# 1

## Statistiques Spatiales en Écologie, Épidémiologie et Environnement

Session organisée par Rachid SENOUSSI et Joël CHADŒUF

Les recherches actuelles en environnement, écologie et épidémiologie pointent de façon cruciale sur la nécessité d'introduire la composante spatiale et dynamique pour mieux appréhender les phénomènes. Le besoin se fait sentir aussi pour l'exploitation de données de plus en plus abondantes recueillies par des systèmes de surveillance (épidémies, pollution, biodiversité, couvert végétal,etc).

L'objectif de cette session est double. D'une part, on entend montrer par l'exemple, l'intérêt à introduire des concepts de statistiques spatiales pour aborder de façon plus efficace ces applications. D'autre part, on cherche aussi à montrer par l'exemple comment, ces mêmes applications permettent de "faire remonter" des questions théoriques ou méthodologiques nouvelles.

Outre la diversité des applications, le choix des exposés a été guidé aussi par le souci de présenter des cadres de statistiques spatiales un peu divers incluant les processus ponctuels, les champs sur grilles régulières, et les données aggrégées sur tessellations.

*Références :*


[1] *Highly Structured Stochastic Systems,* (2003) Eds P. J. Green, N. L. Hjort and S. Richardson, Oxford University Press.

[2] *Disease Mapping and Risk Assessment for Public Health,* (1999) Eds A. Lawson, John Wiley, Chichester.

[3] *Stochastic Geometry and its Applications,* (1995) D. Stoyan, W.S. Kendall and J. Mecke, John Wiley, Chichester.

[4] *Statistics for Spatial Data,* (1993) N. Cressie, Wiley, NY.



*Adresse de l'organisateur :*

Rachid Senoussi et Joël Chadœuf
Unité de biométrie, INRA
Domaine St Paul, Site Agroparc
84 914 Avignon Cedex 9, France
E-mail : senoussi@avignon.inra.fr et chadoeuf@avignon.inra.f
<http://www.avignon.inra.fr/internet/unites/biometrie/>

## Sur les méthodes de détection de clusters spatio-temporels

par Émilie Gay et **Rachid Senoussi**

Dans un premier temps, nous passons en revue les principales méthodes statistiques de détection de clusters spatiaux (scan statistics,...). Nous tenterons de montrer dans un cadre homogène les spécificités de chacune de méthodes et les contextes théoriques et pratiques de leurs application et ce qu'elles détectent réellement. Les méthodes sont illustrées sur un ensemble de données comprenant les scores des cellules somatiques dans lait (indicatrices biologiques utilisées dans le suivi des mammites bovines) pour 5000 élevages laitiers français et un certain nombre de covariables explicatives.

Dans une seconde partie, nous exposons une contribution propre des auteurs avec 2 méthodes de detection d'aggrégats de variables continues. La première, non paramétrique, s'appuie sur la définition de statistiques pertinentes basées sur les (pseudo)distances entre distributions (Hellinger, Kullback,...) et les procédures de randomisation classiques. La seconde, paramétrique (ou semi-paramétrique), inspirée des méthodes de durées de survie et en particulier du modèle de survie de Cox, propose un modèle d'une fonction du risque spatialisé dans la recherche de zones à haut (ou faible) risque en tenant compte des covariables explicatives locales.

*Adresses :*

Émilie Gay
Unité de biométrie, INRA
Domaine St Paul, Site Agroparc
84 914 Avignon Cedex9 France
E-mail : gay@avignon.inra.fr

Rachid Senoussi
Unité de biométrie,INRA
Domaine St Paul, Site Agroparc
84 914 Avignon Cedex9 France
E-mail : senoussi@avignon.inra.fr

## Spatio temporal pattern of human Campylobacter gastroenteritis

par **Edith Gabriel**

Campylobacter gastroenteritis define a spatio-temporal point pattern displaying a highly aggregated incidence rate. To understand the transmission and the evolution of food-borne zoonotic pathogens, we formulate a spatio-temporal model for the dynamic of the disease (data-base supplied by the Preston Public Health Laboratory Service).

