Groupement De Recherche
MASCOT NUM

Méthodes d'Analyse Stochastique pour les COdes et Traitements NUMériques

SOCIETE FRANÇAISE DE STATISTIQUE

Groupe Environnement et statistique

A l'occasion de la visite de ***Michael Stein****, Professeur à l'Université de Chicago, les groupes "Environnement" du GDR Mascot-Num et de la SFdS organisent une Journée sur:*

Modélisation pour l'environnement et expérimentation numérique

le mardi 21 juin 2011

AgroParisTech, Amphi Riesler
16 rue Claude Bernard, 75005 Paris

Inscription (gratuite) sur le site Mascot-Num: http://www.gdr-mascotnum.fr/juin11.html

Programme :
Accueil à partir de 9h

- 9h30-9h40 : Présentation de la Journée. Clémentine Prieur (LJK/MOISE), Hervé Monod (INRA), Liliane Bel. (AgroParisTech).
- 9h40-10h40 : Michael Stein (Chicago's University, Department of Statistics).
  Statistical Analysis of Large Space-Time Datasets.
- 10h40-11h20 : Clémentine Prieur (LJK/MOISE).
  Spatio-temporal prediction for West African Monsoon.
- 11h20-12h : Philippe Naveau (LSCE).
  Extreme value theory and state space modeling for environmental data.

Pause déjeuner libre

- 14h-14h40 : Christian Lantuéjoul (MinesParisTech).
  Prediction by conditional simulation.
- 14h40-15h20 : Reinhard Furrer (University of Zurich).
  A Spatial Analysis of Multivariate Output from Regional Climate Models.

Pause

- 15h40-16h20 : Marc Bocquet (Université Paris Est et INRIA CEREA).
  Filtrage d'ensemble et assimilation de données en très grande dimension.
- 16h20-17h00 : Denis Allard (Unité BioSP, INRA Avignon).
  A weather generator for generating multivariate climatic series.
- 17h00-17h15 : Conclusion de la Journée

Avec le concours de: AgroParisTech ANR Costa Brava

Résumés :

- **Spatio-temporal prediction for West African Monsoon**
  Clémentine Prieur (Université de Grenoble et INRIA Grenoble Rhône Alpes)

  We are interested in the large-scale environmental phenomenon of West African monsoon. It is the major atmospheric phenomenon which drives the rainfall regime in Western Africa. The causes of spatio-temporal variability in monsoon rainfall have not yet been determined in an unequivocal manner. However, there is a considerable body of evidence suggesting that spatio-temporal changes in sea surface temperatures in the Gulf of Guinea (SST) and changes in the Saharan and sub-Saharan albedo are major factors. To simulate the rainfall, a regional atmospheric model (MAR) is used. An important point in this study is that the numerical storage and processing of model outputs, as far as the statistical description of the data, requires considerable computation resources. Hence an important preliminary step is the construction of a stochastic spatio-temporal metamodel approximating the MAR.

  In this work, we focus on the impact of SST on precipitation. We first present the functional modeling for both SST and precipitation. Then we perform a regularized regression of the precipitation Êon the sea surface temperatures.

  Modélisation spatio-temporelle de la mousson en Afrique de l'Ouest
  Nous nous intéressons au phénomène majeur que représente la mousson en Afrique de l'Ouest. Il s'agit du principal phénomène atmosphérique qui pilote le régime des pluies en Afrique de l'Ouest. Les causes de la variabilité spatio-temporelle des pluies de mousson ne sont pas déterminées de manière univoque. Ceci dit, les physiciens sont d'accord sur le fait que la température en surface de l'océan dans le Golf de Guinée (SST) ainsi que l'albédo saharien et sub-saharien sont des facteurs déterminants. Pour simuler la pluie, un modèle atmosphérique régional (MAR) est utilisé. Le stockage numérique et le traitement des sorties du modèle nécessitent des ressources de calcul considérables. C'est pourquoi, avant d'envisager une analyse de sensibilité sur ce type d'applications, une étape préliminaire est la construction d'un méta-modèle stochastique spatio-temporel reliant les sorties du modèle (précipitations en différents points d'une grille spatio-temporelle) aux entrées (SST en différents points d'une grille spatio-temporelle couvrant le Golf de Guinée sur la période active du phénomène). Ensuite, compte tenu des dimensions du problème, nous adoptons une approche de régression pénalisée.

- **Extreme value theory and state space modeling for environmental data**
  Philippe Naveau (Laboratoire des Sciences du Climat et de l'Environnement, Gif-sur-Yvette), joint work with Anne Sabourin (LSCE) and Gwladys Toulemonde (Montpellier university).

  We study two state-space models that are closely related to extreme value theory. The first one deals with the temporal modeling of maxima that follow a Gumbel distribution for both the space and observational equations. It is applied to daily maxima of CH_4 and CO recorded at Gif-sur-Yvette. Different filtering methods are compared. The second state-space model focuses on the elliptical distribution. It allows the modeling of a wide range of tail behaviors, from bounded to heavy tails. The Kalman filter equations are derived when a Generalized Pareto Distribution generator is used within an elliptical state-space framework. This model performance is assessed via simulated data and applied to real observations. Advantages, limits and drawbacks of these two approaches will also be discussed.

- **Prediction by conditional simulation**
  Christian Lantuéjoul (MinesParisTech).

  Predicting spatial phenomena is a problem commonly encountered in many fields including the earth and environmental sciences, agriculture, forestry, meteorology and telecommunications to name just a few. Standard prediction techniques (e.g. linear regression) are not always satisfactory because they cannot integrate the structural aspects of the phenomenon under study, or account for the complexity of the feature of interest to predict, or even cope with the diversity of the available data.

