Workshop on Extreme Value Theory, with an emphasis on spatial and temporal aspects

November 3–5, 2014
Besançon, France

Organized by

Jean-Noel Bacro	(Univ. Montpellier 2)
Clément Dombry	(Univ. Franche-Comté)
Stéphane Girard	(Inria Grenoble Rhône-Alpes)
Mathieu Ribatet	(Univ. Montpellier 2)
Gwladys Toulemonde	(Univ. Montpellier 2)

Introduction

This workshop will encompass recent research works on extreme value theory with an emphazis on its spatial and temporal aspects. A non exhaustive list of subjects covered will be :
– new advances in univariate or multivariate extreme value theory ;
– extremes of time series and regularly varying time series ;
– max-stable random fields : models, inference, simulation ;
– applications of the above to environmental sciences, actuarial sciences...

Dates and place

The workshop will take place on November 3–5, 2014 at the Centre Diocésain, 20 rue Mégevand, Besançon. The closest tramway/bus stops are : tramway stop Chamars, bus stop Mégevand and bus stop Granvelle.

Organization

The workshop is co-organized by the Laboratoire de Mathématiques de Besançon together with the ANR project McSim and the GICC project MIRACCLE, with the financial support of the GdR GeoSto.

Acknowledgement

The organizers are grateful to :
– the Franche-Comté region for its financial support through the operation « Les trimestres du LMB » ;
– Catherine Vuillemenot, Mahdya Debayle and Sophie Cazanave-Pin for administrative and practical support.

Monday November 3, 2014

10h00–11h00 : Registration & Coffee

11h00–12h30 : Session « Max-stable processes »

- 11h00 : Laurens de Haan, *Extreme value statistics for a stochastic process that is observed at discrete points only*
- 11h45 : Vicky Fasen, *Stable random fields, point processes and large deviations*

12h30–14h00 : Lunch

14h00–15h30 : Session « Multivariate extremes I »

- 14h00 : Armelle Guillou, *Robust and bias-corrected estimation of extreme failure sets*
- 14h45 : Gilles Stupfler, *On the asymptotic behaviour of extreme geometric quantiles*

15h30–16h00 : Coffee-break

16h00–17h30 : Session « Multivariate extremes II »

- 16h00 : Michael Falk, *The Space of D-Norms – An Offspring of Multivariate EVT*
- 16h45 : Anne-Laure Fougères, *Multivariate Archimax copulas*

Tuesday November 4, 2014

9h00–10h30 : Session « Simulation of max-stable processes I »

- 9h00 : Christian Lantuéjoul, *Conditional simulation of max-stable processes using auxiliary variables*
- 9h45 : Marco Oesting, *Sampling from max-stable processes conditional on a homogeneous functional*

10h30–11h00 : Coffee-break

11h00–12h30 : Session « Simulation of max-stable processes II »

- 11h00 : Chen Zhou, *The normalized spectral representation of a max-stable process on a compact set*
- 11h45 : Martin Schlather, *Simulation of mixed moving maxima processes with monotone shape functions*

12h30–13h30 : Lunch

13h30–15h45 : Sight-seeing, visit of the Besançon citadelle

- 13h30 : Group departure from the « centre diocésain » and walk to the citadelle
- 14h00 : Guided tour of the Besançon citadelle (approx. 1 hour) and free tour
- 15h45 : Group departure from the citadelle entrance and return to the « centre diocésain »
- 16h00 : Coffee

16h15–18h30 : Session « Application to environmental statistics »

- 16h15 : Philippe Naveau, *Scoring extremes*
- 17h00 : Benjamin Renard, *Conditioning the distribution of hydrologic extremes on large-scale atmospheric or oceanic fields*
- 17h45 : Nicolas Eckert, *Some extreme value problems in snow avalanche risk assessment*

20h : Conference dinner

Conference dinner at the « Brasserie 1802 », 2 place Granvelle, Besançon.

