Avignon, November 7-9, 2018

Place: Campus Sainte Marthe, Université d'Avignon, 74 Rue Louis Pasteur (modern building, room 126 first floor)

WORKSHOP: THEORY AND PRACTICE OF SPDE MODELS AND INLA

WEDNESDAY

08h30-09h00 Welcome

09h00-09h15 Introduction

09h15-10h00 Janine Illian (Saint Andrews University)
Point processes – abstraction and practical relevance

10h00-11h00 Thomas Opitz (BioSP, INRA)
Tutorial: *INLA and R-INLA: a short introduction*

11h00-11h20 Coffee break

11h20-12h20 Short presentations

Facundo Muñoz (CIRAD)
Accounting for the spatio-temporal spread of a fungal invasion

Yuheng Ling (LISA, Corsica University)
Hedonic Housing Prices in Corsica: A spatiotemporal hierarchical approach

Florent Bonneu (UAPV)
Using INLA to model wildfire occurrences in Southern France

12h20-13h30 Lunch break

13h30-15h15 Haakon Bakka (KAUST)
Tutorial: *A practical introduction to the SPDE approach with INLA*

15h15-16h00 David Bolin (Chalmers University)
The rational SPDE approach for Gaussian random fields with general smoothness

16h00-16h20 Coffee break

16h20-18h00 All Together
Workshop: *Implementing INLA-SPDE models for ecological data*

THURSDAY

09h00-09h45 Finn Lindgren (Edinburgh University)
Recent, current, and future issues for large scale space-time and non-linear INLA

09h45-10h30 Nicolas Desassis (MinesParisTech)
Can the SPDE approach replace traditional geostatistics for industrial applications?

10h30-10h50 Coffee break

10h50-11h30 Short presentations

Ricardo Carrizo (MinesParisTech)
On simulations of SPDE-based stationary random fields

Mike Pereira (MinesParisTech, ESTIMAGES)
Finite element simulations of non-Markovian random fields on Riemannian manifolds

11h30-12h00 Thomas Opitz (BioSP, INRA)
Some topics in INLA practice

12h00-13h30 Lunch break

13h30-14h15 Haakon Bakka (KAUST)
Non-stationary Gaussian models with physical barriers

14h15-14h55 Short presentations

Anastasiia Gorshechnikova (Ca' Foscari, Venice)
Fast Approximation of Covariance Functions Using a Hierarchical Matrices Approach

Thomas Romary (MinesParisTech)
Second order finite elements for the SPDE approach

14h55-15h40 Maxime Beauchamp (PSL University; MinesParisTech)
Introduction of the air pollution dataset and elements of comparison between space-time estimation methods applied to air quality forecasting

15h40-16h00 Coffee break

16h00-18h00 All Together
Workshop: *Implementing INLA-SPDE models for predicting air pollution*

20h00 **Conference dinner at Grand Café Baretta**
(free for all participants registered to the conference)
Address: 14 Rue Prévôt, 84000 Avignon

FRIDAY

09h00-10h00	Debriefing the workshops
10h00-11h00	Joint discussion: • What can we currently do with R-INLA, what could be implemented, what cannot be done? • RINLA vs other Bayesian implementations (*BUGS, Stan, …)
11h00-11h20	Coffee break
11h20-12h30	Free program: revisiting concepts, codes, and data; participants' data
12h30	Lunch break
13h30 - …	Free program (continued): revisiting concepts, codes, and data; participants' data

ABSTRACTS OF TALKS

The following abstracts are in alphabetical order of the presenting speakers.

Haakon Bakka (KAUST)
Non-stationary Gaussian models with physical barriers

When modeling spatial data near the coast, we need to consider which assumptions to make on the Gaussian field with respect to the coastline, i.e. what kind of boundary effect to assume. One possibility is to have no boundary effect, modeling both water and land, but with observation and prediction locations only in water, leading to a model with a stationary correlation structure. However, a stationary field smooths over islands and peninsulas, inappropriately assuming that observations on two sides of land are highly correlated. Other approaches in the literature range from simple use of Dirichlet or Neumann boundary conditions, to being quite complex and costly. In this talk I showcase a new approach, the Barrier model, implemented in R-INLA, that is intuitive in the way correlation follows the coastline, and is as easy to set up and do inference with as a stationary field. I compare this model to two others, showing significant improvement at reconstructing a test function. A real data application shows that the Barrier model smooths around peninsulas, and that inference is numerically stable. I also detail how the stochastic partial differential equations (SPDE) approach was used to construct the Barrier model.

