Statistical Modeling of Extremes in Data Assimilation and Filtering Approaches

Strasbourg, 23-26 June, 2008

CONTENTS

Forward 4

Scientific Programme 5

Lectures

O. Cappé: *An Introduction to Sequential Monte Carlo for Filtering and Smoothing* 8

F. Chevallier: *Data assimilation for large state vectors* 9

R. Davis: *Extreme Value Theory in Times series analysis and Spatial Models with Applications in Computer Experiments* 10

A. L. Fougères: *Extreme Value Theory for multivariate data* 11

Invited Talks

E. Blayo: *Some tools for focusing variational data assimilation in ocean modelling* 12

M. Bocquet: *Singular behaviours in air quality data assimilation and network design* 13

D. Cooley: *Spatial Hierarchical Modeling of Weather Extremes from a Regional Climate Model* 14

G. Desroziers: *Data assimilation schemes in numerical weather forecasting and their link with ensemble forecasting* 15

C. Lantuéjoul, J. N. Bacro & L. Bel: *Max-stable random fields and stochastic geometry* 16

E. Parent, J. Bernier & J.J. Boreux: *Engineering applications of POT, GEV and compound Poisson models: a unified approach via Bayesian thinking* 17

H. Wackernagel: *Extremes and data assimilation in epidemiology* 18

Contributed Talks

J. Carreau: *Statistical rainfall-runoff models based on a mixture with heavy-tailed components* 19

L. de Montera, C. Mallet, L. Barthès & P. Golé: *The effect of the on-off intermittency on rain multifractal parameters* 20

A. Ferreira: *Extremal index and Markov chains* 21

M. Mehdi Gholam-Rezaee: *Spatial modeling of climate extremes: A composite likelihood approach* 22

P. Kalcheva Jordanova: *Functional transfer theorems for maxima of moving average with subexponential noise* 23

E. Masiello, A. Fougères & P. Soulier: *Estimating a discrete spectral measure* 24

T. Meinguet & J. Segers: *Regularly varying time series of random fields* 25

A. Necir & D. Meraghni: *Estimating L-functionals for heavy-tailed distributions* 26

P. Ribereau, A. Guillou & P. Naveau: *Generalized Probability Weighted Moments for the Generalized Extreme Value Distribution* 27

M. Süveges & A. Davison: *A Dirichlet mixture approach to the estimation of clusters of extreme events* 28

G. Toulemonde, A. Guillou, P. Naveau, M. Vrac & F. Chevallier: *Auto-Regressive models for maxima with applications to atmospheric chemistry* 29

M. Werner, K. Ide & D. Sornette: *Earthquake Forecasting Based on Data Assimilation* 30

FORWARD

This issue contains the programme as well as the abstracts of the lectures and oral presentations of the meeting *Statistical Modeling of Extremes in Data Assimilation and Filtering Approaches*, to be held at the University of Louis Pasteur, 23-26 June, 2008. The abstracts were ordered alphabetically.

This meeting is part of the AssimilEx ANR project entitled "Theoretical Developments of Data Assimilation Models for Climate Extremes" which is partly supported by the University Louis Pasteur. We would like to take this opportunity to thank the invited speakers, the participants and all those who through advise, sponsoring and organizational skills have contributed to organize this meeting.

Strasbourg, June 2008

Frédéric Chevallier, Anne-Laure Fougères
Armelle Guillou, Philippe Naveau

SCIENTIFIC PROGRAMME

Monday, June 23th, IRMA Room

- 1pm30-3pm : **Lecture given by Richard Davis**
- 3pm-3pm30 : **Coffee break**
- 3pm30-5pm : **Lecture given by Olivier Cappé**
- 5pm-5pm15 : **Break**
- 5pm15-6pm15 : **Lecture given by Frédéric Chevallier**

Tuesday, June 24th, IRMA Room

- 9am-10am30 : **Lecture given by Richard Davis**
- 10am30-11am : **Coffee break**
- 11am-12am30 : **Lecture given by Olivier Cappé**
- 12am30-2pm : **Lunch (sandwitches and drinks)**
- 2pm-3pm30 : **Lecture given by Anne-Laure Fougères**
- 3pm30-4pm : **Coffee Break**
- 4pm-5pm30 : **Lecture given by Frédéric Chevallier**

