10th French-Danish Workshop on Spatial Statistics and Image Analysis in Biology

Abstracts

List of presentations

A new estimation approach for combining epidemiological data from multiple sources

Yongtao Guan

We propose a novel two-step procedure to combine epidemiological data obtained from diverse sources with the aim to quantify risk factors affecting the probability that an individual develops certain disease such as cancer. In the first step we derive all possible unbiased estimating functions based on a group of cases and a group of controls each time. In the second step, we combine these estimating functions efficiently in order to make full use of the information contained in data. Our approach is computationally simple and flexible. We illustrate its efficacy through simulation and apply it to investigate pancreatic cancer risks based on data obtained from the Connecticut Tumor Registry, a population-based case-control study, and the Behavioral Risk Factor Surveillance System which is a state-based system of health surveys.

On mean and Variance for an estimator of the spatio-temporal product density function

Francisco J. Rodríguez-Cortés[a], Mohammad Ghorbani[b], Jorge Mateu[a], and Dietrich Stoyan[c]
a: Department of Mathematics, Universitat Jaume I, Castellon, Spain
b: Department of Mathematics Sciences, Aalborg University, Aalborg, Denmark
c: Institut für Stochastik, TU Bergakademie Freiberg, Freiberg, Germany

Second-order characteristics are used to analyse the spatio-temporal structure of point processes. A particular rule plays the spatio-temporal product density function, for which we develop a non-parametric edge-corrected kernel estimator under the second-order intensity-reweighted stationary hypothesis. We drive mean and variance of the estimator and give closed form approximation in the Poisson case. We present a detailed simulation study to compare our approximation for the variance with estimated values. Finally, we apply the resulting estimator to data on the spatio-temporal distribution of the invasive meningococcal disease in Germany.

Keywords: Spatio-temporal separability, Second-order product density, Second-order intensity-reweighted stationarity, Variance.

Multivariate Cox processes

Rasmus Waagepetersen
Department of Mathematical Sciences
Aalborg University

Modeling of multivariate point patterns is challening and in particular when moving beyond the bivariate case. In this talk we will discuss two types of models based on multivariate Cox point processes and we will consider applications of these in tropical rain forest ecology. The first type of models are the well-known multivariate log Gaussian Cox processes while the second is new and based on certain product-shot-noise fields. We will address the different biological interpretations of the models, how they can be fitted and how one may extract biologically relevant information from the models.

Spatial statistics for the analysis of polar ice cores

Claudia Redenbach

University of Kaiserslautern

We discuss methods for the analysis of anisotropic marked point processes which are motivated by the investigation of the pore system in ice cores. These cores are drilled from the Antarctic ice shield and consist of compacted snow. During the compression process, air pores are isolated within the ice. The system of these pores can be analysed using tomographic images. In earlier work, directional versions of the K-function and the nearest neighbour distance distribution function were used to analyse the anisotropy of the ice. It was shown that the spatial arrangement of the pores yields information on the compaction and the flow of the ice shield.

However, directional analysis is hampered by the fact that some extra pores form due to relaxation when the ice core is taken out of the drilling hole. These do not carry any information on the movement. Hence, they should be detected and removed prior to the directional analysis. For that purpose, we assume that the pattern of true pore centers is a realization of a regular Strauss point process, while the centers of noise pores can be modelled by a Poisson process. The aim is to decide which of the two processes each pore belongs to. To this end, we construct an MCMC algorithm which estimates the parameters of the mixture model and obtains posterior probabilities for each point of being a Strauss point. Based on these, the points can be classified.

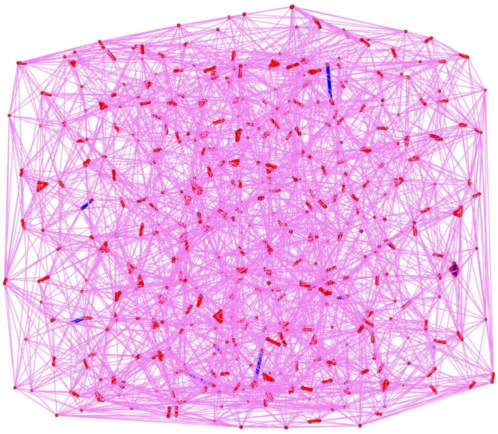
PLAYING WITH EBSPATCGAL A R PACKAGE DEDICATED TO SIMULATION AND ESTIMATION IN THE FRAMEWORK OF GIBBS NEAREST-NEIGHBOUR POINT PROCESSES

Rémy Drouilhet

LJK, 1251 avenue centrale, BP 47, 38040 Grenoble Cedex 9.

The `R` package `EBSpatCGAL` is a complete code rewriting of the `R` package `EBSpat` offering on the one hand simulation tools for the nearest-neighbour point processes ([1,4]) and, on the other hand, estimation tools based on the pseudo-likelihood [2] and Takacs-Fiksel methods [3].

The main difference with `EBSpat` is that, instead of relying on the `C API` of `R`, `EBSpatCGAL` relies on the package `Rcpp` which provides excellent tools to interface `C++` objects inside the `R` system. This allows us to easily interface the `C++` objects of the Computational Geometry Algorithms Library (`http://www.cgal.org`). As a direct consequence, it was possible to envisage the coding in 2D and 3D simultaneously. The following figure represents a realization of a Delaunay Gibbs point process in 3D:



```
> del3 <- Delaunay(3)
> # init gibbs
gd3 <- SimGibbs(
  del3 ~ 5 +
        Del2(
         th[1]*(l<=20)
        + th[2]*(20<l & l<=80),
         th=c(-2,10)
        ),
  domain=c(-350,-350,-350,350,350,350)
)
> run(gd3)
```

The next step will be the creation of a new `R` package, probably named `PoLiTe`, which will result from the merging of this package with the `LiTe` package developed by Kiên Kiêu.

