sim \frac{1}{\pi^{(d-1)/2}}\, I_d\, u^{3d-1} \left(1 - \Phi(u)\right) \quad \text{as} \quad u \to \infty.$$

An explicit expression for the constant $I_d$ may be applied to obtain values of $I_d$ for $d$ small.



## References

[1] J. Bai and P. Perron *Estimating and testing linear models with multiple structural changes*, Econometrica 66, 47 – 78, 1998.

[2] V. I. Piterbarg *Asymptotic methods in the theory of Gaussian processes and fields*, American Mathematical Society, 1996.

[3] D. Jarušková and V. I. Piterbarg *Log-likelihood ratio test for detecting transient change*, Statistics and Probability Letters, 2011, doi:10.1016.


# Max-stable Processes for Threshold Exceedances in Spatial Extremes


Soyoung Jeon – *University of North Carolina at Chapel Hill, U.S.A.*
Email: soyoung@email.unc.edu
Co-author: Richard L. Smith



## Abstract

The analysis of spatial extremes requires the joint modeling of a spatial process at a large number of stations. Multivariate extreme value theory can be used to model the joint extremal behavior of environmental data such as precipitation, snow depths or daily temperatures. Max-stable processes are the natural generalization of extremal dependence structures to infinite dimensions arising from the extension of multivariate extreme value theory. However, there have been few or no works on the threshold approach of max-stable processes.

Padoan, Ribatet and Sisson [1] proposed the maximum composite likelihood approach for fitting max-stable processes to avoid the complexity and unavailability of the multivariate density function. We propose the threshold version of max-stable process estimation and we apply the pairwise composite likelihood method to it. We assume a strict form of condition, so called the second-order regular variation condition, for the distribution satisfying the domain of attraction. It is well known that the condition was used to prove the asymptotic properties of estimators in univariate threshold approach (see Smith [2]) and the second-order condition was studied for bivariate extremes by de Haan and Ferreira [3]. To obtain the limit behavior, we also consider the increasing domain structure with stochastic sampling design based on the setting and conditions in Lahiri [4] and we then establish consistency and asymptotic normality of the estimator for dependence parameter in the threshold method of max-stable processes. The method is studied by simulation and illustrated by the application of temperature data in North Carolina, United States.

### References


[1] S. Padoan, M. Ribatet and S. Sisson, *Likelihood-based inference for max-stable processes*, Journal of the American Statistical Association 105(489), 263–277, 2010.

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

[3] L. de Haan and A. Ferreira, *Extreme Value Theory, An Introduction*, New York: Springer, 2006.

[4] S. N. Lahiri, *Central Limit Theorems for weighted sums under some stochastic and fixed spatial sampling designs*, Sankhya: The indian Journal of Statistics, Ser. A 65, 356–388, 2003.


## Conditional extremes incorporating covariates


Philip Jonathan – *Shell Technology Centre Thornton, U.K.*
Email: philip.jonathan@shell.com
Co-authors: David Randell, Kevin Ewans and Jan Flynn


### Abstract


Reliable design and assessment of marine structures requires estimation of both marginal and joint characteristics of extreme ocean environments. The conditional extremes model of Heffernan and Tawn [1] motivates a particular parametric model for the distribution of variables given a large value of a conditioning variate, for suitably transformed variables. In this work, we extend the conditional extremes approach to incorporate covariate effects in both marginal and conditional extremes. The smoothness of marginal and conditional extremes model parameters with covariate is regulated by roughness-penalised maximum likelihood (see, e.g., [2]). The approach is applied to the estimation of extreme quantiles of ocean storm severity and conditional values of wave peak periods given storm wave direction, for a number of ocean locations, extending [3].


### References


[1] Heffernan, J. E. and J. A. Tawn, *A conditional approach for multivariate extreme values*, J. R. Statist. Soc. B, 2004, **66**, 497.

[2] Ewans, K. C. and P. Jonathan, *The effect of directionality on Northern North Sea extreme wave design criteria*, J. Offshore Mechanics Arctic Engineering, 2008, **130**, 10.

[3] Jonathan, P., J. Flynn and K. C. Ewans, *Joint modelling of wave spectral parameters for extreme sea states*, Ocean Engineering, 2010, **37**, 1070.


## Extreme value statistics of wind speed data by the ACER method


Oleh Karpa – *CeSOS, Norwegian University of Science and Technology, Trondheim, NO-7491, Norway*
Email: oleh.karpa@ntnu.no
Co-author: Arvid Naess


### Abstract


The paper concentrates on the application of a method that appears to be uniquely appropriate for predicting extreme wind speeds. Standard methods for estimating extreme wind speeds from observed data series are commonly based on assuming either that episodical extreme values are Gumbel distributed or by adopting a peaks over threshold (POT) approach, assuming that the exceedances above high thresholds follow a generalized Pareto distribution. A weakness of these approaches is that they depend on adopting asymptotic distributions. However, it is hard to fully verify the applicability or correctness of such procedures. An alternative method for predicting extreme wind speeds that avoids invoking the ultimate asymptotic distributions, but rather try to capture the sub-asymptotic behavior of extreme value data has been developed by Naess and Gaidai [2]. This approach has been used in this work and seems to be the appropriate way to deal with the recorded data time series of, for example, the hourly largest wind speeds observed at a given location.

Assume that the observed data $X_1, ..., X_N$ are allocated to the discrete times $t_1, ..., t_N$. In general, these data may be strongly dependent. Our goal is to accurately determine the distribution function of the extreme value $M_N = \max\{X_j; j = 1, ..., N\}$. Specifically, we want to estimate $P(\eta) = \text{Prob}(M_N \leq \eta) = \text{Prob}(X_1 \leq \eta, \ldots, X_n \leq \eta)$ accurately for large values of $\eta$. The joint distribution function on the right hand side of this equation cannot in general be estimated directly from the data. However, this problem can be solved in practice by introducing a cascade of conditioning approximations $P_k(\eta)$ of $P(\eta)$, where $P_k(\eta) \to P(\eta)$ as $k$ increases, and $P_k(\eta) \approx \exp\big(-\sum_{j=k}^{N} \alpha_{kj}(\eta)\big)$, for $N >> 1$ and $k = 1, 2, ...$, where $\alpha_{kj}(\eta) = \text{Prob}(X_j > \eta | X_{j-1} \leq \eta, ..., X_{j-k+1} \leq \eta)$.

For the empirical estimation of the quantities in the $P_k(\eta)$, average conditional exceedance rates (ACER) are introduced as follows, $\varepsilon_k(\eta) = \sum_{j=k}^{N} \alpha_{kj}(\eta)/(N-k+1)$. Estimation of the ACER function $\varepsilon_k(\eta)$ proceeds by counting the total number of favorable incidents, that is, exceedances conditioned on the requisite number of preceding non-exceedances, for the total data time series.

For the prediction problem, it is argued in [1] that the ACER functions can be represented in the tail as $\varepsilon_k(\eta) \approx q_k \exp\{-a_k(\eta - b_k)^{c_k}\}, \eta \geq \eta_1$, where $a_k, b_k, c_k$ and $q_k$ are suitable constants, that in general will be dependent on $k$. The optimal values of the parameters are obtained by optimizing the fit on the log level by minimizing a mean square error function. Note that the values $c_k = q_k = 1$ correspond to the Gumbel asymptotic form.

Wind speed data, measured at Sula and Torsvåg Fyr weather stations in Norway, were analyzed to obtain numerical results. Hourly maximum wind gust speeds were recorded during 12 years 1998-2010 at the first station and 13 years 1997-2010 at the second.

For the analysis of the data with the ACER method, $\varepsilon_k(\eta)$ for $k = 1, \ldots, 5$ were chosen for both stations. As expected, it was revealed that there is significant dependence between the data. However, this dependence is largely accounted for by $k = 2$; and for $k \geq 5$ full convergence has been achieved for all practical purposes.

Optimal curve fitting and estimation of 100-year return period values were achieved for both stations. In the table 100-year return period values are listed together with 95% confidence intervals.

| | Sula | | Torsvåg | |
|---|---|---|---|---|
| $k$ | $\eta_{100}$, m/s | 95% CI($\eta_{100}$), m/s | $\eta_{100}$, m/s | 95% CI($\eta_{100}$), m/s |
| 1 | 45.27 | (42.68, 47.32) | 49.07 | (43.05, 52.79) |
| 2 | 46.10 | (43.58, 48.60) | 51.11 | (45.66, 56.06) |
| 3 | 47.11 | (44.27, 50.27) | 51.19 | (45.16, 57.28) |
| 4 | 47.28 | (44.35, 50.21) | 50.86 | (46.44, 55.99) |
| 5 | 47.30 | (44.35, 50.00) | 51.13 | (45.52, 55.96) |

The new ACER method offers a unique approach to the prediction of extreme wind speeds. The converged empirical ACER functions provide an estimate of the exact extreme value distribution inherent in the data. It has been observed that the ACER method seems to give consistent and accurate results compared with e.g. the POT method. The ACER method appears to be quite robust with respect to the choice of parameter values. Even if the parameter values may deviate somewhat from the optimal values, the obtained predictions are still good. Under the premise that the estimated ACER function can be used to make extreme value predictions, another advantage of the proposed ACER method over the POT method is also the lack of sensitivity to outliers.