Wednesday, June 25th, IRMA Room

Session	Time	Author	Title
1	9:00-9:40am	M. Bocquet	*Singular behaviours in air quality data assimilation and network design*
1	9:40-10:05am	M. Werner et al.	*Earthquake Forecasting Based on Data Assimilation*
1	10:05-10:30am	G. Toulemonde et al.	*Auto-Regressive models for maxima with applications to atmospheric chemistry*
Coffee break	10:30-11:00am		
2	11:00-11:40am	E. Blayo	*Some tools for focusing variational data assimilation in ocean modelling*
2	11:40-12:05am	J. Carreau	*Statistical rainfall-runoff models based on a mixture with heavy-tailed components*
2	12:05-12:30am	M. Mehdi et al.	*Spatial modeling of climate extremes: A composite likelihood approach*
Lunch	12:30-2:00pm		
3	2:00-2:40pm	H. Wackernagel	*Extremes and data assimilation in epidemiology*
3	2:40-3:05pm	L. de Montera et al.	*The effect of the on-off intermittency on rain multifractal parameters*
3	3:05-3:30pm	M. Süveges et al.	*A Dirichlet mixture approach to the estimation of clusters of extreme events*
Coffee break	3:30-4:00pm		
4	4:00-4:40pm	G. Desroziers	*Data assimilation schemes in numerical weather forecasting and their link with ensemble forecasting*
4	4:40-5:05pm	P. Ribereau	*Generalized Probability Weighted Moments for the Generalized Extreme Value Distribution*
4	5:05-5:45pm	D. Cooley et al.	*Spatial Hierarchical Modeling of Weather Extremes from a Regional Climate Model*

Thursday, June 26th, IRMA Room

Session	Time	Author	Title
5	8:45-9:25am	E. Parent et al.	*Engineering applications of POT, GEV and compound Poisson models: a unified approach via Bayesian thinking*
5	9:25-9:50am	A. Ferreira	*Extremal index and markov chains*
5	9:50-10:15am	E. Masiello et al.	*Estimating a discrete spectral measure*
5	10:15-10:40am	P. Kalcheva Jordanova	*Functional transfer theorems for maxima moving average with subexponential noise*
Coffee break	10:40-11:10am		
6	11:10-11:50am	C. Lantuéjoul et al.	*Max-stable random fields and stochastic geometry*
6	11:50-12:15am	T. Meinguet at al.	*Regularly varying time series of random fields*
6	12:15-12:40am	A. Necir et al.	*Estimating L-functionals for heavy-tailed distributions*
Lunch	12:45-2:00pm		

An Introduction to Sequential Monte Carlo for Filtering and Smoothing

Olivier Cappé

Telecom, ParisTech

Sequential Monte Carlo methods or particle filters are numerical simulation-based approaches to the filtering and smoothing problems in partially-observed discrete-time dynamic state-space models (also sometimes called hidden Markov models). Sequential Monte Carlo is based on importance sampling and resampling and is truly sequential, which makes it suitable for online processing.

In this talk, I will review the classical filtering and smoothing recursions and describe some basic sequential Monte Carlo approaches such as the bootstrap filter and the auxiliary particle filter. More advanced topics to be discussed include smoothing approaches as well as methods that combine Monte Carlo simulations and exact computations, such as mixture Kalman filters.

Data assimilation for large state vectors

Frédéric Chevallier

Laboratoire des Sciences du Climat et de l'Environnement

Bayes' theorem describes how new information improves the knowledge of any phenomenon: it is a mathematical model of learning. Discovered in the 18th century, it provides a generic solution to inverse problems and has become a reference for statistical estimation methods. Its application involves diverse skills, because each component of an inversion or data-assimilation system requires a specific study, about statistics, measurements, modeling and computer engineering.

