The Book of Abstracts

For the 10th International Conference

on

Extreme Value Analysis

Delft University of Technology, The Netherlands
June 26-30, 2017

Session C11
June 27, 11:30
Room: Data

On the relative approximation error of extreme quantiles by the block maxima method

Anne Dutfoy
EDF R&D Département Management des Risques Industriels (France)

Stéphane Girard
Inria Grenoble Rhône-Alpes & Laboratoire Jean Kuntzmann (France)

Clément Albert
Inria Grenoble Rhône-Alpes & Laboratoire Jean Kuntzmann (France)

This study takes place in the context of extreme quantiles estimation by the block maxima method. We investigate the behaviour of the relative approximation error of a quantile estimator dedicated to the Gumbel maximum domain of attraction. Our work is based on a regular variation assumption on the first derivative of the logarithm of the inverse cumulative hazard rate function, introduced by de Valk (2016) [Approximation of high quantiles from intermediate quantiles. *Extremes* **19**(4), 661–686].

Let us denote by $X_{m,m}$ the maximum of m iid observations from a distribution function F, where m is referred to as the block size. We focus on an extreme quantile associated with F defined by $x_{p_m} = F^{-1}(1-p_m) = H^{-1}(-\log p_m)$, where H is the cumulative hazard rate function and $p_m = m^{-\tau_m}$ with $\tau_m \geq 1$. Assume that $(X_{m,m} - b_m)/a_m$ converges to a Gumbel distribution for some normalizing constants $a_m > 0$ and $b_m \in \mathbb{R}$. The approximation $\tilde{x}_{p_m}$ of x_{p_m} by the block maxima method is given by

$$\tilde{x}_{p_m} = b_m - a_m \log(m p_m)$$

and the associated relative approximation error is

$$\epsilon_{app_m} = (x_{p_m} - \tilde{x}_{p_m})/x_{p_m}.$$

Our main result is:

$$\epsilon_{app_m} \overset{m\to+\infty}{\longrightarrow} 0 \iff (\tau_m - 1)^2 K_2(\log m) \overset{m\to+\infty}{\longrightarrow} 0,$$

where $K_2(t) = t^2 (H^{-1})''(t)/H^{-1}(t)$, $t > 0$. This result exhibits three families of distributions according to the limit of K_2 which can be either zero, a constant or infinite. We also provide a first order approximation of the relative approximation error when the latter converges towards zero. Our results are illustrated on simulated data.

Key Words: extreme quantiles estimation, relative approximation error, asymptotic properties, regular variation.

Modeling on maximum daily rainfall in Thailand

Session C2
June 26, 13:30
Room: Chip

Piyapatr Busababodhin
Mahasarakham University (Thailand)

Pimpan Amphanthong
Rajamangala University of Technology Suvarnabhumi (Thailand)

Attempts to use extreme value theory based on univariate and bivariate due to maximum daily rainfall data in Thailand. We analyze annual maxima of daily rainfall for the years 1989 - 2016 which are modeled for 99 meteorological stations that were set on 6 regions; Northern, Northeastern, Center, Western, Eastern, and Southern, in Thailand. Both univariate and bivariate extreme value distributions gave models which fitted to data from each station to describe the extremes of rainfall and to predict its future behavior. The univariate model indicated Trat weather station, in the Eastern region, had highest return level. Furthermore, the bivariate models led to lower estimates of tail probability, except for extreme cases, but taking statistical uncertainty into account the two models lead to qualitatively similar results.

Key Words: extreme value theory, univariate extreme value distribution, bivariate extreme value distribution, return level, copula.

Spectral properties of high-dimensional time series

Session I4
June 27, 9:40
Room: Boole

Haoyang Liu
UC Berkeley (USA)

Debashis Pau
UC Davis (USA)

Alexander Aue
UC Davis (USA)

This talk discusses some of the recent results connecting high-dimensional time series methodology with random matrix theory. In particular, results for the distribution of the eigenvalues of the covariance matrix and symmetrized autocovariance matrices of linear processes are presented in the asymptotic regime of proportionally growing dimension and sample size. This limiting spectral distribution is characterized through a nonlinear integral equation for its Stieltjes transform. Applications of the large-sample results are for example in designing estimation procedures for spectra of coefficient matrices in high-dimensional time series. These may then be used as the basic building block for a new algorithm to aid portfolio optimization, resulting in a correction of finite sample bias known to occur in the estimation of the mean-variance frontier in high-dimensions.

Key Words: Marčenko–Pastur law, mean-variance frontier, portfolio optimization, Stieltjes transform.

Session C18
June 27, 16:00
Room: Chip

On heavy-tailed crack distribution for loss severity modelling

Jingjiao Chen
University of Regina (Canada)

Taehan Bae
University of Regina (Canada)

The three-parameter crack lifetime distribution family which includes the inverse Gaussian, length biased inverse Gaussian and Birnbaum-Saunders distributions as its members, has recently been extended to a larger class of generalized crack distributions. A review of basic distributional properties and the tail relationships between the auxiliary distribution and the resulting generalized crack distribution are discussed in the context of classical extreme value theory. A few concrete examples of heavy-tailed crack distribution are constructed and used for model fitting to some real catastrophic loss data sets. The fitting results show that the heavy-tailed crack distribution with an appropriate choice of auxiliary density function outperforms some commonly used parametric models.

Key Words: Birnbaum-Saunders distribution, generalized Birnbaum-Saunders distribution, three parameter crack distribution, extreme value theory, regular variation.

Session C31
June 29, 11:30
Room: Data

Extremes of Gaussian chaos processes with trend

Krzysztof Dębicki
University of Wrocław (Poland)

Enkelejd Hashorva
University of Lausanne (Switzerland)

Long Bai
University of Lausanne (Switzerland)

Let $\boldsymbol{X}(t) = (X_1(t), \ldots, X_d(t))$ be a Gaussian vector process and $g(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^d$ a homogenous function. In this paper we are concerned with the exact tail asymptotics of the chaos process $g(\boldsymbol{X}(t))$ with trend over $[0, S]$. Both scenarios that $\boldsymbol{X}(t)$ is locally stationary and non-stationary are considered. Important examples include $\Pi_{i=1}^d X_i(t) - ct$ and chi-processes with trend, i.e., $\left(\sum_{i=1}^d b_i X_i^2(t)\right) - ct$.

Key Words: Gaussian chaos, Gaussian vector processes, asymptotic methods, Pickands constant.

Hidden regular variation and limit shape

Session C36
June 29, 16:00
Room: Chip

Natalia Nolde
University of British Columbia (Canada)

Guus Balkema
University of Amsterdam (The Netherlands)

Samples from vectors (X, Y) with light tails often have a well-defined shape. There may exist constants such that the scaled samples converge in probability to a compact set. The limit set S is star-shaped. If the coordinate-wise maximum of the points in S does not lie in S the components X and Y are tail-independent.

Tail-dependence of two heavy-tailed positive variables U and V may be obliterated by the contamination of a distribution with independent components. Is it possible to determine the hidden regular variation by looking at the limit shape of the light-tailed log transform $(X, Y) = (\log U, \log V)$?

Limit laws for a class of multidimensional record values

Session C28
June 28, 11:00
Room: Pi

Margarida Brito
CMUP & FC University of Porto (Portugal)

Marie-Françoise Barme-Delcroix
Lab. Painlevé, University of Lille (France)

We consider the record values of a multidimensional sequence of random variables and study their asymptotic properties. Ordering multivariate data can be done in many different ways and, accordingly, different notions of record values of multivariate observations can be introduced. We develop here the concept of multivariate record values introduced in Barme-Delcroix (2013) [Multivariate extremes: A conditional quantile approach. In: C. Becker et al. (eds), *Robustness and Complex Data Structures: Festschrift in Honour of Ursula Gather*, Springer-Verlag, Berlin Heidelberg] based on the ordering of a multivariate sample through an increasing family of conditional quantile surfaces, called isobars. The order statistics defined in this way may be characterized by a unidimensional approach, see e.g. Barme-Delcroix and Brito (2001) [Multivariate stability and strong limiting behaviour of intermediate statistics. *Journal of Multivariate Analysis* **79**, 157–170] and Barme-Delcroix and Gather (2007) [Limit laws for multidimensional extremes. *Statistics and Probability Letters* **77**, 1750–1755].

The form of the possible limiting record value distributions is given explicitly. The resulting probabilistic properties and the classical unidimensional framework introduced in Resnick (1973) [Limit laws for record values. *Stochastic Processes and Their Applications* **1**, 67–82] provide domain of attraction criteria for the three types of limit laws. Characterizations are derived in term of the distribution function for such record values. Some illustrative examples are also provided.

Key Words: multivariate records, limit laws, domains of attraction, isobars.

Session C37
June 29, 16:00
Room: Data

On stationary regularly varying random fields

Hrvoje Planinić
University of Zagreb (Croatia)

Bojan Basrak
University of Zagreb (Croatia)

We study a stationary and regularly varying stochastic processes on a rectangular section of d-dimensional integer lattice under certain dependence conditions. We analyze behavior of extremes in such a random field when the size of the rectangle tends to infinity. After describing the limiting distribution of individual clusters of extremes and corresponding notion of extremal index, we present a point process convergence theorem in this context. We discuss applications of our results to problems in theory of random matrices and other fields.

The talk is based on an extension of the earlier joint work with Hrvoje Planinić and Philippe Soulier (2016) [An invariance principle for sums and record times of regularly varying stationary sequences. *arXiv:1609.00687*].

Key Words: regular variation, point processes, limit theorems, integer lattice.

Session C7
June 26, 16:30
Room: Data

Sample path large deviations for random walks and Lévy processes with Weibullian tails

Chang-Han Rhee, Bert Zwart
Centrum Wiskunde & Informatica (The Netherlands)

Mihail Bazhba
Centrum Wiskunde & Informatica (The Netherlands)

While sample path large deviations with light tailed increments can be obtained due to the existence of the moment generating function, the number of similar results for heavy tails are scarce. In this paper, we consider sample path large deviations for a Lévy process with heavy tailed semi-exponential increments. We develop a proof approach based on an appropriate representation, a suitable normalization,

Bryc's inverse lemma, and the use of concentration inequalities. Our result yields an LDP for Weibullian tails in the Shorokhod space with the $J1$ topology, improving on a result in the $L2$ topology due to Gantert (1998) [Functional Erdős-Renyi laws for semiexponential random variables. *The Annals of Probability* **26**(3), 1356–1369]. We obtain a similar result for random walks and outline several applications of this result.

Key Words: sample path large deviations, Lévy processes, heavy tails, semiexponential increments.

Bias reduced tail modelling and modelling a full data set

Session C29
June 29, 11:30
Room: Boole

Gao Maribe, Andrehette Verster
University of the Free State (South Africa)

Jan Beirlant
KU Leuven (Belgium) & University of the Free State (South Africa)

In recent years several attempts have been made to model both the modal and tail part of the data. Frigessi et al. (2002) [A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes* **5**, 219–235] proposed dynamic mixtures of two components with a weight function $\pi = \pi(x)$ smoothly connecting the bulk and the tail of the distribution. Several other papers have then provided other mixture models. Recently, Naveau et al. (2016) [Modeling jointly low, moderate and heavy rainfall intensities without a threshold selection. *Water Resources Research* **52**, 2753–2769] propose a nice review on this topic, and, continuing on the work by Papastathopoulos and Tawn (2013) [Extended generalized Pareto models for tail estimation. *Journal of Statistical Planning and Inference* **143**, 131–143], propose a statistical model which is in compliance with extreme value theory and allows for a smooth transition between the modal and tail part using the inverse probability integral transform $\sigma H^{-1}(G^{-1}(U))$ where H denotes the distribution function of the generalized Pareto distribution and G is an appropriate distribution function that preserves the tail characteristics, and U a uniform (0,1) random variable.

Incorporating second order rates of convergence for distributions of peaks over thresholds, Beirlant et al. (2002) [Modelling excesses over high thresholds by perturbed generalized Pareto distributions. *Eurandom technical report*] and Beirlant et al. (2009) [Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. *Journal of Statistical Planning and Inference* **139**(8), 2800–2815] constructed models that can be viewed as special cases from both approaches discussed above. When fitting such second order models it turns out that the bias of the resulting extreme value estimators is significantly reduced compared to the fit with one Pareto component. Such mixture models also provide an appropriate tail fit over a much larger range of data than simple tail models. Recently, Beirlant et al. (2017) [Using shrinkage estimators to reduce bias and MSE in estimation of heavy tails. *Revstat*, to appear] showed that using penalized likelihood methods on the

weight parameter π one can obtain good bias and mean squared error properties for tail estimators, while modelling modal and tail parts of the data. We consider such solutions in the simple i.i.d. case and under random censoring.

Key Words: bias reduction, mixture models, shrinkage estimators, penalized likelihood.

Session C1
June 26, 13:30
Room: Boole

Extremal attractors of Liouville copulas

Johanna G. Nešlehová
McGill University (Canada)

Léo Belzile
École Polytechnique Fédérale de Lausanne (Switzerland)

The Liouville copulas introduced in McNeil and Nešlehová (2010) [From Archimedean to Liouville copulas. *Journal of Multivariate Analysis* **101**, 1772–1790] are asymmetric generalizations of the ubiquitous Archimedean copula class. They are the dependence structures of scale mixtures of Dirichlet distributions, also called Liouville distributions. The limiting extreme-value attractors of Liouville copulas and of their survival counterparts are derived. These max-stable models, termed scaled extremal Dirichlet, are new and encompass several existing classes of multivariate max-stable distributions, including the logistic, negative logistic and extremal Dirichlet. The stable tail dependence function and the angular density of the scaled extremal Dirichlet model have a tractable form, which in turn leads to a simple de Haan representation. The latter is used to design efficient algorithms for unconditional simulations and to derive tractable formulas for maximum-likelihood inference. We illustrate the model on river flow data of the river Isar in southern Germany.

Key Words: Liouville copulas, scaled Dirichlet, multivariate extreme value distribution.

Session C14
June 27, 14:00
Room: Chip

A composite likelihood based approach for max-stable processes using histogram-valued variables

Scott Sisson, Thomas Whitaker
UNSW Sydney (Australia)

Boris Beranger
UNSW Sydney (Australia)

The analysis of Spatial Extremes has been given a growing interest over the last decade in a broad range of areas and especially in climatology with the goal to better

understand the behaviour of events such as floods, heat waves or storms. Max-stable processes are a convenient and widely used tool to model such phenomena. In recent years, composite likelihood methods have appeared to bypass the intractability of the multivariate density function of such processes. However, the computational cost of these methods explodes as the number of temporal observations gets large. This is even more noticeable when working with a large number of spatial locations across a study region. To bypass this issue we introduce a symbolic data analysis (SDA) based approach which consists in aggregating data into histograms leading to a reduction of the complexity of the data. A symbolic version of the composite likelihood approach where observations are multivariate histogram-valued is provided and the classical results from Padoan et al. (2010) [Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association* **105**, 263–277] are shown to be recovered as a limiting case. The performance of our procedure in terms of inferential and computational efficiency is studied in an extensive simulation study and the impact of coarsening the data and the design of the symbols (histograms) is discussed. Finally, the utility of the method is illustrated through the analysis of fortnightly maximum temperatures at 105 locations across Australia using historical data and simulated data from two climate models.

Key Words: spatial extremes, composite likelihood, symbolic data analysis, climate.

On trimming of the Hill estimator: robustness, optimality and adaptivity

Session C16
June 27, 14:30
Room: Pi

Stilian Stoev
University of Michigan (USA)

Shrijita Bhattacharya
University of Michigan (USA)

Motivated by the approach in Brazauskas and Serfling (2000) [Robust estimation of tail parameters for two-parameter Pareto and exponential models via generalized quantile statistics. *Extremes* **3**, 231–349], we focus on the problem of the robust estimation of the tail index of a regularly varying distribution. We propose and study a trimmed version of the Hill estimator introduced in Hill (1975) [A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* **3**, 1163–1174]. We show that the estimator is asymptotically and nearly finite-sample efficient among all unbiased estimators with a given strong upper break-down point. We also develop an automatic, data-driven procedure for the choice of trimming, which results in an adaptive estimator with nearly optimal asymptotic and finite sample properties. As a by-product we also obtain a methodology for identifying extreme outliers in heavy tailed Pareto-like data. The performance of the new estimators is illustrated with simulations and applied to detect DDoS attacks from Internet traffic traces as well as the analysis of stock trading volume time series.

Key Words: robust, break down point, data driven trimming.

Session C13
June 27, 14:00
Room: Boole

Optimal insurance contract under ambiguity: applications in extreme events

Georg Ch. Pflug
University of Vienna (Austria)

Corina Birghila
University of Vienna (Austria)

In case of extreme events, high ambiguity concerning the occurrence and the magnitude of losses increases the difficulty of managing and estimating risk. In order to incorporate robustness into risk quantification, our approach includes the baseline model of losses, as well as all the models that are at a specified ε-distance from the baseline model and can explain the data in a similar way. In this work we propose a modified Wasserstein distance, which more accurately incorporates the differences between models, with an emphasis on the tail of the extreme value distributions. To this end, a more robust technique of tail estimation is required. Based on this metric, we solve a minimax problem consisting of finding an optimal insurance contract designed to protect against extreme losses. This insurance contract is robust against model misspecification. The current approach is a generalization of ideas presented in Pflug et al. (2017) [Incorporating model uncertainty into optimal insurance contract design. *Insurance: Mathematics and Economics*, accepted for publication], regarding the design of optimal insurance contract for extreme events.

Key Words: ambiguity, tail estimation, robustness, insurance.

Session C5
June 26, 16:00
Room: Boole

Controlling the time discretization error

Daan Crommelin, Michel Mandjes
Centrum Wiskunde & Informatica, University of Amsterdam (The Netherlands)

Krzysztof Bisewski
Centrum Wiskunde & Informatica, University of Amsterdam (The Netherlands)

It is often of interest to find the threshold-crossing probability

$$w(b) := \mathbb{P}\Big(\sup_{t\in[0,1]} X_t > b \Big)$$

of a real stochastic process $(X_t)_{t\in[0,1]}$. When an explicit expression for $w(b)$ is unavailable it is estimated numerically using rare event simulation techniques. For most of the available methods the underlying process needs to be discretized in

time, which induces a *relative bias* on the estimator. Typically the *discretization grid* is chosen to be *equidistant*, i.e. $T_n := \{\frac{1}{n}, \frac{2}{n}, \ldots, 1\}$, and in that case the bias can be reduced only by increasing the grid-size n. The finer the grid, the smaller the bias, but also, the larger the computational effort to estimate $w(b)$.

To study the effect of the choice of the grid on the bias we focus on a standard Brownian Motion. It is well known that when b is fixed, the bias induced by an equidistant grid decays proportionally to $n^{-1/2}$. However when n is fixed and $b \to \infty$, the bias tends to its maximal value. We prove that if the discretization is equidistant, then in order to control the bias as b grows large, n has to grow at least quadratically in b. As an alternative to equidistant discretization, we derive an explicit family of grids, with grid-points adaptive in b. For the *adaptive family* the required number of grid-points to control the bias is independent of b, providing a significant computational improvement.

Key Words: discretization error, rare event simulation, hitting time, Brownian motion, optimal discretization.

Optimal transport and risk analysis (OTRA)

Session I5
June 27, 9:00
Room: Chip

Yang Kang, Karthyek Murthy
Columbia University (USA)

Jose Blanchet
Columbia University and Stanford University (USA)

We will discuss several results at the intersection of risk theory, robust optimization, non-parametric statistics and optimal transport. The results that we will present are motivated by questions such as: What is the impact of model error when quantifying risk in terms of statistics of interest? We discuss various robust optimization formulations based on optimal transport and discuss how these formulations, when applied to classical inference problems, retrieve well-known estimators in the machine learning community (such as LASSO and Support Vector Machines). We also discuss the elements of a novel inference approach built on optimal transport. These results provide the foundations for a robust data-driven approach to stress testing.

Key Words: insurance modeling, robust optimization, model risk, nonparametric statistics.

Gradient sampling algorithm for POT

Session C44
June 30, 11:00
Room: Pi

Valérie Chavez–Demoulin
University of Lausanne (Switzerland)

Marc-Olivier Boldi
University of Lausanne (Switzerland)

This paper introduces the use of the gradient sampling algorithm for several cases of likelihood type estimations. In its essence, the gradient sampling algorithm, introduced by Burke et al. (2002) [Approximating subdifferentials by random sampling of gradients. *Mathematics of Operations Research* **27**(3), 567–584], is a descent algorithm where the gradient direction is replaced by a stochastic approximation of the Clarke subdifferential. As a result, the descent algorithm is stabilized and can also be used with concave and non-derivable objective functions. The idea of using the gradient sampling algorithm in descent algorithms can be extended to any optimization method that uses gradient direction. In this paper, we present two of its applications; one is a modification of the algorithm GAMLSS of Rigby and Stasinopoulos (2005) [Generalized additive models for location, scale and shape. *Applied Statistics* **54**(3), 507–554] when applied to a Peaks-Over-Threshold (POT) context of extreme value theory. The other is an additive quantile regression application and an alternative algorithm to Portnoy and Koenker (1997) [The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. *Statistical Science* **12**(4), 279–300] applied in Koenker (2005) [*Quantile Regression*. Cambridge University Press, Cambridge]. From these applications, we see that the gradient sampling brings an appreciable stabilization to the algorithm.

Key Words: peaks-over-threshold, quantile regression, gradient sampling.

Procyclicality of empirical measurements of risk in financial markets

Session C33
June 29, 14:30
Room: Chip

Michel Dacorogna
DEAR Consulting (Switzerland)

Marcel Bräutigam, Marie Kratz
ESSEC CREAR (France)

In this study we empirically explore the capacity of historical VaR to correctly predict the future risk of a financial institution. We observe that rolling samples are better able to capture the dynamics of future risks. We thus introduce another risk measure, the Sample Quantile Process (SQP), which is a generalization of the VaR calculated on a rolling sample, and study its behavior as a predictor by varying its parameters. We analyze various realizations of this process through 11 different stock indices of well developed stock markets. Moreover, we study the behavior of the future risk as a function of past volatility. We show that, in all the cases empirically studied, if the past volatility is low, the historical computation of the risk measure underestimates the future risk, while in period of high volatility, the risk measure overestimates the risk, confirming that the current way financial institutions measure their risk is highly procyclical. This has important consequences for the regulation in times of crisis.

Key Words: risk measure, sample quantile process, value-at-risk, volatility.

Multivariate Cox models and copulas

Session C42
June 30, 11:00
Room: Chip

Boris Beranger, Scott Sisson
UNSW Sydney (Australia)

Michel Broniatowski
Université Pierre and Marie Curie (France)

In a number of areas such as biostatistics and industry, the analysis of risks is of great interest and in particular when focusing on multivariate dependent data. In this work we consider the joint distribution between some random variables to be obtained using the conditioning survival distribution function. The marginal and conditional survival functions are assumed to be dependent on some environmental covariates through a proportional hazard model. Our model is adapted for positive orthant dependence and we prove its mathematical validity under some restrictions. The dependence structure turns out to be the key for our model to hold, no marginal constraints are required. The levels of dependence vary with those of the covariates and their features are captured via the copula function of the multivariate baseline. The class of extreme value copulas is stable under our model and some examples are presented. Finally the use of our model is illustrated in an application to competing risks.