Spatio-temporal point patterns are usually modeled by inhomogeneous Poisson processes or Cox processes. Because these processes provide very similar point patterns, the following question arises in practice : Given a single realisation of a point process, how can we distinguish between an inhomogeneous Poisson process and a Cox process with a stochastic intensity ? On the other hand, the analysis of spatio-temporal point processes raises the question of testing the existence of space-time interaction.

To distinguish between an inhomogeneous Poisson process and a Cox process as well as testing the space-time interaction, we developed and applied a consistent non parametric estimate of the space-time inhomogeneous K-function under the assumption of second order intensity re-weighted stationarity.

*Adresse :*

Edith Gabriel
Medical Statistics Unit, Dpt Math. Stat
Fylde College, Lancaster University
Lancaster LA1 4YF, United Kingdom
E-mail : e.gabriel@lancaster.ac.uk

## Modélisation de populations spatialement structurées et inférence à partir de génotypes multi-locus : variations autour de l'équilibre de Hardy-Weinberg

par **Gilles Guillot**

We propose a spatial statistical model aimed at describing sub-populations structure in term of variations of allele frequencies across space. The overall population is modelled as a spatial mixture of sub-populations, the spatial organisation of sub-populations being modelled with a colored Voronoi tesselation. Correlations of allele frequencies accross sub-populations are taken into account through the introduction of an ancestral population from which all observed sub-populations are assumed to derive with specific drifts. Then, Hardy-Weinberg equilibrium and linkage equilibrium between loci is assumed within each sub-population. We also take into account uncertainty on the location of individuals. Inference is carried out by investigating the posterior distribution with MCMC simulations. Results from simulated data sets suggest that our method improve two other non spatial competing methods in term of rate of misclassification. We illustrate how our method can be used to delinate borders between sub-populations (which can be interpreted as barriers to gene flow), and to identify migrants.

In further recent developments, we also model and discuss two kinds of departure from Hardy-Weinberg equilibrium.

*Adresse :*

Gilles Guillot
UMR Math. Inf. Appli. INRA-ENGREF-INAPG , Paris
16, rue Claude Bernard
75231 Paris cedex 5, France
E-mail : gilles.guillot@inapg.inra.fr

## Modélisation de données océanographiques par ARMA spatiaux

par **Aude Illig**

Nous montrons que les modèles autorégressifs spatiaux peuvent constituer des modèles adéquats à l'étude de données spatio-temporelles de températures de surface dans les océans (données sur l'Océan Pacifique dans le cas présent). Dans le cadre d'un modèle avec des innovations de type différences de martingale spatiale, nous identifierons les paramètres à l'aide des équations de Yule-Walker. Nous utiliserons ensuite les propriétés asymptotiques des estimateurs pour proposer un test statistique pour la détermination de l'ordre de l'auroregressif. Nous montrons ainsi que de tels modèles permettent de construire des prédicteurs simples et faciles à calculer pour certaines variables environnementales en océanographie.

*Adresse :*

Aude Illig
CEREMADE, Université Paris Dauphine
Place du Maréchal de Lattre de Tassigny
75775 Paris Cedex 16, France
E-mail : illig@ceremade.dauphine.fr

## Prise en compte de l'hétéroscédasticité et de la dépendance dans l'analyse de données spatiales agrégées

par **Olivier Perrin**

En statistique spatiale, nous disposons souvent de données agrégées par domaine. Ces données peuvent être transformées en taux pour lesquels l'analyse statistique doit tenir compte de l'hétéroscédasticité inhérente à de telles variables. De plus, lorsque nous souhaitons régresser ces taux par d'autres variables explicatives à l'aide d'un modèle linéaire, il semble a priori nécessaire de tenir compte de la dépendance spatiale, même si celle-ci peut provenir d'une mauvaise spécification du modèle. En utilisant un jeu de données, où les taux sont des montants de prescription par consultation des cantons de Midi-Pyrénées, nous montrons que la prise en compte de la dépendance spatiale associée à une pondération ad hoc des variables (permettant de se ramener à de l'homoscédasticité conditionnelle) améliore le modèle en terme de prévision. Le choix final des variables explicatives introduites dans le modèle de régression provient d'une analyse exploratoire approfondie de cet ensemble de données.

*Adresse :*

Olivier Perrin
GREMAQ
Université des Sciences Sociales
21 Allée de Brienne
31000 Toulouse, France
E-mail : `operrin@univ-tlse1.fr`