[1] Bertin, E., Billiot, J.M. et Drouilhet, R. (1999) *Existence of "Nearest-Neighbour" Gibbs Point Models*, Adv. Appl. Prob., 31, 895–909.

[2] Billiot, J.-M., Coeurjolly, J.-F and Drouilhet, R. (2008) *Maximum pseudolikelihood estimator for exponential family models of marked Gibbs point processes*, Electronic Journal of Statistics, 2 234-264.

[3] Coeurjolly J.-F., Dereudre, D., Drouilhet, R. et Lavancier, F. (2010) *Takacs Fiksel method for stationary marked Gibbs point processes*, HAL, numéro hal-00502004 (To appear in Scandinavian Journal of Statistics).

[4] Dereudre, D., Drouilhet, R. et Georgii, H.-0. (2010) *Existence of Gibbsian point processes with geometry-dependent interactions*, Proba. Theory Related Fields 150 (2011).

Modelling spatio-temporal multi-structured patterns of forest fires

Florent Bonneu & Edith Gabriel

Avignon University, LMA EA 2151, 84000 Avignon, France

Wildfires have been studied in many ways and point processes have found interest in modelling their occurrence. Space-time point process models are particularly useful to predict forest fire occurrence in order to establish both enhanced fire prevention and fire fighting plans. The distribution of ignition points can be modelled as a function of time, location, and background factors: topographic variables as slope and altitude; proximity to roads and urban areas; land use; . . . Most of works perform spatial and temporal analyses separately and there are few spatio-temporal models (Serra *et al.*, 2013 ; Møller and Diaz-Avaloz, 2010).

We consider forest fires in south of France from 1993 to 2009. The probability of occurrence at given location and time must be based on the past and we particularly show that it is weaker within recent burned areas. This remark has methodological consequences because we have to model both spatio-temporal clustering partially related to the covariates and spatio-temporal inhibition due to the past. Hence, to model the spatio-temporal distribution of ignitions points, we take into account the size of fires (burned areas). Such model is realistic, predictive and explains the observed spatio-temporal interaction at fine scale.

Møller and Diaz-Avaloz (2010) Structured spatio-temporal shot-noise Cox point process models, with a view to modelling forest fires. *Scandinavian Journal of Statistics*, 37(1), 2–25.

Serra *e*t al. (2013) A spatio-temporal Poisson hurdle point process to model wildfires. Stochastic Environmental Research and Risk Assessment, 1–14.

MCEM algorithm for the log-Gaussian Cox process

Céline Delmas, Nathalie Peyrard, Régis Sabbadin

UR 875 - Unité de Mathématiques et Informatique Appliquées, Toulouse
INRA, 24 Chemin de Borde Rouge - Auzeville - CS 52627
31326 Castanet Tolosan cedex

Abstract Log-Gaussian Cox processes are an important class of models for aggregated point patterns. They have been largely used in spatial epidemiology (Diggle et al., 2005), in agronomy (Bourgeois et al., 2012), in forestry (Moller et al.), in ecology (sightings of wild animals) or in environmental sciences (radioactivity counts).

A log-Gaussian Cox process is a Poisson process with a stochastic intensity depending on a Gaussian random field. We consider the case where this Gaussian random field is centered with exponential covariance function. Moreover we assume that we observe count variables in disjoint quadrats assumed to be the same up to a translation. In that special case, we consider the problem of the estimation of the parameters of the log-Gaussian Cox process. We propose a MCEM algorithm and compare this algorithm to a moments method. Our MCEM algorithm considers the non observed count variables and the Gaussian random field as hidden variables. Simulations show that, in most cases, this MCEM algorithm improves the precision of the estimations in comparison to the moments method.

As a future direction of this work, we will discuss the problem of the optimal sampling for estimation and prediction. We will also discuss the problem of hypothesis testing.

Bibliography

BOURGEOIS, A., GABA, S., MUNIER-JOLAIN, N., BORGY, B., MONESTIEZ, P., SOUBEYRAND, S. (2012) Inferring weed spatial distribution from multi-type data. *Ecological Modelling*, 226, 92-98.

DIGGLE, P.J., ROWLINGSON, B., SU, T.L. (2005) Point process methodology for on-line spatio-temporal disease surveillance *Environmetrics*, 16, 423-434.

EMERY, X., HERNANDEZ, J. (2010) A computer package for modeling and simulating regionalized count variables. *Computers & Geosciences*, 36, 24-33.

MOLLER, J., SYVERSVEEN, A.R., WAAGEPETERSEN, R.P. Log Gaussian Cox Processes: A Statistical Model for Analyzing stand structural heterogeneity in Forestry.

VBEM algorithm for the Log Gaussian Cox Process

Julia Radoszycki, Nathalie Peyrard, Régis Sabbadin

UR 875 - Unité de Mathématiques et Informatique Appliquées, Toulouse
INRA, 24 Chemin de Borde Rouge - Auzeville - CS 52627
31326 Castanet Tolosan cedex

Abstract. The log Gaussian Cox process [4] is a classical model to represent spatial interactions in count-based maps. For example, it has been used to model weeds counts in crop fields divided into quadrats [2,3]. A MCMC algorithm has been proposed by [2] to estimate jointly the a posteriori mode of the parameters vector and of the hidden Gaussian field which captures the spatial structure of weed repartition.