Wednesday November 5, 2014

9h00–10h30 : Session « Univariate and implicit extremes »

– 9h00 : Ana Ferreira, *On the block maxima method in EVT : PWM estimators*
– 9h45 : Stilian Stoev, *Implicit extremes and implicit max-stable laws*

10h30–11h00 : Coffee-break

11h00–12h30 : Session « Inference for tail processes »

– 11h00 : Michal Warchol, *Statistics for tail processes of Markov chains*
– 11h45 : Anna Kiriliouk, *An M-estimator of spatial tail dependence*

12h30–14h00 : Lunch

14h00–15h30 : Session « Inference for max-stable processes »

– 14h00 : Jenny Wadsworth, *Full likelihood-based inference for max-stable processes : issues of bias and how they may be overcome*
– 14h45 : Christian Robert, *Likelihood based inference for high-dimensional extreme value distributions*

Extreme value statistics for a stochastic process that is observed at discrete points only

Laurens de Haan (Erasmus University Rotterdam and CEAUL)

When dealing with spatial extremes (extreme heat, rainfall, wind speed) the proper tool is a max-stable process i.e. the infinite-dimensional extension of extreme value theory. The main assumption is that the underlying is in the domain of attraction of a (continuous) max-stable process. On the basis of this assumption the main features of the limiting max-stable process can be estimated. Asymptotic properties of the estimators have been derived (T.Lin e.a. Ann Statist. 2003, 2006) under the assumption that there are repeated observations of the entire underlying process. If the process is observed only at a limited number of points one can merely fit a max-stable process depending on finitely many parameters. We show that the general non-parametric case can be recovered if the observation points are close together. The repeated observations do not need to be at the same points. A similar result holds for independent random variables with smoothly changing limit distributions.

Stable random fields, point processes and large deviations

Vicky Fasen (Karlsruhe Institute of Technology) & Parthanil Roy (Indian Statistical Institute)

We consider the large deviation behavior of point process sequences based on stationary symmetric α-stable ($0 < \alpha < 2$) discrete-time parameter random fields. If the random field is generated by a dissipative group action then we will show that the large deviation of the point process sequence converges in the sense of Hult-Lindskog-Samorodnitsky (HLS) to a Borel measure induced by a cluster Poisson process. For the conservative case, no general result exists even in the one-dimensional case. We look at a specific class of stable random fields generated by a conservative action whose effective dimension can be computed using the structure theorem of finitely generated abelian groups. The corresponding large deviation behavior of the point process sequence needs to be properly normalized in order to ensure HLS-convergence. Several other large deviation results can be followed as special cases.

Robust and bias-corrected estimation of extreme failure sets

Armelle Guillou (Université de Strasbourg)

In multivariate extreme value statistics, the estimation of probabilities of extreme failure sets is an important problem, with practical relevance for applications in several scientific disciplines. Some estimators have been introduced in the literature, though so far the typical bias issues that arise in application of extreme value methods and the non-robustness of such methods with respect to outliers were not addressed. We introduce a bias-corrected and robust estimator for small tail probabilities. The estimator is obtained from a second order model that is fitted to properly transformed bivariate observations by means of the minimum density power divergence technique. The asymptotic properties are derived under some mild regularity conditions and the finite sample performance is evaluated by a small simulation experiment. We illustrate the practical applicability of the method on a dataset from the actuarial context. This is a joint work with Christophe Dutang and Yuri Goegebeur.

On the asymptotic behaviour of extreme geometric quantiles

Gilles Stupfler (Université Aix-Marseille) & Stéphane Girard (Inria Grenoble Rhône-Alpes)

A popular way to study the tail of a distribution is to consider its high or extreme quantiles. While this is a standard procedure for univariate distributions, it is harder for multivariate ones, primarily because there is no universally accepted definition of what a multivariate quantile should be. In this talk, we focus on extreme geometric quantiles. We discuss their asymptotics, both in direction and magnitude, when the norm of the associated index vector tends to one. In particular, it appears that if a random vector X has a finite covariance matrix M, then the magnitude of its extreme geometric quantiles grows at a fixed rate and is asymptotically characterised by M. The case when X does not have a finite covariance matrix is tackled in a multivariate regular variation framework. We conclude by some numerical illustrations of our results.

The Space of D-Norms – An Offspring of Multivariate EVT

Stefan Aulbach, Michael Falk and Maximilian Zott (University of Würzburg, Institute of Mathematics)

The theory of D-norms is an offspring of multivariate extreme value theory (EVT). We present some recent results. The space of D-norms can be equipped with a multiplication such that it becomes a commutative semigroup with an identity element. This leads to idempotent D-norms. We characterize the set of idempotent D-norms. Iterating the multiplication provides a track of D-norms, whose limit exists and is again a D-norm. If this iteration is repeatedly done on the same D-norm, then the limit of the track is idempotent. If equipped with the Wasserstein-metric in a suitable way, the space of D-norms is a complete separable metric space. Multiplying a generator Z of a D-norm with a doubly stochastic matrix yields another generator. An iteration of this multiplication provides again a sequence of D-norms and we compute its limit. Finally, we consider a parametric family of D-norms, where we assume that Z follows a symmetric Dirichlet distribution with parameter $\alpha > 0$. This family covers the whole range between complete dependence and independence. The theory of D-norms can be extended to functional extreme value theory as well.