This talk focuses on the application of Bayes' theorem to optimize a large state vector, with typically millions of variables. It will be illustrated with two broad examples. One is the assimilation of atmospheric observations for numerical weather prediction, the other ones is the estimation of global CO2 surface fluxes from atmospheric CO2 concentrations. Various formulations of the theorem will be explored in order to tackle the dimension issue: analytic, variational and by ensembles (weighted, like particle filters, or not, like ensemble Kalman filters). The last part of the talk will be devoted to the assignment of error statistics for observations and for prior information.

Talk 1: Extreme Value Theory in Times series analysis

Talk 2: Spatial Models with Applications in Computer Experiments

Richard Davis

Columbia University

Abstract 1: A broad overview of Extreme Value Theory (EVT) for times series models will be presented. This review will include the study of temporal dependences for extremes (extreme index, clustering, ...) and other topics dealing with heavy and light tailed distributions. Applications will be discussed.

Abstract 2: Often, a deterministic computer response is modeled as a realization from a stochastic process such as a Gaussian random field. In this talk, we describe this stochastic modeling approach and will focus on the use of a stochastic heteroskedastic process (SHP). a stationary non-Gaussian process with non-stationary covariance function. By conditioning on a latent process, the SHP is a non-stationary Gaussian process. As such, the sample paths of this process exhibit greater variability and hence offer more modeling flexibility than those produced by a traditional Gaussian process (GP) model. By introducing a spatial stochastic volatility component into the Gaussian process model, the SHP models the spatial heteroskedasticity jointly with the spatial correlation in the computer code outputs. The SHP model can also recover Gaussian-like sample paths for certain model parameter values. We use maximum likelihood for inference, which is complicated by the high dimensionality of the latent process. Accordingly, we develop an importance sampling method for likelihood computation and use a low-rank kriging approximation to reconstruct the latent process. Responses at unobserved locations can be predicted using empirical best predictors or by empirical best linear unbiased predictors. Prediction error variances are also obtained. In examples with simulated and real computer experiment data, the SHP model is superior to traditional Gaussian process models. In addition, the SHP model can be used in an active learning context to select new locations that provide improved estimates of the response surface. Implementing active learning via the SHP model appears to work better than other traditional approaches.

(This work is joint with Ke Wang, Wenying Huang and Jay Breidt.)

Extreme Value Theory for multivariate data

Anne-Laure Fougères

MODAL'X, Nanterre

As a complement to Richard Davis' presentation, an overview of Multivariate Extreme Value Theory (MEVT) will be given. We will see in particular how the spatial dependence can be handled in some ways using MEVT. Applications will be discussed. We will also exhibit that things might be different depending if the probability that two or more components are large at the same time is negligible with respect to the probability that one component is large.

Some tools for "focusing" variational data assimilation in ocean modelling

Eric Blayo

University of Grenoble and INRIA

Some tools are presented, which aim is to concentrate the effect of data assimilation on particular aspects of interest.

In a first part, a reduced order approach for 4D-Var data assimilation is described. The control space is defined as the span of a few vectors representing a significant part of the system variability. It is shown that such an approach can lead to significant improvements, both in terms of the quality of the solution and of the computational efficiency, with regard to data assimilation with a full control vector. However this approach presents also several limitations. In particular, the choice and the evolution of the reduced basis are discussed.

In a second part, we address the problem of variational data assimilation for nested models. An adjoint formulation is derived, and it is shown in particular that assimilation considering the whole multigrid system can lead to improved results with regard to assimilation in the high resolution model only.

Examples of applications will be given in the context of ocean modelling.

Singular behaviours in air quality data assimilation and network design

Marc Bocquet

CEREA, joint laboratory ENPC/EDF R&D Université Paris-Est and INRIA

With increasingly powerful computing means, applied mathematical techniques are more and more used in high dimensional geophysical applications. Following meteorology and oceanography, data assimilation is now used in atmospheric chemistry and air quality. Other optimisation issues, such as the design of monitoring station network is emerging in air quality. In air quality, these tools (data assimilation, network design) have been so far mostly dedicated to the physics of ozone and its precursors.