Maxime Beauchamp (PSL University; MinesParisTech)
Introduction of the air pollution dataset and elements of comparison between space-time estimation methods applied to air quality forecasting

The national PREV'AIR system (www2.prevair.org) delivers daily analyses and forecasts of different atmospheric pollutant concentrations over Europe and France. Forecast maps for the current and next two days (D+0, D+1, D+2) are computed by kriging statistical forecasts at the monitoring sites obtained by station specific multilinear regression models. Output data from the chemistry-transport model (CMT) CHIMERE are also used as an external drift in the kriging. In this study two kriging competitors are used: the usual space-time covariance-based kriging with external drift (KED) involving some appropriate neighbourhood to deal with reasonable CPU time and the SPDE-based kriging approach (SPDE). The performance is assessed using 2013 daily data and CMT simulations over France. It will be shown that both local fitting of the drift by (KED) and more global estimation made by (SPDE) can be a good alternative to the former (SA) framework.

David Bolin
The rational SPDE approach for Gaussian random fields with general smoothness

A popular approach for modeling and inference in spatial statistics is to represent Gaussian random fields as solutions to stochastic partial differential equations (SPDEs) of the form $L^{\beta} u = W$, where W is Gaussian white noise, L is a second-order differential operator, and $\beta>0$ is a parameter that determines the smoothness of u. However, this approach has been limited to the case $2\beta\in\mathbb{N}$, which excludes several important covariance models and makes it necessary to keep β fixed during inference.

We introduce a new method, the rational SPDE approach, which is applicable for any $\beta>0$ and therefore remedies the mentioned limitation. The presented scheme combines a finite element discretization in space with a rational approximation of the function $x^{-\beta}$ to approximate u. For the resulting approximation, an explicit rate of strong convergence to u is derived and we show that the method has the same computational benefits as in the restricted case $2\beta\in\mathbb{N}$ when used for statistical inference and prediction. Numerical experiments are performed to illustrate the accuracy of the method, and an application to climate reanalysis data is presented.

Florent Bonneu (UAPV)
Using INLA to model wildfire occurrences in Southern France

Wildfires can cause important economic and ecological disasters. Their prevention begins with understanding the stochastic mechanisms governing the intensity of occurrences and the severity of fires. We focus on wildfires in the Mediterranean region Bouches-du-Rhone (South of France) observed since 1981, with burnt area larger than one hectare. Occurrences depend on the presence and concomitance of several factors: climatic (temperature, humidity, wind speed), environmental (vegetation types, urbanization, road network) and human activity. Whilst human activity is the main direct cause of wildfires, climatic and environmental conditions are a prior condition to their outbreak and propagation. Therefore, the structure of relative risk of wildfires is highly complex and shows strong variation over space and time and is driven by numerous covariates. Statistical challenges arise from the multi-scale spatio-temporal structure of data defined over various supports like fine grids for land use, coarse grids for fire position leading to positional uncertainty, and meteorological series observed at irregularly spaced measurement sites. The spatial heterogeneity of wildfires depends on the spatial distribution of current land use like vegetation type, urban zones or wetlands. We also show that changes in vegetation due to past fires affect the probability of wildfire occurrence during a regeneration period. Log-Gaussian Cox processes, along with the INLA method for inference and prediction, are particularly useful to model clustered events. Here, we show that they can also deal with more complex structures, allowing us to include the temporal inhibition at small spatial scales and thus providing more accurate predictions.

Ricardo Carrizo (MinesParisTech)
On simulations of SPDE-based stationary random fields

We study a simulation method which is particularly adapted for the non-conditional simulation of stationary random fields related to a wide-class of linear SPDEs. This method, which is already existent in the literature, can be catalogued as a spectral method. It is based on an approximation of the Fourier transform of the field, which is an orthogonal random measure. The result is an approximation of the target stationary random field, whose convergence in suitable weak and strong senses when increasing the approximation order can be proven. Computations can be efficiently performed using the FFT algorithm. The class of SPDEs which can be treated with this method consists in linear equations defined through a continuous symbol function. Many of the most used stationary models can be expressed as solution for such kinds of equations. We are then able to simulate approximations for a wide variety of stationary models presenting interesting properties, such as separated arbitrary fractional regularities along the axes or with anisotropies defined through advections. We are also able to apply this method in a spatio-temporal framework in order to simulate stationary models with non-trivial properties such as separated regularity in space and time with non-symmetrical conditions. Finally, this method can be adapted for conveniently simulating solutions to physically driven equations. We present, for instance, simulations for solutions to first order evolution equations, and for waving models, which are stationary solutions to the homogeneous wave equation following any desired arbitrary spatial stationary covariance model.