The drawback of MCMC algorithms is the required computational time. Hence, we propose here a faster algorithm based on a variational principle. The generic VBEM (*Variational Bayesian Expectation Maximization*, [1]) algorithm has been proposed recently. In practice, it is necessary to specify the E and the M steps of the VBEM algorithm for the log gaussian Cox process. We propose a specification based on a mean field hypothesis and on Monte-Carlo simulations in the case of an exponential covariance function. Experiments on simulated data show that the proposed VBEM algorithm is as efficient (except for the estimation of the covariance parameter) and much faster than the MCMC algorithm presented in [2].

Weeds counts are usually only available for a sample of quadrats, since observations are costly to acquire. A future direction of this work will be to consider estimation of the log Gaussian Cox process' parameters from a limited size sample.

Bibliography.

[1] Beal, M. J. (2003), Variational algorithms for approximate bayesian inference, PhD thesis, Gatsby Computational Neuroscience Unit, University College London.
[2] Bourgeois, A., Gaba, S., Munier-Jolain, N., Borgy, B., Monestiez, P. et Soubeyrand, S. (2012), Inferring weed spatial distribution from multi-type data, *Ecological Modelling*, 226, 92–98.
[3] Brix, A. et Møller, J. (2001), Space-time multi type log gaussian Cox processes with a view to modeling weeds, *Scandinavian Journal of Statistics*, 28, 471–488.
[4] Møller, J., Syversveen, A. R. et Waagepetersen, R. P. (1998), Log gaussian Cox processes, *Scandinavian Journal of Statistics*, 25, 451–482.

Inhomogeneous space-time shot-noise Cox point processes – asymptotics for estimators based on projection processes

Jiří Dvořák (Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University in Prague; dvorak@karlin.mff.cuni.cz)

Shot-noise Cox point processes constitute an important class of models for spatial or space-time cluster processes. For space-time point processes parameter estimation methodology has not been fully developed yet. In general, for Cox point processes finding maximum likelihood estimates is typically computationally too demanding and inhomogeneous space-time shot-noise Cox processes are no exception. Thus, development of faster, simulation-free methods is desirable.

In this contribution we will discuss the possibility of using projections of the space-time process into the spatial and temporal domain for parameter estimation, assuming appropriate non-trivial model structure. Showing consistency and asymptotic normality of the estimators is a challenging task in this setting due to the different normalization needed in different estimation steps. We will discuss the main obstacles and formulate the conditions under which these asymptotic properties hold.

Stein estimation of the intensity of a stationary spatial Poisson point process

Jean-François Coeurjolly
Univ. Grenoble Alpes, LJK, F-38000, Grenoble, France
`Jean-Francois.Coeurjolly@upmf-grenoble.fr`

We revisit the problem of estimating the intensity parameter of a homogeneous Poisson point process observed in a bounded window of $\mathbb{R}^d$ making use of a (now) old idea of James and Stein. For this, we prove an integration by parts formula for functionals on the Poisson space. This formula, extends the one obtained by Privault and Réveillac (Statistical inference for Stochastic Processes, 2009) in the one-dimensional case. As in Privault-Réveillac, this formula is adapted to a notion of gradient of a Poisson functional satisfying the chain rule which is the key ingredient to propose new estimators able to outperform the maximum likelihood estimator (MLE) in terms of mean squared error (MSE). The new estimator can be viewed as biased version of the MLE with a well-constructed bias reducing the variance. We study a large class of examples and show that with a controlled probability the Stein estimator outperforms the MLE. We will illustrate in a simulation study that for very reasonable practical cases (like an intensity of 10 or 20 of a Poisson point process observed in the euclidean ball of dimension between 1 and 5) we can obtain a gain (in terms of MSE with respect to the MLE) larger than 30%.

This is a joint work with Marianne Clausel and Jérôme Lelong (Grenoble Alpes University).

Defining a variographic approach from the characteristics of a point process to estimate and predict the local intensity of partially observed data

Edith Gabriel* & Joël Chadœuf**

* *Avignon University, LMA EA 2151, 84000 Avignon, France*
** *Statistics, GAFL UR 1052, INRA, 84000 Avignon, France*

Biodiversity is at the heart of many debates and a common stake is to know its changes. Spatial dependence in biology is complex as it can be expressed at several scales of concern and should be taken into account in the estimation of local intensity in ecology (when looking at a species distribution for example) or environment. Usually, ecological data are clustered and clusters correspond to small geographical areas (quadrats) on which they are summarized (abundance, richness, . . .). Such data are characterized by excess zeros due to an heterogeneous spatial distribution of species. Moreover, a trade-off between the area to be surveyed and the observation time may lead to incompleteness of observations within regions of interest. Estimating and predicting the local intensity is thus a well-known challenging issue.

Here we consider a spatial point pattern observed within a large window W. We assume that the process is stationary, isotropic and ergodic, and that it is obtained by a weak dependent process with a parameter driven by a stationary random field at a larger scale. In order to estimate the local intensity, we propose to define the first- and second-order characteristics of the random field (i.e mean and variogram) from the first- and second-characteristics of the point process and to interpolate the local intensity by using ordinary kriging. Our aim is to answer the following issues:

1. Observing the full pattern within W which method, amongst (a) kernel smoothing, (b) counting and kriging and (c) our approach, is the most efficient? the faster? And is there an optimal grid size for the interpolation?
2. Having a partially observed pattern within W, how to estimate the local intensity?