## References


[1] Naess, A., Gaidai, O., 2009. *Estimation of extreme values from sampled time series.*, Structural Safety 31, 325-334.

[2] Naess, A., Gaidai, O., 2008. *Monte Carlo Methods for Estimating the Extreme Response of Dynamical Systems.* Journal of Engineering Mechanics. ASCE 134(8), 628-636.


# Extremal dependence and the ACE algorithm


Keith Knight – *University of Toronto, Canada*
Email: keith@utstat.toronto.edu


## Abstract


The maximal correlation between two random variables $X$ and $Y$ was defined by Gebelein in [3] as the supremum of $E[\psi(X)\phi(Y)]$ over functions $\psi$ and $\phi$ satisfying $E[\psi(X)] = E[\phi(Y)] = 0$ and $E[\psi^2(X)] = E[\phi^2(Y)] = 1$. Given a sample from a joint distribution, the maximal correlation (as well as the optimal functions $\psi$ and $\phi$) can be estimated via the ACE algorithm introduced in [1]; see also [2]. In this talk, we will define a version of maximal correlation for bivariate extremal dependence, which depends on the random variables only via their copula, and discuss its estimation using an appropriate modification of the ACE algorithm.

## References

[1] L. Breiman and J. Friedman (1985) Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association.* **80**, 580–598.

[2] A. Buja (1990) Remarks on functional canonical variates, alternating least squares methods and ACE. *Annals of Statistics.* **18**, 1032–1069.

[3] H. Gebelein (1941) Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung. *Zeitscrift für Angewandte Mathematik und Mechanik.* **21**, 364–379.


# Multi–element generalized Polynomial Chaos for Extreme Quantile Estimation


Jordan Ko – *Laboratoire de Probabilités et Modèles Aléatoires, Université de Paris 7, France*
Email: jordan.ko@mac.com
Co-author: Josselin Garnier



## Abstract

The estimation of extreme quantile is a challenging numerical topic and has received a considerable amount of attention in many research disciplines. In the context of this paper, we wish to evaluate the $\alpha$–quantile, $Y_\alpha$, of the solution of a numerical function, $f(\boldsymbol{\xi})$, whose input, $\boldsymbol{\xi}$, is a multi–dimensional random vector. Monte Carlo (MC) method and its variants are the traditional approach for quantile approximation but large number of MC samples are needed to accurately determine extreme quantiles. Therefore, it may not be practical in cases where the numerical function is computational costly to evaluate. Importance sampling (IS) reduces the numerical cost by concentrating MC samples near design points, *i.e.* input values associated with the quantiles; however, determining the new IS distribution can be costly. Within the context of structural reliability, First-Order Reliability Method (FORM) and Second-Order Reliability Method (SORM) have been developed to estimate quantiles but the methods are accurate only for moderate quantiles and the error is difficult to estimate [1].

In the current study, we investigate the quantile estimation by multi–element generalized Polynomial Chaos (gPC) metamodels. gPC is a generalization of the original Homogeneous Chaos introduced by Wiener [2] and it is an expansion of the function solution with multivariate orthogonal polynomials, *i.e.*

$$f_r(\boldsymbol{\xi}) = \sum_{m=0}^{M} f_m \phi_m(\boldsymbol{\xi}),$$

where $r$ denotes that the function is a metamodel and $\phi_m(\boldsymbol{\xi})$ is an orthogonal polynomial whose weight function is similar to the *pdf* of $\boldsymbol{\xi}$. The original Wiener polynomial chaos used the Hermite polynomials in terms of the Gaussian random variables [2]. The coefficients of expansion $f_m$ are determined through a collocation method where the Galerkin projection of $f_r(\boldsymbol{\xi})$ with respect to $\phi_m(\boldsymbol{\xi})$ is approximated with numerical quadrature. The gPC has recently been applied to many uncertainty quantification studies [3, 4]. From the spectral gPC metamodel, different statistical measures such as mean, variance and sensitivity indices can be readily computed [5]. Function evaluation on the gPC metamodel can be considered as essentially free; thus, large number of MC samples from the metamodel can be used to estimate the $Y_\alpha$, for moderate $\alpha$.

As the gPC metamodels are expansions about the means of the inputs, their accuracy may worsen away from these mean values where the extreme events may occur. By increasing the quadrature accuracy, we may eventually improve accuracy of the quantile but it can be very expensive. Thus, a multi–element approach is used by combining the global metamodel with supplementary local metamodels centered at the design points. The design point can be sought by solving a minimum constraint problem where the limit–state function is the feasibility surface, $f(\boldsymbol{\xi}) = Y_\alpha$, and the objective function is the Euclidean norm of $\boldsymbol{\xi}$. It is solved with the Lagrange multiplier algorithm where the gPC metamodel is used in the limit–state function.

Supplemental local refinement metamodels are constructed in a bounded domain centered on the design point. To improve the accuracy and to minimize the sampling cost, sparse–tensor and anisotropic quadratres are tested in addition to the full–tensor Gauss quadrature in the computation of local metamodel. The global and local metamodels are combined in the multi–element gPC (MEgPC) approach for extreme quantile estimation. After the MEgPC quantile estimation approach is validated with some test functions, it is applied to estimation extreme quantiles in the context of a flooding model, a General Circulation Model for precipitation prediction and a turbulence model. It is shown that MEgPC can be more accurate than MC or IS for extreme quantile estimations for input dimensions less than $N = 5$ to $N = 7$.



## References

[1] R. Rackwitz, *Reliability analysis - a review and some perspectives*, Structural Safety 23 (2001) 365395.

[2] N. Wiener, *The homogeneous chaos*, American Journal of Mathematics 60 (1938) 897936.

[3] R. Ghanem, P. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Springer, Berlin, 1991.

[4] D. Xiu, D. Lucor, C.-H. Su, G. Karniadakis, *Stochastic modeling of flow-structure interactions using Generalized Polynomial Chaos*, Journal of Fluids Engineering 124 (2001) 5159.

[5] J. Ko, D. Lucor, P. Sagaut, *Sensitivity of two-dimensional spatially developing mixing layers with respect to uncertain inflow conditions*, Physics of Fluids 20 (2008) 077102.

# Estimation of High Conditional Quantiles

Deyuan Li – *Fudan University, China*
Email: deyuanli@fudan.edu.cn
Co-authors: Xuming He and Huixia Wang

## Abstract

Estimation of conditional quantiles at very high or low tails is of interest in numerous applications such as climate studies, finance and economics, medical cost studies, studies of infant birthweights. Quantile regression provides a convenient and direct tool to assess the impact of covariates at different tails of the response distribution. However, due to lack of information in the tail areas, estimations from quantile regression are often not precise at tails especially for heavy-tailed distributions. In this paper, we develop two new estimation methods for high conditional quantiles based on the extreme value theory. The new methods operate by first estimating the intermediate conditional quantiles using quantile regression, and then extrapolating these estimates to the high tails using different assumptions on the tail behavior. We rigorously establish the asymptotic properties of the proposed conditional quantile estimators and the associated tail index estimators. Compared to the conventional quantile regression, the proposed extrapolation methods demonstrate higher accuracy in a simulation study, and they lead to more insightful estimation of high conditional quantiles of precipitation in a downscaling analysis of daily precipitation in Chicago urban area.

## References

[1] Drees, H. (1998). On smooth statistical tail functionals. *Scand. J. Statis.* **25**, 187-210.

[2] de Haan, L and Ferreira, A. (2006). *Extreme value theory: an introduction.* Springer.

[3] Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis* 91, 74-89.

[4] Koenker, R. (2005). *Quantile Regression.* Cambridge University Press, Cambridge.

[5] Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, 36, 1108-1126.

# Asymptotic Approximations of Integrals of Gaussian Random Fields

Jingchen Liu – *Columbia University*
Email: jcliu@stat.columbia.edu

## Abstract

In this talk, we present asymptotic approximations of $P(\int_T e^{f(t)}dt > u)$ as $u \to \infty$ for a smooth Gaussian random field, $f$, living on a $d$-dimensional compact set $T$. The results for the homogeneous cases are provided by [1]. Further results when $f$ has a spatially varying mean are given in [2]. From a theoretical point of view, this analysis is a generalization of the study of the probability $P(\sup_T f(t) > u)$. From an application point view, the integral of exponent of Gaussian random field is an important random variable for many generic models in spatial point processes, portfolio risk analysis, asset pricing, material failure theory, and so forth.

The analysis technique consists of two steps: 1. evaluate the tail probability $P(\int_\Xi e^{f(t)}dt > u)$ over a small domain $\Xi$ depending on $u$, where $mes(\Xi) \to 0$ as $b \to \infty$ and $mes(\cdot)$ is the Lebesgue measure; 2. with $\Xi$ appropriately chosen, we show that $P(\int_T e^{f(t)}dt > b) = (1 + o(1))mes(T)mes^{-1}(\Xi)P(\int_\Xi e^{f(t)}dt > u)$.