Anastasia Gorshechnikova (Ca' Foscari, Venice, Italy)
Fast Approximation of Covariance Functions Using a Hierarchical Matrices Approach

Large datasets with n irregularly sited locations are difficult to handle for several applications of Gaussian random fields due to the computational complexity of order Opn3q. For example, a Gaussian field xpuq with the Matérn covariance, known as a solution to the linear fractional stochastic partial differential equations (SPDE), can be written through the kernel of the in- verse of the differential operator (Green function) which is usually dense. Under the regularity conditions, the hierarchical (H-) matrices approach affords a separable approximation of the kernel functions which significantly reduces the matrix operations cost. We focus on the techni- cal details of the H-matrices technique and describe the results obtained in the spatial settings. Approximation of the covariance matrix in the H-format allowed for the fast computation of matrix factorisations followed by the efficient maximum likelihood estimation and kriging pre- diction. As a part of the ongoing PhD thesis, we are extending the application of the H-matrix approach to the spatio-temporal context.

Janine Illian (University of St Andrews)
Point processes – abstraction and practical relevance

All statistical modelling of complex data structures involves an abstraction to the essential properties of interest into quantifiable units and associated random variables. In addition, it also often goes along with simplifying assumptions as part of the abstraction process, typically for practical reasons. As a result, methodology can tend to be far removed from reality and hence be of little practical relevance.
In the context of point process modelling, the usual abstraction reduces the available information to locations of points in space, whose spatial structure is analysed. Classical simplifications often concern assumptions of homogeneity, isotropy and known detection probabilities, often for computational reasons. Recent computational improvement however, allows us to relax some of these assumptions.
This talk discusses these classical assumptions along with associated abstractions may be relaxed, leading to increased practical relevance. In particular, I will discuss how this increased practical relevance has also caused an increasing demand for the development of new methodology that has previously played a rather minor role.

Finn Lindgren (Edinburgh University)
Recent, current, and future issues for large scale space-time and non-linear INLA

When working with real world problems, we don't have the luxury of restricting our models to nicely linear predictors and computationally convenient sizes. I will discuss some issues and partial solutions, arising from climate change estimation and ecological point process models with non-trivial observation structures. The main themes are large scale SPDE discretisations with hierarchical multi-timescale structure and multi-billion vector dimensions, generalising INLA to (slightly) non-linear predictors, and providing a more user-friendly interface.

Yuheng Ling (LISA, Corsica University)
Hedonic Housing Prices in Corsica: A spatiotemporal hierarchical approach

In the last decades, property prices analyses have captured the interest of both policy-makers and economists (OECD, 2017). In many countries, houses are the most valuable family's asset (Hui et al., 2009). Thus, the understanding of house price and market determinants is still up to date. With the development of spatial econometrics and with the emerging of large geo-referenced database, economists are now able to better gauge the impact of potentially spatial latent structures on home price markets. Based on hedonic price theory, several economists (Kolbe et al., 2012; Koschinsky et al., 2012) have confirmed the key role of housing location in determining house prices. However, merely considering the spatial latent structures is not enough. Other scholars (Iacoviello, 2002; Robstad, 2017) underline that temporal effects also play an important role in determining house prices. Shortly, first, house markets have cycles; second, it is the fact that a geo-referenced database may consist of spatial data at different time points. Lastly, it's hard to know whether houses are sold repeatedly. Therefore, neglecting the temporal effects on housing prices could result in biased coefficients. A number of researchers (Baltagi et al. , 2014, 2015; Dubé and Legros, 2013) have put their efforts to model both spatial effects and temporal effects simultaneously. However, they choose either nested error component models or spatial econometric approaches in which a spatial weighted matrix is trickily identified.
Our approach is quite different to the abovementioned studies; we propose a hierarchical spatio-temporal model referring mainly to the latent Gaussian models (LGM) framework.