Spatial normalisation of replicated point patterns for the estimation of average intensity map

David Legland[1,2], Marie-Françoise Devaux[3], Fabienne Guillon[3]

1: UMR0782 GMPA Génie et Microbiologie des Procédés Alimentaires, INRA, F-78850 Thiverval-Grignon.
2: UMR1318 IJPB Institut Jean-Pierre Bourgin, INRA-AgroParisTech, F-78026 Versailles.
3: UR1268 Biopolymères Interactions Assemblages, INRA, F-44300 Nantes

The cellular structure of plant tissues is a key parameter for determining their properties, and few studies have tried to quantify the spatial distribution of different types of tissues within an organ. As plants have various shapes and sizes, the integration of several individuals for statistical analysis of tissues distribution is not an obvious task. The aim of this study is to propose a generic method that quantifies the average intensity distribution of vascular bundles within plant stems, by integrating information obtained from replicated images.

Internodes from two maize genotypes were studied. In order to compare observations, a spatial normalisation strategy was applied. Outlines of stem cross-sections were automatically digitised and registered. Statistical analysis on outlines produced an average model of stem that was used as reference contour, as well as fitted contour for each section. The cellular structure was observed using macroscopy imaging on 150-thick stem cross sections. The position of vascular bundles was automatically extracted using image analysis, and projected into the coordinate system of the reference contour. Bundles intensity was estimated on each normalised slice using kernel density estimators. As the same sampling grid was used for all slices it was possible to superimpose intensity maps, and to compute an average intensity map. This average intensity map showed heterogeneity in the repartition of vascular bundles within the reference stem. By computing average intensity maps by genotype, it was also possible to explore differences between genotypes.

Future work will focus on the analysis of the spatial interaction of vascular bundles, using for example pair correlation function. Tissue distribution modelling through point pattern models is also envisioned.

U-statistics in stochastic geometry and their characteristics

Viktor Beneš and Markéta Zikmundová

U-statistics of finite spatial point processes given by a density with respect to the Poisson process are investigated. General formulas for moments are derived using the expansion of functionals of a Poisson process. The role of conditional intensities of arbitrary order is emphasized. In the logarithmic form functionals are connected to Gibbs models.

More explicit results are obtained for some parametric models in stochastic geometry suitable for applications, namely interacting segments in the plane or circular plates in the three-dimensional space. Limitations on the parameter space are indicated which allow for repulsive interactions. Natural systems of U-statistics are studied which correspond e.g. for plates to the total surface area, total length of intersection lines and total number of intersections points of triplets.

Global envelope tests, part I: test principle

Tomáš Mrkvička, Department of Applied Mathematics and Informatics, Faculty of Economics, University of South Bohemia, České Budějovice, Czech Republic,

Ute Hahn, Centre for Stochastic Geometry and Advanced Bioimaging, Department of Mathematics, University of Aarhus, Aarhus, Denmark.

There are two options for simulation based tests of hypotheses for spatial point processes: the so-called deviation test or the popular, but informal, envelope test. Both tests are based on a spatial summary function $S(r)$, like the K-, L- or J-function, and compare the observed estimate of $S(r)$ with simulated counterparts on a range of distances r simultaneously. The deviation test is a proper test with a prescribed significance level, but it does not have a direct visual interpretation that is often desired by researchers and it says only little about the reason of rejection of the null hypothesis. These are the reasons why the informal envelope test is so popular. Grabarnik et al. (2011) showed how the envelope can be refined in order to correct for simultaneous inference, but the proposed procedure requires adjusting the number of simulations from the null hypothesis together with evaluation of the type I error of the test.

In this talk, we present an improvement to the envelope test that allows to fix the number s of simulations in advance and to estimate the p-value of the test (Myllymäki et al., 2013). This so-called global envelope test has a direct graphical interpretation: if the data function is outside the constructed envelopes, the null hypothesis can be rejected. The method is based on ranking of the observed summary function among the summary functions of simulated patterns from the null model. Since the proposed ranking is not unique, there exists ties, thus the p-interval is given instead of a single p-value. Simulation results suggest that the width of this p-interval is of order s^{-1}. We then introduce a new idea how to break these ties, which we call "gold first ordering" in analogy to ranking olympic achievements. In combination with gold first ordering, the probability for ties is minimized, and the envelope test virtually always yields a unique p-value.

References

P. Grabarnik, M. Myllymäki, and D. Stoyan (2011) Correct testing of mark independence for marked point patterns. *Ecological Modelling*, **222** , 3888–3894.

M. Myllymäki, T. Mrkvička, H. Seijo and P. Grabarnik (2013) Global envelope tests for spatial point patterns. `arXiv:1307.0239`

Coauthors

This talk is based on previous and ongoing joint work with **Mari Myllymäki**, Department of Biomedical Engineering and Computational Science, Aalto University School of Science, Espoo, Finland, and **Pavel Grabarnik**, Institute of Physico-Chemical and Biological Problems in Soil Science, the Russian Academy of Sciences, Pushchino, Russia.

Global envelope tests, part II: extensions

Ute Hahn, Centre for Stochastic Geometry and Advanced Bioimaging, Department of Mathematics, University of Aarhus, Aarhus, Denmark,

Tomáš Mrkvička, Department of Applied Mathematics and Informatics, Faculty of Economics, University of South Bohemia, České Budějovice, Czech Republic.

In the second part of our joint talk on global envelope tests, we propose generalisations of the test procedure and extend the range of the test.
Applications of the rank envelope test are not limited to summary functions of point processes, but also include more general functional or multivariate statistics. Examples discussed in this talk cover combination of multiple tests using several summary statistics of point processes, and simultaneous tests using multiple statistics for more general random closed sets.
Moreover, the scope of the test can be broadened by using a bootstrap or permutation approach instead of simulations. We will show how to use the idea of the envelope test for the comparison of groups of point patterns and contrast it to permutation tests that are based on integral deviation measures (Hahn, 2012).