  Under such circumstances, conditional simulations offer an interesting alternative. These are realizations of a random field honoring the conditions or constraints prescribed by the data. By replicating conditional simulations (Monte Carlo procedure), one can derive not only a predictor of any feature of interest, but also its precision.

  In this presentation, various algorithms will be presented for conditionally simulating various random fields (Gaussian random fields, their excursion sets, Boolean models, Cox process). Specific techniques will be mentioned to accommodate large multisupport data sets, where standard techniques fail.

- **A Spatial Analysis of Multivariate Output from Regional Climate Models**
  Reinhard Furrer (University of Zurich, Institute of Mathematics), joint work with Steve Sain (NCAR) and Noël Cressie (OSU).

  Climate models have become an important tool in the study of climate and climate change, and ensemble experiments consisting of multiple climate-model runs are used in studying and quantifying the uncertainty in climate-model output. However, there are often only a limited number of model runs available for a particular experiment, and one of the statistical challenges is to characterize the distribution of the model output. To that end, we have developed a multivariate hierarchical approach, at the heart of which is a new representation of a multivariate Markov random field. This approach allows for flexible modeling of the multivariate spatial dependencies, including the cross-dependencies between variables. We demonstrate this statistical model on an ensemble arising from a regional-climate-model experiment over the western United States, and we focus on the projected change in seasonal temperature and precipitation over the next 50 years.

- **Filtrage d'ensemble et assimilation de données en très grande dimension**
  Marc Bocquet (Université Paris-Est et INRIA CEREA, laboratoire commun École des Ponts ParisTech et EdF R&D)

  Suite aux efforts de la communauté météo dans les années 1980 et 1990, les méthodes variationnelles pour l'assimilation des observations se sont brillamment imposées, et ont été validées de façon opérationnelle. Elles sont dérivées de la théorie du contrôle optimal et permettent d'initialiser la prévision numérique du temps. Dans un contexte de forte non-linéarité des modèles et de la très grande dimensionalité des systèmes physiques (typiquement jusqu'à 10^9 variables), les méthodes de filtrage de

Kalman sont inadaptées, et ne peuvent rivaliser. Cependant des méthodes réduites ou bien des méthodes d'ensemble, et notamment le filtre de Kalman d'ensemble, s'imposent de plus en plus comme des concurrents sérieux des méthodes variationnelles. Toutefois le succès de ces méthodes repose sur un certain nombre d'astuces, l'inflation et la localisation en particulier, pas toujours bien comprises ni justifiées.

Lors de ce séminaire, je ferai un point sur l'état de l'art sur les questions de filtrage d'ensemble et de filtrage particulaire telles que perçues par la communauté océan-atmosphère (météorologie, océanographie, chimie atmosphérique, etc.). Dans un deuxième temps, j'évoquerai un travail personnel récent destiné à comprendre l'impact des erreurs d'échantillonnage dans le filtre de Kalman d'ensemble, et qui conduit à construire un filtre d'ensemble permettant de s'affranchir partiellement de l'inflation.

- **A weather generator for generating multivariate climatic series** *Denis Allard* (Unité Biostatistique et Processus Spatiaux, INRA Avignon), en collaboration avec Cédric Flecher et Philippe Naveau (LSCE, CNRS Gif-sur-Yvette)

  Stochastic weather generators (Kats, 1996; Semenov, 1997; Qian et al., 2005) aim at reproducing the statistical distributional properties of meteorological variables. They have been applied to a wide range of hydrological, ecological, and agricultural studies. Agronomical models and more specifically crop models need a large variety of daily or hourly weather data as inputs (Brisson et al., 2003; Brisson et al., 2009), to model past, present and future variability for yields. In impact studies, such inputs have to be simulated quickly and easily for long time periods at a given station.

  In this paper we focus on five variables: minimum and maximum temperatures Tn and Tx, precipitation P, wind speed at two meters V and radiation R. The choice of these variables was motivated by the inputs required for the crop models used in a research project (CLIMATOR project) aimed at exploring the impact of climate change on agriculture in the XXIst century.

  We will first present WACS-gen (Flecher et al., in press), a stochastic weather generator at the daily time lag; we will then show how daily precipitation data can be disaggregated at the hourly time lag, using a stochastic disaggregation technique.

  When a seasonal trend is detected on a variable, the median and the average absolute deviation (defined as the mean of absolute difference between the variable and its median) are computed for each day and smoothed by a spline function. This smoothed median is then subtracted to the studied variable and the difference is rescaled by the smoothed average absolute deviation. These residuals are then assumed to be stationary; to account for seasonal differences in the weather state regimes, the standardized variables are studied independently within the four following seasons: December-January-February (DJF), March-April-May (MAM), June-July-August (JJA) and September-October-November (SON) (Semenov et al., 1998).

  In contrast to most previous weather generators (e.g., Semenov et al., 1998) for which only two states (dry days and wet days) were defined, we have chosen to extend the number of daily states in WACS-gen. This strategy allows us to better capture the complexity of weather changes. Weather states (and their number) are obtained by running the clustering algorithm Mclust (Fraley, 2003) separately on wet days and on dry days. The transitions between successive weather states are modelled by a first-order homogeneous Markov chain.

Conditionally on the precipitation state, the other meteorological variables are modelled using the family of multivariate closed skew-normal (CSN) distributions (Allard and Naveau, 2007; Genton, 2004) . This class of distribution offers a general framework to fit both non-Gaussian and Gaussian variables. Unconditionally to the weather state the climatic variables are thus a mixture of CSN distributions.

It will be illustrated on climatic series recorded in France between 1970 and 1999.