$$\mathrm{y}(s_i,t) = \mathrm{z}(s_i,t)\beta + \xi(s_i,t) + \varepsilon(s_i,t)$$
$$\xi(s_i,t) = a\xi(s_i,t-1) + \omega(s_i,t)$$
$$\varepsilon(s_i,t) \sim N(0,\sigma_\varepsilon^2 I_d)$$
$$\omega(s_i,t) \sim N\left(0, \textstyle\sum = \sigma_\omega^2 \widetilde{\sum}\right)$$

where $\mathrm{y}(s_i,t)$ is an observed value of an independent variable at location s at time t (for instance, a house price observed at location s_i and at time t. $\mathrm{z}(s_i,t)\beta$ contains covariates information. $\varepsilon(s_i,t)$ is the error term. $\xi(s_i,t)$ is the so-called state process which is assumed to be an autoregressive process with a spatio-temporal Gaussian kernel. To estimate the model, an integrated nested Laplace approximation (INLA) approach is used. The combination of Laplace approximation and numerical integration allows INLA to have few good properties such as fast, accurate and efficient estimates. In addition, following Finn Lindgren's work (2011), a stochastic partial differential equation approach is used to model geo-referenced point data on a continuous plane and to estimate the impact of locations on prices.
Based on hedonic price theory, we perform an empirical study using the house price market in Corsica (France) as a case study. Using PERVAL database, we study 7634 apartment transactions between 2006 and 2017. To address the possible spatial and temporal biases, we apply the INLA-SPDE approach (Blangiardo and Cameletti, 2015). The empirical results demonstrated that the hedonic housing estimates and the corresponding marginal effects are affected by the spatial effects as well as the temporal effects. Ignoring the spatiotemporal effects could result in serious forecasting issues.

Facundo Muñoz (CIRAD)
Accounting for the spatio-temporal spread of a fungal invasion

Ash-dieback is an invasive fungal disease of ash trees characterised by leaf loss and crown dieback in infected trees. First detected in Poland in the early 1990s, it rapidly progressed throughout Europe causing large numbers of deaths and threatening today the survival of the species. Using annual observations from 2010 to 2014 within a field experiment located in northern France, I was asked to assess the genetic (co)variation associated with two different symptoms: Crown Dieback (CD) and Collar Lesion (CL). However, separating the progress of the disease from the genetic effects was of paramount importance for the estimation accuracy.
I modelled the regionalized relative exposure to pathogen agent using a latent spatio-temporal effect constructed as the tensor product of a Matérn spatial process and a exchangeable temporal structure using the SPDE approach for the spatial component and the `group` feature of INLA. This approach allowed to leverage all the available data to

identify families with higher resistance, and a association between resistance and early-flushing. Furthermore, the high estimated heritabilities imply that breeding is feasible and the disease is manageable.

Mike Pereira (MinesParisTech, ESTIMAGES)
Finite element simulations of non-Markovian random fields on Riemannian manifolds

The “SPDE approach” consists in characterizing continuous Markov random fields as solutions of stochastic partial differential equations (SPDE). Discrete Markovian approximations of the solutions of these SPDEs, obtained using Galerkin methods such as the Finite Element method, are used in place of the original field in numerical computations, thus yielding significant improvements on the associated computational costs. This same approach is also generalized to other fields characterized by SPDEs including oscillating and non-stationary random fields, but also random fields defined on manifolds.
In this work, the SPDE approach is generalized to non-Markovian fields characterized by a radial spectral density. First, an explicit formula for the covariance matrix of the weights of the finite element representation of such fields is provided. An algorithm based on a Chebyshev polynomial approximation and allowing to compute simulations of these weights with a linear computational complexity is then introduced. To derive an order of approximation that would yield good quality simulations, a criterion based on the principle of statistical tests is exposed. Finally, this approach is illustrated by a few examples. Joint work with Nicolas Desassis.

Thomas Romary (MinesParisTech)
Second order finite elements for the SPDE approach

The SPDE approach has become extremely popular in the past few years. This approach enables to work with Gaussian Markov random field approximations of a large family of random fields and opens a large range of new possibilities in the analysis and the processing of spatially referenced data. Among its many practical advantages, it allows to handle large datasets, on differentiable manifolds and to include easily possible second order non stationarities. The implementation of SPDE is currently based on first order finite elements methods. In this work, we investigate the use of second order Lagrange finite elements. The motivation for this is that using more accurate finite elements should allow to work with less dense target matrices thus lowering the computational burden, especially for 3D applications, without loss of accuracy. It occurs that the direct formulation of Lagrange P2 elements does not allow for mass lumping, that is the substitution of the mass matrix by a diagonal counterpart. This property is however necessary in practice to provide a fast computation of the precision matrix. Insights from the finite element literature show that mass lumping is available for Lagrange P2 elements using a slightly larger space of polynomials. We show how to derive these new finite elements and the associated basis functions. Illustration are provided on synthetic data. Joint work with Nicolas Desassis.