## References

[1] Jingchen Liu. Tail Approximations of Integrals of Gaussian Random Fields. *Annals of Probability.* Accepted.

[2] Jingchen Liu and Gongjun Xu. Some Asymptotic Results of Gaussian Random Fields with Varying Mean Functions and the Associated Processes. Preprint.

# Asymptotic portfolio risk in multivariate regularly varying models

Georg Mainik – *ETH Zürich, Switzerland*
Email: georg.mainik@math.ethz.ch

## Abstract

This talk comprises a series of results on the extremal behaviour of portfolio losses in multivariate regularly varying models. Asymptotic distributions of extreme portfolio losses are characterized by a functional of the portfolio weights and the intrinsic parameters of the multivariate regular variation, given by the tail index and the spectral measure.

Existence, uniqueness, and location of the optimal portfolio are analysed and applied to risk minimization. The analysis of the optimization problem shows that the asymptotic portfolio risk is a convex function of the portfolio vector if the tail index is higher than or equal to 1. On the other hand, if the components of the multivariate regularly varying vector are non-negative and the tail index is lower than 1, the asymptotic portfolio risk is concave.

Estimation of the asymptotic portfolio risk factor in the i.i.d. setting is approached by a semiparametric method. Strong consistency and asymptotic normality are established as a functional law of large numbers and a functional central limit theorem. The estimated portfolio risk factor converges uniformly on compact portfolio sets.

To compare the diversificatio effecst across different models, the notion of asymptotic portfolio loss ordering is introduced. The initial definiton is stated in terms of uniform ordering of asymptotic portfolio losses for all portfolios with non-negative weights. Under the assumption of multivariate regular variation this ordering has equivalent criteria in terms of marginal distributions and spectral measures. These results are applied to copula models and to the characterization of the worst and the best dependence structures for the diversification of multivariate regularly varying risks.

## References

[1] G. Mainik and L. Rüschendorf, *On optimal portfolio diversification with respect to extreme risks*, Finance and Stochastics, 2010.

[2] G. Mainik, *Estimating asymptotic dependence functionals in multivariate regularly varying models*, preprint, 2011

[3] G. Mainik and L. Rüschendorf, *Ordering of multivariate risk models with respect to extreme portfolio losses*, preprint, 2011

# Estimating bivarite tails, a copula based approach

Véronique Maume-Deschamps – *Université de Lyon, Université Lyon 1, ISFA, Laboratoire SAF, 69366, France*
Email: veronique.maume@univ-lyon1.fr
Co-authors: Elena Di Bernardino and Clémentine Prieur.

## Abstract

We are interested in the problem of fitting the joint distribution of bivariate observations exceeding high thresholds. To this end, we develop a two-dimensional extension of the POT method, mainly based on a version of the Pickands-Balkema-de Haan Theorem in dimension 2. This extension allows to consider a two-dimensional structure of dependence between both continuous random components $X$ and $Y$. The modelling of this dependence is done via a copula $C$, which is supposed to be unknown. Asymptotic dependence as well as asymptotic independence are considered.

Starting from the work by Juri and Wüthrich ([4]) and Charpentier and Juri ([2]), we construct a two-dimensional tail estimator and study its asymptotic properties. A parameter that describes the nature of the tail dependence is also introduced and estimated. We also present real data examples which illustrate our theoretical results.

Modeling the dependence by using copulas leads to exploit the one dimensional generalized Pareto distribution of the excesses of the marginal laws. Using directely the multivariate generalized Pareto distribution developed by Falk and Reiss ([3]) and Rootzen and Tajvidi ([5]) could be another attempt to estimating the bivariate tail bistribution. Nevertheless, estimation of the scaling parameters of the multivariate generalized Pareto distribution have to be adressed first. The model proposed by Ledford and Tawn provides a way to take into account the dependence structure in the tail (see [1] for a recent work). We propose an aternative model based on regularity conditions of the copula and on the explicit description and estimation of the dependence structure in the joint tail.

## References

[1] Beirlant J., Dierckx G. and Guillou A., *Reduced bias estimators for bivariate tail modelling* Insurance: Mathematics and Economics, in press (2011).

[2] Charpentier A. and Juri A. *Limiting dependence structures for tail events, with applications to credit derivatives.*, J. Appl. Probab., **43**, 2, 563–586, (2006).

[3] Falk, M. and Reiss, R.-D., *On the distribution of Pickands coordinates in bivariate EV and GP models*, J. Multivariate Anal., **93**, 2, 267–295, (2005).

[4] Juri A. and Wüthrich M.V., Tail Dependence From a Distributional Point of View, *Extremes*, **3**, 213-246, (2003).

[5] Rootzén H. , Tajvidi N., *Multivariate generalized Pareto distributions*, Bernoulli, **12**, 5, 917–930, (2006).

# POT methods: a new insight into the estimation of extreme value distributions

Franck Mazas – *Sogreah Maritime, ARTELIA Group, France.*
Email: Franck.Mazas@arteliagroup.com
Co-authors: Philippe Garat and Luc Hamm

## Abstract

For a decade, a broad consensus has emerged in favor of a general methodology for determining extreme values of environmental data. Roughly speaking, this methodology consists in the five following steps. First, an independent and identically distributed (i.i.d.) sample is extracted from the raw time series. For this purpose, the time series is homogenized in order to identify the different populations it may contain, then declustering is applied with a Peaks-Over-Threshold (POT) approach to extract the independent extreme data from these homogeneous sub-samples. Second, a statistically meaningful threshold is set, usually by theoretical properties of the Generalized Pareto Distribution (GPD). Third, a two-parameter GPD is fit to the threshold excesses using the Maximum Likelihood Estimator (MLE). Fourth, quantiles (return values) are calculated for the return periods of interest (assuming the number of events per year is a Poisson process). Fifth, confidence intervals are computed.

This methodology, often referred to as the GPD-Poisson model, has sound theoretical justifications (see for instance [1]). The GPD is the asymptotic law for the threshold excesses. As for the MLE, it is particularly valued by statisticians for its asymptotic properties: consistency, efficiency, asymptotic normality, and hence is widely used among analysts.

However, we wish to discuss in this communication two questionable points of this methodology: one concerning the general form of the GPD-Poisson model itself, the other one concerning the estimation techniques.

First, one should keep in mind that the GPD is an asymptotic law, meaning that it is an approximation of the true law of excesses over a threshold valid only when this threshold is high enough. Consequently, if the asymptotic domain is not reached, other distributions may fit the data better. As presented in [2], we extend the univariate and stationary GPD-Poisson model by fitting two additional distributions (namely the Weibull and Gamma distributions), though others could of course be tested. Objective criteria are then used to select the best-fitting law.

Such a multi-distribution approach is useful for analysts and/or engineers who wish to cover a wide range of situations and is justified since the GPD is outperformed in several practical cases.

Second, MLE exhibits strange behavior when the threshold is allowed to vary between two consecutive data values [3]: the estimated parameters of the distribution (scale and shape) and consequently the estimated quantiles do not remain constant. Thus, a slight translation of the sample leads to a significant change in the estimation, although the sample size remains identical. This phenomenon is particularly noteworthy for the Weibull and Gamma distributions, but also exists for the GPD. For the former distributions, the likelihood tends to infinity in many cases when the threshold tends to the open upper bound of its interval of validity, i.e. the first data. For the latter, the maximum of the likelihood is reached at this open upper bound, but with a non-nil derivate. Still, it must be kept in mind that the asymptotic properties of the MLE are valid only when the maximum is reached on an interior point of an open set [4].

Furthermore, resampling techniques from real extreme data show that when MLE is applied to 2-parameter GPD, the empirical density (histogram) of estimated parameters exhibits 2 or 3 peaks, resulting in many peaks in the density of quantiles.

Actually, the widely spread 2-parameter GPD-Poisson model makes a confusion between the threshold, whose role should be limited to data selection/censorship, and the location parameter, whose role is to accurately set the origin of the distributions. Hence it is necessary to fit 3-parameter distributions and MLE must be rejected since it is not valid in many practical cases. We use the L-moments estimator [5] which allows a proper estimation of the three parameters for the studied distributions. This estimator is unbiased, although its variance is still significant.

In conclusion: i) a multi-distribution approach to POT methods is recommended since in practical applications it is unsure whether the asymptotic domain for the GPD is reached; ii) we strongly favor including a location parameter, clearly distinct from the threshold, in the distributions for extreme values, and iii) MLE should not be used without a close examination of its validity conditions. Our tests with real environmental data (extreme wave heights for the design of coastal structures) show that in this case, MLE should be rejected whereas the 3-parameter L-moments estimator behaves well.

## References

[1] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*, Springer, London, 2001.

[2] F. Mazas, L. Hamm, A multi-distribution approach of POT methods for determining extreme wave heights, *Coastal Engineering*, **58**, 385-394, 2011.