On a pseudo-maximum likelihood estimator for the extremal index

Session I10
June 30, 14:00
Room: Boole

Betina Berghaus
Ruhr-Universität Bochum (Germany)

Axel Bücher
Ruhr-Universität Bochum (Germany)

The extremes of a stationary time series typically occur in clusters. A primary measure for this phenomenon is the extremal index, representing the reciprocal of the expected cluster size. Both a disjoint and a sliding blocks estimator for the extremal index, essentially due to Northrop (2015) [An efficient semiparametric maxima estimator of the extremal index. *Extremes* **18**, 585–603], are analyzed in detail. In contrast to many competitors, the estimators only depend on the choice of one parameter sequence. We derive an asymptotic expansion, prove asymptotic normality and show consistency of an estimator for the asymptotic variance. Explicit calculations in certain models and a finite-sample Monte Carlo simulation study reveal that the sliding blocks estimator outperforms other blocks estimators, and that it is competitive to runs- and inter-exceedance estimators in various models. The methods are applied to a variety of financial time series.

Key Words: clusters of extremes, extremal index, stationary time series, mixing coefficients, block maxima.

Session C20
June 27, 16:00
Room: Pi

Modeling on maximum rainfall and temperature based on extreme value copula analyses

Piyapatr Busababodhin
Mahasarakham University (Thailand)

The modeling of the extremes of a natural process such as rainfall or temperature is important in environmental sciences, so a variety of statistical tools have been used for the modeling of extremes. Especially, how to due to possible dependence between the joint distribution of rainfall and temperature are difficult. The purpose of this article is to investigate the dependence structure between extreme rainfall and temperature, using parametric and nonparametric models of extreme value. The selection and estimation of the copula extreme value theory is based on maximum likelihood estimation (MLE) method, and a behavior of dependence was determined by the dependence function. This work is to examine the maximum daily rainfall and temperature at Kalasin province, a leading agricultural province in the northeast of Thailand, for the period from 1985-2016. The historical climatic data for Kalasin is used to demonstrate the modeling process. The results indicate that for Kalasin province, there are high correlations between rainfall and temperature for the summer and winter seasons. Finally, we believe that the dependence model provided the most realistic picture of the real uncertainties and our models can be integrated with agricultural productions research for studying the effects of climate change on crop yields.

Key Words: bivariate extreme value theory, extreme value copulas, dependence function.

Session C16
June 27, 14:00
Room: Pi

A bias corrected generalized Hill estimator

Ivanilda Cabral
Universidade de Cabo Verde (Cabo Verde)

M. Ivette Gomes
Universidade de Lisboa (Portugal)

Frederico Caeiro
Universidade Nova de Lisboa (Portugal)

The extreme value index (EVI) characterizes the tail behaviour of a distribution, and indicates the size and frequency of certain extreme events. For heavy tails, i.e.

for a positive EVI, a commonly used estimator is the one proposed by Hill (1975) [A simple general approach to inference about the tail of a distribution. *Annals of Statistics* **3**, 1163–1174]. The Hill estimator is the average of the k log-excesses over the $(k+1)$-th upper order statistic. Since this estimator can often have a high bias, the development of new EVI-estimators and bias reduction have recently been important topics of research in statistics of extremes. The Lehmer mean of the k log-excesses over the $(k+1)$-th upper order statistic has been recently considered in the literature for the estimation of a positive EVI. This class of estimators is a generalization of the classic Hill estimator and it has a parameter (the order of the mean) that allows us to eventually achieve a smaller mean squared error. In this work, we are interested in the bias reduction of such EVI estimator.

Key Words: bias reduction, extreme value index, heavy tails, semi-parametric estimation.

On the modelling of the temporal and spatial evolution of extreme hydraulic events

Session I3
June 26, 10:10
Room: Pi

Jacco Groeneweg, Joana van Nieuwkoop
Deltares (Netherlands)

Sofia Caires
Deltares (Netherlands)

A semi-parametric method based on the theory of max-stable processes has been developed for the determination of the spatial and temporal evolution of extreme wind events. These are obtained by lifting observed wind fields and allow the determination of extreme wind-driven hydraulic loads. However, before and after the peak of a storm, when the observed conditions are not extreme the lifting produced by the method is considered to be excessive. Therefore, an adjustment of the method, consisting of the use of a scaling factor when the original conditions are not extreme is proposed and assessed. The assessment is carried out in terms of wave loads along the eastern banks of Lake IJssel and water levels in Hoek van Holland in the Netherlands and by considering other lifting alternatives and other ways to obtain such estimates. Although all the lifting methods considered lead to acceptable results, those of the adjusted method, the Combined Max-Stable lifting method, are judged to be the most reliable.

Key Words: extreme value analysis, time and space evolution of extremes, wind speed, hydraulic loads, assessment of water defences.

Session C42
June 30, 11:30
Room: Chip

Asymptotically distribution-free goodness-of-fit testing for copulas

John H.J. Einmahl
Tilburg University (The Netherlands)

Roger J.A. Laeven
University of Amsterdam (The Netherlands)

Sami Umut Can
University of Amsterdam (The Netherlands)

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be an i.i.d. sample from an unknown d-variate ($d \geq 2$) distribution function F with continuous marginal distributions that belong to some parametric family $\mathscr{F}$. We propose a procedure for constructing asymptotically distribution-free goodness-of-fit tests for the copula C associated with F. The procedure is based on a transformation of a suitable empirical process derived from semi-parametric and parametric estimators of C. The transformed empirical process converges weakly to a standard d-variate Wiener process, paving the way for a multitude of asymptotically distribution-free goodness-of-fit tests. We investigate the finite-sample performance of our approach through a simulation study and data analysis.

Key Words: copula, goodness-of-fit testing, martingale transform.

Session C20
June 27, 16:30
Room: Pi

Local likelihood inference for high dimensional spatial extremes applied to US precipitation data

Raphaël Huser
King Abdullah University of Science and Technology (Saudi Arabia)

Daniela Castro Camilo
King Abdullah University of Science and Technology (Saudi Arabia)

In order to model complex dependence structures in spatial extremes, we propose an approach based on factor copula models. The latter, which can be seen as Gaussian location mixture processes, assume the presence of a common factor affecting the joint dependence of all measurements. When the common factor is exponentially distributed, the resulting copula is asymptotically equivalent to the Husler-Reiss copula; therefore, the so-called exponential factor model is suitable to capture tail dependence. Under the assumption of local stationarity, the exponential factor model is used to model non-stationary extreme measurements over high thresholds. Inference is performed using a censored local likelihood. Performance is assessed using simulation experiments, and illustrated using a daily rainfall dataset.

Key Words: spatial extremes, factor copula models, local likelihood, non-stationarity, threshold-based inference.

Asymptotic behavior of Gaussian minima

Session C15
June 27, 14:00
Room: Data

Gennady Samorodnitsky
Cornell University (USA)

Arijit Chakrabarty
Indian Statistical Institute (India)

In this work, we investigate what happens when an entire sample path of a smooth Gaussian process on a compact interval lies above a high level. Specifically, we determine the precise asymptotic probability of such an event, the extent to which the high level is exceeded, the conditional shape of the process above the high level, and the location of the minimum of the process given that the sample path is above a high level.

Key Words: Gaussian processes, minima, smooth paths.

Inference for multivariate Archimax copulas

Session C42
June 30, 12:00
Room: Chip

Anne-Laure Fougères
Université Claude Bernard Lyon 1 (France)

Johanna G. Nešlehová
McGill University (Canada)

Simon Chatelain
Université Claude Bernard Lyon 1 (France) and McGill University (Canada)

Archimax copulas, which generalize both extreme-value and Archimedean families, were investigated in arbitrary dimensions in Charpentier et al. (2014) [Multivariate Archimax copulas. *Journal of Multivariate Analysis* **126**, 118–136]. This flexible family is a candidate to model data belonging to a maximum domain of attraction while not exhibiting asymptotic extremal behavior. In this talk, we will consider the statistical problem of fitting Archimax distributions to data in full generality. The estimation procedure extends the ideas of Pickands (1981) [Multivariate extreme value distributions. *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2* **49**, 859–878, 894–902] and Capéraà et al. (1997) [A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* **84**, 567–577] developed for multivariate extreme-value copulas. Under certain regularity assumptions, these estimators are shown to converge to a centered Gaussian process. Their

small sample size behavior is studied through simulation and an application to rainfall data.

Key Words: multivariate modeling, Archimax copulas, pre-asymptotic inference.

Efficient rare event simulation for multiple jump events in regularly varying random walks and compound Poisson processes

Session C5
June 26, 16:30
Room: Boole

Jose H. Blanchet
Columbia University (USA)

Chang-Han Rhee, Bert Zwart
Centrum Wiskunde & Informatica (The Netherlands)

Bohan Chen
Centrum Wiskunde & Informatica (The Netherlands)

We propose a class of strongly efficient rare event simulation estimators for random walks and compound Poisson processes with a regularly varying increment/jump-size distribution in a general large deviations regime. Our estimator is based on an importance sampling strategy that hinges on the heavy-tailed sample path large deviations result recently established in Rhee et al. (2016) [Sample path large deviations for heavy-tailed Lévy processes and random walks. *arXiv:1606.02795*]. The new estimators are straightforward to implement and can be used to systematically evaluate the probability of a wide range of rare events with bounded relative error. They are "universal" in the sense that a single importance sampling scheme applies to a very general class of rare events that arise in heavy-tailed systems. In particular, our estimators can deal with rare events that are provoked by multiple big jumps (therefore, beyond the usual principle of single big jump) as well as multidimensional processes such as the buffer content process of a queueing network. We illustrate the versatility of our approach with several applications that arise in the context of mathematical finance, actuarial science, and queueing theory.

Key Words: rare event simulation, regularly varying increment, strongly efficient estimator.

Cluster size distributions of extreme values for the Poisson-Voronoi tessellation

Session C19
June 27, 16:30
Room: Data

Christian Y. Robert
Université Claude Bernard Lyon 1 (France)

Nicolas Chenavier
Université du Littoral Côte d'Opale (France)

We consider the Voronoi tessellation based on a homogeneous Poisson point process in $\mathbf{R}^d$. For a geometric characteristic of the cells (e.g. the inradius, the circumradius, the volume), we investigate the point process of the nuclei of the cells with large values. Conditions are obtained for the convergence in distribution of this point process of exceedances to a homogeneous compound Poisson point process. We provide a characterization of the asymptotic cluster size distribution which is based on the Palm version of the point process of exceedances. This characterization allows us to compute efficiently the values of the extremal index and the cluster size probabilities by simulation for various geometric characteristics. The extension to the Poisson-Delaunay tessellation is also discussed.

Key Words: extreme values, Voronoi tessellations, exceedance point processes.

Exact distributions of the multinomial extremes

Session C44
June 30, 11:30
Room: Pi

Marco Bonetti
Bocconi University (Italy)

Anton Ogay
Delft University of Technology (The Netherlands)

Pasquale Cirillo
Delft University of Technology (The Netherlands)

Starting from an unpublished work by Rappeport (1968) [Algorithms and computational procedures for the application of order statistics to queing problems. *Unpublished dissertation*], we re-propose (and improve) an exact algorithm for the multinomial maximum under the hypothesis of equiprobability. We then show how to exactly compute the probabilities of the sum of the k largest order statistics, correcting the errors available in the same work. Finally we introduce novel ways of computing the exact probabilities of the multinomial minimum and of the multinomial range.

In developing our algorithms, we represent the multinomial as the distribution arising from throwing N independent balls into M urns. We show how our exact probabilities outperform the common approximations available in the literature, especially for small values of N and M, when the approximation error is proven to be not at all negligible.

The exact probabilities we derive can be used in all those situations in which the multinomial extremes plays an important role, from goodness-of-fit tests to the study of Poisson processes, from clinical trials to finance.

Session I2
June 26, 11:30
Room: Chip

Critical regions for anomaly detection via empirical MV-sets on the sphere

Alexandre Gramfort, Anne Sabourin, Albert Thomas
Telecom ParisTech (France)

Stephan Clémençon
Telecom ParisTech (France)

Extreme regions in the feature space are of particular concern for anomaly detection: anomalies are likely to be located in the tails, whereas data scarcity in such regions makes it difficult to distinguish between large normal instances and anomalies. Preprocessing steps may considerably alter the geometric properties of the extremal parts of the scatterplot, with a significant impact on the output of an unsupervised algorithm. This paper presents an unsupervised algorithm for anomaly detection in extreme regions. We propose a Minimum Volume set (MV-set) approach relying on multivariate extreme value theory. This framework includes a canonical pre-processing step, which addresses the issue of output sensitivity to standardization choices. The resulting data representation on the sphere highlights the dependence structure of the extremal observations. Anomaly detection is then cast as a MV-set estimation problem on the sphere, where volume is measured by the spherical measure and mass refers to the angular measure. An anomaly then corresponds to an unusual observation given that one of its variables is large. A preliminary rate bound analysis is carried out for the learning method we introduce and its computational advantages are discussed and illustrated by numerical experiments.

Key Words: angular measure, anomaly detection, minimum volume set, multivariate extreme value theory.

Session I7
June 29, 9:00
Room: Boole

Principal component decomposition and completely positive decomposition of dependence for multivariate extremes

Emeric Thibaud
École Polytechnique Fédérale de Lausanne (Switzerland)

Daniel Cooley
Colorado State University (USA)

Tail dependence in high dimensions is difficult to summarize and model. Via the framework of regular variation, we propose two decompositions which help to summarize and model high-dimensional tail dependence.

We start by defining, via transformation, a vector space on the positive orthant, yielding the notion of basis. With a suitably-chosen transformation, we show that transformed-linear operations applied to regulary-varying random vectors preserve regular variation.

Rather than model regular-variation's angular measure, we summarize tail dependence via a matrix of pairwise tail dependence metrics. Because this matrix is positive semidefinite, standard eigendecomposition allows one to interpret tail dependence via the eigenbasis. Additionally this matrix is completely positive, and a resulting decomposition allows one to easily construct regularly varying random vectors which share the same pairwise tail dependencies.

We apply our methods to extreme rainfall data in Switzerland, using the eigendecomposition to understand the modes of extremal dependence and the completely positive decomposition to estimate probabilities of jointly extreme events. Applying to financial data, we use the eigendecomposition to compare and contrast the dependence structure in the center and tail of the distribution.

Key Words: regular variation, tail dependence, dimension reduction, angular measure.

On the behavior of common connections in social networks

Session I1
June 26, 11:30
Room: Boole

Souvik Ghosh
LinkedIn (USA)

Bikramjit Das
Singapore University of Technology and Design (Singapore)

Studies regarding the generation and structure of social network connections have been ubiquitous. Understanding relationships in the web ensues how we could optimally dissipate important information, be it the occurrence of an extreme event (flood, disease, etc.), or the release of a new drug or the availability of a new job, as exemplified by Twitter, Facebook, LinkedIn in the past decade. An important aspect of such relationships is the number of common connections between two individuals. Recent work has exhibited power-law behavior of the degree distribution (joint distribution of in and out degree for undirected graphs) on linear preferential attachment models for network growth. Restricting to preferential attachment models we show that a power law behavior also holds for the growth of number of common connections among two individual nodes. We exhibit this where the edges in the preferential attachment graph grow linearly with the nodes (sparse case). We also indicate methods to approach this problem where the growth is more than linear (dense case), the techniques used are different.

Key Words: preferential attachment model, power-law tail, asymptotic limit.

Session C3
June 26, 13:30
Room: Data

Limiting joint distribution of sums and maxima: semistability and max-semistability

Maria de Graça Temido
CMUC, University of Coimbra (Portugal)

Sandra Dias
Pólo CMAT-UTAD, CEMAT, University of Trás-os-Montes e Alto Douro (Portugal)

The limiting joint distribution of the sum and maximum of n independent and identically distributed (i.i.d.) random variables, S_n and M_n, was introduced in Chow and Teugels (1978) [The sum and the maximum of i.i.d. random variables. In: Mandel and Hušková (eds), *Proceedings of the Second Prague Symposium on Asymptotic Statistics*, North-Holland, Amsterdam-New York]. Considering a stationary strong mixing sequence with finite variance, Anderson and Turkman (1991) [The joint limiting distribution of sums and maxima of stationary sequences. *Journal of Applied Probability* **28**, 33–44] studied the same problem and the asymptotic independence between S_n and M_n was deduced. In this work we obtain, as well, the limiting joint distribution of the sum and of the maximum of k_n random variables, where $\{k_n\}$ is an integer-valued geometric growing sequence and the underlying marginal distributions are not attracted to the stable and max-stable usual limits. In both i.i.d. and stationary approaches we obtain the desired extensions, considering that S_{k_n} and M_{k_n}, under linear normalization, are attracted to a semistable and to a max-semistable distribution, respectively. Particular attention is given to the case of non-negative integer-valued stationary sequences.

Key Words: sum and maximum, domain of attraction, semistability, max-semistability.

Session I3
June 26, 10:50
Room: Pi

Applications in extreme river discharge modelling

Ferdinand Diermanse
Deltares, The Netherlands

In traditional flood frequency analysis, an extreme value distribution function is fitted through a selection of observed discharges at the site of interest. Extrapolation to extreme events is often influenced by the lack of such events in the data record. Consequently, these methods may fail to fully capture the impact of processes that are only relevant during extreme events. An alternative to such purely statistical methods is offered by rainfall based methods that involve rainfall statistics and modeling of the catchment response to rainfall events. Traditionally, such analysis involved the simulation of only a few rainfall events, representative for a set of return periods. Over the last years, the ever continuing advances in computation power have led to an increase in popularity of probabilistic methods for

extreme value statistics. These probabilistic methods can be grouped into time series methods and event-based methods. In the former, a long synthetic time series of rainfall with similar statistical features as the observed series is generated. In the event based methods, a large number of synthetic flood events are simulated and flood probabilities are estimated with joint-probability methods such as Monte Carlo Simulation or numerical integration. We will present two examples of probabilistic methods that have been applied in practice for the design and assessment of flood mitigating measures. We will show how these methods compare to the more traditional approaches and discuss how combinations of both approaches can be expected to result in more reliable design flow estimates.

Key Words: probabilistic, time series, event based, catchment response.

Conditionally max-stable random fields

Session C37
June 29, 16:30
Room: Data

Martin Schlather, Kirstin Strokorb
Universität Mannheim (Germany)

Martin Dirrler
Universität Mannheim (Germany)

We introduce a class of spatial stochastic processes in the max-domain of attraction of familiar max-stable processes. The new class is based on Cox processes and comprises models with short range dependence. We show that statistical inference is possible within the given framework, at least under some reasonable restrictions.

Key Words: max-stable, max-domain of attraction, Cox processes.

Asymptotic properties of likelihood estimators in multivariate extremes

Session C17
June 27, 16:00
Room: Boole

Sebastian Engelke
École Polytechnique Fédérale de Lausanne (Switzerland)

Marco Oesting
Universität Siegen (Germany)

Clément Dombry
Université de Franche-Comté (France)

Max-stable distributions and processes are important models for extreme events and the assessment of tail risks. The full, multivariate likelihood of a parametric max-stable distribution is complicated and only recent advances enable its use in both frequentist and Bayesian statistics. The asymptotic properties of the maximum

likelihood estimator and the median of the posterior distribution in multivariate extremes are mostly unknown. In this paper we provide natural conditions on the exponent function and the angular measure of the max-stable distribution that ensure asymptotic normality of these likelihood estimators. We show the effectiveness of this result by applying it to popular parametric models in multivariate extreme value statistics and to the most commonly used families of spatial max-stable processes.

Key Words: max-stable distribution, asymptotic normality, differentiability in quadratic mean.

Analyzing the extremal dependence of multivariate time series

Session I10
June 30, 14:40
Room: Boole

Anja Janßen
University of Copenhagen (Denmark)

Holger Drees
University of Hamburg (Germany)

We discuss the statistical analysis of the extremal dependence structure of a stationary multivariate regularly varying time series. In particular, we examine the asymptotic behavior of estimators of the distribution of the pertaining spectral process under suitable mixing conditions. It is shown that in some cases the time change formula established by Basrak and Segers (2009) [Regularly varying multivariate time series. *Stochastic Processes and Their Applications* **119**, 1055–1080] can be used to improve on the estimation accuracy. Furthermore, multiplier block bootstrap enables us to construct confidence sets.

Key Words: extremal dependence, stationary time series, spectral process, time change formula.

Realizing the extremes: estimation of tail-risk measures from a high-frequency perspective

Session C33
June 29, 14:00
Room: Chip

Marco Bee
University of Trento (Italy)

Luca Trapin
Scuola Normale Superiore (Italy)

Debbie J. Dupuis
HEC Montréal (Canada)

In this talk, we apply realized volatility forecasting to Extreme Value Theory (EVT). We propose a two-step approach where returns are first pre-whitened with a high-frequency based volatility model, and then an EVT based model is fitted to the tails of the standardized residuals. This realized EVT approach is compared to the conditional EVT of McNeil and Frey (2000) [Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* **7**, 271–300]. We assess both approaches' ability to filter the dependence in the extremes and to produce stable out-of-sample VaR and ES estimates for one-day and ten-day time horizons. The main finding is that GARCH-type models perform well in filtering the dependence, while the realized EVT approach seems preferable in forecasting, especially at longer time horizons.

Key Words: realized volatility, high-frequency data, extreme value theory, value-at-risk, expected shortfall.

Limits to human life span through extreme value theory

Session C12
June 27, 12:00
Room: Pi

Jesson J. Einmahl
Tilburg University (The Netherlands)

Laurens de Haan
Erasmus University (The Netherlands)

John H.J. Einmahl
Tilburg University (The Netherlands)

There is no scientific consensus on the fundamental question whether the probability distribution of the human life span has a finite endpoint or not and, if so, whether this upper limit changes over time. Crucially, the limit is not defined as the highest observed age at death but the highest age that possibly could be reached. Our approach to the problem is to concentrate on precisely observed mortality data. The study is based on a unique dataset of the ages at death in days of all (about 285,000) Dutch residents, born in the Netherlands, who died in the years 1986-2015 at a minimum age of 92 years. Unlike most other studies we use extreme value theory and base our analysis on the configuration of thousands of mortality data of old people, not just the few oldest old. The existence of a finite upper limit to the life span will follow from the fact that the 30 annual extreme value indices take on only negative values. We prove that there is indeed an upper limit to the life span of men and to that of women and, moreover, that there are no indications of trends in these upper limits over the last 30 years, despite the fact that the number of people reaching high age (say 95 years) is increasing rapidly, almost tripling, in these 30 years.

Key Words: endpoint of distribution, extreme value index, human life span.

Session C34
June 29, 14:00
Room: Data

Functional location, scale and shape parameters of extremal distributions

Anthony C. Davison
École Polytechnique Fédérale de Lausanne (Switzerland)

Yousra El-Bachir
École Polytechnique Fédérale de Lausanne (Switzerland)

We present general statistical methodology for fitting distributions with several functional parameters, each of which may depend smoothly on covariates through a separate generalized additive model (GAM). The multiple GAMs are estimated simultaneously, and choice of the smoothing parameters is incorporated automatically through a likelihood-based approach. The proposed method is statistically efficient, outperforms the gold standard while being simpler; and the algorithm is fast and numerically stable. For illustration, we apply the method to capture non-stationarity not only in the location, but also in the scale and the shape parameters of spatio-temporal extremal distributions.

Key Words: generic smoothing, multiple generalized additive models, non-stationary extremes.