Multivariate Archimax copulas

Anne-Laure Fougères (Université de Lyon)

A multivariate extension of the bivariate class of Archimax copulas was recently proposed by Mesiar and Jagr (2013), who asked under which conditions it holds. In a joint paper with A. Charpentier, C. Genest and J.G. Neslehova, we answer their question and provide a stochastic representation of multivariate Archimax copulas. The aim of my talk will be to present this family and explore some nice properties of these copulas. Several examples will also be provided.

Conditional simulation of max-stable processes using auxiliary variables

Christian Lantuéjoul (MinesParisTech, Centre de Géosciences) & Liliane Bel (AgroParisTech/INRA)

Spatial prediction of max-stable processes cannot be performed by standard regression techniques (kriging) as these models have their first two moments infinite. To overcome this problem, Monte Carlo techniques based on conditional simulations can be considered. Starting from the general theoretical expression of the conditional distribution of a max-stable process by Dombry and Eyi-Minko (Electron. J. Probab. (2013), 18(7) :1-21), a conditional simulation algorithm was developed by Dombry, Eyi-Minko and Ribatet (Biometrika (2013) 100(1) :111-124). Yet, its first step requires the computation of multidimensional integrals, which may be the source of numerical instabilities and thus limits the number of conditioning data points to respect. In this paper, an alternative algorithm is proposed to bypass the computation of those integrals. The first step is replaced by a conditional simulation of a random matrix, the dimension of which is upper-bounded by the number of data points. As a byproduct, the second step simplifies significantly. In the case of the extremal Gaussian, t-extremal and Brown-Resnick processes, it amounts to simulating independent Gaussian random fields conditionally. A consequence is that conditional simulations can be performed much faster using many more conditioning data points.

Sampling from max-stable processes conditional on a homogeneous functional

Marco Oesting (INRA, AgroParisTech)

Max-stable processes have become frequently used models for spatial and spatio-temporal extremes. When some information on the process is available, conditional simulations allow for evaluating these extremes more precisely. Besides observed values of the process at some sites, often aggregated data are given. For example, climate models provide spatial averages on grid cells for some variables of interest. In this talk, we deal with the problem of simulation of a max-stable process with respect to a single condition given by a positively homogeneous functional like the integral or the maximum of the process. Due to the analytic intractability of the involved distributions, we propose a sampling algorithm based on MCMC techniques. The procedure consists of two steps where the second step is based on conditional sampling from a max-linear model. We illustrate the performance of the proposed algorithms in a simulation study and in an example of a real dataset of precipitation observations.

The normalized spectral representation of a max-stable process on a compact set

Chen Zhou (Nederlandsche Bank, Erasmus University Rotterdam & Tinbergen Institute)

WThe normalized spectral representation of a max-stable process on a compact set is the unique representation where all spectral functions share the same supremum. Among the class of equivalent spectral representations of a process, the normalized spectral representation plays a distinctive role as a solution of two optimization problems in the context of an efficient simulation of max-stable processes. Our approach has the potential of considerably reducing the simulation time of max-stable processes.

Simulation of mixed moving maxima processes with monotone shape functions

Martin Schlather (Institute for Mathematical Stochastics, Georgia Augusta University)

The normalized spectral representation presented by Chen Zhou is used in the latest versions of the R package RandomFields to simulate max-stable random fields on a grid. The algorithm was implemented in a rather general way, so that the normalized spectral representation had to be modified for practical reasons, including a particular family of multivariate distribution functions, which seems to be new. As another peculiarity, the algorithm is based on two nested MCMC steps. Finally, we present the new user interface of RandomFields 3.0 that allows for a general, but rather simple access.

Scoring extremes

Philippe Naveau (Laboratoire des Sciences du Climat et l'Environnement (LSCE) & CNRS)

Assessing the quality of forecasts is of primary interest for meteorologists and a number of scoring rules are routinely used to compare forecasts. In the first part of this talk, we will give a brief overview of classical scores, e.g.. the Continuous Ranked Probability Score (CRPS) that can be easily interpreted and inferred (see, Gneiting and Raftery, 2005 ; Gneiting and Ranjan, 2011). We will discuss about the performances of such classical scores with respect to extremes. In the second part of the talk, we will focus on the multivariate case. On one hand, developing proper scores within a multivariate context remains a statistical challenge. On the other hand, judging the forecasts of maxima is of primary interest for many end-users in risk analysis. In this context, we focus on the special case of multivariate distributions tailored to model the behaviour of component-wise maxima. Multivariate extreme value theory indicates that max-stable random vectors provide a mathematically sound framework to model the dependence among multivariate maxima. It is then legitimate to investigate how it is possible to adapt a few classical scores for such max-stable vectors. If time allowed, we will also speak about the scores for excesses. This is joint work with R. de Fondeville and D. Cooley.