[3] F. Mazas, P. Garat, L. Hamm, Questioning MLE: a new insight into the estimation of extreme wave distributions. Submitted to *Coastal Engineering*, 2011.

[4] E.L. Lehmann, *Theory of point estimation*, Wiley & Sons, Australia, 1983.

[5] J.R.M. Hosking, L-moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics, *J. R. Statist. Soc.*, **52**(1), 105-124, 1990.

# Limit theorems for coupon collector's problem and generalized Pareto distributions

Pavle Mladenović – *University of Belgrade, Serbia*
Email: paja@matf.bg.ac.rs
Co-author: Jelena Jocković

## Abstract

Limit theorems for waiting times that arise in coupon collector's problem will be presented. New results that give connection with generalized Pareto distributions are also included.

## References

[1] J. Jocković and P. Mladenović, Coupon collector's problem and generalized Pareto distributions, *Journal of Statistical Planning and Inference*, **141**, 2348-2352, 2011.

[2] P. Mladenović and J. Vukmirović, Rates of convergence in certain limit theorems for extreme values, *Journal of Mathematical Analysis and Applications*, **363**, 287-295, 2010.

[3] P. Mladenović, Limit distributions for the problem of collecting pairs, *Bernoulli*, **14**(2), 419-439, 2008.

# Estimation of parameters of extreme events for randomly censored data

M. Manuela Neves – *CEAUL, University of Lisbon and ISA, Technical University of Lisbon, Portugal*
Email: manela@isa.utl.pt
Co-author: M. Ivette Gomes

## Abstract

In *Statistics of Extremes* we deal essentially with the estimation of parameters of extreme or even rare events. The most common assumption on any set of univariate data, $(X_1, X_2, \ldots, X_n)$, is to consider them as a *complete* sample of size $n$, with observations either independent and identically distributed or weakly dependent and stationary, from an unknown distribution function $F = F_X$. There is a large variety of parameters of extreme events, but in all applications of extreme value theory (EVT), the estimation of the *extreme value index* (EVI), denoted $\gamma_X$, is of primordial importance and the basis for the estimation of all other parameters of extreme events. Among the most relevant parameters of extreme events, and assuming that we are interested in large values, i.e., in the right tail of the underlying model $F$, we mention:

- the *probability of exceedance* of a high level $x \equiv x_H$, $p_x := \mathbb{P}(X > x) = 1 - F(x)$,
- the *return period* of a high level $x$, which is given by $r_x := 1/(1 - F(x))$, in an i.i.d. scheme,
- the *right endpoint* of an underlying model $F$, $x^* \equiv x^F := \sup\{x : F(x) < 1\}$, and
- a *high quantile* of probability $1 - p$, $p$ small, situated in the border or even beyond the range of the available data, defined as $\chi_{1-p} := \inf\{x : F(x) \geq 1 - p\} =: F^{\leftarrow}(1 - p),\ p < 1/n$.

But in many real situations, censored observations can occur. For example,

- in the analysis of lifetime data or reliability data,
- in the analysis of some physical phenomena such as wind speeds, earthquake intensities or floods, where extreme measurements are sometimes not available due to damage in the instruments.

We shall give here special attention to the estimation of the extreme value index $\gamma_X$, under *random censorship*, where apart from two recent papers by Einmahl *et al.* (2008) and Gomes and Neves (2010), there is only, as far as we know, a brief reference to the topic in Reiss and Thomas (1997, Section 6.1) and a paper by Beirlant *et al.* (2007). We first provide a few details on the EVI and max-domains of attraction. Next, we introduce a set of semi-parametric EVI estimators, valid for complete samples, providing some details on their asymptotic non-degenerate behaviour, and we illustrate the effect of random censorship on the EVI of the potential, non-available sample $\underline{\mathbf{X}} = (X_1, \ldots, X_n)$. This is done in order to motivate the functional expression of the EVI estimators for randomly censored data. Indeed, if we are under a random censor scheme, any of the common EVI estimators needs to be slightly modified in order to be consistent. We shall also present the results of a small-scale Monte-Carlo simulation devised to obtain the behaviour of the EVI-estimators under study. Finally, we illustrate the behaviour of the same EVI-estimators for a few sets of survival data, available in Klein and Moeschberger (2005).

## References

[1] Beirlant, J., Guillou, A., Dierckx, G. and Fils-Viletard, A. (2007). Estimation of the extreme value index and extreme quantiles under random censoring. *Extremes* **10**:3, 151-174.

[2] Einmahl, J.H.J, Fils-Villetard, A. and Guillou, A. (2008). Statistics of extremes under random censoring. *Bernoulli* **14**:1, 207-227.

[3] Gomes, M.I. and Neves, M.M. (2010). A note on statistics of extremes for censoring schemes on a heavy right tail. In Luzar-Stiffler, V., Jarec, I. and Bekic, Z. (eds.), *Proceedings of ITI* 2010, SRCE Univ. Computing Centre Editions, 539-544.

[4] Klein, J.P. and Moeschberger, M.L. (2005). *Datasets for Survival Analysis – Techniques for Censored and Truncated Data*, Second Edition, Springer.

[5] Reiss. R.-D. and Thomas, M. (1997). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and other Fields*, Birkhäuser Verlag, Basel.


# Dense classes of multivariate max stable distributions


John P. Nolan – *American University, Washington, DC, USA*
Email: jpnolan@american.edu
Co-authors: Anne-Laure Fougères and Cécile Mercadier



## Abstract

Our goal is to describe flexible max-stable models in high dimensions. The first class is the max stable distributions with discrete angular measures. It is tractable in all dimensions, with simple exact methods to simulate and compute the distribution function. The second class is max stable models with piecewise polynomial density for the angular measure. Currently, these are only accessible in two dimensions. A third class of models is the family of generalized asymmetric logistic models. We detail the connection between this class of models and the closely related class of generalized stable mixtures. It is shown that all three classes are dense in the space of max stable models.



## References

[1] Fougères, A.-L., Nolan, J. P. and Rootzén, H., Models for dependent extremes using stable mixtures, *Scand. J. Stat.*, **36**, pp. 42–59, 2009.

[2] J.P. Nolan, *Metrics for multivariate stable distributions*, To appear in Banach Center Publications, 2011.


# Conditional sampling for mixed moving maxima on the real axis


Marco Oesting – *Georg-August-Universität Göttingen, Germany*
Email: oesting@math.uni-goettingen.de
Co-author: Martin Schlather



## Abstract

Max-stable processes are frequently used for modelling extreme events with spatial or temporal dependencies. In this context the question of conditional sampling and prediction turns out to be a challenging problem. Recently, an exact algorithm for the max-linear model

$$X_i = \max_j a_{i,j} Z_j, \quad Z_j \text{ i.i.d. Fréchet},$$

has been introduced [5].

Here, we consider the class of stationary max-stable processes which allow for a mixed moving maxima representation, see [3], i.e.

$$Z(t) = \max_{(s,u)\in\Pi} u F_{s,u}(t-s), \quad t \in \mathbb{R}^d,$$

where $\Pi$ is a Poisson point process on $\mathbb{R}^d \times (0,\infty)$ with intensity measure $\mathrm{d}s \times u^{-2}\mathrm{d}u$, and $F_{s,u}$ are independent copies of a random measurable "shape function" $F : \mathbb{R}^d \to (0,\infty)$.

We develop an exact procedure for conditional sampling using of the Poisson point process structure. Since explicit calculations turn out to be quite sophisticated, we restrict ourselves to the case where $d = 1$ and use a finite number of shape functions satisfying some regularity conditions. For more general shape functions approximation techniques using MCMC methods are presented.

Our algorithm is applied to the Gaussian extreme value process where the shape function is the Gaussian density ([4]) and, secondly, to a Brown-Resnick process with a mixed moving maxima representation [1], [2]. Finally, we compare our computational results to other approaches including the algorithm of [5].

## References

[1] B. M. Brown and S. I. Resnick, Extreme Values of Independent Stochastic Processes, *J. Appl. Probab.*, 14(4): 732–739, 1977.

[2] S. Engelke, Z. Kabluchko and M. Schlather, An equivalent representation of the Brown-Resnick process, *Statist. Probab. Lett.*, To appear, 2011.

[3] M. Schlather, Models for Stationary Max-Stable Random Fields, *Extremes*, 5(1):33–44, 2002.

[4] R. L. Smith, Max-stable processes and spatial extremes, Unpublished manuscript, 1990.

[5] Y. Wang and S. A. Stoev, Conditional sampling for spectrally discrete max-stable random fields, *ArXiv e-prints*, 1005.0312, 2010.


# New inferential tools for max-stable processes


Thomas Opitz – *University Montpellier 2, France*
Email: Thomas.Opitz@math.univ-montp2.fr
Co-authors: Jean-Noël Bacro and Pierre Ribereau



## Abstract

Modelling the extreme behaviour of random spatial phenomena is indispensable for the management of risks related to climate and environment. Phenomena of interest include precipitation, snowfall, storm tides/surges, wind and temperatures. Max-stable fields arise as limits of adequately normalised iid sequences of spatial processes and thus are obvious candidats for the modelisation of spatial dependence between extreme events. We recall some of their properties, the current principal inferential methods and some common models, see for example [1], [2] and [3]. We propose new inferential approaches to characterize the extremal dependence structure, based on the so-called spectral measure of the bivariate distributions for site pairs. In particular, a novel graphical tool and a threshold-based (composite) likelihood approach are presented.



