Air Quality @ ARPAE Emilia-Romagna

Roberta Amorati e Chiara Agostini

The principal activity of our group at ARPAE deals with air quality. One of the main purpose is to ensure a proper flux of information to the public and to authorities about the concentration of the main pollutants. In particular, the evaluation of the spatial distribution of pollutants over the whole Emilia-Romagna territory is a key issue. For that reason, a post processing of a chemical transport model output is accomplished to merge data from the observational network. A regression kriging is applied to the point-referenced data, getting emissions, altitude and the chemical transport model output as proxy variables. A gridded spatial distribution of pollutants concentration is obtained.

Another key issue is providing decision makers with air quality forecasts, associated with uncertainties and with the probability of exceedance of admitted pollution thresholds. A probabilistic prediction model is then required. At ARPAE a Bayesian hierarchical spatio-temporal model based on the spTimer R package, has been implemented and is now under test. It produces a three days probabilistic forecast, combining point-referenced data and the chemical transport model concentrations.

In the next future it is planned to extend these data fusion techniques to larger areas and with other covariates, such as satellite observations.

Feature selection for spatial point processes intensity estimation via regularization methods

Achmad Choiruddin

Abstract: Many methods for estimating parametrically the intensity function for inhomogeneous spatial point processes are available in the literature. Almost all papers consider the setting where the number of covariates is moderate. More and more applications involve a large number of covariates. Our paper considers feature selection procedures based on convex and non-convex regularization techniques to deal with such data. We investigate theoretical and computational aspects. We provide general conditions on the spatial point processes and on penalty functions which ensure consistency, sparsity, and asymptotic normality. In particular, from a theoretical point of view, we consider asymptotic properties which make our results available for several penalty functions and large classes of spatial point processes.

Joint work with Jean-François Coeurjolly and Frédérique Letué.

Passenger flows through airport terminals

Anthony Ebert

My supervisors and I use queueing theory and discrete event simulation to understand study passenger flows through airport terminals. The aim of my PhD project is to forecast passenger

congestion and to provide decision support for managing this congestion based on the latest information in the flight schedule.

Parameter estimation for such a complex queueing system is hard. However, if we can simulate data from a generative model of the system quickly then we can use likelihood-free techniques such as approximate Bayesian computation (ABC) to estimate parameters. If we can estimate parameters then we can predict performance. If we can predict performance then we can use the model for planning and even real-time decision making.

To simulate data from the generative model quickly we have developed the R package *queuecomputer*. Speedups of more than 2 orders of magnitude are observed compared to existing discrete event simulation (DES) packages *simmer* and *simpy*. This allows queueing system parameters to be estimated with existing ABC techniques.

TBA

Francesco Gabriele

Bayesian hierarchical modelling to estimate the risk of psychiatric disease in teachers

Maria Lodolo D'Oria

Some professional categories are at risk of burnout due to the high levels of stress that characterise their job. Symptoms of burnout are physical and psychological exhaustion, cynicism, irritability, apathy, low personal accomplishment and self-control. Several studies point out that who works in the helping professions, especially teachers, experiences burnout more frequently than any other job. Currently, burnout still not officially recognised as a disease by the two major American and European psychiatric manuals DSM V and ICD 10, respectively. Therefore, in order to achieve a scientific result about the mental disease risk in teachers data exclusively based on psychiatric diagnosis have been used. Data analysed are clinical diagnosis related to the 1993 - 2003 decade performed by the Medical Commission of Milan's Health District in order to evaluate workers disability. For the empirical analysis Bayesian hierarchical logistic regression has been implemented in order to investigate the association between the development of psychiatric diseases and job categories, after statistically controlling for gender and age. All the empirical analysis has been carried out using the R software. In our findings show a significantly higher risk of psychiatric disease in teachers compared with all other categories investigated (white collars, blue collars and healthcare professionals). Furthermore, the specific risk of developing depression or anxiety disorders has been assessed. Once again, results reveal that those mental illnesses are more likely diagnosed with teachers, after statistical controlling for gender and age. In the light of these findings, an epidemiological study on psychiatric disorders among teachers on a national basis is as much desirable as urgent. Moreover, state interventions are needed in order to prevent teachers mental disorders, preserving their teaching skills and safeguarding their crucial role in society.

Formal language and spatio-temporal analysis for ecological systems

Ludovica Luisa Vissat

In this talk I will present a novel formal framework to model and study spatio-temporal properties of stochastic models of ecological systems. This formal framework consists of a new stochastic process algebra MELA, for Modelling in Ecology with Location Attributes, and suitable techniques to analyse properties expressed through suitable spatio-temporal logics. The aim of MELA is to provide ecologists with a straightforward yet flexible modelling tool, allowing the modeller to create stochastic, individual-based and spatially explicit models. Using stochastic simulations describing the dynamic evolution of the systems, we perform statistical model checking to analyse properties of the models, formally expressed through the recently proposed Signal Spatio-Temporal Logic (SSTL). I will explain how this has been extended to widen the analysis of stochastic systems, introducing a case-study to show interesting examples of applications of this new logic.