Conditioning the distribution of hydrologic extremes on large-scale atmospheric or oceanic fields

Benjamin Renard (Irstea - National Research Institute of Science and Technology for Environment and Agriculture)

Many studies report that hydrologic regimes are modulated by large-scale modes of climate variability such as the El Niño Southern Oscillation (ENSO) or the North Atlantic Oscillation (NAO). In particular, the occurrence of hydrologic extremes varies according to specific states of the oceanic/atmospheric system. Hydrologists have therefore developed "climate-informed" probabilistic models where the distribution of the target hydrologic variable is conditioned on standard climate indices (such as ENSO or NAO). This yields a dynamic, time-varying characterization of extremes that has applications in terms of e.g. downscaling or seasonal forecasting. A limitation of this approach is that standard climate indices such as NAO or ENSO have been primarily derived because they explain a significant part of the internal variability of climate. However, these standard indices turn out to be poor predictors to describe the external variability of hydrologic extremes in some regions or seasons. This presentation therefore describes a probabilistic framework that conditions the distribution of hydrologic variables directly on atmospheric or oceanic fields, as opposed to predefined climate indices. This framework is based on a 2-level probabilistic model describing both climate and hydrologic data. The climate dataset is typically a time series of atmospheric or oceanic fields defined on a grid over some area, while the hydrologic dataset is typically a regional dataset of station data (e.g. annual maximum flow at several gauging stations). In the first level, the probabilistic model postulates the existence of a latent temporal pattern that explains the temporal variability of hydrologic data. In the second level, the same latent temporal pattern is used to decompose the climate dataset into space and time components, therefore identifying the climate spatial pattern associated with the temporal variability of hydrologic data. A Bayesian estimation framework is used, so that a natural quantification of uncertainties affecting hydrologic predictions is available. The application of this model will be illustrated through a case study, with a focus on the predictive performances of such a dynamic model compared to baseline predictions that ignore climate information.

Some extreme value problems in snow avalanche risk assessment

Nicolas Eckert (Irstea - National Research Institute of Science and Technology for Environment and Agriculture)

Mountain communities are exposed to snow avalanches in winter. Hence, evaluating risk to extreme avalanches realistically is a crucial question for land use planning and the design of appropriate defense structures in these areas. For evaluating high return levels, the use of univariate extreme value models remains cumbersome because the most critical variable, the travelled distance, strongly depends on topography. Also, different damageable quantities (travelled distance, impact pressure, flow depth, deposit volumes) have to be considered and, for these, field data are generally not available, making the use of multivariate extremes practically impossible. The existing alternative is the combination of a mechanical model for flow propagation with a stochastic model describing the variability of the different inputs/outputs. High avalanche magnitude and frequency are modeled independently, leading to a sort of multivariate Peak Over Threshold (POT) model where the correlation between the different magnitude variables is constrained by the physical rules describing avalanche propagation. The problem is that the consistency of such an approach with extreme value theory in terms of attraction domain and asymptotic dependence of the different variables is then not guaranteed, making the question of model predictions critical. Also, in practice, return period/level approaches, purely hazard-oriented, have the drawback to not considering elements at risk explicitly (buildings, people inside, etc.), and to neglect possible budgetary constraints. To overcome these, risk based zoning methods and cost-benefit analyses have emerged recently. They combining hazard distribution and vulnerability relations (damage susceptibility functions) for various elements at risk. Here as well, computations can be made with standard extreme value models, or with a statistical numerical model, with advantages / drawbacks for each option. These points are discussed with reference to a real case study in the French Alps for which different kind of data are available.

On the block maxima method in EVT : PWM estimators

Ana Ferreira (ISA and CEAUL, University of Lisbon & Laurens de Haan (Erasmus University Rotterdam and CEAUL)

In extreme value theory (EVT) there are two fundamental approaches both widely used : the block maxima (BM) method and the peaks-over-threshold (POT) method. Whereas much theoretical research has gone into the POT method, the BM method has not been studied thoroughly. We shall discuss conditions under which the BM method can be justified. We shall also provide a theoretical comparative study of the methods through their probability weighted moment (PWM) estimators, which is in general consistent with the vast literature on comparing the methods all based on simulated data and fully parametric models. The results indicate that the BM method is a rather efficient method under usual practical conditions. We restrict attention to the independent and identically distributed case.