We have recently used and extended these methods in the context of an accidental release of pollutant, from short, through mesoscale, to continental scale. I will show that, due to the physics of dispersion, new mathematical and physical problems are emerging in this context. This will be illustrated on two problems: the control space resolution choice in inverse modelling of tracers, and the design of a network of monitoring stations in case of an accidental release of pollutant.

In particular, I will show that the analysis is affected by events that could be considered as rare (high concentrations areas in the wake of the accidental plume or near the sources). Singular behaviours may then occur: improper high resolution limit in one hand and collapse of the optimal network in the other hand.

Spatial Hierarchical Modeling of Weather Extremes from a Regional Climate Model

Dan Cooley

Colorado State University

Regional climate models (RCMs) are tools which allow scientists to begin to understand how different forcings may affect climate. There has been some statistical work done to characterize the difference in mean behavior between control and future scenarios as predicted by RCMs. The goal of this work is to characterize the extremes as produced by a RCM and to additionally examine the difference in extremes between a control and future scenario.

To characterize the spatial behavior of extremes we construct a hierarchical model. The data level is formed by the point process representation of extremes, and the process level is based on a conditional autoregressive (CAR) model since our data are on a regular lattice. To our knowledge, this is the first work which spatially models the shape parameter of the extreme value distribution.

Data assimilation schemes in numerical weather forecasting and their link with ensemble forecasting

Gérald Desroziers

Météo France

Modern operational data assimilation schemes rely today on the variational formalism. This makes possible the use of a large range of types of observations, via the implementation of observation operators that allow to go from model space to observation space and that especially permit the direct assimilation of satellite radiances. Such assimilation schemes are also directly linked with the Kalman filter formalism, which combines observations with a background given by a forecast issued from a previous analysis. As a consequence, there is a need to diagnose and specify observation and background error covariances. Because of the large size of the state vectors used in operational weather forecasting, the determination and representation of flow-dependent background error covariances is a difficult problem. Interestingly, small size ensembles of perturbed assimilations are more and more implemented in order to document these background error covariances for a reasonable cost. Such an use of ensemble approaches makes the link between data assimilation and ensemble forecasting more and more evident.

Max-stable random fields and stochastic geometry

Christian Lantuéjoul[1], and Noël Bacro[2] & Liliane Bel[3]

[1] École des Mines de Paris, France

[2] Université de Montpellier II, France

[3] AgroParisTech, France

Stochastic geometry provides a number of analytical tools that can help to comprehend the statistical properties of d-dimensional random fields with extreme value distributions. As a first example, all compatibility relationships between the multivariate extremal coefficients of a max-stable random field with unit Fréchet marginals are given by a positive and σ-finite measure on the closed subsets of $\mathbb{R}^d$. In a second example, a variation of the storm process is designed to have its spatial distribution analytically tractable; moreover this process can be exactly simulated in any bounded domain of $\mathbb{R}^d$. It can be used not only to describe physical phenomena, but also to assess the performance of procedures for estimating extremal coefficients.

Engineering applications of POT, GEV and compound Poisson models: a unified approach via Bayesian thinking

Eric Parent[(1)], Jacques Bernier[(1,2)], Jean Jacques Boreux[(2)]

Laboratoire de Modélisation, Risque, Statistique, Environnement, France
Département des Sciences et Gestion de l'Environnement, ULg, Belgique

Conditional reasoning is the cornerstone of both steps of Bayesian analysis : stochastic models can be naturally specified by assembling conditionnal layers of variables and inference- via MCMC algorithms- also relies on probabilistic conditionning. Conditional reasoning casts the stage ligths onto latent variables. It may be useful, even in a frequentist perspective. We emphasize the use of latent variables in the analysis of three different marked points models applied to meteorogical data for extreme value analysis. The explicit introduction of such hidden variables helps for a better conceptualisation of a realistic model. Together with the "augmentation data algorithms" it also provides a straightforward treatment of the inferential problems.