Session I8
June 29, 9:40
Room: Chip

Extremal behavior of aggregated data with an application to downscaling

Marco Oesting
Universität Siegen (Germany)

Sebastian Engelke, Raphaël de Fondeville
Ecole Polytechnique Fédérale de Lausanne (Switzerland)

Meteorological gridded data sets exhibit different statistical behavior than measurements at gauging stations. A possible explanation is that these data result from some form of aggregation, for instance averages or maxima over a geographical region. In order to use this data for accurate risk assessment, downscaling to point locations is required. Methods have been developed for averages, but there is only little work for extremes. In this paper, we derive the joint tail behavior of general aggregation functionals for data in the domain of attraction of a max-stable process. Particularly simple, explicit formulas arise for the important case of Brown-Resnick processes. We further show that these asymptotic results provide the bases for statistical downscaling of extremes.

Key Words: aggregated data, downscaling, spatial extremes.

Local robust estimation of the Pickands dependence function

Session C30
June 29, 12:00
Room: Chip

Yuri Goegebeur
Syddansk Universitet (Denmark)

Armelle Guillou
Université de Strasbourg (France)

Mikael Escobar-Bach
Syddansk Universitet (Denmark)

Modelling dependence among extremes is of primary importance in practical applications where extreme phenomena occur. Similarly to classical statistics, one can summarize extremal dependency in a number of well-chosen coefficients that give a representative picture of the dependency structure. The prime example of such a dependency measure is the coefficient of tail dependence. Alternatively, a full characterization of the extremal dependence between variables can be obtained from functions like e.g. the stable tail dependence function, the spectral distribution function or the Pickands dependence function. In this talk, we consider the robust estimation of the Pickands dependence function in the covariate framework. Our estimator is based on a local estimation with the minimum density power divergence criterion. We provide its main asymptotic properties, in particular the convergence of the stochastic process, correctly normalized, towards a tight centered Gaussian process. The finite sample performance of our estimator is illustrated on simulation studies involving both uncontaminated and contaminated samples.

Key Words: conditional Pickands dependence function, robustness, stochastic convergence.

Estimating trend and dependence for extremal space-time processes

Session C14
June 27, 14:30
Room: Chip

Petra Friederichs
Meteorologisches Institut, Universität Bonn (Germany)

Laurens de Haan
Erasmus Universiteit Rotterdam (The Netherlands)

Cláudia Neves
University of Reading (UK)

Martin Schlather
Universität Mannheim (Germany)

Ana Ferreira
Instituto Superior Técnico da Universidade de Lisboa (Portugal)

A new approach for evaluating time-trends in extreme values accounting also for spatial dependence is proposed. Based on exceedances over a space-time threshold, estimators for a trend function and for extreme value parameters are given, leading to a homogenization procedure for then applying stationary extreme value processes. Extremal dependence over space is further evaluated through variogram analysis including anisotropy.

The methodology is applied to precipitation data, over a time period of 84 years, and from 68 observational weather stations from North-West Germany. In particular the procedure is able to identify significant stations for extremal behaviour including trends.

Key Words: peaks-over-threshold, max-stable processes, non-identical distributions, trend, variogram.

Session C14
June 27, 15:00
Room: Chip

A simultaneous autoregressive model for spatial extremes

Daniel Cooley
Colorado State University (USA)

Emeric Thibaud
École Polytechnique Fédérale de Lausanne (Switzerland)

Miranda J. Fix
Colorado State University (USA)

Areal data is an important subclass of spatial data, e.g. county-level public health data or gridded climate model output. Current models for spatial extremes which characterize spatial tail dependence, such as max-stable models, are geostatistical in nature and thus best suited for point-referenced data. There also exist challenges to fitting such models to datasets with a large number of spatial locations. The objective of our work is to develop a simple spatial model for areal extremes that is computationally feasible for large datasets. In classical spatial statistics, the simultaneous autoregressive (SAR) model for areal data, first introduced by Whittle (1954) [On stationary processes in the plane. *Biometrika* **41**, 434–449], constructs a simple model that captures spatial dependence given a neighborhood structure. We apply recent results from Cooley and Thibaud (2016) [Principal component decomposition and completely positive decomposition of dependence for multivariate extremes. *arXiv:1612.07190*] on transformed linear operations for regularly varying random vectors with tail index $\alpha = 2$ to develop an analogous SAR model for extremes. We will describe the model and how to simulate from it, the resulting tail pairwise dependence matrix which depends on the neighborhood structure and autoregressive parameter, and discuss preliminary methods for estimation and inference.

Key Words: spatial extremes, areal data, lattice data, autoregressive model, regular variation.

Generalized peaks-over-threshold inference for high-dimensional spatial processes

Session C1
June 26, 14:00
Room: Boole

Anthony C. Davison
École Polytechnique Fédérale de Lausanne (Switzerland)

Raphaël de Fondeville
École Polytechnique Fédérale de Lausanne (Switzerland)

Classical spatial models for extremes rely on block maxima, but this approach is limited by computational considerations to a few dozen variables. In order to get a better understanding of extremal dependence and reduce model uncertainties, exploitation of gridded datasets, such as climate models, is necessary. Generalized Pareto processes based on a peaks-over-threshold approach, use single extreme events, generalize the notion of exceedance, and have simpler mathematical expressions.

For spatial modelling, we focus on the Brown–Resnick model, which relies on classical Gaussian models widely used in applications. An efficient algorithm for censored likelihood allows us to perform inference with hundreds of locations. For higher dimensions and generalized risk functionals, we develop an estimator based on the gradient score (Hyvarinen, 2005) [Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research* **6**(4), 695–708], with a complexity similar to likelihood-based inference methods for Gaussian field.

We apply our method to fit a model for extreme rainfall over Florida on a grid with 3600 locations for two types of exceedances: locally intense and areal cumulated rainfall. We can then use the model to generate new extreme events with unobserved intensity and spatial pattern.

Key Words: Brown–Resnick, extreme rainfall, high-dimensional extremes, peaks-over-threshold, $\mathcal{R}$-Pareto process.

Tail dimension reduction for extreme quantile estimation

Session I11
June 30, 14:40
Room: Chip

Laurent Gardes
Université de Strasbourg (France)

In a regression context where a response variable $Y \in \mathbb{R}$ is recorded with a covariate $X \in \mathbb{R}^p$, two situations can occur simultaneously in some applications: (a) we are interested in the tail of the conditional distribution and not on the central part of the distribution and (b) the number p of regressors is large. Up to our knowledge, these two situations have only been considered separately in the literature. The aim of this paper is to propose a new dimension reduction approach adapted to the tail of the distribution and to introduce an extreme conditional quantile estimator. A simulation experiment and an illustration on a real data set are presented.

Key Words: regression, extreme quantile, dimension reduction, kernel smoothing.

Session I3
June 26, 11:30
Room: Pi

Bayesian networks for modelling the impact of extreme weather events on the safety, reliability and availability of infrastructural networks

Noel van Erp
Delft University of Technology, The Netherlands

Pieter van Gelder
Delft University of Technology, The Netherlands

Infrastructural networks, such as road and rail transportation networks, as well as electricity transmission networks, can be modelled by a system of nodes and edges, where the nodes represent the infrastructural objects such as tunnels, bridges, pylons, and where the edges represent the connections between the objects, such as road sections, rail tracks, transmission lines, etc. Performance indicators of infrastructural networks can often be expressed in terms of safety, reliability and availability levels, and they depend on the hazards exposed to the system, such as extreme weather events and geological hazards, as well as the resistance characteristics of the system, such as the network density and the fragility curves of the components. A Bayesian network is developed to model the relationship between hazards and infrastructure performance. The network is able to exhibit cascading effects and to incorporate mitigating actions to reduce the overall risk levels. A Bayesian decision rule is derived to judge the performance of the network in relation to the investment costs of mitigating measures.

Key Words: extreme weather events, Bayesian networks, cascading effects, Bayesian decision theory, performance indicators.

Session I11
June 30, 14:00
Room: Chip

Estimation of the functional Weibull tail-coefficient

Laurent Gardes
Université de Strasbourg (France)

Stéphane Girard
Inria Grenoble Rhône-Alpes & Laboratoire Jean Kuntzmann (France)

We present a nonparametric family of estimators for the tail coefficient of a Weibull tail-distribution when a functional covariate is available. Our estimators are based on a kernel estimator of extreme conditional quantiles, extending a previous work of Daouia et al. (2013) [On kernel smoothing for extremal quantile regression.

Bernoulli **19**, 2557–2589] to the infinite dimensional case. Asymptotic normality of the estimators is proved under mild regularity conditions. Their finite sample performances are illustrated both on simulated and real data. We refer to Gardes and Girard (2016) [On the estimation of the functional Weibull tail-coefficient. *Journal of Multivariate Analysis* **146**, 24–45] for further details.

Key Words: conditional Weibull tail-coefficient, extreme quantiles, nonparametric estimation.

Extreme values of the uniform order 1 autoregressive processes and missing observations

Session C39
June 30, 9:00
Room: Boole

Pavle Mladenović
University of Belgrade (Serbia)

Gennady Samorodnitsky
Cornell University (USA)

Lenka Glavaš
University of Belgrade (Serbia)

We investigate partial maxima of the uniform $AR(1)$ processes with parameter $r \geq 2$. Positively and negatively correlated processes are considered. New limit theorems for maxima in complete and incomplete samples are obtained.

Key Words: extreme values, missing observations, partial maxima, uniform autoregressive processes.

Improving extremal index blocks estimators through resampling procedures

Session C22
June 28, 9:00
Room: Chip

M. Manuela Neves
CEAUL & Universidade de Lisboa (Portugal)

Dora Prata Gomes
CMA & Universidade Nova de Lisboa (Portugal)

The main objective of statistics of extremes is the estimation of parameters of rare events. One of these parameters is the extremal index, θ, that measures the degree of local dependence in the extremes of a stationary process. Clusters of extreme values are linked with incidences and durations of catastrophic phenomena, an important issue in areas like environment, finance, insurance among others. The extremal index is a key parameter and its estimation has been greatly addressed in literature. Here we focus on the estimation of θ using sliding blocks estimators,

that has revealed to be more efficient than the disjoint version and has a smaller asymptotic bias. However blocks estimators require the choice of a threshold u and a block length b. The choice of the threshold is made through stability criteria and b is chosen using resampling techniques in a mean squared error minimization criteria. We have conducted a simulation study to compare the performances of those estimators, discussing also a procedure for the choice of the threshold u. An application to daily mean flow discharge rate in the hydrometric station of Fragas da Torre in Paiva river, data collected from 1 October 1946 to 30 September 2006 is shown.

This research was partially funded by FCT–Fundação para a Ciência e a Tecnologia, Portugal, through the projects UID/MAT/00297/2013 (CMA) and UID/MAT/00006/2013 (CEAUL).

Key Words: blocks estimators, block size, extremal index, mean squared error, resampling techniques.

PORT estimation of parameters of extreme events through generalized means

Session C16
June 27, 15:00
Room: Pi

Fernanda Figueiredo
CEAUL & Universidade do Porto (Portugal)

Lígia Henriques-Rodrigues
CEAUL & Universidade de São Paulo (Brazil)

M. Ivette Gomes
CEAUL & Universidade de Lisboa (Portugal)

In many areas of application, like environment, finance, insurance, statistical quality control, and on the basis of a transformed sample, which can be considered weakly dependent and stationary from an unknown model F, it is a common practice to estimate different parameters of extreme events. Among them, we refer the *value-at-risk* (VaR) at a small level q, a high quantile of probability $1-q$, with q often smaller than $1/n$, where n is the size of the available sample. The semi-parametric estimation of these parameters depends heavily on a reliable and adequate estimation of the *extreme value index* (EVI), one of the primary parameters of extreme events. It happens that most of those semi-parametric estimators do not enjoy such adequate behavior. For instance, most EVI-estimators are not location-invariant, an EVI property, and most VaR-estimators do not suffer the appropriate linear shift in the presence of linear transformations of the data, as does any theoretical quantile. For heavy tails, i.e. for a positive EVI, new VaR-estimators were introduced with such a behavior, the so-called PORT VaR-estimators, with PORT standing for *peaks over a random threshold*. Regarding EVI-estimation, new classes of PORT-EVI estimators, based on powerful generalizations of the Hill EVI-estimator were recently introduced. Now, also for heavy tails, we discuss the use of new classes of estimators with the

aforementioned adequate behavior, using classes of EVI-estimators related to the Hill EVI-estimators, but based on adequate generalized means.

Key Words: extreme value theory, generalized means, heavy tails, Monte Carlo simulation, semi-parametric estimation.

Pareto-type models for precipitation in Madeira Island

Session C2
June 26, 14:00
Room: Chip

Sandra Mendonça
University of Madeira (Portugal)

Délia Gouveia-Reis
University of Madeira (Portugal)

Tail index estimation under Pareto-type models of the extreme value index has been challenging statisticians for a while and by now many proposals exist in the literature, including some versatile ones that address the possibility of adjusting truncated Pareto-type models. In this work some of the known estimators are applied to daily precipitation data collected at 18 stations located in Madeira Island for a period of more than 30 hydrological years. Madeira's history includes a significant number of flash floods and this island, although with an area of only 737 km^2, presents a variety of micro-climates due to its location, topography and natural vegetation. Also, the parameters and return levels estimates obtained in previous works suggest a complex extreme precipitation behaviour in this island.

Key Words: Pareto-type models, extreme value index, estimators, precipitation.

On Turing's formula and the estimation of the missing mass

Session C39
June 30, 9:30
Room: Boole

Michael Grabchak
UNC Charlotte (USA)

The missing mass is the total probability of observing something that has not been observed before. For instance, in the context of ecological applications, it corresponds to the probability of observing a new species, while, in the context of authorship attribution studies, it corresponds to the probability that an author will use a word that he or she has not used before. Perhaps the most famous estimator of the missing mass is Turing's formula, sometimes also called the Good-Turing formula, which was developed by Alan Turing. In this talk, we discuss theoretical properties of Turing's formula and, in particular, give conditions for the consistency and asymptotic normality of its relative error. We then show that these conditions always hold when the distribution is regularly varying with index $\alpha \in (0, 1]$. This is based on joint work with Dr. Zhiyi Zhang.

Key Words: missing mass, regular variation, Turing's formula.

Session C30
June 29, 11:30
Room: Chip

Local estimation of the conditional stable tail dependence function

Mikael Escobar-Bach, Yuri Goegebeur
Syddansk Universitet (Denmark)

Armelle Guillou
Université de Strasbourg & CNRS (France)

We consider the local estimation of the stable tail dependence function when a random covariate is observed together with the variables of main interest. Our estimator is a weighted version of the empirical estimator adapted to the covariate framework, assuming we have data available from a distribution in the max-domain of attraction of a multivariate extreme value distribution. We provide the main asymptotic properties of our estimator, when properly normalized, in particular the convergence of the empirical process towards a tight centered Gaussian process. The finite sample performance of our estimator is illustrated on a small simulation study.

Key Words: conditional stable tail dependence function, empirical process, stochastic convergence.

Session C26
June 28, 11:00
Room: Chip

The bootstrap in extreme value theory

Laurens de Haan
Erasmus Universiteit Rotterdam (The Netherlands)

Alex Koning
Erasmus Universiteit Rotterdam (The Netherlands)

Chen Zhou
Erasmus Universiteit Rotterdam & De Nederlandsche Bank (The Netherlands)

A well-known example where the usual bootstrap is inconsistent, is in the case of sample maxima (Bickel and Freedman, 1981) [Some asymptotic theory for the bootstrap. *Annals of Statistics* **9**(6), 1196-1217]. The solution is to make the bootstrap sample size much smaller than the original sample size. If one considers not only the maximum of a sample but also intermediate order statistics . this problem does not occur. We prove a bootstrap version of Drees well-known approximation of the tail quantile process by a Brownian motion (useful for the peaks-over-threshold method). We also prove a bootstrap version of the approximation of the quantile process of block maxima by a Brownian bridge (Ferreira and de Haan, 2015) [On the

block maxima method in extreme value theory: PWM estimators. *Annals of Statistics* **43**(1), 276-298]. The latter is particularly useful since the asymptotic variance of the PWM estimator for block maxima does not have a closed form.

What are the economic determinants of operational losses severity? A regularized generalized Pareto regression approach.

Session C18
June 27, 16:30
Room: Chip

Andreas Groll, Thomas Kneib
Georg-August-Universität Göttingen (Germany)

Julien Hambuckers
Georg-August-Universität Göttingen (Germany)

We investigate a database of 40,871 extreme operational losses from the Italian bank UniCredit. These data cover a period of 10 years and 7 different event types. We study the dependence between a set of macroeconomic, financial and firm-specific factors with the severity distribution of these losses, assumed to be Generalized Pareto. Answering to this question is of particular interest for banks and regulators to define a risk capital charge in line with the economic situation. To perform a selection of the relevant explanatory variables, we use a penalized-likelihood approach based on a local quadratic approximation of L_1 penalty terms. The finite-sample properties of this estimation technique is briefly investigated in a simulation study. Then, we perform the regression analysis with up to 33 candidate terms for each parameters, including a mix of categorical, continuous and interaction variables, and estimate their coefficients with the proposed approach. Our results suggest that, contrary to what usual unpenalized approaches would suggest, only a small subset of the covariates are deemed relevant. Among other, the GDP growth rate, the VIX index, the leverage ratio and the proportion of non-interest income are found to be good explanatory variables of the severity distribution. Last, we illustrate the impact of several economic scenarios on the requested capital if the total loss distribution (assumed to be a compound Poisson-GPD process) is conditioned on these scenarios.

Key Words: operational loss, generalized Pareto, penalized likelihood.

Generalised Pickands and Piterbarg constants

Session C15
June 27, 14:30
Room: Data

Krzysztof Dębicki
University of Wrocław (Poland)

Enkelejd Hashorva
University of Lausanne (Switzerland)

The classical Pickands and Piterbarg constants appear in numerous asymptotic problems related to extremes of Gaussian processes and random fields. In this talk we shall discuss various generalisations of these constants which naturally relate to stationary max-stable random fields. Our main results concern new formulas and bounds as well as examples where these constants appear.

Session C1
June 26, 14:30
Room: Boole

Asymptotics for extreme depth-based quantile region estimation

Yi He
Monash University (Australia)

Consider the small-probability quantile region in arbitrary dimensions consisting of extremely outlying points with nearly zero data depth value. Using extreme value statistics, we extend the semiparametric estimation procedures proposed in Cai et al. (2011) [Estimation of extreme risk regions under multivariate regular variation. *Annals of Statistics* **39**, 1803–1826] and He and Einmahl (2016+) [Estimation of extreme depth-based quantile regions. *Journal of the Royal Statistical Society: Series B*, forthcoming, DOI: 10.1111/rssb.12163] to incorporate various depth functions. Under weak regular variation conditions, a general consistency result is derived. The main novelty of this paper is the construction of confidence sets that asymptotically cover the extreme quantile region or/and its complement with a pre-specified probability. Refined asymptotic confidence sets are derived particularly for the half-space depth to include the shape estimation uncertainty. The finite-sample coverage probabilities of our asymptotic confidence sets are evaluated in a simulation study for the half-space depth and the projection depth. We use the procedure for risk management by applying it to stock market returns.

Key Words: data depth, multivariate quantile, multivariate tail empirical process, extreme value statistics, outlier detection.

Session I4
June 27, 9:00
Room: Boole

Extreme eigenvalues of sample correlation matrices

Thomas Mikosch
University of Copenhagen (Denmark)

Johannes Heiny
University of Aarhus (Denmark)

In Principal Component Analysis one studies the sample covariance or sample correlation matrix, both of which often lead to the same result. We compare the behavior of the eigenvalues of the sample covariance and sample correlation matrices and argue that the latter seems more robust, in particular in the case of infinite fourth moment.

We show that the largest and smallest eigenvalues of a sample correlation matrix stemming from n independent observations of a p-dimensional time series with iid components converge almost surely to $(1+\sqrt{\gamma})^2$ and $(1-\sqrt{\gamma})^2$, respectively, as $n \to \infty$, if $p/n \to \gamma \in (0,1]$ and the truncated variance of the entry distribution is "almost slowly varying", a condition we describe via moment properties of self-normalized sums. Moreover, the empirical spectral distributions of these sample correlation matrices converge weakly, with probability 1, to the Marčenko–Pastur law.

Key Words: PCA, sample correlation matrix, extreme eigenvalues, spectral distribution, regular variation.

Discrete extremes

Session C9
June 27, 11:30
Room: Boole

Adrien Hitz
University of Oxford (UK)

In this talk, I will present results that are widening the scope of extreme value analysis applied to discrete-valued data containing many tied observations. Extreme values of a random variable X are commonly modeled using the generalized Pareto distribution, a method that often gives good results in practice. When X is discrete, we propose two other methods using a discrete generalized Pareto and a generalized Zipf distribution respectively. The latter are not only motivated for a more general class of discrete distributions on theoretical grounds, but they perform well for estimating rare events in several simulated data. I will end by illustrating the methods on two discrete-valued data sets consisting of word frequency and multiple births. This talk is based on a joint work with Profs. Gennady Samorodnitsky and Richard Davis.

Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures

Session I8
June 29, 10:20
Room: Chip

Thomas Opitz
BioSP, INRA (France)

Emeric Thibaud
École Polytechnique Fédérale de Lausanne (Switzerland)

Raphaël Huser
King Abdullah University of Science and Technology (Saudi Arabia)

Gaussian scale mixtures are constructed as Gaussian processes with a random variance. They have non-Gaussian marginals and can exhibit asymptotic dependence unlike Gaussian processes, which are always asymptotically independent except for perfect dependence. Motivated by the analysis of spatial extremes, we

propose a flexible but parsimonious Gaussian scale mixture copula model, which smoothly interpolates from asymptotic dependence to independence. We show how this new model can be fitted to high threshold exceedances using a censored likelihood approach, and we demonstrate that it provides valuable information about tail characteristics. The methodology will then be illustrated with an application to wind speed data in the Pacific Northwest, US, showing that it adequately captures the data's extremal properties.

Key Words: asymptotic dependence and independence, censored likelihood, Gaussian scale mixture, spatial extreme, threshold exceedance.

A bootstrap test for the tail index of autoregressive models with heavy-tailed innovations

Session C26
June 28, 11:30
Room: Chip

Won-Tak Hong
Gachon University (South Korea)

Eunju Hwang
Gachon University (South Korea)

In this work we consider an autoregressive model with heavy-tailed innovations and propose a bootstrap test for the tail index of the innovation distribution, based on the nonparametric test of Jureckova et al. (2009) [Testing on the tail index in autoregressive models. *Annals of the Institute of Statistical Mathematics* **61**, 579–598]. We show the first order validity of the bootstrap for the Jureckova-type nonparametric test. We construct a bootstrap test statistic for testing whether the right tail of the innovation distribution has the same as or heavier than a given Pareto tail, against alternatively the right tail is lighter. We establish the limiting distribution of the bootstrap test statistic in the null hypothesis and verify the consistency of the bootstrap test. A simulation study is conducted to see the performance of the bootstrap test and to compare with the existing results allowing various tail indices.

Key Words: Pareto tail index, bootstrap test, empirical process, autoregressive model.