Implicit extremes and implicit max-Stable laws

Stilian Stoev (University of Michigan)

Let $X_1, \cdots, X_n$ be iid random vectors and $f \geq 0$ be a non–negative function. Let also $k(n) = \text{Argmax}_{i=1,\cdots,n} f(X_i)$. We are interested in the distribution of $X_{k(n)}$ and their limit theorems. In other words, what is the distribution the random vector where a function of its components is extreme. This question is motivated by a kind of inverse problem where one wants to determine the extremal behavior of X when only explicitly observing $f(X)$. We shall refer to such types of results as to *implicit extremes.* It turns out that as in the usual case of explicit extremes, all limit *implicit extreme value* laws are *implicit max–stable.* We characterize the regularly varying implicit max–stable laws in terms of their spectral and stochastic representations and illustrate the theory with examples drawing connections to *hidden regular variation* and regular variation on general cones. (joint work with Hans-Peter Scheffler)

Statistics for tail processes of Markov chains

Holger Drees (University of Hamburg), Michal Warchol & Johan Segers (Université Catholique de Louvain)

At high levels, the asymptotic distribution of a stationary, regularly varying Markov chain is conveniently given by its tail process. The latter takes the form of a geometric random walk, the increment distribution depending on the sign of the process at the current state and on the flow of time, either forward or backward. Estimation of the tail process provides a nonparametric approach to analyze extreme values. A duality between the distributions of the forward and backward increments provides additional information that can be exploited in the construction of more efficient estimators. The large-sample distribution of such estimators is derived via empirical process theory for cluster functionals. Their finite-sample performance is evaluated via Monte Carlo simulations involving copula Markov models and solutions to stochastic recurrence equations. The estimators are applied to stock market data to study the absence or presence of symmetries in the succession of large losses and gains.

An M-estimator of spatial tail dependence

Anna Kiriliouk (Université Catholique de Louvain)

Tail dependence models for distributions attracted to a max-stable law are fitted using observations above a high threshold. To cope with spatial, high-dimensional data, a rank-based M-estimator is proposed relying on bivariate margins only. A data-driven weight matrix is used to minimize the asymptotic variance. Empirical process arguments show that the estimator is consistent and asymptotically normal. Its finite-sample performance is assessed in simulation experiments involving popular max-stable processes perturbed with additive noise. An analysis of wind speed data from the Netherlands illustrates the method.

Full likelihood-based inference for max-stable processes : issues of bias and how they may be overcome

Jenny Wadsworth (University of Cambridge)

Recently full likelihood inference has become available for certain classes of max-stable (and related) processes. The inclusion of information on which extremes occurred in which event is fundamental to the viability of the full likelihood. This either entails taking an event-based approach, where inference is based on Poisson or Pareto process likelihoods, or including occurrence time information in the max-stable process likelihood. However, whilst event information makes full inference feasible by simplification of the likelihood, it can also cause or accentuate bias in inference for weakly-dependent processes. In this talk I will describe the ideas behind full likelihood inference for max-stable processes, and discuss how this bias can occur. Understanding of the bias issue helps to identify potential solutions, and I will illustrate one possibility that has been successful in a high-dimensional logistic model.

Likelihood based inference for high-dimensional extreme value distributions

Christian Robert (Institut de Science Financière et d'Assurances)

Abstract : Multivariate extreme value statistical analysis is concerned with observations on several variables which are thought to possess some degree of tail-dependence. In areas such as the modeling of financial and insurance risks, or as the modeling of spatial variables, extreme value models in high dimensions (up to fifty or more) with their statistical inference procedures are needed. In this paper, we consider max-stable distributions whose spectral random variables have absolutely continuous distributions. For random samples with max-stable distributions we provide quasi-explicit analytical expressions of the full likelihoods. When the full likelihood becomes numerically intractable, it is however necessary to split the variables into subgroups and to consider a composite likelihood approach. For random samples in the max-domain of attraction of a max-stable distribution, two approaches for which the likelihoods dramatically simplify are possible : (i) a threshold approach combined with a censoring scheme, (ii) a block maxima approach when adding information on the occurence times of extreme events. We also provide a method to identify the sprectal structure of the max-stable distribution that is all the more efficient as the dimension is large. The asymptotic properties of the estimators are given and the utility of the methods is examined via simulation. The estimators are also compared with those derived from the pairwise composite likelihood method which has been previously proposed in the spatial extreme value literature. Finally, a real data application on financial extreme events is presented.