References

U. Hahn (2012) Studentized Permutation Test for the Comparison of Spatial Point Patterns. *Journal of the American Statistical Association*, **107**, 754–764.

Coauthors

This part of the talk represents ongoing joint work with **Mari Myllymäki**, Department of Biomedical Engineering and Computational Science, Aalto University School of Science, Espoo, Finland.

Exposing the spatio-temporal nature of eye movement data

Anna-Kaisa Ylitalo

University of Jyväskylä, Finland

Co-author: Antti Penttinen, University of Jyväskylä

Electronic eye trackers record spatial locations of the gaze, as well as the corresponding time instances, when the gaze moves while person is looking at a target. Ideally, such data can be expressed as a sequence of jumps in two-dimensional space with associated times of occurrence. The gaze locations are called fixations and the jumps between them saccades. The objectives of the study are to develop statistical analysis methods and to model the process evolution in order to deal with changes in the eye movement process during the experiment.

In this work summary statistics from stochastic geometry for the eye movement process are studied. The focus is on coverage statistics and their properties, e. g. whether they can distinguish the global and local viewing behaviors. Those summary statistics are functions of time and hence they capture information about the sequential nature of the process. An open question is, can we detect extreme behavior of the eye movement process using these summary statistics and how this should be done. A data set belonging to the field of art study is used as an empirical example.

Stochastic modeling of eye movements

Aila Särkkä

Mathematical Sciences, Chalmers and University of Gothenburg, Sweden

Abstract: Eye movements are outcomes of cognitive processes in the human brain, and can be recorded with high spatial and temporal resolution by computerized eye trackers. Eye movements give valuable information on cognitive processes and are therefore, used in many areas. Here, the question of interest is how people look at art. The data come from a cognitive art research experiment, where the eye movements of twenty test subjects were recorded while they were looking at six paintings, each painting during three minutes. We will concentrate on studying the eye movements on one of the six paintings, namely Koli landscape by Eero Järnefelt.

Eye movements can be represented as an alternating sequence of fixations (periods in which the gaze is staying relatively still around a location of the target space) and saccades (rapid movements between the fixations). We view the process of fixations as a spatio-temporal point process and introduce models for the spatio-temporal eye movement process including fixation locations, fixation durations, and saccade durations and lengths.

The work has been done jointly with Anna-Kaisa Ylitalo (University of Jyväskylä) and Peter Guttorp (University of Washington).

ON THE USE OF PERCENT MODEL AFFINITY INDEX IN BIOMONITORING

Johanna Ärje[1], Kwok-Pui Choi[2], Fabio Divino[3], Salme Kärkkäinen[1] &
Kristian Meissner[4]
[1] University of Jyväskylä, Finland
[2] National University of Singapore, Republic of Singapore
[3] University of Molise, Italy
[4] Finnish Environment Institute, Finland

Biomonitoring of benthic macroinvertebrates is an important part of the ecological status assessment of rivers and lakes. It is known that changes in the relative abundances of different species compared to a reference population may indicate changes in the environment. One index for measuring the changes is the percent model affinity (PMA) index defined as follows

$$PMA = 1 - 0.5 \sum_{h=1}^{C} |p_h - q_h|, \tag{1}$$

where q_h is the unknown proportion of taxonomical group h, $h = 1, \ldots, C$, in the reference population and p_h the unknown proportion of taxonomical group h in the population of interest. In practise, both parameter sets are usually estimated: the first one from the samples from a reference population and the latter one from the sample(s) in question. There exist some studies on the properties of the estimator of PMA index in several set-ups. Some results are shown.

Macroinvertebrate samples are traditionally identified by taxonomic experts but there are recent studies on the classification of macroinvertebrates based on images. Whether the samples are classified automatically or manually, classification errors affect PMA index calculated from the samples. An advantage of automated classification is that the information on classification errors is stored in a confusion matrix. Using simulation experiments, we study the effect of classification errors of automatic classification and the effect of the existing confusion matrix correction on the values of PMA index.

Quermass-interaction process with non-circular grains

Kateřina Helisová

Consider a random set observed in a bounded window $S \subset \mathbf{R}^2$. The set is given by a union of interacting objects. Recently, there was studied a random-disc Quermass-interaction process (see [1], [2] or [3]), i.e. the process with probability density of any finite configuration $\mathbf{x} = (x_1, ..., x_n)$ of the discs $x_1, ..., x_n$ with respect to a stationary random-disc Boolean model given by

$$f_\theta(\mathbf{x}) = \frac{\exp\{\theta_1 A(U_\mathbf{x}) + \theta_2 L(U_\mathbf{x}) + \theta_3 \chi(U_\mathbf{x})\}}{c_\theta},$$

where $A(U_\mathbf{x})$ denotes the area, $L(U_\mathbf{x})$ the perimeter and $\chi(U_\mathbf{x})$ the Euler-Poincare characteristic of the union $U_\mathbf{x}$ of the configuration $\mathbf{x}$, $\theta = (\theta_1, \theta_2, \theta_3)$ is the vector of parameters and c_θ is the normalizing constant.

This contribution will present the generalization of this process to the process with non-circular grains. Especially, the simulation algorithm will be shown and compared to that one described in [2].

References

[1] Kendall W.S., van Lieshout M.N.M., Baddeley A.J. (1999): Quermass-interaction processes: conditions for stability. *Advances in Applied Probability* **31**, 315–342.