Phantom distribution functions for weakly dependent random vectors

Session C19
June 27, 16:00
Room: Data

Natalia Soja-Kukieła
Nicolaus Copernicus University (Poland)

Adam Jakubowski
Nicolaus Copernicus University (Poland)

The notion of a phantom distribution (phdf) function was introduced by O'Brien (1987) [Extreme values for stationary and Markov sequences. *Annals of Probability* **15**, 281–292]. Let $\{X_j\}$ be a stationary sequence of *real* random variables with partial maxima

$$M_n = \max_{1 \le j \le n} X_j.$$

The sequence $\{X_n\}$ is said to admit a phantom distribution function G if

$$\sup_{u \in \mathbb{R}} \left| P(M_n \le u) - G^n(u) \right| \to 0, \text{ as } n \to \infty.$$

For Markov chains with a regenerative structure Rootzén (1988) [Maxima and exceedances of stationary Markov chains. *Advances in Applied Probability* **20**, 371–390] gave an interpretation of G in terms of maxima over the regeneration cycle.

In general it is known from Doukhan et al. (2015) [Phantom distribution functions for some stationary sequences. *Extremes* **18**, 697–725] that the existence of a phdf is a quite common phenomenon for stationary weakly dependent sequences. In particular, any α-mixing stationary sequence with continuous marginals admits a continuous phdf.

In the talk we discuss the corresponding notion of a phantom distribution function for stationary random vectors and provide some sufficient conditions for its existence. The machinery is much in the spirit of Perfekt (1997) [Extreme value theory for a class of Markov chains with values in $\mathbb{R}^d$, *Advances in Applied Probability* **29**, 138–164].

Key Words: maxima of random vectors, phantom distribution function, stationary sequences, weakly dependent random vectors.

Estimating the return level of the 2011 Lake Champlain flood considering the clustering of precipitation extremes

Session C2
June 26, 14:30
Room: Chip

Christian Genest, Orla Murphy, Johanna Nešlehová
McGill University (Canada)

Jonathan Jalbert
McGill University (Canada)

Lake Champlain is a natural freshwater lake that straddles the Canadian-US boarder in Eastern North America. In the spring of 2011, its water level was at a record high, causing massive floods in the surrounding valley and along the Richelieu River (Québec, Canada). The severity of this flood was largely due to heavy rainfalls occurring in clusters over several days. For this reason, it has proved difficult to estimate the return period of this unprecedented event using standard extreme-value methods.

A cluster of high precipitation can be defined as a streak of rainy days, at least one of which results in extreme rainfall. A common way of declustering the time

series is to retain only the cluster maxima. However, this approach is not suitable in the present case since the flood was largely due to the presence of clusters of precipitation rather than single days with extreme rainfall.

The purpose of this talk is to show how one can successfully estimate the return period of the spring precipitation that triggered the 2011 flood in the Richelieu River. To this end, a statistical model that takes into account clusters of extreme precipitation must be used. One such modeling strategy will be proposed, based on an adaptation of the work by Süveges and Davison (2012) [A case study of a "Dragon-King": The 1999 Venezuelan catastrophe. *European Physical Journal Special Topics* **205**, 131-146]. The model relies on a decomposition of clusters into polar coordinates. Extreme-value distributions are then used to model both the radial component and the angular component of the clusters.

Key Words: clusters, M3 process, precipitation.

Spectral tail processes of stationary regularly varying multivariate time series

Session I9
June 29, 9:00
Room: Pi

Anja Janßen
Royal Institute of Technology, Stockholm (Sweden)

A regularly varying time series as introduced in Basrak and Segers (2009) [Regularly varying multivariate time series. *Stochastic Processes and Their Applications* **119**, 1055–1080] is a (multivariate) time series such that all finite dimensional distributions are multivariate regularly varying. The extremal behavior of such a process can then be described by the index of regular variation and the so-called spectral tail process.

As shown in Basrak and Segers (2009), the stationarity of the underlying time series implies a certain structure of the spectral tail process, which is sometimes called the "time change formula". In this talk, we analyze whether in turn also every process which satisfies this property is the spectral tail process of an underlying stationary time series. To this end, we construct suitable max-stable processes in order to act as a proper underlying process for the realization of the spectral tail process. Furthermore, we discuss some statistical applications of our results.

Key Words: max-stable processes, regularly varying time series, spectral tail process, stationary processes.

Asymptotic properties of max-AR(1) sequences of the Kendall type

Session C25
June 28, 11:00
Room: Boole

Marek Arendarczyk, Władysław Szczotka
University of Wrocław (Poland)

Barbara Jasiulis-Gołdyn
University of Wrocław (Poland)

Following Borowiecka-Olszewska at al. (2015) [Lévy processes and stochastic integral in the sense of generalized convolution. *Bernoulli* **24**(1), 2513–2551] we consider extremal Markovian sequences of the Kendall type introduced by Jasiulis-Gołdyn (2016) [Kendall random walks. *Probability and Mathematical Statistics* **36**(1), 165-185], i.e. stochastic processes defined by:

$$X_0 = 1, \quad X_1 = Y_1, \quad X_{n+1} = M_{n+1}\left[\mathbf{1}(\xi_n < \varrho_{n+1}) + \theta_{n+1}\mathbf{1}(\xi_n > \varrho_{n+1})\right],$$

where

$$M_{n+1} = \max\{X_n, Y_{n+1}\}, \quad \varrho_{n+1} = \frac{\min\{X_n, Y_{n+1}\}^\alpha}{\max\{X_n, Y_{n+1}\}^\alpha},$$

and sequences $(Y_k) \sim i.i.d.(\nu)$, $(\xi_k) \sim i.i.d.(U[0,1])$, $(\theta_k) \sim i.i.d.(\pi_{2\alpha})$ are independent.

Recall that the Kendall convolution algebra is the world, where:

$$\delta_x \triangle_\alpha \delta_1 = x^\alpha \pi_{2\alpha} + (1 - x^\alpha)\delta_1, \quad x \in [0,1],$$
$$\pi_{2\alpha}(dx) = 2\alpha x^{-2\alpha-1}\mathbf{1}_{[1,\infty)}(x)dx.$$

We prove that one dimensional distributions of Kendall random walks are regularly varying. The Central Limit Theorem in the Kendall convolution algebra will be showed using the Williamson transform; see Jasiulis-Gołdyn and Misiewicz (2016) [Kendall random walk, Williamson transform and the corresponding Wiener-Hopf factorization. *Lithuanian Mathematical Journal*, in press]. We notice that obtained stable distributions belong to maximal domain of attraction of the Fréchet distribution. We also prove convergence of finite dimensional distributions of the following process:

$$Z_n(t) = n^{-1/\alpha} X_{[nt]}.$$

The results one can find in Arendarczyk et al. (2017) [Asymptotic properties of Kendall random walks, *in preparation*].

Notice that limit distributions of max-AR(1) sequences of the Kendall type describe some indicators of air pollution.

Key Words: extremal Markov chain, Kendall random walk, regular variation, Williamson transform, max-autoregressive sequences.

Extremes of locally stationary chi-square processes with trend

Session C31
June 29, 12:00
Room: Data

Peng Liu
University of Lausanne (Switzerland)

Lanpeng Ji
University of Applied Sciences of Western Switzerland (Switzerland)

Chi-square processes with trend appear naturally as limiting processes in various statistical models. Let $\chi^2(t), t \geq 0$ be a locally stationary chi-square process, and let $g(t), t \geq 0$ be a nonnegative continuous function. In this talk, we shall discuss the exact asymptotics of

$$\mathbb{P}\Big(\sup_{t\in(0,1)} \big(\chi^2(t) - g(t)\big) > u\Big), \quad u \to \infty$$

for particular admissible trend functions g. The difficulty of the study comes from the fact that the above supremum is taken over an open interval other than a usually considered compact interval, for which we derive an admissibility condition on function g which is a generalized version of the Kolmogorov-Dvoretsky-Erdős integral test of the underlying Gaussian processes. For some special cases, such as Brownian bridge and Bessel process, the relation between the admissibility condition and the Kolmogorov-Dvoretsky-Erdős integral test (or law of iterated logarithm) will be discussed in detail. Another important application of our main results is the analysis of the critical values of the limit of some test statistics.

Key Words: chi-square process, Brownian bridge, Bessel process, generalized Kolmogorov-Dvoretsky-Erdős integral test, law of iterated logarithm.

Regression based clustering of spatio-temporal excesses with systematically missing covariates

Session C41
June 30, 9:00
Room: Pi

Illia Horenko
Università della Svizzera Italiana (Switzerland)

Olivia Martius
University of Bern (Switzerland)

Olga Kaiser
Università della Svizzera Italiana (Switzerland)

Regression based Generalized Pareto Distribution (GPD) is often used for statistical analysis of threshold excesses. Regression based models rely on the explicit availability of all of the relevant covariates. In cases, when the complete set of relevant covariates is not available it was shown that under weak assumptions their influence can be reflected by a nonstationary and nonhomogenous dynamics. We present a semiparametric and an adaptive framework for regression based clustering of spatio-temporal threshold excesses in a presence of systematically missing covariates. The nonstationary and nonhomogenous behavior of threshold excesses is described by a set of local stationary GPD models, where the parameters are expressed as regression models, and a non-parametric spatio-temporal hidden switching process. Nonparametric Finite Element time-series analysis Methodology (FEM) with Bounded Variation of the model parameters (BV) is deployed for describing the spatio-temporal switching process. The presented approach goes beyond strong a

priori assumptions made is standard latent class models like Mixture Models and Hidden Markov Models. Additionally, the resulting FEM-BV-GPD framework provides pragmatic insights into the underlying spatial dependence structure by grouping together all locations that exhibit similar behavior of the switching process. The performance of the framework is demonstrated on threshold excesses extracted from daily accumulated precipitation series over 17 different locations in Switzerland from 1981 till 2013 – showing that the introduced approach allows for a better description of the data.

Key Words: spatio-temporal clustering, regression analysis, systematically missing covariates.

Stochastic simulation of joint fields of daily precipitation and river flow

Session C32
June 29, 11:30
Room: Pi

Vasily Ogorodnikov, Olga Sereseva
Institute of Computational Mathematics and Mathematical Geophysics (Russia)

Nina Kargapolova
Institute of Computational Mathematics and Mathematical Geophysics (Russia)

We propose a numerical parametric stochastic model of joint random fields of daily precipitation and river flow. One-dimensional distributions of precipitation and river flow, their auto- and cross-correlation functions are input parameters of the model, that is constructed on the assumption that random fields are homogeneous. Parameters of marginal and correlation functions are chosen on the basis of long-term observations on weather and hydrologic stations situated in considered area. The model is based on a special modification of the inverse distribution function method. This modification let to simulate both unconditional joint random fields and conditional fields of river flow when precipitation amount in some points of the area is given. With the use of the model statistical characteristics of precipitation field and river flow are studied. Properties of extreme river flow under various precipitation scenarios are analyzed.

This research was supported by the Russian Foundation for Basis Research (grants No 15-01-01458-a, 16-01-00145-a, 16-31-00123-mol-a, 16-31-00038-mol-a) and the President of the Russian Federation grant (No MK-659.2017.1).

Key Words: stochastic simulation, daily precipitation fields, daily river flow, extreme river flow.

Some results on joint record events

Session C28
June 28, 11:30
Room: Pi

Michael Falk
University of Würzburg (Germany)

Simone Padoan
Bocconi University (Italy)

Amir Khorrami Chokami
Bocconi University (Italy)

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ be independent and identically distributed (i.i.d.) copies of a random vector (r.v.) $\boldsymbol{X} \in \mathbb{R}^d$ with a continuous joint distribution function F. For $n \in \mathbb{N}$, let $\boldsymbol{M}_j := \max_{1 \leq i \leq j} \boldsymbol{X}_i$ be the r.v. of componentwise maxima. Given a sequence of rvs $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, we call $\boldsymbol{X}_n$ a *simple record* when $\boldsymbol{X}_n \not\leq \boldsymbol{M}_{n-1}$, while we call it a *complete record* when $\boldsymbol{X}_n > \boldsymbol{M}_{n-1}$. When $d = 1$, the two definitions coincide.

In the univariate case it is a well known fact that the events (X_j is a record) and (X_k is a record), for indices $k > j$, $j = 1, 2, \ldots$, are independent. We provide additional asymptotic results on the conditional bivariate survival and distribution functions of two joint records. From the limiting conditional joint distribution function, we then derive some measures of dependence between records such as the correlation.

In the multivariate case ($d \geq 2$), much less is known on records. We investigate the problem of establishing which results in univariate records *do not* carry over to the multivariate case. For complete records, we compute useful quantities such as the probability that $\boldsymbol{X}_n$ is the final complete record in the case of independence of margins and we find the distribution function of the final complete record in such a case.

By means of a counterexample we show that the result of independence mentioned above for univariate records does not hold anymore in a multivariate setting. Finally, we show our findings obtained with simple records.

Key Words: complete records, simple records, max-stable distribution, terminal records.

Estimation of the upper limit of distribution and its application to assessment of the maximum earthquake magnitude m_{max}

Session C24
June 28, 9:00
Room: Pi

Ansie Smit
University of Pretoria (South Africa)

Andrzej Kijko
University of Pretoria (South Africa)

This presentation provides an alternative to the classic extreme value theory approach for evaluation of the upper limit of the distribution. The proposed procedure is generic and is applied to the estimation of the maximum earthquake magnitude m_{max}. The procedure is capable of generating solutions in different forms,

depending on assumptions on the distribution model. It includes the cases when (i) seismic event magnitudes are distributed according to the truncated exponential (Gutenberg-Richter) relation, (ii) when the empirical magnitude distribution deviates moderately from the Gutenberg-Richter relation, and (iii) no specific type of magnitude distribution is assumed. Both synthetic, Monte Carlo simulated seismic event catalogues and actual data from Southern California are used to demonstrate the performance of the procedure. The three estimates of m_{max} for California, obtained by the three forms of the generic equation are 8.32 ± 0.43, 8.31 ± 0.42 and 8.34 ± 0.45 for the three cases respectively. The three estimates are nearly identical, although higher than the value obtained by Field et al. (1999) [A mutually consistent seismic-hazard source model for Southern California. *Bulletin of the Seismological Society of America* **89**, 559–578].

Key Words: maximum earthquake magnitude.

A continuous updating weighted least squares estimator of tail dependence in high dimensions

Session I2
June 26, 10:10
Room: Chip

John H.J. Einmahl
Tilburg University (The Netherlands)

Johan Segers
Université catholique de Louvain (Belgium)

Anna Kiriliouk
Université catholique de Louvain (Belgium)

Likelihood-based procedures are a common way to estimate tail dependence parameters. They are not applicable, however, in non-differentiable models such as those arising from recent max-linear structural equation models. Moreover, they can be hard to compute in higher dimensions. An adaptive weighted least-squares procedure matching nonparametric estimates of the stable tail dependence function with the corresponding values of a parametrically specified proposal yields a novel minimum-distance estimator. The estimator is easy to calculate and applies to a wide range of sampling schemes and tail dependence models. In large samples, it is asymptotically normal with an explicit and estimable covariance matrix. The minimum distance obtained forms the basis of a goodness-of-fit statistic whose asymptotic distribution is chi-square. Extensive Monte Carlo simulations confirm the excellent finite-sample performance of the estimator and demonstrate that it is a strong competitor to currently available methods. The estimator is then applied to disentangle sources of tail dependence in European stock markets.

Key Words: directed acyclic graph, extremal coefficient, max-linear model, multivariate extremes, stable tail dependence function.

Session I7
June 29, 9:40
Room: Boole

Can we identify a max-linear model on a DAG by the tail dependence coefficient matrix?

Nadine Gissibl
Technical University of Munich (Germany)

Moritz Otto
Karlsruhe Institute of Technology (Germany)

Claudia Klüppelberg
Technical University of Munich (Germany)

We investigate multivariate regularly varying random vectors with discrete spectral measure induced by a directed acyclic graph (DAG). The tail dependence coefficient measures extreme dependence between two vector components, and we investigate how the matrix of tail dependence coefficients can be used to identify the full dependence structure of the random vector on a DAG or even the DAG itself. Furthermore, we estimate the distributional model by the matrix of empirical tail dependence coefficients. From these observations we want to infer the causal dependence structure in the data.

Key Words: directed acyclic graph, max-linear model, regular variation, structural equation model, tail dependence coefficient.

Session I6
June 27, 9:00
Room: Pi

Space-time max-stable models with spectral separability

Paul Embrechts
ETH Zurich, RiskLab (Switzerland)

Christian Yann Robert
Université Claude Bernard Lyon 1 (France)

Erwan Koch
ETH Zurich, RiskLab (Switzerland)

Natural disasters may have considerable impact on society as well as on the (re-)insurance industry. Max-stable processes are ideally suited for the modelling of the spatial extent of such extreme events, but it is often assumed that there is no temporal dependence. Only a few papers have introduced spatiotemporal max-stable models, extending the Smith, Schlather and BrownResnick spatial processes. These models suffer from two major drawbacks: time plays a similar role to space and the temporal dynamics are not explicit. In order to overcome these defects, we introduce spatiotemporal max-stable models where we partly decouple the influence of time and space in their spectral representations. We introduce both continuous- and discrete-time versions. We then consider particular Markovian cases with a

max-autoregressive representation and discuss their properties. Finally,we briefly propose an inference methodology which is tested through a simulation study.

Key Words: extreme value theory, spatiotemporal max-stable process, spectral separability, temporal dependence.

Empirical likelihood based testing for multivariate regular variation

Session C40
June 30, 9:00
Room: Chip

John H.J. Einmahl
Tilburg University (The Netherlands)

Andrea Krajina
Georg-August-Universität Göttingen (Germany)

Multivariate regular variation is a property ensuring regularity in the tail of a multivariate distribution function and it is often encountered in the field of multivariate extreme value theory. We introduce a localised empirical likelihood based test for multivariate regular variation. We provide the asymptotic distribution of the test statistic, the critical values of the test and illustrate its finite-sample behaviour on simulated and real data examples.

Key Words: asymptotic theory, empirical processes, extreme value index, multivariate tail.

Functional convergence of multivariate partial maxima processes

Session C4
June 26, 13:30
Room: Pi

Danijel Krizmanić
University of Rijeka (Croatia)

For a strictly stationary sequence of $\mathbb{R}_+^d$–valued random vectors (X_n) we derive functional convergence of the partial maxima stochastic process

$$M_n(t) = \bigvee_{i=1}^{\lfloor nt \rfloor} \frac{X_i}{a_n}, \quad t \in [0,1],$$

under joint regular variation and weak dependence conditions, where (a_n) is a sequence of positive real numbers such that $n\mathrm{P}(||X_1|| > a_n) \to 1$ as $n \to \infty$. The limit process is an extremal process, and the convergence takes place in the space of $\mathbb{R}_+^d$–valued càdlàg functions on $[0,1]$, with the standard (or strong) Skorohod M_1 topology when $d = 1$ and weak Skorohod M_1 topology when $d \geq 2$.

Key Words: functional limit theorem, regular variation, M_1 topology, extremal process.

Session C44
June 30, 12:00
Room: Pi

Cumulant estimators for stable laws in Nolan's 0-parametrization

Tõnu Kollo
University of Tartu (Estonia)

Annika Krutto
University of Tartu (Estonia)

The four-parameter stable laws form a flexible class of distributions that can allow heavy tails as well as skewness. However, modelling via stable laws is complicated due to the lack of a closed form expression for the density function. For estimating the parameters several methods, e.g. maximum likelihood based, quantile based and characteristic function based, have been proposed. An alternative method, based on the cumulant function, i.e. the logarithm of the characteristic function, yields explicit point estimators for the four parameters of a stable law at two arbitrary real valued arguments of the empirical cumulant function. Estimation procedure is computationally simple while the optimal selection of the two arguments still is an open question. The method was introduced by Press (1975) [Estimation in univariate and multivariate stable distributions. *JASA* **67**, 842-846] and more recently studied in Krutto (2016) [Parameter estimation in stable laws. *Risks* **4**, 43]. In this study we search for the optimal arguments of cumulant estimators in Nolan's 0-parametrization under several data standardizations. To assess the effectiveness of estimators, we perform simulations over the parameter space, measure the quality of cumulant estimates by the means of squared errors, and compare our results with the maximum likelihood estimates.

Key Words: empirical cumulant function, generalized method of moments, parameter estimation, standardization.

Session C18
June 27, 17:00
Room: Chip

Expected utility and catastrophic risk in a stochastic economy-climate model

Masako Ikefuji
Syddansk Universitet (Denmark)

Jan R. Magnus
Vrije Universiteit Amsterdam & Tilburg University (The Netherlands)

Chris Muris
Simon Fraser University (USA)

Roger J.A. Laeven
University of Amsterdam (The Netherlands)

We analyze a stochastic dynamic finite-horizon economic model with climate change, in which the social planner faces uncertainty about future climate change and its economic damages. Our model (SSICE) is a simplified version of Nordhaus' deterministic DICE model, but it incorporates, possibly heavy-tailed, stochasticity. We develop a regression-based numerical method for solving a general class of dynamic finite-horizon economy-climate models with potentially heavy-tailed uncertainty and general utility functions. We then apply this method to SSICE and examine the effects of light- and heavy-tailed uncertainty. The results indicate that the effects can be substantial.

Robust risk analysis in insurance applications

Session I5
June 27, 9:40
Room: Chip

Jose Blanchet
Columbia University (USA)

Qihe Tang
University of Iowa (USA)

Zhongyi Yuan
Pennsylvania State University (USA)

Henry Lam
University of Michigan (USA)

Risk analysis in insurance and financial applications often involves estimating extremal quantities. The typical estimation approach is to build distributions or probabilistic models justified via extreme value theory or other statistical tools. However, under insufficient data or high-complexity situations, these models often possess discrepancies from the real-world behaviors (nonparametrically). This talk will discuss approaches using ideas from robust optimization to alleviate these model errors, where the uncertainties of the models are encapsulated as constraints in the optimization formulations. The discussion will include the specifications of these constraints, their integration with data, and also the tractability and solution techniques of the resulting optimizations.

Key Words: insurance modeling, robust optimization, model risk, nonparametric.

Session C3
June 26, 14:00
Room: Data

Random max-closure property of heavy-tailed random variables

Jonas Šiaulys
Vilnius University (Lithuania)

Remigijus Leipus
Vilnius University (Lithuania)

In this talk we present some results on the random max-closure property for independent (not necessarily identically distributed) real-valued random variables $X_1, X_2, \dots$, which states that, given distributions $F_{X_1}, F_{X_2}, \dots$ from some class of heavy-tailed distributions, the distribution of random maximum $X_{(\eta)} := \max\{0, X_1, \dots, X_\eta\}$ or random maximum $S_{(\eta)} := \max\{0, S_1, \dots, S_\eta\}$ belongs to the same class of heavy-tailed distributions. Here, $S_n := X_1 + \cdots + X_n$, $n \geq 1$, η is a counting random variable, independent of $\{X_1, X_2, \dots\}$. We provide the conditions for the random max-closure property in the case of classes $\mathscr{D}$ and $\mathscr{L}$.

Key Words: heavy tails, random maximum, closure property, long tail, dominatedly varying tail.

Session C10
June 27, 11:30
Room: Chip

Endpoint estimation for observations with normal measurement errors

Liang Peng, Xing Wang
Georgia State University (USA)

Chen Zhou
Erasmus Universiteit Rotterdam (The Netherlands)

Xuan Leng
Erasmus Universiteit Rotterdam (The Netherlands)

This paper investigates the estimation of the finite endpoint of a distribution function when the observations are contaminated by normally distributed measurement errors. Under the framework of Extreme Value Theory, we propose a class of estimators for the standard deviation of the measurement errors as well as for the endpoint. Asymptotic theories for the proposed estimators are established while their finite sample performance are demonstrated by simulations. In addition, we apply the proposed methods to the outdoor long jump data to estimate the ultimate limit for human beings in the long jump.