Extremes and data assimilation in epidemiology

Hans Wackernagel

École des Mines de Paris, France

Past influenza pandemics, notably the Spanish influenza (1918-1919) with a death toll of more than 40 million people, and other pandemics such as Asian influenza (1957), Hong Kong influenza (1968), Russian influenza (1977), have greatly impacted humans all around the world. Influenza leads to pneumonia in the more serious cases, and most influenza deaths result from secondary bacterial pneumonia.

In this paper for the Assimilex workshop we deal with two topics: the use of extreme value theory to model series of pneumonia and influenza (P&I) mortality data from the US, the use of data assimilation by particle filtering for the early detection of the outbreak of the yearly major influenza epidemic in France.

Statistical rainfall-runoff models based on a mixture with heavy-tailed components

Julie Carreau

Laboratoire des Sciences du Climat et de l'Environnement

Rainfall-runoff models are used to predict the distribution of river flows given appropriate meteorological variables. We propose a class of stochastic latent models that better accounts for extreme floods while still giving a sensible model for small to medium discharge. Latent states can capture the different modes of runoff processes. Given the latent state value, the runoff is modelled with a flexible distribution, based on the Generalized Pareto distribution. We evaluate the proposed class of models on the Orgeval basin data. The latent state and the runoff distributions depend on rainfall measurements from the surrounding area.

The effect of the on-off intermittency on rain multifractal parameters

L. de Montera, C. Mallet, L. Barthès and P. Golé

Centre d'étude des Environnements Terrestres et Planétaires,
Vélizy-Villacoublay, France

In this paper we investigate the multifractal properties of rain within the Universal Multifractal framework developed by Schertzer & Lovejoy (1987). The database includes measurements performed over several months in different locations (USA, Bénin, and France) by means of a dual-beam spectropluviometer. An assessment of the effect of the on-off intermittency shows that the analysis of long time series may lead to an artificial scaling and consequently to erroneous parameter estimates. The estimation of rain multifractal parameters is therefore performed on an event-by-event basis and they are found to be significantly different from those proposed in the literature. In particular, the non-conservation parameter H, that has often been estimated to be 0, is more likely to be 0.53, meaning that rain is a non-conservative fractionally integrated multiplicative cascade (Fractionally Integrated Flux model). Another important result is that rain multifractal parameters are found to be the same regardless of the measurement location, thus indicating that they could be constant worldwide. Finally, simulations of synthetic time series with these new parameters show that the model is capable of reproducing the on-off intermittency and especially the fractal dimension of the rain periods.

Extremal index and Markov chains

Ana Ferreira

Instituto Superior de Agronomia, da Universidade Técnica de Lisboa

We present a class of consistent estimators for the extremal index, a well known dependence coefficient for extreme values of stationary sequences (Leadbetter, Lindgren and Rootzén, 1983), under clear and flexible conditions. Our motivation started from looking at alternative approaches to deal with dependence, particularly for Markov chain sequences. We complement our theoretical results with some simulations.

Spatial modeling of climate extremes: A composite likelihood approach

M. Mehdi Gholam-Rezaee   Supervised by Prof. A. C. Davison

École Polytechnique Fédérale de Lausanne

Modelling of spatial extremes is increasingly recognised as crucial to environmental risk estimation. In this talk parametric models for the spatial dependence structure of climate extremes are revisited, and a composite likelihood approach is employed to estimate their parameters. The occurrence times of extremes are also utilized to improve the efficiency of estimation, and some diagnostic methods are proposed to validate the results. The approach is applied to multivariate series of extreme temperatures in Switzerland, for which results are presented.

Functional transfer theorems for maxima moving average with subexponential noise

Pavlina Kalcheva Jordanova

Faculty of Mathematics and Informatics, Bulgaria

In this paper we investigate the asymptotic behaviour of sequences of random processes, which time-intersections are maxima of random number of stationary moving average. The distribution function of the noise components has regularly varying tails or is subexponential and belongs to the domain of attraction of Gumbel distribution. The time points are almost surely increasing to infinity. The max-increments of these processes are not independent. Here is proved that such a sequence of random processes converges weakly to a compound extremal process. In particular we consider cases when the counting process is a mixed Poisson, or when the time points constitute a renewal process.