[2] Møller J., Helisová K. (2008): Power diagrams and interaction processes for unions of discs. *Advances in Applied Probability* **40**, 321–347.

[3] Møller J., Helisová K. (2010): Likelihood inference for unions of interacting discs. *Scandinavian Journal of Statistics* **37**, 365–381.

A Spectral Mean for Point Sampled Closed Curves

Marie-Colette van Lieshout
CWI Amsterdam

We propose a spectral mean for closed curves described by sample points on its boundary subject to mis-alignment and noise. First, we ignore mis-alignment and derive maximum likelihood estimators of the model and noise parameters in the Fourier domain. We estimate the unknown curve by back-transformation and derive the distribution of the integrated squared error. Then, we model mis-alignment by means of a shifted parametric diffeomorphism and minimise a suitable objective function simultaneously over the unknown curve and the mis-alignment parameters. Finally, the method is illustrated on simulated data as well as on photographs of Lake Tana taken by astronauts during a Shuttle mission.

Towards pseudo-likelihood inference for Gibbsian T-tessellations

Kiên Kiêu
UR341 Mathématiques et Informatique Appliquées
INRA Jouy-en-Josas

Recently a new class of Gibbs models for planar T-tessellations of bounded planar domains has been proposed [3]. The so-called *completely random T-tessellation model* introduced in that paper is closely related to former work by Arak, Clifford and Surgailis [1]. These models are characterized by Campbell measures and Papangelou intensities that can be computed explicitly. An approach for deriving a pseudo-likelihood will be presented. It is based on Campbell measures and Kullback-Leibler divergence and leads to a contrast function. The proposed pseudo-likelihood is an empirical estimator of the constrast function. It can be seen as an adaptation of the now classical pseudo-likelihood for point processes [2, 4] to random T-tessellations. Preliminary numerical experiments will be presented.

References

[1] Arak, T., Clifford, P. and Surgailis, D. Point-based polygonal models for random graphs. *Adv. Appl. Probab.*, **25**, 348–372, 1993.

[2] Besag, J. Some methods of statistical analysis for spatial data. *Bull. Internat. Statist. Inst.*, **47**, 77–92, 1977.

[3] Kiêu, K., Adamczyk-Chauvat, K., Stoica, R. S. and Monod, H. A completely random T-tessellation model and Gibbsian extensions. *Spatial Statistics*, **6**, 118–138, 2013.

[4] Jensen, J. L. and Møller, J. Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab.*, **1**, 445–461.

Some specific problems of dynamics recovery in cell biology

A. Trubuil* (*INRA/MIA)*

alain.trubuil@jouy.inra.fr

*Domaine de Vilvert, INRA, Unité MIA, 78350 Jouy-en-Josas

In this talk, I will highlight several biological studies for which we are faced with dynamics recovery at the cell scale. I will start with early plant embryogenesis and explain how we can extract dynamical information despite the fact that no time lapse data is available but observations of a collection of embryos at different stages of development. In the second part of the talk, I will address a new study we are involved in and concerns the dynamics of swimmers bacteria inside biofilms. In this context two image channels could be combined in order to better understand the interaction between swimmers bacteria and biofilm.

A general procedure to combine estimators

Frederic Lavancier
University of Nantes

We propose a general method to combine several estimators of the same quantity in order to produce a better estimate. In the spirit of model and forecast averaging, the final estimator is computed as a weighted average of the initial ones, where the weights are constrained to sum to one. In this framework, the optimal weights, minimizing the quadratic loss, are entirely determined by the mean square error matrix of the vector of initial estimators. The averaging estimator is derived using an estimation of this matrix, which can be computed from the same dataset. We show that the solution satisfies a non-asymptotic oracle inequality and is asymptotically optimal, provided the mean square error matrix is suitably estimated. This method is illustrated on standard statistical problems (estimation of the position of a symmetric distribution, kernel density estimation) and on some spatial statistical parametric models, where our averaging estimator outperforms the initial estimators in most cases. This work has been carried out with Paul Rochet (University of Nantes) and is available on http://arxiv.org/abs/1401.6371.

Prediction of L*a*b* values for meat

Line Clemmensen

Color is an important quality parameter of food products such as meat. Objective and non-intrusive measurements of the color (often measured as L*a*b* values) are desirable, as they are repetitive and do not contaminate the food. We propose to predict L*a*b* values for a variety of meat products, both raw and processed, using multispectral imaging in the visual and NIR area. The imaging technique provides spatial information of the meat samples, information not present in traditionally used chroma meter techniques. We examine both lienar and non-linear models to make the link between the multispectral images and the reference L*a*b* values. As training data, we use chroma meter L*a*b* values of meat samples, as opposed to a color checker which traditionally is used. This accommodates for training data including the translucency and fiber structures of the meat samples. The predictions based on the multispectral images with models trained from meat samples provide a substantial improvement compared to those trained on color checkers.

Cross-Covariance Functions for Multivariate Geostatistics

Marc G. Genton

King Abdullah University of Science and Technology

Continuously indexed datasets with multiple variables have become ubiquitous in the geophysical, ecological, environmental and climate sciences, and pose substantial analysis challenges to scientists and statisticians. For many years, scientists developed models that aimed at capturing the spatial behavior for an individual process; only within the last few decades has it become commonplace to model multiple processes jointly. The key difficulty is in specifying the cross-covariance function, that is, the function responsible for the relationship between distinct variables. Indeed, these cross-covariance functions must be chosen to be consistent with marginal covariance functions in such a way that the second order structure always yields a nonnegative definite covariance matrix. We review the main approaches to building cross-covariance models, including the linear model of coregionalization, convolution methods, the multivariate Matérn, and nonstationary and space-time extensions of these among others. We additionally cover specialized constructions, including those designed for asymmetry, compact support and spherical domains, with a review of physics-constrained models. We illustrate select models on a bivariate regional climate model output example for temperature and pressure, along with a bivariate minimum and maximum temperature observational dataset; we compare models by likelihood value as well as via cross-validation co-kriging studies. We close with a discussion of unsolved problems.