## References

[1] S.A. Padoan, M. Ribatet and S.A. Sisson. Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105:263-277, 2010.

[2] Z. Kabluchko, M. Schlather and L. de Haan. Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, 37:2042-2065, 2009.

[3] R.L. Smith. Max-stable processes and spatial extremes. *Unpublished Manuscript*, 1990.


# A diffusion model for the simulation of observed temperature series


Sylvie Parey – *EDF R&D, France*
Email: sylvie.parey@edf.fr
Co-authors: Thi Tu Huong Hoang and Didier Dacunha-Castelle



## Abstract

Stochastic simulation provides a tool to study the occurrence of hazards potentially penalizing for industrial activities. Extreme low or extreme high temperatures are one of those hazards regarding electricity production and consumption. A methodology is thus proposed to simulate observed daily minimum or maximum temperature series. In a first step, the observed temperature is analyzed in order to remove low frequency processes, like trends and seasonality. The non parametric LOESS technique is used to derive trends in mean and variance, and a modified cross-validation technique is proposed to select the smooth parameter. Then, seasonality of the mean and variance of the detrended reduced variable is computed as a trigonometric polynomial and removed. The remaining process is checked regarding possible remaining trends or seasonality in different moments, as well as in its extremes. Finally, a diffusion process is designed to simulate it in the best way, especially regarding extreme events. Examples will be given for temperature series in Europe or in the United States.



## References

[1] Dacunha-Castelle D. and Florens-Zmirou D.: Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, 19 :263-284,1986.

[2] Hoang T.T.H: Modlisation de sries chronologiques non stationnaires, non linaires: application la dfinition des tendances sur la moyenne, la variabilit et les extrmes de la temprature de l?air en Europe. *Thse de Doctorat de l?universit Paris 11 Orsay*, 2010.

[3] Hoang T.T.H., Parey S., and Dacunha-Castelle D.: Multidimensional trends : the example of temperature. *European Physical Journal-Special Topics*, 174 :113-124, 2009.

[4] Klein Tank, A.M.G., J.B. Wijngaard, G.P. Knnen, R. Bhm, G. Demare, A. Gocheva, M. Mileta, S. Pashiardis, L. Hejkrlik, C. Kern-Hansen, R. Heino, P. Bessemoulin, G. Mller-Westermeier, M. Tzanakou, S. Szalai, T. Plsdttir, D. Fitzgerald, S. Rubin, M. Capaldo, M. Maugeri, A. Leitass, A. Bukantis, R. Aberfeld, A.F.V. van Engelen, E. Forland, M. Mietus, F. Coelho, C. Mares, V. Razuvaev, E. Nieplova, T. Cegnar, J. Antonio Lpez, B. Dahlstrm, A. Moberg, W. Kirchhofer, A. Ceylan, O. Pachaliuk, L.V. Alexander, and P. Petrovic. Daily datasets of 20th-century surface air temperature and precipitation series for the European climate assessment. *Int. J. of Clim.*, 22 :1441-1453, 2002


.

# Modelling extreme values of processes observed at irregular time step

Nicolas Raillard – *University Brest, France*
Email: nicolas.raillard@univ-brest.fr
Co-authors: Pierre Ailliot and Jian-Feng Yao

## Abstract

When modelling the extremal properties of a given phenomenon is needed, two main approaches are under statistical interest: modelling the maxima over a large block of consecutive observations or modelling the excesses over a high threshold.

The first approach, denoted as the 'block maxima' method relies on results that describe the acceptable distributions for the maximum of a random variable. The main drawback of this approach is the waste of data induced by taking the maximum over a large block (e.g. a year). Hence another approach, denoted **POT** for Peaks Over Threshold, consists in taking into account any value that is high enough to lie in the tail of the distribution, i.e that exceeds a sufficiently high threshold.

A specific problem related to the **POT** approach is that consecutive data are dependent. Indeed, every data that exceeds a threshold has to be taken into account, but it is very likely that clusters of consecutive excess are being observed, even more if the time-lag between observations is smaller than the characteristic duration of an extremal event. This is a real problem since the theory only deals with independent observations and forgetting this dependence leads to under- of over-estimation of the extremes and despite its importance, this fact is often forgotten in applications. To deal with this dependence, two main approaches are now well-developed: declustering and markov-chains models. The first scheme is based on a probabilistic result stated by Leadbetter and well explained by Coles [1], Chapter 5: the extremes of a dependant stationary sequence show clusters of consecutive large values, and there exists a parameter, named the **extremal index** which summarize this dependence and is found to be, under some assumptions, the reciprocal of the mean cluster size. This result induces a filtering of the dependent observations to obtain nearly independent excesses: once the clusters are identified, the parameters can be estimated thanks to the maxima over the clusters and the clusters length.

As seen above, two specific problems arise: first, a definition of a cluster of consecutive exeedances is needed, and this scheme leads to a waste of data, since only maxima within each clusters are used to estimate the parameters. Hence another approach is to keep all excess but with the introduction of an appropriate dependence structure. This has been used with a **first-order Markov-chain** by Smith [4, 5] with tools to compute several features of the fitted model. More recently, Ribatet [3] applied this approach to several flood characteristic along with diagnostic tools.

The present study takes place within the latter framework. The idea guiding this work was the modelling of extremes of the significant wave height along satellite tracks: indeed, those data exhibits a complex structure, both in space and time, in addition to the presence of missing values along a track. It may thus be quite difficult to apply the usual declustering since identifying clusters when there is missing values might be a hard task. The need for extension come from the irregular time sampling on the data: the complex repartition of the satellite induce a complex recurrence time of passage at a given location, hence there is no natural time-lag between consecutive points, and a markovian structure is quite inadequate. So extensions of the aforementioned method to this irregular sampling in time will be under study in this talk. The main idea is to approximate the data above a threshold by a **censored max-stable process**, whereas previous works used a bivariate extreme value distribution. Based on work of Padoan [2], an estimation procedure is provided thanks to a composite likelihood method. Performance of this new estimator will be assessed both on synthetic data and on significant wave height data that comes from buoy and from numerical models.

## References

[1] Stuart Coles. *An introduction to statistical modeling of extreme values.* Springer Series in Statistics. Springer-Verlag London Ltd., London, 2001.

[2] S. A. Padoan, M. Ribatet, and S. A. Sisson. *Likelihood-based inference for max-stable processes.* Journal of the American Statistical Association, 105(489):263-277, 2010.

[3] M. Ribatet, T. B. M. J. Ouarda, E. Sauquet, and J. M. Gresillon. *Modeling all exceedances above a threshold using an extremal dependence structure: Inferences on several flood characteristics.* Water Resources Research, 45(3), March 2009.

[4] Richard L. Smith. *The extremal index for a markov chain.* J. Appl. Probab., 29(1):37-45, 1992.

[5] Richard L. Smith, Jonathan A. Tawn, and Stuart G. Coles. *Markov chain models for threshold exceedances.* Biometrika, 84(2):249-2681997

# Conditioning Exceedances Processes on Covariate Information

Rolf-Dieter Reiss – *University Siegen, Germany*
Email: rolf.reiss@gmx.net
Co-author: Ulf Cormann

## Abstract

This talk concerns the statistical inference for the upper tail of the conditional distribution of a response variable $Y$ given a covariate $X = x$ based on $n$ random vectors within the parametric extreme value framework. Pioneering work in this field was done by Smith [2] and Smith and Shively [3]. We propose to base the inference on a conditional distribution of the point process of exceedances given the point process of covariates. It is of importance that the conditional distribution merely depends on the conditional distribution of the response variable given the covariates. In the special case of Poisson processes such a result may be found in a book by Reiss [1]. Our results are valid within the broader model where the response variables are conditionally independent given the covariates.

## References

[1] Reiss, R.-D., A Course on Point Processes. Springer, New York (1993).

[2] Smith, R.L., Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-Level Ozone. Statistical Science 4, 367-393 (1989).

[3] Smith, R.L. and Shively, T.S., Point process approach to modeling trends in tropospheric ozone based on exceedances of high threshold. Atmospheric Environment 29, 3489-3499 (1995).

# On the comparison of methods for predicting extremes: a data-based methodological framework.

Benjamin Renard – *Cemagref, UR HHLY Hydrology-Hydraulics, Lyon, France*
Email: benjamin.renard@cemagref.fr
Co-authors: Krzysztof Kochanek, Michel Lang, Federico Garavaglia, Emmanuel Paquet, Luc Neppel, Kenza Najib, Julie Carreau, Patrick Arnaud, Yoann Aubert, Franois Borchi, Jean-Michel Soubeyroux, Sylvie Jourdain, Jean-Michel Veysseire, Eric Sauquet, Annick Auffray

## Abstract

Extreme value analysis is one of the cornerstones of hazard quantification and risk assessment. Its basic objective is to estimate the distribution of some environmental variable $X$, e.g. annual maximum of the areal rainfall over some catchment, annual maximum flood, etc. This distribution can be used to estimate the exceedance probability of a given value of $X$ (often expressed in terms of return period), or alternatively, to estimate the $p$-quantile of $X$. The estimation of quantiles is of primary importance since they are used to design civil engineering structures (e.g. dams, reservoirs, bridges) or to map hazard-prone areas where restrictions may be enforced (e.g. building restrictions in flood zones).