Estimating a discrete spectral measure

Esterina Masiello, Anne-Laure Fougères & Philippe Soulier

Modal'X, Université Paris Ouest-Nanterre

Let $\|\cdot\|$ denote any norm on $\mathbb{R}^2$ and $\mathcal{S}_1$ denote the unit sphere for this norm. Consider a bivariate random vector $\mathbf{X}$ whose distribution is regularly varying at infinity, i.e. such that there exists a probability measure ν on $\mathcal{S}_1$ satisfying

$$\lim_{r\to\infty} P\left(\frac{\mathbf{X}}{\|\mathbf{X}\|} \in A \mid \ \|\mathbf{X}\| > r\right) = \nu(A)$$

for any Borel set A. Under the assumption that the spectral measure ν is a point measure, how can we estimate ν ? We propose a simple estimator, study some of its theoretical properties and provide a simulation study to illustrate its small sample behaviour.

Regularly varying time series of random fields

Thomas Meinguet & Johan Segers

Université catholique de Louvain, Institut de statistique, Belgique

The natural way to model spatial phenomena is to resort to continuous functions. As for random vectors, regularly varying continuous random fields can be characterized by a spectral measure or, equivalently, an exponent measure. A stationary time series of continuous random fields is said to be jointly regularly varying if all of its finite stretches are regularly varying random fields with the same index. This property is shown to be equivalent to the existence of a limit process, called the tail process, which admits a regularly varying radial component and a sequence-valued angular component. As a first example of application, an explicit formula for the extremal index is obtained thanks to the spectral process, which is a renormalization of the tail process.

Estimating L-functionals for heavy-tailed distributions

Abdelhakim Necir & Djamel Meraghni

Mohamed Khider University, Biskra, Algeria

L-functionals summarize numerous statistical parameters and actuarial risk measures. Their sample estimators are linear combinations of order statistics ($L-$statistics). There exists a class of heavy-tailed distributions for which the asymptotic normality of these estimators cannot be obtained by classical results. In this paper we propose, by means of extreme value theory, alternative estimators for $L-$functionals and establish their asymptotic normality. Our results may be applied to estimate the trimmed $L-$moments and financial risk measures for heavy-tailed distributions.

Generalized Probability Weighted Moments for the Generalized Extreme Value Distribution

Pierre Ribereau[(1)], Armelle Guillou[(2)] & Philippe Naveau[(3)]

[(1)] Université de Montpellier II, France

[(2)] Université de Strasbourg I, France

[(3)] Laboratoire des Sciences du Climat et de l'Environnement

The Probability Weighted Moments (PWM) of a random variable Z with distribution function F were introduced by Greenwood *et al.* (1979) and are defined by :

$$M_{p,r,s} = \mathbb{E}\left[Z^p (F(Z))^r (1 - F(Z))^s\right]$$

for real numbers p, r and s. The specific case of PWM parameter estimation for the $GEV(\sigma, \gamma, \mu)$ was extensively studied in Hosking *et al.* (1985).

However, the main problem with the PWM method is the range of validity ($\gamma < 1/2$) in order to derive the asymptotic normality of their estimators. In fact this restriction is quite restrictive for many applications. So, in order to solve this problem, we propose to use the Generalized Probability Weighted Moments (GPWM) method recently introduced by Diebolt *et al.* in the case of the Generalized Pareto Distribution.

A generalized probability weighted moment is the quantity defined by

$$\nu_\omega = \mathbb{E}\left[Zw(F)\right]$$

where ω is a suitable function.

We investigate the asymptotic properties of the estimators obtained and we extend the use of the GPWM method in a non stationary case.