Key Words: convolution, extreme value theory, ultimate world record, Weibull domain of attraction.

Asymptotics and statistical inference on independent and non-identically distributed bivariate Gaussian triangular arrays

Session C4
June 26, 14:00
Room: Pi

Zuoxiang Peng
Southwest University (China)

Xin Liao
University of Shanghai for Science and Technology (China)

Let $\{(X_{ni}, Y_{ni}), 1 \leq i \leq n, n \geq 1\}$ be independent bivariate Gaussian triangular arrays, and let ρ_{ni} denote the correlation coefficient of (X_{ni}, Y_{ni}), $1 \leq i \leq n$. For $\rho_{ni} = \rho \in (-1, 1)$, Sibuya (1960) [Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics* **11**, 195-210] showed that the normalized maxima are asymptotically independent, which may seriously underestimate extreme probabilities in practice. To remove this drawback, Hüsler and Reiss (1989) [Maxima of normal random vectors: between independence and complete dependence. *Statistics and Probability Letters* **7**, 283-286] provided a Hüsler-Reiss condition, which lets $\rho_{ni} = \rho_n$ depend on the sample size n and tend to one with a certain rate. Under the so-called Hüsler-Reiss model, Hüsler and Reiss (1989) presented that the normalized maxima become asymptotically dependent.

In this paper, we extend the Hüsler-Reiss model by allowing ρ_{ni} to depend on both i and n, and focus on the independent and non-identically distributed bivariate Gaussian triangular arrays with ρ_{ni} being a monotone continuous function of i/n. For the extended model, we establish the first and the second-order asymptotics of distributions of normalized maxima. Furthermore, parametric inference for this unknown function is studied. Some simulation study and real data sets analysis are also presented.

Key Words: bivariate Gaussian random vector, maximum, limiting distribution, second-order expansion, estimation.

Extreme values and nearest neighbor degrees in large networks

Session I4
June 27, 10:20
Room: Boole

Pim van der Hoorn
Northeastern University (USA)

Dong Yao
University of Science and Technology of China (China)

Nelly Litvak
University of Twente (The Netherlands)

The hubs – nodes with extremely large number of connections – play a crucial role in the analysis of complex networks, such as social networks or the Word Wide

Web. We analyze the average nearest neighbor degree (ANND) of a node of degree k, as a function of k. The ANND is often used to characterize dependencies between degrees of a node and its neighbors. If ANND is increasing (decreasing) in k, then the dependencies are positive (negative). We study the limiting behavior of the ANND when the degree sequence in the network is sampled from a regularly varying distribution. In particular, we study the configuration model (CM), where the nodes connect to each other at random. When degrees have finite variance, we prove the ANND converges to a deterministic function in a large class of random graphs. When degrees have infinite variance, we establish the scaling of the ANND in CM and prove the CLT for any fixed k. When k scales as a power of the graph size, the behavior of the ANND in the infinite variance scenario is unknown, and we encounter that the ANND has very large fluctuations. To remedy the disadvantages of the ANND, we introduce the average nearest neighbor rank (ANNR) measure, which is much more stable. If time permits, I will mention several other problems in network sampling, that are related to the nodes of extremely large degrees.

Session C21
June 28, 9:00
Room: Boole

Ruin with dependent insurance and financial risks in a discrete-time risk model with investment

Yiqing Chen
Drake University (USA)

Yang Yang
Nanjing Audit University (China)

Jiajun Liu
Xi'an Jiao-tong Liverpool University (China)

Consider a discrete-time annuity-immediate risk model in which the insurer is allowed to invest its wealth into a risk-free or a risky portfolio under a certain regulation. Then the insurer is said to be exposed to a stochastic economic environment that contains two kinds of risks, the insurance risk and the financial risk. We are interested in the ruin probability and the tail behaviour of maximum of the stochastic discounted values of aggregate net loss with Sarmanov or Farlie-Gumbel-Morgenstern dependent insurance and financial risks. We derive some asymptotic formulas for the finite-time ruin probability with lighted-tailed or moderately heavy-tailed insurance risk in a discrete-time risk model with a risk-free or risky investment. As an extension, we improve our results to the case of extreme risks, which arise from rare events, by combining some simulation with asymptotics, to compute the ruin probabilities more efficiently.

Key Words: asymptotics, heavy-tailed distribution, light-tailed distribution, financial risk, insurance risk.

Extremes of transient Gaussian fluid queues

Session C23
June 28, 9:00
Room: Data

Krzysztof Dębicki
University of Wrocław (Poland)

Peng Liu
University of Lausanne (Switzerland)

This contribution investigates asymptotic properties of transient queue length process

$$Q(t) = \max\left(x + X(t) - ct, \sup_{0\le s\le t}(X(t) - X(s) - c(t-s))\right), \quad t \ge 0$$

in Gaussian fluid queueing model, where input process X is modeled by a centered Gaussian process with stationary increments, $c > 0$ is the output rate and $x = Q(0) \ge 0$. More specifically, under some mild conditions on X, exact asymptotics of

$$\mathbb{P}\left(Q(T_u) > u\right) \quad \text{and} \quad \mathbb{P}\left(\sup_{[0,T_u]} Q(t) > u\right),$$

as $u \to \infty$, are derived. The play between u and T_u leads to two qualitatively different regimes: (A) short-time horizon when T_u is relatively small with respect to u; (B) moderate- or long-time horizon when T_u is asymptotically much larger than u. As a by-product, some implications for the speed of convergence to stationarity of the considered model are discussed.

Key Words: transient queue, overload probability, exact asymptotics, Gaussian process, generalized Pickands-Piterbarg constant.

Modelling extremes of Markov chains

Session C43
June 30, 11:00
Room: Data

Anthony C. Davison
École Polytechnique Fédérale de Lausanne (Switzerland)

Jonathan A. Tawn
Lancaster University (UK)

Thomas Lugrin
École Polytechnique Fédérale de Lausanne (Switzerland)

Standard approaches to modelling extremes of stationary time series with short-range dependence typically involve pre-processing of the series and marginal modelling of subjectively selected peaks. Such methods heavily rely on a pre-processing stage, which can yield badly biased estimates of related risk measures. Assuming that the series is stationary and Markovian, we use an all-in-one approach where

the marginal distribution is fitted simultaneously with a conditional tail model for the Markov kernel. The dependence model covers a broad class of extremal dependence structures and includes information from non-extreme events which contribute through a censored likelihood. The model provides a unified framework in which marginal and dependence features can be estimated simultaneously, yielding efficient estimates and providing a natural assessment of the full uncertainty of the model. An application to river flows illustrates our approach.

Session C7
June 26, 16:00
Room: Data

Renewal theory for extremal Markovian sequences of the Kendall type

Barbara Jasiulis-Gołdyn
University of Wrocław

Jolanta Misiewicz
Warsaw University of Technology (Poland)

Karolina Łukaszewicz
University of Wrocław (Poland)

We deal with the renewal theory for a class of extremal Markov chains with respect to a generelized convolution. Borowiecka-Olszewska et al. (2015) [Lévy processes and stochastic integral in the sense of generalized convolution. *Bernoulli* **24**(1), 2513–2551] defined Markov processes generated by generalized convolutions. We consider a generalized convolution called Kendall convolution defined by:

$$\delta_x \triangle_\alpha \delta_1 = x^\alpha \pi_{2\alpha} + (1 - x^\alpha)\delta_1, \quad x \in [0, 1],$$

where $\pi_{2\alpha}(dx) = 2\alpha x^{-2\alpha-1}\mathbf{1}_{[1,\infty)}(x)dx$.

Jasiulis-Gołdyn and Misiewicz (2015) [Classical definitions of the Poisson process do not coincide in the case of weak generalized convolutions. *Lithuanian Mathematical Journal* **55**(4), 518–542] introduced Poisson processes in the generalized convolution algebras. Here we consider renewal processes for any unit step distributions.

For Kendall random walk $\{X_n : n \in \mathbb{N}_0\}$ we define Δ_α - renewal process $\{N(t) : t \geq 0\}$ as follows:

$$N(t) = \begin{cases} \inf\{n \colon X_{n+1} > t\} \\ \infty \ \text{ if such } n \text{ does not exist.} \end{cases}$$

We are interested in Δ_α - renewal function, which is defined as $R(t) = EN(t)$. We examine its properties in the Williamson transform terms. We describe it in Jasiulis-Gołdyn at al. (2017) [Renewal theory for extremal Markov sequences of the Kendall type, *in preparation*].

Key Words: Kendall random walk, extremal Markov chain, Williamson transform, renewal theory.

Estimation of uncertainties in intensity duration frequency curves of extreme rainfall - a regional analysis

Session C2
June 26, 15:00
Room: Chip

Juliette Blanchet
Université Grenoble Alpes, CNRS, IRD, Grenoble INP, IGE (France)

Victor Mélèse
Université Grenoble Alpes, CNRS, IRD, Grenoble INP, IGE (France)

In recent years, Mediterranean storms have caused serious damages to people and infrastructure in southeastern France. One way to prevent damages from those storms is to compute their probability of occurrence, i.e. their return period. Nevertheless, return periods of interest for risk evaluation amount usually to several hundreds of years, while series at disposal are most of the time much shorter. This is for example the case of the hourly rainfall data in our study region, which are measured since the 80's. Thus estimation of a 100-year return level, for instance, relies on extrapolation using statistical models, which makes results uncertain. Furthermore different temporal scales are usually involved in extreme rainfall events. For example the location of rainfall maxima usually depends on the considered duration. Thus, complete risk evaluation in a region involves the computation of return levels as a function of duration. This is precisely what provide the Intensity-Duration-Frequency (IDF) curves. We propose in this presentation a regional study of uncertainties in IDF curves derived from point-rainfall maxima. We develop two generalized extreme value models based on the simple scaling assumption, first in the frequentist framework and second in the Bayesian framework. We analyze the impact of these frameworks on the estimation of uncertainties. Within the frequentist framework uncertainties are obtained (i) with the normal asymptotic theorem and (ii) with a bootstrap method. Within the Bayesian framework uncertainties are obtained from the posterior densities. This analysis is applied to the Mediterranean region of southeastern France. A network of 420 hourly raingages covering the last 30 years and spanning the range one hour – one week is used to analyze the impact of both frameworks on the estimation of uncertainties. It is shown that the Bayesian framework is more suitable than the frequentist one to evaluate uncertainties.

Key Words: rainfall, extreme values, Bayesian method, uncertainty.

Testing independence (and asymptotic independence)

Session C36
June 29, 16:30
Room: Chip

Christian Genest
McGill University (Canada)

Olivier Roustant
École des Mines de Saint-Étienne (France)

Cécile Mercadier
Institut Camille Jordan, Université Claude Bernard Lyon 1 (France)

A new, powerful rank-based procedure is proposed for testing independence between continuous random variables. The procedure, which is consistent, involves a functional decomposition of the empirical copula process under independence which is akin to, but different from, the Moebius decomposition successfully used for the same purpose, e.g., in Genest and Rémillard (2004) [Tests of independence and randomness based on the empirical copula process. *Test* **13**, 335–369], in Kojadinovic and Holmes (2009) [Tests of independence among continuous random vectors based on Cramér-von Mises functionals of the empirical copula process. *Journal of Multivariate Analysis* **100**, 1137–1154] and references therein. The limiting distribution of the test statistic is determined under the null and the finite-sample power of the test is assessed via simulation.

An extension of the procedure will also be considered in which the hypothesis that the variables are asymptotically independent is assessed by checking whether the true underlying copula belongs to the domain of attraction of independence.

Key Words: testing (asymptotic) independence, functional decomposition, empirical copula process.

Session C41
June 30, 9:30
Room: Pi

Regression type models for extremal dependence

Miguel de Carvalho
University of Edinburgh (UK)

Valérie Chavez-Demoulin
University of Lausanne (Switzerland)

Linda Mhalla
University of Geneva (Switzerland)

We propose a vector generalised additive modelling framework for taking into account the effect of covariates on angular density functions in a multivariate extreme value context. Our model is motivated by the need of quantifying how joint extreme losses in a stock market may change as a function of covariates.

Our approach is constructed from covariate-dependent versions of well-known parametric models for the spectral density of a multivariate extreme value distribution, such as the logistic, Dirichlet, and the Hüsler–Reiss models. Fitting is conducted by a maximum penalized likelihood estimator based on the Poisson Point Process approach. The methods are applied to simulated data, and a case study in finance is used to illustrate the main concepts introduced along the talk.

Key Words: angular density, covariate-adjustment, extremal dependence, penalised log-likelihood, regression splines.

The eigenstructure of sample covariance matrices for high-dimensional heavy-tailed stochastic volatility models

Session I2
June 26, 10:50
Room: Chip

Johannes Heiny
University of Aarhus (Denmark)

Thomas Mikosch
University of Copenhagen (Denmark)

We are interested in the asymptotic behavior of the eigenvalues of the sample covariance matrix $\mathbf{X}\mathbf{X}'$ where the data matrix $\mathbf{X} = \mathbf{X}_n = (X_{it})_{i=1,\ldots,p;t=1,\ldots,n}$ has the structure of a stochastic volatility model, i.e., $X_{it} = \sigma_{it}Z_{it}$ and the stationary volatility field (σ_{it}) is independent of the iid field (Z_{it}). We assume that Z_{it} has regularly varying tails with index $\alpha \in (0, 4)$ and that σ_{it} has lighter tails. We are interested in the case when the dimension $p = p_n \to \infty$. Then the eigenvalues of $\mathbf{X}\mathbf{X}'$ are essentially determined by the diagonal elements $\mathbf{X}_{ii}$, $i = 1, \ldots, n$. We consider limit theory of Poisson-type for the largest eigenvalues and also discuss the structure of the corresponding eigenvectors.

Key Words: regular variation, stochastic volatility, sample covariance matrix, eigenvalues, high dimension.

Extreme events in a changing world

Session C34
June 29, 14:30
Room: Data

Holger Kantz
Max Planck Institute for the Physics of Complex Systems (Germany)

Philipp Müller
Max Planck Institute for the Physics of Complex Systems (Germany)

From biological systems to financial markets to the earth's climate: natural processes are in constant change. This intrinsic non-stationarity makes them hard to describe and our forecasts in terms of return levels less accurate and reliable.

In my talk I will present the benefits and pitfalls of different methods to deal with non-stationarities in time series in the context of the extreme value theory. I will do so from a numerical perspective, point out different ways to improve the fitting procedure, and give a best-practice guide using the example of daily maximum temperatures time series.

Key Words: GEV, GP, fitting, trends, non-stationary.

The asymptotics of the maxima for the distribution of the generalized St. Petersburg game

Session C3
June 26, 14:30
Room: Data

Toshio Nakata
University of Teacher Education Fukuoka (Japan)

In this talk, we consider the maxima of payoffs for the generalized St. Petersburg game. The maxima for the original St. Petersburg game cannot be normalized to converge to a non-degenerate limit distribution. However, tuning the parameters appearing in the generalization of the game, we show the normalized maxima converge to the Fréchet distribution.

Key Words: St. Petersburg game, maxima, Fréchet distribution.

Session I7
June 29, 9:00
Room: Boole

A Bayesian error-in-variable framework to study climate extremes in a detection and attribution context

Brian Reich
North Carolina State University (USA)

Philippe Naveau
Laboratoire des Sciences du Climat et l'Environnement, CNRS (France)

The field of statistics has become one of the mathematical foundations in detection and attribution (D&A) studies, especially with regard to assessing uncertainties in climate change problems. The classical paradigm in D&A is to infer regression coefficients in order to quantity expected response patterns to different external forcings (e.g. greenhouse gases). Although convenient, this approach has a few shortcomings. For example, how to interpret regression coefficients if observations and forced climate runs are both tainted by a large error. To bypass this hurdle, Ribes et al. (2016) [A new statistical approach to climate change detection and attribution. *Climate Dynamics* **48**, 367–386] recently proposed an Error-In-Variable (EIV) framework where regression coefficients were removed from the analysis. Still, this setup based on the Gaussian assumption is not appropriate to handle extremes like annual temperatures maxima. As the key objective of D&A is to discriminate between different causes, we propose, study and discuss how to estimate and compare relevant posterior probabilities to compare different EIV models based on the Generalized Extreme Value distributions within a Bayesian hierarchical framework within a EIV model.

Key Words: Bayesian hierarchical model, error-in-variables, causality theory, GEV, block maxima.

Session C35
June 29, 14:00
Room: Pi

Modelling average minimum daily temperature using extreme value theory with a time varying threshold

Caston Sigauke
University of Venda (South Africa)

Murendeni Maurel Nemukula
University of Limpopo (South Africa)

In this paper we present an application of the Generalized Pareto Distribution (GPD) in the modelling of average minimum daily temperature in South Africa for the period January 2000 to August 2010. A penalized cubic smoothing spline is used as a time varying threshold as well as to cater for seasonality. We then extract excesses (residuals) above the cubic spline and fit a non-parametric mixture model to get a sufficiently high threshold. The data exhibit evidence of short-range dependence and high seasonality which lead to the declustering of the excesses above the sufficiently high threshold and fit the GPD to cluster maxima. The parameters are estimated using the maximum likelihood method. The estimate of the shape parameter shows that the Weibull family of distributions is appropriate in modelling the upper tail of the distribution of average minimum daily temperature in South Africa. The bootstrap resampling method is used as an assessment tool for uncertainty in the parameter estimation. This study has shown that the use of the penalized cubic smoothing spline as a time varying threshold to time series data which exhibits strong seasonality provides a good fit of the GPD to cluster maxima. This results in accurate estimates of return levels.

Key Words: declustering, extreme value theory, time varying threshold, temperature.

The block maxima method in extremum estimation

Session C11
June 27, 12:00
Room: Data

Christopher Jeffree
University of Reading (UK)

Cláudia Neves
University of Reading (UK)

Over the last decade there has been an astonishing growth in the statistical techniques to analyse extreme values. Nevertheless, it is customary to regard the Generalised Extreme Value distribution as the probabilistic instrument fitting a sample of block maxima, typified by the classical maximum likelihood (ML) approach for inference in extreme values. The more general maximum Lq-likelihood method has a notable advantage to the usual ML estimation ascribed to $q = 1$: with small up to moderate sample sizes, a proper choice for the distortion parameter $q > 0$ can deploy the variance in mitigating the mean squared error, swamping the role of bias. By the same token, an alternative class of parametric estimators stems from the maximum product of spacings (MSP) method with their obvious extension to the MSPq class, drawing on a similar setting.

In this talk, we will assess the current state of development and usage of these two classes of estimators and outline a semi-parametric approach to both methods by assuming that the distortion parameter $q = q_m$ depends on the size of blocks m rather than the sample size n. We will proceed via simulation, addressing how the choice of q_m crosses over to the estimation of high quantiles, including the finite right endpoint. The simulation results will be partially mirrored in the practical

application to the annual maxima of Lowestoft sea levels.

Session I1
June 26, 10:10
Room: Boole

PCA and ICA for multivariate stable distributions

Tuncay Alparslan
Department of Energy (USA)

Richard Davis
Columbia University (USA)

Sidney Resnick
Cornell University (USA)

John Nolan
American University (USA)

Principal Component Analysis (PCA) is a standard tool for feature extraction from multivariate data. We show that in the presence of heavy tailed data, the usual method of PCA can give poor results. A new method is presented that more accurately recovers the structure when the data is stably distributed. Independent Component Analysis (ICA) is another technique for determining significant factors in multivariate data. In the presence of heavy tailed data standard ICA can be misleading and we propose a new method of determining factors.

Key Words: PCA, ICA, multivariate stable laws.

Session C4
June 26, 14:30
Room: Pi

Extremes of risk functionals for samples with a limit shape

Guus Balkema
University of Amsterdam (The Netherlands)

Natalia Nolde
University of British Columbia (Canada)

Questions related to risk assessment are often formulated in terms of functionals of a random vector. We investigate the extremal behaviour of homogeneous functionals. It is assumed that samples of independent and identically distributed vectors from the underlying distribution can be scaled to converge onto a deterministic limit set. As a consequence of the limit set existence, we show that the tails of linear and non-linear homogeneous functionals of the components exhibit remarkable similarity. The $(1-\alpha)$ quantiles of these functionals are asymptotically equal, for $\alpha \to 0$, up to a constant determined by the limit set. As multivariate light-tailed distributions often have tail independence, asymptotic dependence in such cases can

be measured by means of the coefficient of residual tail dependence. Our results suggest a simple geometric procedure which in many cases will yield the residual dependence coefficient between homogeneous functionals of the underlying random vector.

Key Words: homogeneous functionals, tail behaviour, asymptotic independence, coefficient of residual tail dependence, limit set.

On a class of max-stable processes represented via ℓ^p norms

Session I6
June 27, 9:40
Room: Pi

Marco Oesting
Universität Siegen (Germany)

Reich and Shaby (2012) [A hierarchical max-stable spatial model for extreme precipitation. *The Annals of Applied Statistics* **6**, 1430–1451] introduced a hierarchical max-stable model which enjoys growing popularity as a model for spatial extremes with a nugget effect. While the common spectral representation of max-stable processes includes the maximum over an infinite number of processes, this model is represented as a finite sum. In this talk, we present a more general representation of max-stable models based on ℓ^p norms. We give a full characterization of the resulting class of processes and provide formulae to switch between different representations. Further, we discuss properties of processes with an ℓ^p norm based representation in comparison with their denoised counterparts.

Key Words: Bayesian hierarchical model, nugget effect, spatial model, spectral representation.

Penultimate modeling of spatial extremes through max-infinitely divisible processes

Session I12
June 30, 14:00
Room: Pi

Raphaël Huser
King Abdullah University of Science and Technology (Saudi Arabia)

Emeric Thibaud
École Polytechnique Fédérale de Lausanne (Switzerland)

Thomas Opitz
BioSP, INRA (France)

Extreme value limit theory for stochastic processes has motivated the use of max-stable models for the statistical analysis of spatial extremes. However, fitting such asymptotic models to the observed maxima (taken over blocks of time) may be problematic when the asymptotic stability in the dependence structure is not yet achieved in finite-sample data. This issue is particularly serious when data are

asymptotically independent, which leads to full independence in the corresponding max-stable limit. We here aim to provide practically useful models that are more flexible and more realistically account for such discrepancies between the data and asymptotic theory. In this work, we develop models pertaining to the wider class of max-infinitely divisible processes, which possess properties that are natural for maxima data by imposing certain restrictions on the dependence structure like positive association. In particular, such models may be adapted in a coherent way to a change of the size of the blocks over which maxima are calculated. We propose a parametric point process construction for spatial max-infinitely divisible models allowing for asymptotic independence while keeping the flexible max-stable extremal-t model as a special case. After developing efficient parameter estimation through pairwise likelihood, we demonstrate the usefulness of this new modeling framework with meteorological data.

Key Words: asymptotic dependence and independence, blockwise maxima, max-id processes, spatial extremes, spatial dependence.