Extreme value analysis has been the subject of extensive research, yielding an abundance of approaches. In Hydrology, several families of methods exist, including (but not limited to):

- Standard application of extreme value theory (EVT), i.e. estimation of an extreme value distribution based on a sample of block maxima or peaks over a high threshold
- Climate/Weather-informed application of EVT. This family of methods uses additional meteorological (e.g., weather type, [1]) or climatic (e.g. Interdecadal Pacific Oscillation IPO, [2]) information.
- Regional approaches, conjointly using data from several sites to perform the inference, which may improve the precision of estimates.
- Model-based approaches, using a simulation model reproducing the main characteristics of the environmental variable (e.g. [3]).

In practice, users and practitioners of extreme value analyses may feel lost facing such an abundance of methods. Consequently, it is necessary to provide them with practical guidelines to choose and implement adequate methods, depending on the conditions of application (e.g. availability of long series, geographical area, type of hydrological regime, etc.).

This presentation describes a methodological framework to perform a data-based comparison of competing approaches for predicting extremes. This framework is based on the following principles:

- The objective is to assess the *predictive* performance of competing methods (as opposed to standard goodness-of-fit evaluations). This requires decomposing the available dataset into estimation / validation sub-samples.
- Reliably quantifying uncertainties is recognized as a primary objective, and the issue of scrutinizing uncertainty estimates is discussed. To this aim, we make use of predictive distributions for extremes, obtained by integrating out parameter uncertainty. Such predictive distributions are standard in a Bayesian context [4] but can also be derived in a frequentist context [5].

- Reliability indices are derived in order to compare the performances of competing methods on an objective basis.

In a second step, this framework is used to perform a thorough comparison between approaches currently used in France for extreme prediction. The comparison is based on an extensive dataset of long series of rainfall and runoff (about 40-50 years of daily data), available for hundreds of sites over France. Results demonstrate the ability of the comparison framework to distinguish between "good" and "bad" approaches, and yield valuable insights into the optimal ambit of each approach.

## References


[1] F. Garavaglia, M. Lang, E. Paquet, J. Gailhard, R. Garcon, and B. Renard, *Robustness and reliability of a rainfall compound distribution model based on weather pattern sub-sampling*, Hydrology and Earth System Sciences, 15, 519-532, 2011.

[2] T. Micevski, S. W. Franks, and G. Kuczera, *Multidecadal variability in coastal eastern Australian flood data*, J. Hydrol., 327(1-2), 219-225, 2006.

[3] P. Arnaud, and J. Lavabre, *Using a stochastic model for generating hourly hyetographs to study extreme rainfalls*, Hydrol. Sci. J., 44(3), 433-446, 1999.

[4] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian data analysis*, 2 ed., 696 pp., Chapman & Hall, 2004.

[5] I. R. Harris, *Predictive Fit for Natural Exponential-Families*, Biometrika, 76(4), 675-684, 1989.


# Tail estimation methods for the number of false positives in high-throughput testing


Holger Rootzén – *Chalmers, Sweden*
Email: hrootzen@chalmers.se
Co-author: Dmitrii Zholud


## Abstract


This paper develops methods to handle false rejections in high-throughput screening experiments. The setting is very highly multiple testing problems where testing is done at extreme significance levels and with low degrees of freedom, and where the true null distribution may differ from the theoretical one. We show that the conditional distribution of the number of false positives, given that there is in all $r$ positives, approximately has a binomial distribution, and develop efficient and accurate methods to estimate its success probability parameter. Furthermore we provide efficient and accurate methods for estimation of the true null distribution resulting from a preprocessing method, and techniques to compare it with the theoretical null distribution. Extreme Value Statistics provides the natural analysis tools, a simple polynomial model for the tail of the distribution of p-values. We provide asymptotics which motivate this model, exhibit properties of estimators of the parameters of the model, and point to model checking tools, both for independent data and for dependent data. The methods are tried out on two large scale genomic studies and on an fMRI brain scan experiment. A software implementation, SmartTail, may be downloaded from the web.


## References


[1] K. F. Kerr. Comments on the analysis of unbalanced microarray data. *Bioinformatics* **25**(16):20352041, 2009.

[2] T. A. Knijnenburg, L. F. A. Wessels, J. T. M. Reinders, and I. Shmulevich. Fewer permutations, more accurate P-values. *Bioinformatics* **25** (12):161168, 2009.


# Ergodic Properties of Stable Random Fields via Group Actions


Parthanil Roy – *Michigan State University, USA*
Email: parthanil.roy@gmail.com
Co-authors: Stilian A. Stoev and Yizao Wang


## Abstract


Ergodic properties of stochastic processes play a very important role in the investigation of their extremal behaviors (e.g., estimation of ruin probability, large deviations, etc.). In this work, we establish characterization results for ergodicity of symmetric $\alpha$–stable (S$\alpha$S) stationary random fields. We first show that the result of Samorodnitsky [1] remains valid in the multiparameter setting, i.e., a stationary S$\alpha$S ($0 < \alpha < 2$) random field is ergodic (or equivalently, weakly mixing) if and only if it is generated by a null group action. We also give a criterion for ergodicity of these random fields which is valid for all dimensions and new even in the one–dimensional case. The similarity of the spectral representations for sum– and max–stable random fields yields parallel characterization results in the max–stable setting.

## References

[1] G. Samorodnitsky. *Null flows, positive flows and the structure of stationary symmetric stable processes*, Annals of Probability, 33:1782–1803, 2005.

# Estimating an endpoint using high order moments

Gilles Stupfler – *Université de Strasbourg, France*
Email: stupfler@math.unistra.fr
Co-authors: Stéphane Girard and Armelle Guillou

## Abstract

In 2008, Girard and Jacob [1] presented a new method to estimate the frontier of a multidimensional sample. This method relies on a kernel regression on high order moments of the data, and was further used by the same authors in [2] to develop local polynomial estimators of a frontier. Using this approach, we present a new way to estimate the endpoint of a unidimensional sample when the distribution function belongs to the Weibull domain of attraction. The estimator is based on computing high order empirical moments of the variable of interest. Contrary to most popular methods, this one is not threshold-based. It is assumed that the order of the moments goes to infinity, and provided a second-order assumption holds, we give conditions on its rate of divergence to get the asymptotic normality of the estimator. The good performance of the estimator is illustrated on some finite sample situations.

## References

[1] Girard, S., Jacob, P. (2008). Frontier estimation via kernel regression on high power-transformed data, *Journal of Multivariate Analysis*, **99**(3):403–420.

[2] Girard, S., Jacob, P. (2009). Frontier estimation with local polynomials and high power-transformed data, *Journal of Multivariate Analysis*, **100**(8):1691–1705.

# Ionosphere Severe Storms and Occurrence Risk Estimation

Norbert Suard – *CNES, Toulouse, France*
Email: Norbert.Suard@cnes.fr
Co-authors: Jean-Marc Azaïs, Sébastien Gadat and Dominique Debailleux

## Abstract

Even if a lot of analyses, often based on last solar cycle data have been done to preserve the integrity of navigation systems such as Satellite Based Augmentation System (SBAS) or Ground Based Augmentation System (GBAS) in case of ionosphere severe storm threat, the in place integrity preservation means have often a drawback in the continuity availability domain. In order to be able to assess the continuity (and availability) risk that can be caused by such ionosphere severe storm, CNES has on one hand, contracted a study led by TASF and supported by IMT (Institute of Mathematics in Toulouse) to assess the occurrence risk of severe/extreme ionosphere storms taking into account the solar cycle profile and based on innovative statistical methods and has on an other hand, established a classical risk by using a min, max, average approach along a solar cycle based on several solar cycle information. In addition to that and in order to support different other activities, CNES has already analysed different historical data sets of solar activity indices and the resulting earth geomagnetism/ionosphere activities over a period covering several solar cycles. This includes the SSN, 10.7cm flux, Dst, Kp and Ap indexes, going back to as far as 1932 for those available, and an assessment of the relationships between these different indexes. Looking into the future, several predictions of the new solar cycle activity are also available, for example from the Marshall Space Flight Centre. Among all the analyses indices, the Ap index (daily or 3 hours), measuring the earth geomagnetism/ionosphere activities and the monthly SSN measuring the solar activity have been used in this ionosphere storm risk occurrence assessment work. More in details, the innovative statistical methods were based on rare event theory, firstly by a global and simple approach over the 80 years, considering a stationary system and resulting in an average risk. This approach is de facto not satisfactory due the evident dependence between geomagnetic/ionosphere activity and solar one that is a cyclic system. So, secondly a different approach, the Cox Model (Proportional Hazard Model with Time-Dependent Covariates), better known in the epidemiology and medical treatment domain has been experienced. Basically, the Cox model provides an estimate of the treatment effect on survival after adjustment for other explanatory variables. It allows to estimate the hazard (or risk) of death, or other event of interest, for individuals, given their prognostic variables. Our interest was driven by the proportional hazards assumption. this assumption is that covariates multiply hazard. In other terms, the hazard function for an observation depends on the values of the covariates and the value of the baseline hazard. Given two observations with particular values for the covariates, the ratio of the estimated hazards over time will be constant - hence the name of the method: the proportional hazard model as in the Cox model [1]. This gives an example on non-stationary extreme model