The talk is based on joint work with William Kleiber.

Covariance models for the analysis of genetic variation

Gilles Guillot
Technical University of Denmark
Joint work with
Filippo Botta, Casper Eriksen, Michael Fontaine,
René Schilling, Emilio Porcu and Moreno Bevilacqua.

May 2014

We propose a model aimed at quantifying the relative effect of isolation by distance and isolation by environmental difference on spatial genetic variation.

If the data consist of counts $(z_i)_{i=1,\ldots,n}$ of a reference allele for n samples over the study domain, we assume that z_i is a binomial sample $B(n_i, f_i)$ where n_i is the sample size and f_i is an unknown local population allele frequency. We assume that allele frequencies at a site indexed by s and e relate to a Gaussian random field $y(s, e)$ via a logistic transform. Denoting by s a generic spatial coordinate and e an environmental variable, we assume that

$$Cor\left[y(s,e), y(s',e')\right] = \exp\left[-\left(\alpha_G|s-s'| + \alpha_E|e-e'|\right)\right] \tag{1}$$

We discuss issues related to positive-definiteness and give conditions under which this model is well-behaved mathematically. We propose an MCMC algorithm to perform inference and model selection. We also report results with this model for the analysis of various marine and terrestrial organisms.

References

- Guillot G., R. Schilling, E. Porcu, M. Bevilacqua. Validity of covariance models for the analysis of geographical variation. 5(4), 329–335, 2014.
- Botta F., C. Eriksen, M. Fontaine, G. Guillot. Bayesian computation to detect effect of ecological isolation on genetic differentiation, Research Report, Technical University of Denmark, May 2014.

Characterization and new models of anisotropy for spatial data

D. Allard[1], E. Porcu[2] and R. Senoussi[1]
[1]Biostatistique et Processus Spatiaux (BioSP), INRA, Site Agroparc,84914 Avignon, France.
[2]Departamento de Matematica, Universida Santa Maria, Valparaso, Chile.
{Denis.Allard,Rachid.Senoussi}@avignon.inra.fr
Emilio.Porcu@usm.cl

In spatial statistics the assumption of isotropy is very common, despite being very restrictive in describing the rich variety of interactions that can characterise spatial processes. Such an ubiquity is probably due to a combination of two main reasons: isotropic models are mathematically easier to build than anisotropic ones and their parameters, being less numerous, are easier to estimate – in particular when data are scarce. In the vast majority of cases, the modelling of anisotropy is restricted to the classical geometric and zonal anisotropies, which in essence amount to apply an isotropic variogram to the modulus of the transformation of the lag vector with the help of a transformation matrix. The literature on anisotropy models evading from these two classical models is very scarce, at the exception of the approaches based on componentwise anisotropy.

The present work was motivated by the need to answer a simple yet crucial question in spatial statistics. Suppose the empirical variograms computed in different directions show not only varying ranges and/or sills, but that they also show different behaviours at the origin, e.g. a close to quadratic behaviour in one direction and a close to linear behaviour in the perpendicular direction. Should the variation of the behaviour at the origin be modelled as a fixed effect (in the trend) or as a random effect through anisotropic regularity parameter? Is such a model admissible? More generally, can we find necessary and sufficient conditions for anisotropy models? Can we exhibit a full characterisation of admissible anisotropy models, of which zonal and geometrical models are special cases? How can we simulate from these models?

Covariance functions and variograms are characterised through Bochner and Schoenberg theorems as being the Fourier transforms of positive measures with specific integrability constraints. We make use here of these Fourier representations to revisit the main examples of anisotropic covariance/variogram structures known in literature: geometric and zonal anisotropy and separable models.

We first establish that power variograms with a power varying continuously with the direction must necessarily have a constant power. This result is then generalised to any variograms, using the concept of principal irregular term. These properties allow us to fully characterise anisotropic models through an integral representation that offers parametric as well as nonparametric anisotropic constructions. We finally illustrate the richness of the different behaviours of these models by some simulations of the associated Gaussian random fields.

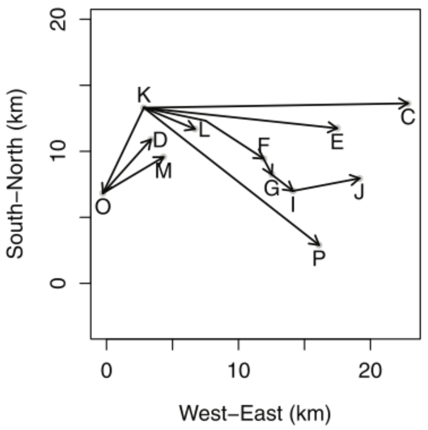
Handling dependencies in spatial, temporal and genetic variables to reconstruct routes of transmissions of pathogens

Samuel Soubeyrand – INRA PACA – Biostatistics and Spatial Processes

The accurate identification of the routes of transmissions taken by an infectious agent through a host population is critical to understanding its epidemiology and informing measures for its control. Infected hosts close together in their locations and timing are often thought to be linked, but timing and locations alone are usually consistent with many different scenarios of who-infected-whom. The genome of many pathogens evolves so quickly relative to the rate that they are transmitted, that we can identify which hosts contain pathogens that are most closely related to each other. Thus, genetic data combined with spatial and temporal data should help us infer more reliably who-transmitted-to-whom over the course of a disease outbreak. However, doing this so that these three different lines of evidence are appropriately weighted and interpreted remains a major statistical challenge.