Distribution of maximal deviation for Lévy density estimators

Session C23
June 28, 9:30
Room: Data

Valentin Konakov
Higher School of Economics (Russia)

Vladimir Panov
Higher School of Economics (Russia)

This talk is essentially based on Konakov and Panov (2016) [Sup-norm convergence rates for Lévy density estimation. *Extremes* **19**(3), 371-403]. In the paper, we consider a one-dimensional Lévy process X_t with Lévy measure ν, which is assumed to have a density $s(\cdot)$. Nonparametric statistical inference for $s(\cdot)$ from the observations of X_t has got much attention in the literature, and typical results concern the quality of the estimators in terms of quadratic risk and minimax convergence rates. The aim of our research is slightly different: we focus on the asymptotic distribution of the maximal deviation and establish the rates of convergence to this distribution.

More precisely, we consider the projection estimates of $s(\cdot)$:

$$\hat{s}_n(x) = \frac{1}{n\Delta} \sum_{r=1}^{d} \left[\sum_{k=1}^{n} \varphi_r \left(X_{k\Delta} - X_{(k-1)\Delta} \right) \right] \varphi_r(x), \quad x \in D,$$

where $X_0, \dots, X_{n\Delta}$ are available observations of the process X_t, and $\{\varphi_r(x) : D \to \mathbb{R}, r = 1, \dots, d\}$ is an orthonormal collection of functions (e.g. trigonometric basis, Legendre polynomials, wavelets).

The contribution of this paper is twofold. First, we derive the asymptotic behaviour of the distribution of the maximal deviation

$$\mathcal{D}_n := \sup_{x \in D} \left(\frac{|\hat{s}_n(x) - s(x)|}{\sqrt{s(x)}} \right)$$

for a broad class of $\{\varphi_r\}$. Second, we show that the exact rates of convergence presented in previous articles on this topic are logarithmic and construct the sequence of accompanying laws, which approximate the deviation distribution with polynomial rate. Finally, we provide the sequence of accompanying laws with power rate of convergence. Our results are based on the idea to reformulate the problems in terms of Gaussian processes of some special type and to further analyze these Gaussian processes using the techniques from Piterbarg (1996) [*Asymptotic Methods in the Theory of Gaussian Processes and Fields*. AMS, Provence].

Key Words: Lévy density, maximal deviation, nonparametric inference, projection estimates.

Modelling time series extremes

Session C22
June 28, 9:30
Room: Chip

Ioannis Papastathopoulos
University of Edinburgh (UK)

Weak long-range dependence of random processes leads to clusters of extreme values. In particular, the generating mechanism of extreme events in time, has a nonhomogeneous Poisson process character that describes the peaks of the events together with a cluster distribution that can also vary. This representation provides an insightful modelling basis according to which asymptotically motivated statistical models can be fitted to observed data in order to infer summaries of the cluster distribution such as the mean extent or maximum severity of the event. We present a broad framework for modelling clusters of extreme events based on limiting forms of Markov processes with higher order memory. Emphasis is placed on both asymptotically dependent and asymptotically independent stochastic processes and their corresponding limiting dynamics described by specific stochastic difference equations. Solutions result in new forms of tail chains that encompass a very rich structure and motivate a parsimonious statistical model by exploiting homogeneity properties of drift and volatility functions. Statistical inference is discussed via a penalised likelihood approach that facilitates temporal coherence and estimation of the order of the Markov chain.

Key Words: higher-order memory, Markov chains, stochastic difference equations, tail chain, penalised likelihood.

Session C6
June 26, 16:30
Room: Chip

Multivariate extremal analysis methodology in function of structural response

Hélène Pineau
ACTIMAR (France)

Marc Prevosto
IFREMER (France)

Nicolas Raillard
IFREMER (France)

Design optimisation is crucial as offshore structures are exposed to deeper and harsher marine conditions. The structure behaviour is dependent on several joint environmental parameters (wind, wave, currents, etc). The environmental contours, i.e. the 2-D or 3-D curves where the N-year return level should lie, are useful representations to provide multivariate design criteria. However, these contours may provide rather different design points according to their calculation process.

In this work, we have proposed to use response meta-model for the inter-comparison of some state-of-the-art methods available for modelling multivariate extremes. In a first step, the key metocean parameters and the associated simplified load models of response for selected application cases (tensions in mooring lines, FPSO roll, ...) were set-up. Several multivariate extremal analysis methods were then applied to derive the environmental contours. Among the conditional models, joint parametric laws (such as Weibull, log-normal, ...), Heffernan and Tawn simulations and several extreme value dependence function model were investigated. Two ways of contour building were applied: Inverse First Order Reliability Model (iFORM) and physical-space Huseby contouring method.

Applications are presented for the selection of design point associating wind, wave and current parameters (including directions) considering tension in a mooring line and FPSO roll. The comparison between the resulting design points for several configurations (using 2 parameters, 3 parameters, per directional sector, ...) will be shown, in addition to the extreme loads obtained from the meta-models. A special attention was also paid to the selection of storm data.

Key Words: extremal analysis, multivariate, environmental contours, structural safety.

Efficient estimation of return value distributions from non-stationary marginal extreme value models using Bayesian inference

Session C41
June 30, 10:00
Room: Pi

Philip Jonathan, Emma Ross
Shell Global Solutions (UK)

David Randell
Shell Global Solutions (UK)

Extreme values of an environmental variable can be estimated by fitting the generalised Pareto (GP) distribution to a sample of exceedances of a high threshold. In oceanographic applications, threshold and GP model parameters are typically functions of physical covariates. A fundamental difficulty is selection or estimation of an appropriate threshold or interval of thresholds, of particular concern since inferences for return values vary with threshold choice.

Hence, following Randell et al. (2016) [Bayesian inference for non-stationary marginal extremes. *Environmetrics* **27**, 439–450], a piecewise gamma-GP model for a sample of storm peak significant wave height, non-stationary with respect to storm directional and seasonal covariates, is estimated here using Bayesian inference. Quantile regression is used to partition the sample prior to independent gamma (body) and GP (tail) estimation. An ensemble of independent models, each member of which corresponds to a choice of quantile probability form a wide interval of quantile threshold probabilities (essentially $(0, 1)$), is estimated. The methodology is illustrated at a South China Sea location.

The estimated posterior predictive return value distribution (for a long return period of the order of 10,000 years) is a particularly useful diagnostic tool for threshold selection, since this return value is a key deliverable in metocean design. Estimating the distribution using Monte Carlo simulation becomes computationally demanding as return period increases. We present an alternative numerical integration scheme, the computation time for which is effectively independent of return period, dramatically improving computational efficiency for longer return periods.

Key Words: Bayesian, covariate, spline, return value.

Session C24
June 28, 9:30
Room: Pi

Estimating the maximum earthquake magnitude

Jan Beirlant
KU Leuven (Belgium)

John H.J. Einmahl
Tilburg University (The Netherlands)

Andrzej Kijko
University of Pretoria (South Africa)

Tom Reynkens
KU Leuven (Belgium)

The area-characteristic, maximum possible earthquake magnitude T_M is required by the earthquake engineering community, disaster management agencies and the insurance industry. The Gutenberg-Richter law predicts that earthquake magnitudes M follow a truncated exponential distribution with survival function

$$\mathbb{P}(M > m) = \frac{e^{-\beta m} - e^{-\beta T_M}}{e^{-\beta t_M} - e^{-\beta T_M}}, \quad t_M < m < T_M.$$

In the geophysical literature several estimation procedures were proposed. See for instance Kijko and Singh (2011) [Statistical tools for maximum possible earthquake estimation. *Acta Geophysica* **59**, 674–700] and the references therein.

Estimation of T_M is of course an extreme value problem to which the classical methods for endpoint estimation could be applied. We argue that recent methods on truncated tails at high levels discussed in Beirlant et al. (2016) [Tail fitting for truncated and non-truncated Pareto-type distributions. *Extremes* **19**, 429–462] and Beirlant et al. (2017) [Fitting tails affected by truncation. *arXiv:1606.02090*] constitute an appropriate setting for this estimation problem. Next to estimates for the endpoint, we also consider upper confidence bounds to quantify uncertainty of the point estimates. We compare methods from the extreme value and geophysical literature through simulations. Finally, we apply the different methods to the magnitude data for the earthquakes induced by gas extraction in the Groningen area (the Netherlands).

Key Words: Gutenberg-Richter law, truncated distributions, endpoint estimation.

Sample path large deviations for heavy tails: the principle of multiple big jumps

Session C25
June 28, 11:30
Room: Boole

Jose H. Blanchet
Columbia University (USA)

Bert Zwart
Centrum Wiskunde & Informatica (The Netherlands)

Chang-Han Rhee
Centrum Wiskunde & Informatica (The Netherlands)

Many rare events in man-made networks exhibit heavy-tailed phenomena: for example, file sizes and delays in communication networks, financial losses, and magnitudes of systemic events such as the size of a blackout in a power grid. While the theory of large deviations has been wildly successful in providing systematic tools for understanding rare events in light-tailed settings, the theory developed in the heavy-tailed setting has been mostly restricted to model-specific results or results pertaining to events that are caused by a single big jump. In this talk, we present our recent result that goes beyond such restrictions and establish sample-path large deviations for a very general class of rare events associated with heavy-tailed random walks and Lévy processes. We will illustrate the implications of our results in the analysis of rare events that arise in mathematical finance, actuarial science, and queueing theory.

Key Words: sample path large deviations, regular variation, Lévy processes, random walks, $\mathbb{M}$-convergence.

Extreme value modelling of water-related insurance claims

Session C13
June 27, 14:30
Room: Boole

Emma F. Eastoe, Jonathan A. Tawn
Lancaster University (UK)

Arnoldo Frigessi
Oslo Centre for Biostatistics and Epidemiology (Norway)

Christian Rohrbeck
Lancaster University (UK)

This talk considers the dependence between weather events, e.g. rainfall or snowmelt, and the number of water-related property insurance claims. Weather events which cause severe damages are of general interest, decision makers want to take efficient actions against them while the insurance companies want to set adequate premiums. The modelling is challenging since the underlying dynamics vary across geographical regions due to differences in topology, construction designs and climate. We develop new methodology to improve the existing models which fail to model

high numbers of claims. The statistical framework is based on both mixture and extremal mixture modelling, with the latter being based on a discretized generalized Pareto distribution. Furthermore, we propose a temporal clustering algorithm and derive new covariates which lead to a better understanding of the association between claims and weather events. The modelling of the claims, conditional on the locally observed weather events, both fits the marginal distributions well and captures the spatial dependence between locations. At the end of the talk, our methodology is applied to three cities across Norway to demonstrate its benefits.

Key Words: mixture modelling, extremal mixture, extremal dependence, insurance claims, spatio-temporal modelling.

Session C12
June 26, 11:30
Room: Pi

Human life is unlimited – but short

Holger Rootzén
Chalmers University of Technology (Sweden)

Does the human lifespan have an impenetrable biological upper limit which ultimately will stop further increase in life lengths? Answers to this question are important for our understanding of the ageing process, and for the organization of society, and have led to intense controversies. Demographic data for humans have been interpreted as showing existence of a limit close to the age, 122.45 years, of the longest living documented human, Jeanne Calment, or even as indication of a decreasing limit, but also as evidence that a limit does not exist. This paper studies what data says about human mortality after age 110. We show that in north America, western Europe, and Japan the yearly probability of dying after age 110 is constant and about 51% per year. Hence there is no finite limit to the human lifespan. Still, given the present stage of biotechnology, it is unlikely that during the next 25 years anyone will live longer than 130 years in these countries. Data, remarkably, show little difference in mortality after age 110 between men and women, between earlier and later periods, between ages, or between persons with different lifestyles or genetic backgrounds. These results can help testing biological theories of aging and aid early confirmation of success of efforts to find a cure for ageing.

Key Words: supercentenarians, extreme human lifelengths, cure for ageing, left truncation, right truncation.

Session C32
June 29, 12:00
Room: Pi

An Australia wide model for rainfall extremes

David Karoly, Alec Stephenson, Peter Taylor
University of Melbourne (Australia)

Kate Saunders
University of Melbourne (Australia)

Extreme rainfall events have major societal impacts including flash flooding, crop destruction and infrastructure damage. To mitigate the potential consequences of these extreme events, we require accurate statistical models. As Australia is one of the largest countries by area, with a varied climate and complex topography, a one size fits all model for rainfall extremes is problematic.

To address this, we have clustered the station data. A regional model is then produced using the cluster centres, while local models for rainfall extremes are fit to the individual clusters. For the local model, we fit a max-stable process to model spatial extremes with dependence. The discretisation of the domain could be perceived as a model limitation, but we exploit it for simulation. Example simulations of the annual maxima extreme rainfall field are given on a country wide scale.

Key to the model is the clustering method, see Bernard et al. (2013) [Clustering of maxima: spatial dependencies among heavy rainfall in France. *Journal of Climate* **26**(20), 7929–7937]. The distance measure used in clustering is based on the extremal coefficient, so clusters capture regions with similar extremal dependence. The distance measure is estimated non-parametrically from the F-madogram; see Cooley et al. (2006) [Variograms for spatial max-stable random fields. In: P. Bertail et al. (eds), *Dependence in Probability and Statistics*, Springer, New York]. Non-parametric estimation of the distance measure is advantageous as no information about climate or topography is required to form geographically homogeneous clusters. The parsimony of our fitted models is therefore improved through identification of homogeneous regions.

Key Words: rainfall, spatial extremes, max-stable, clusters.

Marked point process adjusted tail dependence analysis for high-frequency financial data

Session C43
June 30, 11:30
Room: Data

Alexander Malinowski
Georg-August-Universität Göttingen (Germany)

Zhengjun Zhang
University of Wisconsin-Madison (USA)

Martin Schlather
Universität Mannheim (Germany)

Although the extremes of high-frequency financial transaction data have a huge economic impact, basic characteristics of the data have not been addressed up to now. To capture dependence between the tail behavior of inter-transaction returns and the pattern of transaction times, this paper combines marked point process (MPP) theory with extreme value analysis. Suitable measures of interaction are provided, based on second-order moments of MPPs. Applying these measures to financial transaction data, it is verified that the extreme value index of the return distribution is indeed locally increased, i.e. on the scale of minutes, by the existence

of surrounding transactions. A simulation study underpins the observed effects and enables assessing the finite sample properties of the respective estimators. Further, asymptotic results on the estimators are given.

Session I10
June 30, 15:20
Room: Boole

Regularly varying Markov trees

Gildas Mazo
Université Grenoble Alpes (France)

Johan Segers
Université catholique de Louvain (Belgium)

Extreme values of regularly varying Markov chains can be described in terms of the limiting conditional distribution of the normalized chain given that it is large at a particular time instant. The limit distributions are called forward and backward tail chains, according to the time direction considered. Viewing a chain as a tree consisting of a single, long branch, we seek for generalizations to general Markov trees, i.e., random vectors whose dependence structure is governed by a tree representing a set of conditional independence relations together with a collection of bivariate distributions along the tree edges. As for Markov chains, we find that extremal dependence of such Markov trees can be described in terms of a collection of tail trees, each tree describing the limit distribution of the rescaled Markov tree given that its value at a particular node is large. Moreover, the time-change formula for tail chains generalizes to a relation between these tail trees. Tail trees can be used to compute quantities such as the number of nodes in the graph affected by a shock at a particular node or the probability that a particular part of the graph will be affected by a shock in another part of the graph. Moreover, specifying the graph structure and the bivariate distributions along the edges provides a construction method for max-stable models.

Key Words: regular variation, graphical model, conditional independence, tail chain.

Session C5
June 26, 17:00
Room: Boole

A conditional limit theorem for a bivariate representation of a univariate random variable and conditional extreme values

Philippe Barbe
CNRS Paris (France)

Miriam Isabel Seifert
Ruhr-Universität Bochum (Germany)

We first consider a real random variable X represented through a random pair (R, T) as $X = R \cdot u(T)$ with a deterministic function u. Under quite weak assumptions we prove a limit theorem for (R, T) given $X > x$, as x tends to infinity. The novelty of our approach is that our limit theorem for the representation of the univariate random variable X permits to obtain in an elegant manner conditional limit theorems for random pairs $(X, Y) = R \cdot (u(T), v(T))$ given that X is large. It allows to deduce new results as well as to recover (under considerably weaker assumptions) results obtained previously in the literature. Consequently, we provide the way for a better understanding and systematization of limit statements for the conditional extreme value models.

Key Words: representation of random variables, conditional extreme value model, elliptical distributions, limit theorems, regular variation.

Asymptotic behavior of maximum for sequences of Gaussian random fields

Session C15
June 27, 15:00
Room: Data

Enkelejd Hashorva, Zhongquan Tan
University of Lausanne (Switzerland)

Oleg Seleznjev
Umeå University (Sweden)

We consider the large values and the mean of the uniform norms for a sequence of Gaussian random fields with continuous sample paths. The convergence of the normalized uniform norm to the standard Gumbel (or *double exponential*) law is derived for distributions and means. The main results are obtained from the Poisson convergence of the associated point process of exceedances for a general class of Gaussian random fields. Both isotropic and non-homogeneous Gaussian random fields are investigated. These results extend previous results obtained for Gaussian processes in Hüsler at al. (2003) [On convergence of the uniform norms for Gaussian processes and linear approximation problems. *Annals of Applied Probability* **13**, 1615–1653], Seleznjev (1991) [Limit theorems for maxima and crossings of a sequence of Gaussian processes and approximation of random processes. *Journal of Applied Probability* **28**, 17–32], and Seleznjev (2005) [Asymptotic behavior of mean uniform norms for sequences of Gaussian processes and fields. *Extremes* **8**, 161–169]. As an application, the asymptotic distribution is obtained for a sequence of tensor product Gaussian random fields. More possible applications are related with investigation of the approximation rate and simulation accuracy in the uniform norm for a Gaussian random field.

Key Words: maxima of Gaussian random fields, Gumbel limit theorem, pth-moment convergence.

Session C29
June 29, 12:00
Room: Boole

Unbiased inference for spectral distributions

Anthony C. Davison
École Polytechnique Fédérale de Lausanne (Switzerland)

Claudio Semadeni
École Polytechnique Fédérale de Lausanne (Switzerland)

Several parametric and non-parametric approaches to estimate the spectral distribution of max-stable distributions have been proposed in the bivariate setting, some of which can also be used in higher dimensions. These approaches often rely on the pseudo-angles of observations of large but finite magnitude which yields biased estimates. We introduce a transformation that yields bias-reduced estimates of the spectral distribution in the bivariate setting and which can be used in both the parametric and the non-parametric settings. The approach will be illustrated with applications to environmental and financial datasets.

Key Words: multivariate extremes, spectral distribution, unbiased estimate.

Session C4
June 26, 15:00
Room: Pi

The bivariate power-law distribution in complex systems

Isabel Serra
Centre de Recerca Matemàtica (Spain)

The free-scale assumption in several properties of complex systems requires the use of power-law distribution for univariate measures. We propose a model for describing the joint probability distribution of two magnitudes, each one of them with a free-scale distribution. The result provides the bivariate power-law that we will introduce. This natural definition holds the most important properties required in complex systems analysis: the model has to be closed by truncation, the model has to hold general free-scale assumption and the model has to describe the joint distribution thought relation between the two sets of units of measure.

The model was motivated working on avalanches processes. The marginal described by the energy in an acoustic emission experiments shows a law for extreme events, which is produced by the avalanches. This phenomenological model aimed to describe the joint probability distribution of two magnitudes: energy and duration of an avalanche. That is the main approach for modeling the bivariate law by extreme value distribution. Further analysis comparing the fit of real data has been developed. We compare the results with a several catalog of bivariate distributions, from the approach of extreme value copulas until several parametric forms proposed on literature.

Hierarchical scale mixtures for flexible spatial modeling of extremes

Session I12
June 30, 15:20
Room: Pi

Ben Shaby
Pennsylvania State University (USA)

Scale mixtures of Gaussian processes have emerged as desirable candidates for modeling extremal phenomena in space. They are intuitive, simple to describe constructively, and flexible in the types of extremal dependence that they they can represent. Inference for these models using censored likelihoods has been limited to very small datasets due to the presence of a high-dimensional Gaussian integral that must be evaluated numerically. Rather than integrating over a latent Gaussian process, we condition on it, expressing the model hierarchically. This way, we allow Markov chain Monte Carlo to do the hard integration, and open the door to inference on much larger datasets than were previously possible.

Modelling extremes arising from extratropical cyclones

Session C43
June 30, 12:00
Room: Data

Simon Brown
Met Office Hadley Centre (UK)

Jonathan A. Tawn
Lancaster University (UK)

Paul Sharkey
Lancaster University (UK)

Extratropical cyclones are dominant features of the European weather landscape that are responsible for strong winds and heavy rainfall. Accurate statistical modelling of weather extremes related to cyclones is essential to aid the design of robust infrastructure for defence against natural hazards. In recent years, such activity has been linked to widespread flooding events, such as Storm Desmond in December 2015.

Extreme value analysis has long been used to model severe weather events, but largely in a context that ignores the spatial and temporal aspects of the cyclone itself and the physics that generate the extremes of interest. In order to characterise the behaviour of these extremes in relation to the storm systems that drive them, we first construct a model that captures the spatial extent and severity of these weather systems. We adopt a simulation based approach that allows us to infer the long-term characteristics of cyclones based on analysis of observed data. Within this approach, we construct storm tracks that incorporate the history and locality of the storm process, while also accounting for the tail behaviour of storm severity.

However, we are ultimately interested in the behaviour of wind speeds relative to the storm centre. The second part of the talk will detail some exploratory results regarding this behaviour as well as a strategy for simulating wind fields conditional

on the properties of the storm track. We will also discuss some modelling issues, including land/sea differentiation, data contamination by convective events and incorporating modes of climatic variability.

Key Words: cyclone modelling, extreme value analysis, nonparametric methods, Markov chains, serial dependence.

Session C13
June 27, 15:00
Room: Boole

The devil is in the tails: regression discontinuity design with measurement error in the assignment variable

Zhuan Pei
Cornell University (USA)

Yi Shen
University of Waterloo (Canada)

Identification in a regression discontinuity (RD) design hinges on the discontinuity in the probability of treatment when a covariate (assignment variable) exceeds a known threshold. If the assignment variable is measured with error, however, the discontinuity in the relationship between the probability of treatment and the observed mismeasured assignment variable may disappear. Therefore, the presence of measurement error in the assignment variable poses a challenge to treatment effect identification. In this work we provide sufficient conditions to identify the RD treatment effect using very general information about the tail behavior of the measurement error. We prove identification separately for discrete and continuous assignment variables and study the properties of various estimation procedures. We illustrate the proposed methods in an empirical application, where we estimate Medicaid take-up and its crowd-out effect on private health insurance coverage.

Key Words: regression discontinuity design, measurement error, tail behavior.

Session C9
June 27, 12:00
Room: Boole

Discrete distributions whose truncated means have logarithmic order

Toshio Nakata
University of Teacher Education Fukuoka (Japan)

Takaaki Shimura
The Institute of Statistical Mathematics (Japan)

The St. Petersburg game is a well-known paradox caused by infinite mean. Although the distribution of this game does not satisfy the law of large numbers in the usual sense, it is relatively stable because the truncated mean has logarithmic order. We focus on the class of discrete distributions whose truncated means are

asymptotically logarithmic. It is shown that various asymptotic tail behaviors could appear in this class. Also, the number of random variables which attain the maxima is investigated according to the tail behavior.

Key Words: St. Petersburg game, truncated mean, tail behavior, maxima of i.i.d. random variables.