that is different from that of [2]. The validity of this assumption may often be questionable. Nevertheless, the covariate influence create a relative risk that is constant in regards to the reference and this risk can be expressed exponentially as a function of a linear covariates combination. In our approach one of these covariates is a solar activity index. The baseline hazard is estimated by a kernel-smoothing estimator and the coefficients of the covariate by maximum likelihood. After having determine the different parameters of our Cox model and having at disposal some predictions for the new solar cycle, it is then possible to extrapolate the ionosphere severe storm occurrence risk over the new solar cycle and to introduce it in a system continuity risk assessment. The classical method was just based for each solar cycle on a counting year per year of the number of days in regards to a daily Ap ionosphere activity ranking and then to determine the min, max, average values year per year having at our disposal seven complete solar cycles in the analysed dataset.

The paper will present a brief overview of relations between solar events, earth geomagnetism and ionosphere status the data sets that were used in our analyses the different approach to determine an ionosphere severe storm risk of occurrence a focus on our Cox model with its current description, the estimation process and questions that are still open for future work and the already achieved results and comparison between them.

## References


[1] O. Aalen, O. Borgan, H. Gjessing, *Survival and event history analysis*, Springer, 2008

[2] G. Toulemonde , A.Guillou , P. Naveau , M. Vrac, and F. Chevallier. Autoregressive models for maxima and their applications to CH4, *Environmetrics* 21,2,189-207. (2009),


# A Bayesian analysis of extreme precipitation in Mediterranean France using non-stationary GEV models


Xun Sun – *Cemagref Lyon, France*
Email: xun.sun@cemagref.fr


## Abstract


Extreme value theory becomes increasingly important for hydrologists in order to assess the hazard related to extreme rainfall and flooding. The annual maximum daily precipitation at one site is often modelled with a GEV distribution[1].
The objective of the present study is to describe the extreme rainfall in French Mediterranean region with different covariate models[2]. 92 precipitation gauges within this region are utilized. Based on different covariates (like time, weather type and climate indices) and regression models (e.g. linearity on location parameter, scale parameter or both), several non-stationary GEV models are constructed. The Bayesian approach and the Markov chain Monte Carlo (MCMC) methods are used to infer the posterior parameter distributions[3], to assess uncertainties and to examine the performance of different probability models. Moreover, the predictive distributions are computed to make predictions on future observations.
Due to the limited information content of at-site data, the identification of at-site non-stationary models becomes challenging. The construction of a general modelling framework for regional non-stationary models will thus be discussed.


## References


[1] Coles, S.; Pericchi, L. R. and Sisson, S. *A fully probabilistic approach to extreme rainfall modeling*, Journal of Hydrology, 273, 35 - 50, 2003

[2] Katz, R. W., Parlange, M. B. and Naveau, P. *Statistics of extremes in hydrology*, Advances in Water Resources, 25, 1287 - 1304 , 2002

[3] Renard, B.; Garreta, V. and Lang, M. *An application of Bayesian analysis and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima* Water Resources Research, Amer Geophysical Union, 42, W12422, 2006


# Extremes of locally self-similar Gaussian processes


Kamil Tabiś – *University of Wrocław, Poland*
Email: kamil.Tabis@math.uni.wroc.pl
Co-author: Krzysztof Dębicki


## Abstract


*Pickands' double-sum* method allows us to obtain exact asymptotics for the supremum distribution of Gaussian processes ([1], [2]). The classical application of this method is based on the use of stationary or locally stationary structure of the analyzed process $X(\cdot)$, i.e. the assumption that

$$\mathbf{Var}(X(t) - X(s)) \;=\; \text{Const} \cdot |t-s|^{\alpha}(1+o(1)) \qquad \text{as} \quad t,s \to t^{*}\,,$$

for some $\alpha \in (0,2]$, where $t^*$ is a point in which variance function of $X(\cdot)$ attains its maximum.

In the talk we focus on the exact asymptotics of supremum distribution of Gaussian processes such that

$$\mathbf{Var}(X(t) - X(s)) \;=\; \text{Const} \cdot \mathbf{Var}\,(Y(t) - Y(s))\,(1 + o(1)) \qquad \text{as} \quad t, s \to t^* \;,$$

where $Y(\cdot)$ is a self-similar Gaussian process (with not necessarily stationary increments).

We will point out difficulties of the application of the double-sum method to our problem and we present a new approach, that allows us to determine the exact asymptotics of $\mathbf{P}\left(\sup_{t\in[0,T]} X(t) > u\right)$ as $u \to \infty$. In the obtained asymptotics there appear new analogous of Pickands' constants with interesting properties. The theory will be illustrated by some examples.

## References


[1] J. Pickands III, (1969) *Upcrossing probabilities for stationary Gaussian processes*, *Trans. Amer. Math. Soc.* **145**, 51–73.

[2] V. I. Piterbarg, *Asymptotic methods in the theory of Gaussian processes and fields*, Translations of Mathematical Monographs 148, AMS, Providence, 1996.

[3] K. Dębicki, K. Tabiś, (2011) *Extremes of the time-average of stationary Gaussian processes*, In revision.


# Copula, EVT and DCC-MGARCH: an application to portfolio optimization


Tahar Tayachi – *University of Monastir , Tunisia*
Email: tahar_tayachi@yahoo.fr
Co-author: Fathi Abid


## Abstract


Portfolio selection based on Mean-Variance approach used financial data, generally characterized by stylized facts: departure from normality, asymmetry, volatility clustering and fat tails. In a risk management framework, an investor can underestimates risk and results become inefficacy. Evolution of statistics and computer sciences can improve portfolio selection process by taking into account these stylized facts. We propose, in this paper, an alternative to Markowitz's portfolio selection by introducing recent technologies recommended by Basle II. First DCC-MGARCH is used to time series and second Extreme Value Theory to assess fat tails. Third, Copula theory is applied to measure non linear dependence between extreme risks and finally, Monte Carlo simulation is used to simulate returns in order to minimize CVaR of a portfolio of financial assets. We applied this approach to a portfolio of three assets and results, compared to classical model, confirm the necessity of recent technologies to risk management.


## References


[1] Zong-Run Wang, Xiao-Hong Chen, Yan-Bo Jin and Yan-Ju Zhou, *Estimating risk of foreign exchange portfolio: Using VaR and CVaR based on GARCH-EVT-Copula model* , Physica A: Statistical Mechanics and its Applications, Volume 389, Issue 21, 1 November 2010, Pages 4918-4928.


# Probabilistic wind gust forecasting using non-homogeneous Gaussian regression


Thordis L. Thorarinsdottir – *Heidelberg University, Germany*
Email: thordis@uni-heidelberg.de
Co-author: Matthew S. Johnson


## Abstract


Skillful probabilistic forecasts for wind speed and wind gust are called for both in specialized weather related risk estimation and in every-day situations in aviation, agriculture, and energy production. Probabilistic forecasts are usually obtained with statistical post-processing of an ensemble and recently, an ensemble post-processing method for wind speed based on non-homogeneous Gaussian regression (NGR) has been proposed [1]. However, this method does not apply directly to wind gust as gust speed forecasts are not a standard output from numerical weather prediction models. We propose a solution to this, where we combine the NGR probabilistic forecasts for wind speed with a gust factor to obtain a probabilistic forecast for wind gust at a very low additional computational cost. We apply our framework to 48-hour ahead forecasts of wind speed over the North American Pacific Northwest obtained from the University of Washington mesoscale ensemble. The resulting density forecasts for gust speed are calibrated and sharp and offer substantial improvement in predictive performance over the raw ensemble multiplied with a gust factor or climatological reference forecasts.

## References

[1] T. L. Thorarinsdottir and T. Gneiting, Probabilistic forecasts of wind speed: Ensemble model output statistics using heteroskedastic censored regression, *Journal of the Royal Statistical Society, Series A*, **173**, pp. 371-388, 1999.