A Dirichlet mixture approach to the estimation of clusters of extreme events

Mária Süveges & Anthony Davison

École Polytechnique Fédérale de Lausanne, Suisse

We propose a new way of dealing with extreme clusters and estimating their characteristics, by approximating the distribution of the relative sizes of observations within a cluster by a mixture of Dirichlet densities. Smith and Weissman (1996, unpublished) argued that the extremes of a large class of multivariate stationary dependent sequences can be approximated by a multivariate maxima of moving maxima (M4) process, which models the observed neighbourhood of the extreme events as the result from the filtering of a collection of independent, identically distributed unit Fréchet sequences. The "signatures", the images of the filter matrices in the observed series are blurred by finite threshold choice, measurement errors, and other effects; we model this uncertainty by the assumption that the observed signatures are a sample from a mixture of Dirichlet distributions with the true M4 signatures as their means. Estimation is performed by the EM algorithm. Some diagnostic methods to assess the quality of the fit are also proposed, and a data example, the bivariate set of daily summer maximum temperatures from two Swiss sites, Arosa and Bern is shown.

Auto-Regressive models for maxima with applications to atmospheric chemistry

Gwladys Toulemonde[1], Armelle Guillou[2], Philippe Naveau[3]
Mathieu Vrac[3] & Frédéric Chevallier[3]

[1] Université Paris VI, France

[2] Université de Strasbourg I, France

[3] Laboratoire des Sciences du Climat et de l'Environnement, France

With the increasing number of extreme value theory applications particularly in meteorological problems, it is important to take into account the complexity of certain non independent and temporal observations in statistical models.
In time series analysis, Markov Auto Regressive (AR) processes are convenient because they consider the dependence between the observed value at time t and values at earlier neighboring time points. By this way, AR processes describe persistence among observations. But unfortunately there is no linear AR model dealing with the generalized extreme value distribution.
In this talk we propose an Autoregressive Markov model of order 1 adapted to the Gumbel distribution. We derive specific properties on such a process and we give general properties in the multivariate case. Then, we fit the model on daily and weekly maxima of methane (CH4) and carbon dioxide (CO2) in Gif-sur-Yvette (France). Finally, simulation results are presented in order to assess the quality of the parameter estimations and of our model on finite samples.

Earthquake Forecasting Based on Data Assimilation

Maximilian Werner[1], Kayo Ide[2], Didier Sornette[1]

[1] ETHZ; [2] UCLA

The Collaboratory for the Study of Earthquake Predictability (CSEP) aims to prospectively test time-dependent earthquake probability forecasts on their consistency with observations. To compete, time-dependent seismicity models are calibrated on earthquake catalog data. But catalogs contain much observational uncertainty. We study the impact of magnitude uncertainties on rate estimates in models of clustering, on their forecasts and on their evaluation by CSEP's consistency tests. First, we quantify magnitude uncertainties. We find that magnitude uncertainty is more heavy-tailed than a Gaussian, and is distributed at least as broadly as a double-sided exponential distribution, with scale parameter $nu_c = 0.1 - 0.3$. Second, we study the impact of such noise on the forecasts of a simple model of clustering which captures the main ingredients of popular short term models. We prove that the deviations of noisy forecasts from an exact forecast are power law distributed in the tail with exponent $\alpha = 1/(a * nu_c)$, where a is the exponent controlling the productivity law of aftershocks by a given mainshock as a function of its magnitude. We further prove that the typical scale of the fluctuations remains sensitively dependent on the specific catalog. Third, we study how noisy forecasts are evaluated in CSEP consistency tests. As a consequence of the above two points, noisy forecasts are rejected more frequently than expected for a given confidence limit. The Poisson assumption of the consistency tests is inadequate for short-term forecast evaluations. To capture the idiosyncrasies of each model together with any propagating uncertainties, the forecasts need to specify the entire likelihood distribution of seismic rates.

Data assimilation is a framework for accounting for such uncertainties: noisy observations can be assimilated via Bayes' theorem with the model forecast as prior. For nonlinear and non-Gaussian models such as (most) point processes, the full distribution can be evolved using sequential Monte Carlo methods. We present an implementation of a sampling importance particle filter for renewal processes and for noisy temporal observations. We evaluate the improvement of the likelihood scores of the particle filter over the benchmark, which assumes that observations are exact, by numerical identical twin experiments, in which true occurrences and observations are based on the same model, with the forecast model subject to a prescribed observation error uncertainty. We discuss theoretical and practical challenges and opportunities of applying sequential Monte Carlo methods to renewal and more general point processes and, in particular, noisy observations of magnitudes.