Determining the dependence structure of multivariate extremes

Session C36
June 29, 17:00
Room: Chip

Jonathan A. Tawn, Jenny Wadsworth
Lancaster University (UK)

Emma Simpson
STOR-i CDT, Lancaster University (UK)

When modelling multivariate extremes, it is important to consider the asymptotic dependence and asymptotic independence properties of the variables. This can have a complicated structure, with only certain subsets of variables taking their largest values simultaneously. In this talk, I will discuss a method that, given a set of data, aims to establish this asymptotic dependence structure.

It is a common method in multivariate extreme value analysis to consider variables in terms of their radial and angular components $(R, \mathbf{W})$. In this case, the angular component $\mathbf{W}$ takes values in the unit simplex and, conditioning on R being extreme, the position of mass on this simplex can reveal the asymptotic dependence structure of the variables of interest. In the bivariate case, this corresponds to deciding whether or not there is mass on the interior or the edges of the simplex, but in the d-dimensional case there are $2^d - 1$ sub-simplices that could contain mass. In reality, data will not lie exactly on the sub-simplices of the angular unit simplex, so assessing the asymptotic dependence structure is not a straightforward task.

We use the radial and angular components as a basis for our method. By partitioning the angular simplex, we aim to find the conditional probability that a point lies in a certain section given that it is extreme in terms of its radial component. This allows us to determine the asymptotic dependence structure of the variables, as well as the proportion of mass corresponding to each subset of the variables being simultaneously extreme.

Key Words: multivariate extreme value theory, asymptotic dependence, asymptotic independence.

Session C24
June 28, 10:00
Room: Pi

Process characteristic extreme value distributions

Andrzej Kijko
University of Pretoria (South Africa)

Alfred Stein
University of Twente (The Netherlands) & University of Pretoria (South Africa)

Ansie Smit
University of Pretoria (South Africa)

Most natural hazard databases for earthquakes, floods and tsunamis contain three types of data: prehistoric, historic and the most recent, instrumentally recorded events. Extreme Value Theory (EVT) and Generalized Extreme Value Theory (GEVT) are often applied to model these types of data by focusing on the extreme events. Weak events are often ignored but still can provide valuable information on the underlying physical process. In this presentation, we provide an alternative to the EVT and GEVT models for earthquakes. Likelihood functions describing the recurrence parameters of earthquake inducing processes are defined for each of the three data types. They are combined into a single likelihood function for the whole database. Potential incompleteness, uncertainty in size of the events, and uncertainty associated with the applied recurrence model are considered. The likelihood functions for the prehistoric and historic data are modelled using extreme value distributions that are not only characteristic of the underlying process, but also linked to the parent distributions of the process. The advantage of such modelling is that the process-characteristic extreme distribution preserves parameters of the parent distributions, and as a rule, provides more accurate process information than EV or GEVT modelling. During the presentation results and conclusions will be provided.

Key Words: incompleteness, uncertainty, derived extreme value distributions.

Session C22
June 28, 10:00
Room: Chip

Asymptotics of the order statistics for a process with a regenerative structure

Natalia Soja-Kukieła
Nicolaus Copernicus University (Poland)

Let $\{X_n : n \in \mathbb{N}\}$ be a stochastic process with an r-max-regenerative structure for some $r \in \mathbb{N}_+$. This means that for some integer-valued random variables $0 = S_{-1} < S_0 < S_1 < \ldots$ and for $\zeta_n^{(1)} \geqslant \zeta_n^{(2)} \geqslant \ldots \geqslant \zeta_n^{(r)}$ denoting the first r largest values in $\{X_k : S_{n-1} \leqslant k < S_n\}$ (with $\zeta_n^{(q)} = -\infty$ if $S_n - S_{n-1} < q \leqslant r$), the random vectors $\left(S_n - S_{n-1}, \zeta_n^{(1)}, \zeta_n^{(2)}, \ldots, \zeta_n^{(r)}\right)$ are independent for $n \geqslant 0$ and identically distributed for $n \geqslant 1$.

Denote by $M_n^{(q)}$ the qth largest value in $\{X_k : 1 \leqslant k \leqslant n\}$. Recall that Rootzén (1988) [Maxima and exceedances of stationary Markov chains. *Advances in*

Applied Probability **20**, 371–390] has shown that if $\{X_n\}$ with a 1-max-regenerative structure satisfies

$$\mu := \mathrm{E}(S_1 - S_0) < \infty \quad \text{and} \quad \lim_{n\to\infty} \mathrm{P}\left(\zeta_0^{(1)} > \max_{1\leqslant k\leqslant n} \zeta_k^{(1)}\right) = 0, \qquad (\star)$$

then

$$\sup_{x\in\mathbb{R}} \left|\mathrm{P}\left(M_n^{(1)} \leqslant x\right) - G(x)^n\right| \to 0 \quad \text{as} \quad n \to \infty,$$

where $G(x) := \mathrm{P}\left(\zeta_1^{(1)} \leqslant x\right)^{1/\mu}$.

Our aim is to describe the asymptotic behaviour of $M_n^{(q)}$ when $2 \leqslant q \leqslant r$. For $\{X_n\}$ with an r-max-regenerative structure and satisfying $(\star)$, we will prove that

$$\sup_{x\in\mathbb{R}} \left| P\left(M_n^{(q)} \leqslant x\right) - G(x)^n \sum_{k=0}^{q-1} \frac{(-\log G(x)^n)^k}{k!} \gamma_{q,k}(x) \right| \to 0 \quad \text{as} \quad n \to \infty,$$

where

$$\gamma_{q,k}(x) := k! \sum_{\substack{i_1,i_2,\ldots,i_{q-1}\geqslant 0 \\ \sum_{j=1}^{q-1} i_j = k \\ \sum_{j=1}^{q-1} j i_j \leqslant q-1}} \prod_{j=1}^{q-1} \frac{\beta_j(x)^{i_j}}{i_j!}$$

and

$$\beta_j(x) := \mathrm{P}\left(\zeta_1^{(j+1)} \leqslant x < \zeta_1^{(j)} \,\middle|\, \zeta_1^{(1)} > x\right).$$

The theorem will be illustrated with some examples.

One can compare our result with the ones given for stationary sequences by Hsing (1988) [On the extreme order statistics for a stationary sequence. *Stochastic Processes and Their Applications* **29**, 155–169] and Jakubowski (1993) [Asymptotic $(r-1)$-dependent representation for rth order statistic from a stationary sequence. *Stochastic Processes and Their Applications* **46**, 29–46].

Key Words: regenerative process, order statistics, extremal index, phantom distribution function.

The tail process revisited

Session I9
June 29, 9:40
Room: Pi

Philippe Soulier
Université Paris Nanterre (France)

A stationary heavy tailed time series belongs to the domain of attraction of some stationary max-stable process. Therefore strong relationships exist between the tail process and the max-stable process associated with the time series. We use these relations to revisit some well known properties such as the time-change formula, some identities for the extremal and cluster indices and possibly to characterize all spectral tail processes.

Key Words: heavy tailed time series, tail process.

Session C27
June 28, 11:00
Room: Data

Robust prediction of extreme length of stays in intensive care

Peter Ruckdeschel
Oldenburg University (Germany)

Matthias Kohl
Furtwangen University (Germany)

Bernhard Spangl
University of Natural Resources & Life Sciences (Austria)

In the presence of ever growing costs in intensive care, good predictive tools are a major concern in current public health. A considerable part of these costs can be attributed to extremely long length of stays (LOS) in intensive care. We are concerned with finding (better) predictors of these extremely long stays.

Starting from the 1990s, the interdisciplinary intensive care unit (ICU) of the department for anesthesiology and intensive care medicine of the University Hospital Jena introduced and established the electronic patient documentation system COPRA (www.copra-system.de). It includes all relevant vital parameters as well as diagnoses, laboratory results, all medications, LOS and much more of the ICU patients. Over this time period a very comprehensive database for critical ill patients has emerged. We use a dataset consisting of 40 variables of more than 11,000 patients to derive a robust regression model for predicting LOS.

To obtain a parsimonious model we select relevant variables beforehand. This is done by component-wise gradient boosting, which is a machine learning method for optimizing prediction accuracy and for obtaining model estimates via gradient descent techniques. This method carries out variable selection during the fitting process without relying on heuristic techniques such as stepwise variable selection.

For each patient an individual GPD is fitted via a generalized linear model. To avoid overly high influence of single observations on future predictions this estimation is done robustly.

Key Words: robust statistics, generalized linear models, boosting, length of stay, intensive care.

Session C7
June 26, 17:00
Room: Data

Wiener-Hopf factorization for extremal Markovian sequences connected with the Kendall convolution

Barbara Jasiulis-Gołdyn
University of Wrocław (Poland)

Mateusz Staniak
University of Wrocław (Poland)

In Borowiecka-Olszewska et al. (2015) [Lévy processes and stochastic integral in the sense of generalized convolution. *Bernoulli* **24**(1), 2513–2551] the authors constructed Markov chains generated by generalized convolutions.

We focus on the case of the Kendall convolution defined by

$$\delta_x \triangle_\alpha \delta_y := T_M(\rho^\alpha \tilde{\pi}_{2\alpha} + (1-\rho^\alpha)\tilde{\delta}_1),$$

where $\tilde{\pi}_{2\alpha}(dx) = \alpha|x|^{-2\alpha-1}\mathbf{1}_{[1,\infty)}(|x|)dx$, $M = \max\{|x|,|y|\}$, $\rho = m/M$ with $m = \min\{|x|,|y|\}$, and $\tilde{\delta}_1$ is symmetrization of δ_1.

We consider barrier crossing problem for extremal Markov process $\{X_n : n \in \mathbb{N}_0\}$ of the Kendall type with the transition probabilities:

$$\mathbb{P}(X_n \in A | X_0 = x) = \delta_x \triangle_\alpha \nu^{\triangle_\alpha n}(A),$$

for all Borel sets $A \subset \mathbb{R}$.

We prove Wiener-Hopf factorization and Spitzer equality analogues for Kendall random walks in the Williamson transform terms. The talk is based on Jasiulis-Gołdyn and Staniak (2017) [Spitzer identity for extremal Markovian sequences of the Kendall type, *in preparation*] which is a continuation of research from Jasiulis-Gołdyn and Misiewicz (2016) [Kendall random walk, Williamson transform and the corresponding Wiener-Hopf factorization. *Lithuanian Mathematical Journal*, in press].

Distributions emerging from the study of Kendall random walks are heavy tailed and can be used to model air pollution data, for example ozone levels in Poland.

Key Words: air pollution, extremal Markov chain, Kendall random walk.

Towards an early warning system for high-impact coastal events in Pacific island nations

Session C35
June 29, 14:30
Room: Pi

Robert Davy, Ron K. Hoeke, Kathleen McInnes, Julian O'Grady
CSIRO Oceans and Atmosphere (Australia)

Gareth Williams
CSIRO eResearch (Australia)

Alec G. Stephenson
CSIRO Data Analytics (Australia)

When some combination of extreme storm waves, surge, tide and sea-level trigger rapid coastal erosion and inundation of low-lying areas, coastal engineers and scientists are frequently called upon to answer questions where the statistics of extremes plays a pivotal role.

Modelling coastal flooding in Pacific Islands is complex and presents a number of challenges, such as the lack of wave buoy and topographic data, and the importance of wind-waves on total water level. In this collaborative study, we model extreme

coastal events in the Pacific using tide gauge data and wind-wave hindcasts derived form oceanographic numerical simulation. We use the data to identify storm characteristics for the most extreme storm events. Our results provide crucial guidance for the design of early warning systems, showing the benefit and utility of extreme value theory when applied to globally important problems.

This work was performed by a multi-disciplinary team, including oceanographers, climate scientists, and an extreme value statistician.

Key Words: tides, storms, sea-level, waves.

How to simulate a Brown-Resnick process? Comparison of previous approaches and efficiency gains by using locally equivalent covariances

Session I6
June 27, 10:20
Room: Pi

Marco Oesting
Universität Siegen (Germany)

Kirstin Strokorb
Cardiff University (Wales, UK)

Several approaches have been proposed in order to simulate a max-stable process on finitely many points on a finite simulation domain. Since the class of Brown-Resnick processes seems particularly attractive for practical use, its simulation algorithms have also received more attention and some approaches are tailored to this class. While the efficiency of these algorithms can be assessed by looking at the expected total number of simulated Gaussian processes as a function of the number of points on which the process is simulated, it is more difficult to understand the role of the simulation domain in which they lie on which we focus in this project. We shall compare previous approaches, add another suggestion to the list and give recommendations on when to use which approach. Our suggestion is concerned with the question of improving the efficiency of the simulation, while a potential error made can still be controlled. Thereby, we also show that a relatively simple and non-costly adjustment of previous simulation algorithms could have saved a lot of time in the past.

Key Words: Brown-Resnick process, locally equivalent representation, max-stable process, simulation.

Multivariate subexponential distributions and their applications

Session C21
June 28, 9:30
Room: Boole

Gennady Samorodnitsky
Cornell University (USA)

Julian Sun
Cornell University (USA)

We propose a new definition of a multivariate subexponential distribution. We compare this definition with the two existing notions of multivariate subexponentiality, and compute the asymptotic behaviour of the ruin probability in the context of an insurance portfolio, when multivariate subexponentiality holds. Previously such results were available only in the case of multivariate regularly varying claims.

Key Words: subexponential distribution, multivariate, regular variation, insurance portfolio, ruin probability.

Hypercontractivity for maxima of dependent random sequences

Session C3
June 26, 15:00
Room: Data

Zbigniew S. Szewczak
Nicolaus Copernicus University (Poland)

Let $\{X_k\}$ be a strictly stationary sequence with partial maxima

$$M_n = \max_{1\le k\le n} |X_k| \quad \text{and} \quad Z_n = \max_{1\le k\le n} \left| \sum_{\nu=1}^{k} X_\nu \right|.$$

We say that M_n is hypercontractive if for some $0 < p < q$ and $C < \infty$ and each n we have $\|M_n\|_q \le C\|M_n\|_p$. For i.i.d. sequences it was proved in Hitczenko et al. (1998) [Hypercontractivity and comparison of moments of iterated maxima and minima of independent random variables. *Electronic Journal of Probability* **3**(2), 1–26] that if $E|X_1|^q I_{[|X_1|>x]}$ has positive decrease then M_n is hypercontractive. This was extended by Szewczak (2011) [A weak law of large numbers for maxima. *Extremes* **14**(3), 325–341] on φ-dependent sequences. We discuss relevant results for Z_n also.

Key Words: hypercontractivity, φ-dependence, maxima.

Estimating the stable tail dependence function via the empirical beta copula

Session C17
June 27, 16:30
Room: Boole

Anna Kiriliouk, Johan Segers
Université catholique de Louvain (Belgium)

Laleh Tafakori
University of Melbourne (Australia)

Estimation of the dependence structure still presents a challenge and an interesting problem in the world of multivariate extremes. An adjustable way to model dependence is via copulas and a fundamental method of inference is by means of the empirical copula. The empirical beta copula (Segers et al. (2017) [The empirical beta copula. *Journal of Multivariate Analysis* **155**, 35–51]) arises as a particular case of the empirical Bernstein copula when the degrees of all Bernstein polynomials are set equal to the sample size. It is a smoothed version of the empirical copula which is a genuine copula without requiring the choice of a smoothing parameter. Simulating random samples from it is straightforward and the corresponding empirical process converges weakly to a Gaussian process under standard smoothness conditions on the underlying copula. It is the goal of this work to present and study a nonparametric estimator of the stable tail dependence function based on the empirical beta copula. Under natural conditions, the estimator is shown to be consistent and asymptotically normal. Monte Carlo simulations show improved finite-sample performance in comparison to the usual empirical stable tail dependence function.

Key Words: multivariate extremes, tail dependence, empirical copula, Bernstein polynomial, asymptotic properties.

Session C8
June 26, 16:00
Room: Pi

Quantifying the basis risk of industry loss warranties

Jose H. Blanchet
Columbia University (USA)

Henry Lam
University of Michigan (USA)

Zhongyi Yuan
Pennsylvania State University (USA)

Qihe Tang
University of Iowa (USA)

A well-known problem related to the use of an index-linked catastrophic loss instrument in the context of hedging is basis risk. This arises when the company's loss is not sufficiently dependent on the reference index, and hence the latter is not a good representative of the former. In this talk we discuss quantification of the basis risk of dual-triggered industry loss warranties (ILWs), as well as the sensitivity of the basis risk to the dependence between the company's loss and the industry loss. We employ the central limit theorem (CLT) approach for the case with average sized attachment points while employ the extreme value theory (EVT) approach for the case with large attachment points.

Key Words: basis risk, industry loss warranties, catastrophe risk, hedge effectiveness.

Exploration and inference in spatial extremes using empirical basis functions

Session I8
June 29, 9:00
Room: Chip

Samuel Morris, Brian Reich
North Carolina State University (USA)

Emeric Thibaud
École Polytechnique Fédérale de Lausanne (Switzerland)

Statistical methods for inference on spatial extremes of large datasets are yet to be developed. Motivated by standard dimension reduction techniques used in spatial statistics, we propose an approach based on empirical basis functions to explore and model spatial extremal dependence. Based on a low-rank max-stable model we propose a data-driven approach to estimate meaningful basis functions using empirical pairwise extremal coefficients. These spatial empirical basis functions can be used to visualize the main trends in extremal dependence. In addition to exploratory analysis, we show how these functions can be used in a Bayesian hierarchical model to model spatial extremes of large datasets. We illustrate our method with an application to extreme precipitations in eastern U.S.

Key Words: dimension reduction, max-stable process, non-stationary data analysis.

Predicting extreme influenza epidemics

Session C27
June 28, 11:30
Room: Data

Holger Rootzén
Chalmers University of Technology (Sweden)

Maud Thomas
Université Pierre et Marie Curie (France)

Influenza viruses are responsible for annual epidemics, causing more than 500,000 deaths per year worldwide. A crucial question for resource planning in public health is to predict the morbidity burden of extreme epidemics. We say that an epidemic is extreme whenever the influenza incidence rate exceeds a high threshold for at least one week. Our objective is to predict whether an extreme epidemic will occur in the near future, say the next couple of weeks.

The weekly numbers of influenza-like illness (ILI) incidence rates in France are available from the Sentinel network for the period 1991-2017. ILI incidence rates exhibit two different regimes, an epidemic regime during winter and a non-epidemic regime during the rest of the year. To identify epidemic periods, we use a two-state autoregressive hidden Markov model.

A main goal of Extreme Value Theory is to assess, from a series of observations, the probability of events that are more extreme than those previously recorded. Because of the autoregressive structure of the data, we choose to fit one of the multivariate generalized Pareto distribution models proposed in Rootzén et al. (2016a)

[Multivariate peaks over threshold models. *arXiv:1603.06619v2*]; see also Rootzén et al. (2016b) [Peaks over thresholds modeling with multivariate generalized Pareto distributions. *arXiv:1612.01773v1*]. For these models, explicit densities are given, and formulas for conditional probabilities can then be deduced, from which we can predict if an epidemic will be extreme, given the first weeks of observation.

Key Words: multivariate peaks over threshold models, prediction, influenza.

Session C21
June 28, 10:00
Room: Boole

Regular variation of a random length sequence of random variables and application to risk assessment

Olivier Wintenberger
Université Pierre et Marie Curie (France)

Charles Tillier
Université Paris Nanterre (France)

Risk analyses play a leading role within fields such as dietary risk, hydrology, nuclear security, finance and insurance and is more and more present in the applications of various probability tools and statistical methods. When assessing risks on a finite-time horizon, the problem can often be reduced to the study of a random sequence $C(N) = (C_1, \dots, C_N)$ of random length N, where $C(N)$ comes from the product of a matrix $A(N)$ of random size $N \times N$ and a random sequence $X(N)$ of random length N. In this presentation, I will explain how to build a regular variation framework for such random sequences of random length and how to study their spectral properties. Moreover, I will expose a generalization of Breiman's Lemma that gives way to a tail estimate of $\|C(N)\|$ and I will show how this result provides many risk indicators such as the *ruin probability* and *the expected time over a threshold* on a finite-time horizon. Finally, to illustrate the applicability of our method, I will apply the main result on dietary risk assessment models.

Key Words: regular variation, risk indicators, ruin theory, dietary risk assessment.

Session C35
June 29, 15:00
Room: Pi

Parameter uncertainty in attribution studies for extreme weather events

William Collins, Travis O'Brien, Mark Risser, Dáithí Stone
Lawrence Berkeley National Laboratory (USA)

Ben Timmermans
Lawrence Berkeley National Laboratory (USA)

The attribution of extreme weather events, such a heat waves, to anthropogenic influence involves the analysis of their probability in simulations of climate. The climate models used however, such as the Community Atmosphere Model (CAM5), are affected by "parameter uncertainty"—uncertainty about the most accurate or optimal values of numerical parameters within the model—which can be an important consideration. Parameter uncertainty has been studied in CAM5 but surrogate modelling techniques are necessitated by the cost of running large parameter ensembles. That is, a computationally cheap statistical approximation of the climate model (an "emulator") can be constructed and used in place of the full numerical model.

However, to date, research has focused on emulating averaged output and it is not clear whether existing emulator methods are effective for extremes, particularly noting that tails of output distributions may be poorly characterised due to limited simulation run times. We investigate parameter uncertainty in attribution studies using CAM5 by developing emulators in a systematic way. Firstly, we examine the relationship between sampling duration, output variance and experiment design in idealised cases using both statistical (GARCH) and low complexity numerical models. We then proceed to evaluate possible emulators including generalised linear and Gaussian process models, by emulating output quantities such as high quantiles and parameters from fitted extreme value distributions. Finally, in order to overcome the enormous computational resource requirement, we explore the possibility of a hierarchical approach to emulation by making use of output from CAM5 at low resolution.

Key Words: climate, simulation, emulator, uncertainty, extremes.

High level directional multivariate quantile estimation

Session C38
June 29, 16:00
Room: Pi

Elena Di Bernardino
Conservatoire National des Arts et Métiers (France)

Henry Laniado
Universidad EAFIT Medellín (Colombia)

Rosa E. Lillo
Universidad Carlos III de Madrid (Spain)

Raúl Torres
Universidad Carlos III de Madrid (Spain)

Extreme value theory (EVT) has been applied in fields such as insurance, finance, economy and environmental science. EVT is a useful tool to quantify the multivariate risks outside of the observable sampling zone; that is, the estimation of multivariate quantiles located at high levels. The lack of a total order in the multivariate framework allows quantile concepts with a directional perspective as can be

appreciated in the literature. Combining EVT and a directional approach, this work provides an *out-sample* estimation method for the *directional multivariate quantiles* recently introduced in Torres et al. (2015) [A directional multivariate value at risk. *Insurance: Mathematics and Economics* **65**, 111–123], Torres et al. (2017) [Directional multivariate extremes in environmental phenomena. *Environmetrics.* DOI: 10.1002/env.2428]. The asymptotic normality of the proposed estimator is derived. Finally, the methodology is illustrated with an example for which the theoretical directional multivariate quantiles are known.

Key Words: directional multivariate quantiles, regular variation, tail function.

Session C19
June 27, 17:00
Room: Data

Quenched phantom distribution functions for Markov chains

Adam Jakubowski
Nicolaus Copernicus University (Poland)

Patryk Truszczyński
Nicolaus Copernicus University (Poland)

It is known that random walk Metropolis algorithms with heavy-tailed target densities can model atypical (slow) growth of maxima of weakly dependent sequences (e.g. when the extremal index is zero) by admitting a continuous phantom distribution function. We show that in a large class of positive Harris recurrent Markov chains (containing the above Metropolis chains) a phantom distribution function can be recovered by starting "at the point" rather than from the stationary distribution.