# A characterization of extreme precipitation in the Mediterranean region


Andrea Toreti – *Department of Geography, Climatology, Climate Dynamics and Climate Change, Justus-Liebig University of Giessen, Giessen, Germany and Institute of Geography, Climatology and Meteorology, University of Bern, Bern, Switzerland*
Email: andrea.toreti@giub.unibe.ch
Co-authors: E. Xoplaki and J. Luterbacher



## Abstract

The Mediterranean region has been considered a hot-spot of climate change [1] that is (and will be more) vulnerable and exposed, especially to climate extreme events, such as heat waves and flood events. Therefore, a characterization of extreme precipitation by using observations and global/regional climate models is essential. In this frame, a set of more than 400 daily precipitation series (mostly covering the period 1950-2006) has been collected and quality checked. Then, extended winter (October to March) precipitations have been analyzed by applying a procedure based on a declustered Peak Over Threshold approach and a recently developed minimum density power divergence estimator [2]. Results reveal remarkable spatial differences, in terms of the parameters of the extreme distributions. Besides observations, a new set of climate runs, carried out in the framework of the EU-FP6 'Climate Change and Impact Research: the Mediterranean Environment' CIRCE pr oject by using innovative high resolution global/regional climate models, has been analyzed.



## References

[1] Giorgi, F. 2006. Climate change hot-spots, *Geophys. Res. Lett.*, 33, L08707, doi:10.1029/2006GL025734.

[2] Juárez, S. F., and Schucany, W. R. 2004. Robust and efficient estimation for the Generalized Pareto Distribution, *Extremes*, 7, 237-251.


# Multivariate tail probability representations


Jennifer L. Wadsworth – *Lancaster University, UK*
Email: j.wadsworth@lancaster.ac.uk
Co-author: Jonathan A. Tawn



## Abstract

Existing theory for multivariate extreme values focusses upon characterisations of the distributional tails when all components of a random vector grow at the same rate (for example as presented in [1], [2]). In this talk we consider the effect of allowing the components to grow at different rates, and characterise the link between these marginal growth rates and the multivariate tail probability decay rate. The approach leads to some theoretical results which, for asymptotically independent random vectors, mirror the rich variety of descriptions of max-stable dependence which are available through the exponent measure. In addition we describe a simple inferential approach to joint survivor probability estimation. The key feature of the methodology is that extreme set probabilities can be estimated by extrapolating upon rays emanating from the origin when the margins of the variables are exponential. This offers an appreciable improvement over existing techniques where extrapolation in exponential margins is upon lines parallel to the diagonal.



## References

[1] L. de Haan and J. de Ronde, *Sea and wind: Multivariate extremes at work*, Extremes 1(1) pp.7–45, 1998.

[2] A. W. Ledford and J. A. Tawn, *Modelling dependence within joint tail regions*, J. Roy. Stat. Soc. B 59 pp.475–499, 1997.


# Conditional sampling for spectrally discrete max-stable random fields


Yizao Wang – *University of Michigan, Ann Arbor, USA*
Email: yizwang@umich.edu
Co-author: Stilian A. Stoev

## Abstract

Max-stable random fields play a central role in modeling extreme value phenomena. We will present an explicit formula for the conditional probability in general max-linear models, which include a large class of max-stable random fields. As a consequence, we obtain an algorithm for efficient and exact sampling from the conditional distributions. Our method provides a computational solution to the prediction problem for spectrally discrete max-stable random fields. This work offers new tools and a new perspective to many statistical inference problems for spatial extremes, arising in many applications.

## References

[1] Y. Wang and S. A. Stoev. Conditional sampling for spectrally discrete max-stable random fields. To appear in *Adv. in Appl. Probab.*, 2011. http://arxiv.org/abs/1005.0312

# Empirical likelihood based confidence regions for first order parameters of heavy-tailed distributions

Rym Worms – *LAMA (CNRS UMR 8050), University Paris-Est-Créteil, France*
Email: rym.worms@u-pec.fr
Co-author: Julien Worms

## Abstract

Consider $X_1, \ldots, X_n$ some i.i.d. observations from a heavy tailed distribution $F$, i.e. such that the common distribution of the excesses $Y_1, \ldots, Y_{N_n}$ over a high threshold $u_n$ can be approximated by a Generalized Pareto Distribution $G_{\gamma_0,\sigma_{0n}}$ with $\gamma_0 > 0$. Well-known estimators for the couple $(\gamma_0, \sigma_{0n})$ are the maximum likelihood (ML) estimators of Smith and the probability weighted moments estimators of Hosking & Wallis. However the performance of the confidence regions (for the couple) which are issued from the asymptotic normality of these estimators has never been investigated.

In this work an alternative method is considered, namely the empirical likelihood (EL) method. An empirical likelihood ratio (ELR below) is constructed, starting from the less-known estimating equations proposed by Zhang in [3], which are close to the classical likelihood equations but possess good computational properties. By proceeding so, classical advantages of the EL methodology are expected, namely the avoidance of covariance estimation and the possibility of profiling the likelihood for estimating $\gamma_0$ alone.

The estimating equations of Zhang are based on the following function $g$ (where the constant $r < 0$ is arbitrary)

$$g(y, \gamma, \sigma) := \begin{pmatrix} \log(1 + \gamma y/\sigma) - \gamma \\ (1 + \gamma y/\sigma)^{r/\gamma} - \frac{1}{1-r} \end{pmatrix}.$$

which, when $Z$ is $G_{\gamma,\sigma}$-distributed, satisfies $E[g(Z, \gamma, \sigma)] = 0$. Reminding that the excesses over the high threshold $u_n$ are denoted by $Y_1, \ldots, Y_{N_n}$ (where $N_n$ is therefore random and binomial distributed), the estimators $(\hat{\gamma}, \hat{\sigma})$ resolving in $(\gamma, \sigma)$

$$\frac{1}{N_n} \sum_{i=1}^{N_n} g(Y_i, \gamma, \sigma) = 0$$

are shown to be always well-defined and computable, and therefore to maximize the empirical likelihood ratio

$$ELR(\gamma, \sigma) := \sup \left\{ \prod_{i=1}^{N_n} (N_n p_i) \;\Big/\; \forall i, p_i \geq 0 \;,\; \sum_{i=1}^{N_n} p_i = 1 \;\text{ and }\; \sum_{i=1}^{N_n} p_i g(Y_i, \gamma, \sigma) = 0 \right\}$$

If $c_{k,\alpha}$ denotes the $(1-\alpha)$-quantile of the $\chi^2(k)$ distribution, then a consequence of our asymptotic results is that the Wilks-type region

$$\mathcal{R} \;=\; \{\; (\gamma, \sigma) \;;\; -2 \log ELR(\gamma, \sigma) \leq c_{2,\alpha} \;\}$$

defines a confidence region for $(\gamma_0, \sigma_{0n})$ with asymptotically correct level $1 - \alpha$, and

$$\mathcal{I} \;=\; \{\; \gamma \;;\; -2 \log ELR(\gamma, \hat{\sigma}_\gamma) \leq c_{1,\alpha} \;\}$$

defines a confidence interval for $\gamma_0$ with asymptotically correct level $1 - \alpha$, where $\hat{\sigma}_\gamma$ designs the maximizer, in $\sigma$, of $ELR(\gamma, \sigma)$. Proof of these statements are based on the methodology developed in [2], with difficulties arising from the fact that the distribution of the excesses $Y_i$ depend on the sample size.

Simulations results will be presented, in order to show the performance, in terms of coverage probability: $(i)$ of the confidence regions $\mathcal{R}$ w.r.t. those derived from the gaussian approximation of the ML estimators ; $(ii)$ of the confidence intervals $\mathcal{I}$ for $\gamma$ w.r.t. those proposed by Lu & Peng in [1] by the EL methodology, and connected to the Hill estimator.

## References

[1] J.C. Lu and L. Peng (2002), *Likelihood based confidence intervals for the tail index*, *Extremes* 5, pages 337-352.

[2] Y. S. Qin and J. Lawless (1994), *Empirical likelihood and general estimating equations*, *Annals of Statistics* 22(1), pages 300-325.

[3] J. Zhang (2007), *Likelihood moment estimation for the generalized Pareto distribution*, *Australian & New Zealand J. Stat* 49(1), pages 69-77.


# Systematic risk under extremely adverse market conditions


Chen Zhou – *De Nederlandsche Bank and Erasmus University Rotterdam*
Email: zhou@ese.eur.nl
Co-author: Maarten van Oordt



## Abstract

Extreme losses are the major concern in risk management. The dependence between financial assets and the market portfolio changes under extremely adverse market conditions. We develop a measure of systematic tail risk, the *tail regression beta*, defined by an asset's sensitivity to large negative market shocks, and establish the estimation methodology.

Building on extreme value theory, the estimator of the tail regression beta consists of the asymptotic dependence measure and the marginal risk measures. Theoretically, it has a similar structure as the estimator of the regular beta from regression analysis. Simulations show that our estimation methodology yields an estimator that has a lower mean squared error than performing regressions in the tail.

Empirical results based on analyzing 46 industrial portfolios demonstrate that the regular portfolio sensitivity to the systematic risk is in general different from the sensitivity to systematic risk in severe market downturns. Furthermore, the tail regression beta is a useful instrument in both portfolio risk management and systemic risk management. We demonstrate its applications in analyzing Value-at-Risk (VaR) and Conditional Value-at-Risk (CoVaR).

# Index