(Bayesian) Circular Statistics

Kees Mulder

In usual spatiotemporal models, time is often treated linearly, where data points are close if they occur within a short timeframe of each other. One could also be interested in treating time as periodic, where data points are close if they occur at the same time of day, i.e., 10:00 am. The same problem occurs for observed angles, directions, axes, as well as other periodic timeframes such as days of the week or year. Because the sample space is the circle, distances, means and variances are defined differently. Statistical approaches to circular statistics include models defined directly on the circle such as the famous von Mises density, models defined as wrapped versions of densities on the real line, and models defined as densities in the Euclidean plane projected to the circle.

Investigation of risk factors of chronic airflow obstruction (CAO) in the BOLD study using Bayesian methods

Jaymini Patel

Chronic obstructive pulmonary disease (COPD) is an important cause of morbidity, mortality and increased health care costs. The prevalence of COPD and the exposure to the different risk factors is likely to vary between countries. Thus, explaining the observed differences in disease prevalence, it is important to gain better understanding of the role of risk factors for COPD in different parts of the world. The Burden of Obstructive Lung Disease (BOLD) study was designed to measure the prevalence of COPD and its major risk factors worldwide. The BOLD study is a cross-sectional study on adults of age 40 years and above and so far, has completed in 40 centres in Asia, Europe, North America and Africa. The large BOLD dataset provides a unique opportunity to investigate risk factors for COPD and to explore the heterogeneity in their effects across different geographical areas. During my work, I have come across following three methodological issues in the analysis of the data which limit the scope and efficiency of my investigation of specific hypotheses.
(1) How to deal efficiently with BOLD missing data ***across*** BOLD countries

(2) How to increase the precision of the estimate showing overall impact of risk factor at the population level in the different BOLD countries
(3) How to estimate the effects of candidate risk factors for CAO and evaluate their heterogeneity across BOLD countries
In my presentation, I will discuss the application of the Bayesian model, which is inspired from Jackson's Bayesian graphical model (Biostatistics 2009:10:335) to impute missing variables. The proposed model deals with between centre heterogeneity (random slopes) in the imputation model as well as in the analysis model.
I will also present the method I am using to improve the precision of the measure of impact using the example of the estimate of the population attributable fraction(PAF). Country-specific estimates of PAF obtained using available methods are inaccurate for some of the countries in the BOLD study because of lack of statistical power (due to small sample size; low prevalence of the risk factor; low prevalence of COPD).

Tracking the Evolution of Rainfall Precipitation Fields Using Persistent Maxima

Simone Pittaluga

Presentation of a novel methodology for tracking the maxima of rainfall precipitation fields, whose changes in time may give interesting insights on storm evolution. Our approach is based on a topological analysis of rainfall data allowing for the extraction of the most prominent, and hence meaningful, rainfall field maxima. Then, an ad-hoc bottleneck matching is used to track the evolution of maxima along multiple time instances. This work has been done in the context of the IQmulus project, an EU FP7 research project concluded in 2016.

Future interests: environmental adaptive sampling

Luigi Rocca

A current multi-disciplinary research direction for us is based on the idea of bringing together natural scientists, computer scientists and mathematicians to investigate together new paradigms for environmental sampling and monitoring. In particular, the advent of fast sensors suggests the adoption of adaptive sampling methods. Traditionally, monitoring locations are selected in advance prior to making observations and model fitting. An adaptive approach would take into account prior observations to decide the next sampling points instead using an iterative sequence. Approaches of this kind are known in the literature but have been seldom used in practice, and there are many questions to answer and possible optimizations to perform before adoption in the field.

Fast Bayesian spatial 3D priors for brain imaging

Per Siden

Spatial whole-brain Bayesian modeling of task-related functional magnetic resonance imaging (fMRI) is a serious computational challenge, due to the large amount of data in the brain images,

that are measured over time. In this talk I will present how Gaussian Markov random field (GMRF) models, which have sparse precision (inverse covariance) matrices, can be used for efficient inference schemes with 3D priors on the latent brain activity coefficients, even when the number of coefficients are several hundreds of thousands. We use modern techniques for sampling from high-dimensional Gaussian posterior distribution for both exact inference (MCMC) and approximate inference (VB), with negligible error compared to the MCMC posterior. Also, a method is introduced for computing selected elements of the posterior covariance matrix from the posterior precision matrix, when this is too large for direct inversion.

Joint work with Mattias Villani, David Bolin, Finn Lindgren and Anders Eklund.

The Analysis of Economic Time Series with R

Giuseppe Smigliani

In my M.Sc. I will study the most common statistical methods and tools used to describe economic time series behavior. My analysis will be assisted by the R software environment through which I will perform data calculation and graphical display. The dissertation is composed by two parts.

In the first part I will study the statistical theory which is at the base of the models I will be concerned with. In this respect, to better understand conceptual basics of how models work I will simulate data and I will also provide theoretical examples of them, finally I will analyze some interesting time series from an economic point of view.

The second part of the work is dedicated to forecasting, models building and identification. For the great majority of the work I will be concerned with stationary stochastic models used to describe the typical path dependence of time series data. For, I will study moving-average processes MA, autoregressive processes AR, and autoregressive-moving-average processes, ARMA. Finally, I will also dedicate space to autoregressive-integrated-moving-average models, ARIMA, which is the largest class of models used to describe nonstationary data behavior in economic and financial analysis. Because time series analysis is often classified in time domain analysis and frequency domain analysis, I will provide autocorrelations and spectral densities for each model (data) studied (analyzed).

TBA

Christien Thiart