Key Words: stochastic extremes, Markov chains, phantom distribution function, random walk Metropolis algorithm, coupling.

Session C38
June 29, 16:30
Room: Pi

Estimation of very small probabilities of extreme events: key notions and some applications

Cees de Valk
Université catholique de Louvain (Belgium)

To estimate a very small probability $p_n \ll 1/n$ of an extreme event from a sample of n i.i.d. random elements, a tail large deviation principle (tail LDP) was recently proposed as an alternative to the classical tail limits; see de Valk (2016) [Approximation and estimation of very small probabilities of multivariate extreme events. *Extremes* **19**, 687–717]. Where the latter can be regarded as limits of ratios of probabilities of extreme events, the tail LDP takes the form of asymptotic bounds on ratios of the logarithms of such probabilities. An overview of the main ideas is presented, showing how the tail LDP differs from classical extreme value theory (crude

vs refined, global vs local) and why it is needed. Furthermore, its close connections to a number of concepts developed earlier are indicated, such as convergence of scaled sample clouds to a compact set, residual tail dependence/hidden regular variation and an extension, and the Weibull tail limit and an extension. References will be provided in the talk. To illustrate the practical significance, examples are shown of estimation of probabilities of extreme events and of quantiles and their multivariate generalisations.

Key Words: tail LDP, sample clouds, residual tail dependence, hidden regular variation, Weibull tail limit.

Extreme quantile estimation for the forecast distribution

Session C38
June 29, 17:00
Room: Pi

Juan Juan Cai, Geurt Jongbloed
Delft University of Technology (The Netherlands)

Jasper Velthoen
Delft University of Technology (The Netherlands)

The Numerical Weather Prediction (NWP) models, which approximate the solutions to the physical differential equations describing the atmosphere, provide a basis for weather forecasts. Probabilistic forecasts are derived by estimating the distribution as a function of output of the deterministic NWP model.

Using an extreme quantile estimation approach we attempt to estimate the tail of the forecast distribution, which is not well approximated by current methods. We propose a model for the data in terms of the quantile function. Let Y be the weather quantity of interest and X be the model output, the conditional quantile function is given by $Q_{Y|X}(\tau|X = x) = r(x) + Q_\epsilon(\tau)$, where τ is a probability level and Q_ϵ the quantile function of the noise. Then we construct a two step estimation procedure for the model. First intermediate quantile curves are estimated by a nonparametric quantile estimator that borrows strength from several quantiles. Secondly, the residuals from the estimation of $r(x)$ are used to estimate the Q_ϵ for high percentiles τ.

Key Words: extreme quantile estimation, covariates.

Session C6
June 26, 17:00
Room: Chip

Bayesian clustering and dimension reduction in multivariate extremes

Marc G. Genton, Raphaël Huser
King Abdullah University of Science and Technology (Saudi Arabia)

Johan Segers
Université catholique de Louvain (Belgium)

Sabrina Vettori
King Abdullah University of Science and Technology (Saudi Arabia)

Describing the complex dependence structure of multivariate extremes is particularly challenging and requires very versatile, yet interpretable, models. To tackle this issue we explore two related approaches: clustering and dimension reduction. In particular, we develop a novel statistical algorithm that takes advantage of the inherent hierarchical dependence structure of the max-stable nested logistic distribution and that uses reversible jump Markov chain Monte Carlo techniques to identify homogeneous clusters of variables. Dimension reduction is achieved when clusters are found to be completely independent. We significantly decrease the computational complexity of full likelihood inference by deriving a recursive formula for the nested logistic model likelihood. The algorithm performance is verified through extensive simulation experiments which also consider different likelihood procedures. The new methodology is used to investigate the dependence relationships between extreme concentration of multiple pollutants across a number of sites in California and how these pollutants are related to extreme weather conditions. Overall, we show that our approach allows for the identification of homogeneous clusters of extremes and has valid applications in multivariate data analysis, such as air pollution monitoring, where it can guide policymaking.

Key Words: air pollution, clustering, extreme event, fast likelihood inference, reversible jump Markov chain Monte Carlo.

Session C37
June 29, 17:00
Room: Data

Spatial hybrid downscaling: from large-scale information to high-resolution extreme precipitation fields

Aurelien Bechler
France Telecom Orange Labs (France)

Liliane Bel
AgroParisTech, INRA (France)

Mathieu Vrac
Laboratoire des Sciences du Climat et de l'Environnement, CNRS/IPSL (France)

As extreme climate events are rare and often unexpected, society is not well prepared to face them. It is therefore important to evaluate their evolutions in the future. Global Climate Models (GCMs) are generally used to provide future scenarios of precipitation. However, their resolution (200 km) is not sufficient to describe efficiently what happens at local scales (a few kilometers or less). Hence, dynamical and statistical downscaling methods have been developed to generate local-scale values of climate variables, based on large-scale information. Nevertheless, both the extreme behaviours and the spatial structures are not always well described by these methods. Moreover, most of the statistical downscaling models cannot simulate values at locations where there is no observed data for calibration.

Max-stable processes now emerge as a powerful tool for the statistical modelling of spatial extremes. One recent development in this context is for conditional simulations of spatial fields of maximum values, providing empirical distributions of maximum value at any location in a given region, conditioned by observed values at other locations.

We propose a two-step methodology, called "Spatial Hybrid Downscaling" (SHD), to tackle the problem of spatial downscaling for fields of extreme precipitation. The first step consists in applying a univariate statistical downscaling at given locations. Once this 1d-link is performed, a conditional simulation of max-stable processes is adapted to a flexible process: the extremal t process. This enables us to get conditional distributions and to downscale extreme precipitation values at any point of the region, even in a climate change context.

Key Words: statistical downscaling, precipitation, spatial modelling, extremal t process.

Linking large currency swings to fundamentals' shocks

Session I11
June 30, 15:20
Room: Chip

Phornchanok Cumperayot
Chulalongkorn University (Thailand)

Casper G. de Vries
Erasmus Universiteit Rotterdam (The Netherlands)

The distribution of foreign exchange returns exhibits heavy tails. Under multiplicative parameter uncertainty, the distribution of macroeconomic fundamentals, like money supply and inflation, also exhibit heavy tails, while real income may not. In standard exchange rate models this translates into the tail behavior of exchange rate returns. To provide evidence for this extreme connection, we estimate the asymptotic dependence between the variables, rather than using regression analysis that focusses on average behavior. The strongest links are between Asian and Latin American currencies and fundamentals. The main drivers are monetary and financial macro variables; real income shocks appear disconnected.

Session I12
June 30, 14:40
Room: Pi

Modelling spatial processes with unknown extremal dependence class

Raphaël Huser
King Abdullah University of Science and Technology (Saudi Arabia)

Jenny Wadsworth
Lancaster University (UK)

Many environmental processes exhibit weakening spatial dependence as events become more extreme. Well-known limiting models, such as max-stable or generalized Pareto processes, cannot capture this, which can lead to a preference for models that exhibit a property known as asymptotic independence. However, weakening dependence does not automatically imply asymptotic independence, and whether the process is truly asymptotically (in)dependent is usually far from clear. The distinction is key as it can have a large impact upon extrapolation, i.e., the estimated probabilities of events more extreme than those observed. We present a single spatial model that is able to capture both dependence classes in a parsimonious manner, and with a smooth transition between the two cases. The model covers a wide range of possibilities from asymptotic independence through to complete dependence, and permits weakening dependence of extremes even under asymptotic dependence. Censored likelihood-based inference for the implied copula is feasible in moderate dimensions due to closed-form margins. An application of the model to a dataset of hindcast significant wave heights will be discussed.

Key Words: asymptotic dependence, asymptotic independence, copula, spatial extremes.

Session C1
June 26, 15:00
Room: Boole

Modeling power laws in directed social networks through linear preferential attachment

Richard A. Davis
Columbia University (USA)

Sidney I. Resnick, Tiandong Wang
Cornell University (USA)

Phyllis Wan
Columbia University (USA)

Preferential attachment is an appealing mechanism for modeling the widely observed power-law behavior of the degree distributions in directed social networks. In this talk, we consider fitting a 5-parameter linear preferential model to network data under two data scenarios. In the case where full history of the network formation is available, we derive the maximum likelihood estimators of the parameters and show

that they are strongly consistent and asymptotically normal. In the case where only a single-time snapshot of the network is available, we propose an estimation method that combines method of moments with an approximation to the likelihood. The resulting estimators are also strongly consistent and performs well compared to the MLE estimator based on the full history of the network. We illustrate both estimation procedures through simulated data, and explore the usage of this model in a real data example.

Key Words: network, power laws, multivariate regular variation, preferential attachment, estimation.

An axiomatic theory for measures of tail risk

Session C8
June 26, 16:30
Room: Pi

Fangda Liu
Central University of Finance and Economics (China)

Ruodu Wang
University of Waterloo (Canada)

The notion of tail risk has been a crucial consideration in modern risk management. To achieve a comprehensive understanding of the tail risk, we carry out an axiomatic study for risk measures which quantify the tail risk, that is, the behavior of a risk beyond a certain quantile. Such risk measures are referred to as tail risk measures in this talk. The two popular classes of regulatory risk measures in banking and insurance, the Value-at-Risk (VaR) and the Expected Shortfall (ES), are prominent, yet elementary, examples of tail risk measures. We establish a connection between a tail risk measure and a corresponding law-invariant risk measure, called its generator, and investigate their joint properties. A tail risk measure inherits many properties from its generator, but not subadditivity or convexity; nevertheless, a tail risk measure is coherent if and only if its generator is coherent. We explore further relevant issues on tail risk measures, such as bounds, distortion risk measures, risk aggregation, elicitability, and dual representations. In particular, there is no elicitable tail convex risk measure other than the essential supremum, and under a continuity condition, the only elicitable and positively homogeneous monetary tail risk measures are the VaRs. The study on tail risk measures brings in new tools and insights for prudent risk management as highlighted in the recent Basel documents on financial regulation. This talk focuses on mathematical developments of the theory.

Session C6
June 26, 16:00
Room: Chip

Multivariate regular variation and fitting a preferential attachment network model

Richard Davis, Phyllis Wan
Columbia University (USA)

Sidney I. Resnick
Cornell University (USA)

Tiandong Wang
Cornell University (USA)

Preferential attachment is an appealing mechanism for modeling directed social networks. Both empirical evidence and mathematical analysis reveal that in- and out-degree distributions follow a power law. Other than showing the multivariate regular variation of the joint measure, we prove that the joint mass function of normalized in- and out-degree is also regularly varying. We then turn to the model calibration under different data scenarios. First, using extreme value theory, we present a semi-parametric method to model heavy-tailed features of the degree distributions of the network. Next, in the case where full history of the network formation is given, we derive the maximum likelihood estimator of the parameters. If instead only a single-time snapshot of the network is available, we propose an estimation method through moment matching and approximating the likelihood function. The resulted estimators in the last two cases are shown to be strongly consistent.

Key Words: power laws, multivariate heavy tail statistics, preferential attachment, estimation.

Session C39
June 30, 10:00
Room: Boole

Modelling incomplete record of large volcanic eruptions

Koji Kiyosugi
Kobe University (Japan)

Jiancang Zhuang
Institute of Statistical Mathematics (Japan)

Ting Wang
University of Otago (New Zealand)

Point processes are widely used to model phenomena where events occur at random times or locations, such as volcanic eruptions and earthquakes. The volcanic hazard estimation using point process modelling is strongly influenced by missing events, which usually result in underestimate of hazard. We present a fast approach for replenishing missing data in the record of a temporal point process with time independent marks. If a temporal point process with time-independent marks is

completely observed, then the process can be transformed into a homogeneous Poisson process on the unit square $[0, 1]^2$ by a bi-scale empirical transformation. Based on this, after testing this method on a synthetic data set, we iteratively reconstructed the record of large volcanic eruptions at the Hakone Volcano, Japan. The results show that this algorithm provides a useful way to replenish an incomplete record of a marked point process data. The replenished data provides more robust estimates of the future hazard.

Key Words: large volcanic eruptions, missing data, marked point process.

Spatial statistics for improving collective estimates of extreme precipitation at weather stations and its subsequent gridding

Session C20
June 27, 17:00
Room: Pi

Christopher J. Paciorek
University of California at Berkeley (USA)

Mark D. Risser
Lawrence Berkeley National Laboratory (USA)

Michael F. Wehner
Lawrence Berkeley National Laboratory (USA)

While there are well-established univariate statistical methods for modeling the climatological properties of extreme precipitation, characterizing the heterogeneous properties of multivariate extremes over space for a large geographic area is a more difficult task. We demonstrate the application of spatial statistical analysis tools that incorporate the heterogeneous spatial dependence patterns present in the climatological features of extreme precipitation. We show that the borrowing of strength over space can improve the signal-to-noise ratio in estimates of the climatological features of extreme precipitation (e.g., 20 year return values). We also allow for non-stationarity over time, allowing us to make statements about how extreme precipitation is changing from year to year. Uncertainty estimates are obtained via a block bootstrap technique that requires no assumption of temporal independence within each year of data.

Recent research has called into question the appropriateness and accuracy of using gridded products of observed daily precipitation for comparison to climate model outputs. As an alternative, we demonstrate a new framework for "gridding" extreme precipitation based on our spatial statistical analysis of station data. The essence of our method is to obtain estimates of the climatological features of extreme precipitation based on station data first, followed by a statistical approach (kriging) to interpolate these estimates to a fine grid. Our method is more appealing than current approaches based on extreme value analysis of gridded daily precipitation products, since extreme value functions parameters are likely to be smoother in space than daily precipitation and more amenable to the gridding process.

Key Words: extreme precipitation, spatial statistics, gridded observations.

Session C23
June 28, 10:00
Room: Data

Asymptotic distributions of exceedances point processes in the plane for stationary Gaussian sequences with data missing

Zuoxiang Peng, Jinjun Tong
Southwest University (China)

Zhichao Weng
Fuzhou University (China)

In this paper, we study the limiting distributions of exceedances point processes in the plane of complete and incomplete stationary Gaussian sequences under some weak dependence conditions. One interesting byproduct is to derive the asymptotic joint distribution of the location and height of maxima with data missing.

Key Words: point process in the plane, data missing, location and height of maxima, stationary Gaussian sequence.

Session I9
June 29, 10:20
Room: Pi

Heavy tails for an alternative stochastic perpetuity model

Thomas Mikosch
University of Copenhagen (Denmark)

Mohsen Rezapour
Shahid Bahonar University of Kerman (Iran)

Olivier Wintenberger
Sorbonne Universités (France)

In this talk we consider a stochastic model of perpetuity-type. In contrast to the classical affine perpetuity model of Kesten (1973) [Random difference equations and renewal theory for products of random matrices. *Acta Mathematica* **131**, 207–248] and Goldie (1991) [Implicit renewal theory and tails of solutions of random equations. *Annals of Applied Probability* **1**, 126–166] all discount factors in the model are mutually independent. We prove that the tails of the distributions of this model are regularly varying both in the univariate and multivariate cases. Due to the additional randomness in the model the tails are not pure power laws as in the Kesten-Goldie setting but involve a logarithmic term.

Key Words: Kesten-Goldie theory, heavy tails, large deviations, multivariate regular variation.

Extreme value statistics for censored data with heavy tails under competing risks

Session C10
June 27, 12:00
Room: Chip

Julien Worms
Université de Versailles-Saint-Quentin-en-Yvelines (France)

Rym Worms
Université Paris-Est (France)

The study of the tail of right-censored data has attracted some attention in the recent years. Concerning the estimation of the extreme value index (e.v.i.) of such data, we recently proposed a method which, in the context of heavy tail data, led to an adaptation of the famous Hill estimator taking the form of a Kaplan-Meier functional: it weights the data in the tail in a natural way, in this survival analysis framework.

In this paper, we extend this method to the case where competing risks are present. The target is now the e.v.i. of one of the cause-specific survival functions (also called cumulative incidence functions, assumed here to have heavy tails). We propose an estimator which is a functional of the Aalen-Johansen estimator. Asymptotic properties of the estimator (consistency and asymptotic normality) and finite sample behavior will be presented. Estimation of an extreme quantile of the cause-specific survival function is also adressed.

Key Words: extreme value index, tail inference, random censoring, competing risks, Aalen-Johansen estimator.

Tail indices and scale parameters in financial time series

Session C33
June 29, 15:00
Room: Chip

Thomas Mikosch
University of Copenhagen (Denmark)

Casper G. de Vries
Erasmus Universiteit Rotterdam (The Netherlands)

Xiaolei Xie
University of Copenhagen (Denmark)

We consider an investor with preferences that accord with Generalized Disappointment Aversion. Such an investor cares about downside risk and we assume he recognizes the heavy tail feature of asset return distributions. We argue that when a market is dominated by rational investors of this kind, the return distributions of equities that are actively traded in this market must have very similar tail indices. We also show, in contrast, the scale parameters of the return distributions may differ hugely from one another.

On the other hand, whether or not all equities in a multivariate model have the same tail index is a dividing issue for multivariate GARCH models proposed in the

literature. Therefore, it is important to analyze data of real equity returns and see how close to each other the tail indices actually are.

In this work empirical results are also presented and they appear to support the conclusion that the tail indices are very similar, with respect to the confidence bands of estimation.

Session I5
June 27, 10:20
Room: Chip

CAT bond pricing under a product probability measure with EVT risk characterization

Jose Blanchet
Columbia University (USA)

Henry Lam
University of Michigan (USA)

Qihe Tang
University of Iowa (USA)

Zhongyi Yuan
Pennsylvania State University (USA)

Frequent large losses from recent catastrophes have caused great concerns among insurers/reinsurers, who as a result start to seek mitigations of such catastrophe risks by issuing catastrophe (CAT) bonds and thereby transferring the risks to the bond market. Whereas, the pricing of CAT bonds remains a challenging task, mainly due to the facts that the CAT bond market is incomplete and that their pricing usually requires knowledge about the tail of the risks. In this paper, we propose a general pricing framework that utilizes a product pricing measure in conjunction with extreme value theory (EVT). While the EVT approach is used to uncover the tail risk, the product measure combines a distorted probability measure that prices the catastrophe risks underlying the CAT bond with a risk-neutral probability measure that prices interest rate risk, to provide an integrated pricing framework. A case study using California earthquake data is shown with numerous sensitivity analyses to demonstrate the impact of certain risk parameters on the CAT bond price.

Key Words: CAT bond, distortion, extreme value theory, generalized Pareto distribution, pricing.

Testing the multivariate regular variation model

Session C40
June 30, 9:30
Room: Chip

Chen Zhou
Erasmus Universiteit Rotterdam & De Nederlandsche Bank (The Netherlands)

Fan Yang
University of Waterloo (Canada)

This paper aims at testing the multivariate regular variation model. Let (X, Y) be a random vector that follows a multivariate regular variation (MRV) model with a regularly varying index α. We consider a polar coordinate representation $(X, Y) = R \cdot (1 - \Theta, \Theta)$. A necessary condition for the MRV model is that $R|\Theta \in A$ follows a heavy-tailed distribution with tail index α for any Borel set $A \subset [0, 1]$. We develop a test for the MRV model based on this property. We construct the tail index estimator $\hat{\alpha}(A)$ for any given set A and prove the joint asymptotic property for all such estimators via the tail empirical process. By comparing across the estimates, we construct tests for testing the constancy across the tail indices. This work can be regarded as an extension of the test on whether the tail index is constant for non-identically distributed observations in Einmahl et al. (2016) [Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society: Series B* **78**(1), 31–51].

Semi-parametric models for nonstationary environmental extremes

Session C34
June 29, 15:00
Room: Data

Emma Eastoe, Jonathan A. Tawn
Lancaster University (UK)

Philip Jonathan, David Randell
Shell Technology Centre (UK)

Elena Zanini
STOR-i CDT, Lancaster University (UK)

Realistic and accurate environmental design conditions are fundamental for the construction and maintenance of marine structures. They require an understanding of waves and storm severity, which is often achieved through an extreme value analysis of measured or hind-cast significant wave height. In order to determine as realistic and accurate design criteria as possible, it is fundamental to account for the multiple covariates that characterise the data, particularly wave direction and season. A methodology that accurately models these covariate effects is essential for reliable estimates.

To model the nonstationarity in the Generalised Pareto parameters, we focus on two main formulations, namely P-splines and a proposed linear combinations of Gaussian densities. Inference is carried out using MCMC, and we consider both the

case where the number of components is fixed and when it's unknown. Reversible jump MCMC are used to estimate models with unknown number of components, and Bayesian Additive Regression Splines (BARS) are used for the spline model. Return value estimates and performance for the different formulations are compared, while potential advantages and computational cost issues are considered both for one-dimensional covariates and for extensions to higher dimensions.

Key Words: covariate modelling, environmental, Bayesian, BARS, reversible jump MCMC.

Session C40
June 30, 10:00
Room: Chip

An entropy-based test for multivariate extreme value models

Sebastian Engelke
École Polytechnique Fédérale de Lausanne (Switzerland)

Philippe Naveau
Laboratoire des Sciences du Climat et l'Environnement (France)

Chen Zhou
Erasmus Universiteit Rotterdam & De Nederlandsche Bank (The Netherlands)

Many effects of climate change seem to be reflected in the frequency and severity of the extreme events in the distributional tails. Detecting such changes requires a statistical methodology that efficiently uses the large observations in the sample. We propose a simple, non-parametric test that decides whether two multivariate distributions exhibit the same tail behaviour. The test is based on the entropy, namely Kullback-Leibler divergence, between exceedances over a high threshold of the two multivariate random vectors. We show that such a type of divergence is closely related to the divergence between Bernoulli random variables. We study the properties of the test and further explore its effectiveness for finite sample sizes. As an application we apply the method to precipitation data where we test whether the marginal tails and/or the extreme value dependence structure have changed over time.

Key Words: multivariate extreme value dependence, Kullback-Leibler divergence, asymptotics.

Extreme value analysis without the largest values: what can be done?

Session I1
June 26, 10:50
Room: Boole

Richard A. Davis
Columbia University (USA)

Gennady Samorodnitsky
Cornell University (USA)

Jingjing Zou
Columbia University (USA)

In this paper we are concerned with the analysis of heavy-tailed data when a portion of the extreme values are unavailable. This research was motivated from an analysis of the degree distributions in a large social network. The degree distributions of such networks tend to have power law behavior in the tails. We focus on the Hill estimator, which plays a starring role in heavy-tailed modeling. The Hill estimator for this data exhibited a smooth and increasing "sample path" as a function of the number of upper order statistics used in constructing the estimator. This behavior became more apparent as we artificially removed more of the upper order statistics. Building on this observation, we introduce a new parameterization into the Hill estimator that is a function of δ and θ, that correspond, respectively, to the proportion of extreme values that are unavailable and the index of the sample path of the Hill estimator. As a function of (δ, θ), we establish functional convergence of the normalized Hill estimator to a Gaussian random field. An estimation procedure is developed based on the limit theory to estimate the number of missing extremes and other extreme value parameters including the index of regular variation α and the bias of Hill's estimate. We illustrate this approach in both simulations and real data examples.

Key Words: Hill estimator, heavy-tailed distributions, missing extremes, functional convergence.

Presenting Authors Index