In this context, there are two methodological issues related with statistics. 1. How to model the dependencies in the spatial, temporal and genetic variables that are observed? This was made with a state-space model including a mechanistic model for the epidemiological dynamics. 2. How to estimate model parameters and latent variables of interest given that the state-space model contains a huge amount of latent variables? This was achieved with likelihood approximation and MCMC.

We applied this approach to virus epidemics of foot-and-mouth disease in Great Britain and canine rabies in South Africa. In each case study, we had at our disposal times and locations at which infected cases were observed, and genetic consensus sequences of the virus sampled from the infected cases at the observation times.



Simple example of inferred routes of transmissions (Morelli et al., 2012). Host units are labelled with letters. In the transmission scenario represented here, the outbreak is initiated with the infection of host unit K.

Morelli M. J., Thébaud G., Chadoeuf J., King, D. P., Haydon D. T. and Soubeyrand S. (2012). A Bayesian inference framework to reconstruct transmission trees using epidemiological and genetic data. PLOS Computational Biology 8(11): e1002768.

Mollentze N., Nel L. H., Townsend S., le Roux K., Hampson K., Haydon D. T. and Soubeyrand S. (2014). A Bayesian approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data. Proceedings of the Royal Society B 281: 20133251.

Matérn thinned Cox point processes

Ina Trolle Andersen, Ute Hahn
Aarhus University, Department of Mathematics

This talk introduces a class of point process models that combines short range repulsion with medium range clustering.

The model is motivated by the pattern formed by the centres of particular cells (megakaryocytes) in sections of bone marrow. Certain diseases entail clustering of these cells. However, the cell centres can naturally not come closer to each other than the diameter of the cells. Such a combination of clustering and hard-core behaviour can be achieved by applying a dependent Matérn type II thinning to a Cox point process.

We derive an exact formula for the intensity of Matérn type II thinned shot noise Cox processes, using the reduced Palm distribution found in Møller (2003). The formula can be used by numerical integration. For the more general class of Matérn type II thinned Cox point processes, we derive formulae for the intensity and second-order characteristics using the conditional Poisson assumption. These formulae include more complicated integrals, which make them hardly usable for model fitting, and therefore we propose an approximation. The accuracy of this approximation and implications for model fitting are discussed.

Møller, J. (2003) Shot noise Cox processes. *Advances in Applied Probability 35*, 614-640.

Poisson Line Cluster Point Processes

Farzaneh Safavimanesh

Department of Mathematical Sciences, Aalborg University, Denmark;
e-mail: farzaneh@math.aau.dk

Abstract:
A new class of point processes with columnar structure is introduced. Poisson line cluster point process (PLCPP) models are useful for anisotropy analysis of data with columnar arrangement of points in $\mathcal{R}^d$. The intensity, rose of directions and moments for these models are given. Then the parameters of the model are estimated using a simulation based Bayesian approach. The application of these models both in two- and three- dimensional spaces will be illustrated using two examples from geography and neuroscience.

Spatial cluster point processes related to Poisson-Voronoi tessellations

Jesper Møller and Jakob G. Rasmussen
Department of Mathematical Sciences, Aalborg University

We discuss how to construct models for cluster point processes within territories modelled by d-dimensional Voronoi cells whose nuclei are generated by a latent Poisson process (where the planar case $d = 2$ is of our primary interest). Conditional on the territories/cells, the clusters are independent Poisson processes whose points may be aggregated around or away from the nuclei and along or away from the boundaries of the cells. Observing the superposition of clusters within a bounded region, we discuss how to account for edge effects. Bayesian inference for a particular flexible model is discussed in connection to a botanical example.

Inferring the spatial distribution of intracellular proteins: A Bayesian approach using Fluorescence Resonance Energy Transfer-data

Authors: J.O. Hooghoudt & R. Waagepetersen

The objective of this research is to get a quantitative understanding of the spatial distribution of proteins in living cells. For this purpose we conduct Bayesian inference on Fluorescence Resonance Energy Transfer (FRET) microscopic data.

FRET-microscopy is the preferred tool to obtain indirect information concerning the distribution of intracellular proteins. In a FRET experiment three emission channels are measured as pixel images where one channel is only active if proteins interact at a small scale. Although the interactions between proteins are at the molecular level (1-10 nm), the pixel resolution of FRET microscopy is typical 200x200 nm^2. Therefore no direct information is available concerning the spatial distribution of the proteins on an inner pixel level.

A statistical model for the channel intensities given a protein configuration has been developed. In this model incoming streams of photons are absorbed and redistributed among proteins according to probabilities depending on their spatial configurations. The measured intensities are due to de-excitation of the resulting excited proteins. The photon streams are modeled by Poisson processes. By using the labeling theorem for Poisson processes we are able to derive an explicit model for the intensities in the pixels of each channel image. We further specify a spatial point process model for the distribution of proteins. Combining this with the intensity-model a complete stochastic model for the generation of FRET pixel intensity data is obtained. This allows us, by Bayesian inference, to draw quantitative statements concerning the protein arrangement at an inner pixel level.

In the presentation we will discuss modeling of the channel data, choice of prior distributions and implementation of the Markov chain Monte Carlo algorithm for simulation of the posterior distribution. The performance of the proposed methodology is assessed on synthetic data generated by Poisson, clustered and repulsive point processes. Finally, we discuss results obtained for an empirical data set.