TIES - GRASPA 2017

on

Climate and Environment

Book of Abstracts

The 27th Annual Conference of
The International Environmetrics Society
joint with Biennial GRASPA Conference

24-26 July 2017
Bergamo, Italy

Edited by: Michela Cameletti and Francesco Finazzi

ISSN 2037-7738

TIES-GRASPA 2017 CONFERENCE
On Climate and Environment

The 27th Annual Conference of
The International Environmetrics Society
joint with Biennial GRASPA Conference

Organised by

The International ENVIRONMETRICS Society - TIES

GRASPA

Italian Research Group for Statistical
Applications to Environmental Problems

University of Bergamo
and
Department of Management, Information and Production Engineering

Sponsored by

QUT

Queensland
University of Technology

International Statistical Institute

Società Italiana di Statistica SIS fondata nel 1939

Italian Statistical Society

Welcome Address

The joint TIES-GRASPA 2017 Conference on Climate and Environment is held at the University of Bergamo, Italy from 24th to 26th July and it is a satellite meeting of the 61st World Statistics Congress – ISI2017 in Marrakesh.

The International Environmetrics Society (TIES) is a non-profit organization aimed to foster the development and use of statistical and other quantitative methods in the environmental sciences, environmental engineering and environmental monitoring and protection. The Society promotes the participation of statisticians, mathematicians, scientists and engineers in the solution of environmental problems and emphasizes the need for collaboration and for clear communication between individuals from different disciplines and between researchers and practitioners.

The Italian environmetricians group named GRASPA is active since 1995 and it has become since May 2013 a standing group of the Italian Statistical Society (SIS) for Environmental Statistics, sustainability and territorial safety. GRASPA-SIS promotes statistical and interdisciplinary research in the field of environmental quality, safety and sustainability including air and water quality, epidemiology, earth science and ecology.

The conference is intended to be a bridge for a future TIES conference system after ISI incorporation. On the one side, it renews a fruitful tradition of joint conferences with other scientific societies and associations. On the other side, it explores new ways to collaborate with ISI. From the scientific point of view, the TIES-GRASPA conference on Climate and Environment is focused on a hot topic and has the potential to attract new statisticians and other scientists.

Keynote speakers of the conference are Sudipto Banerjee (University of California), Claudia Tebaldi (National Center for Atmospheric Research), Emilio Porcu (Newcastle University) and Brunero Liseo (University of Rome) while the panel session on "Uncertainty and climate change" has Jochem Marotzke (Max-Planck-Institut für Meteorologie), Peter Thorne (Maynooth University) and Richard Smith (University of North Carolina) as panellists.

This Book of Abstracts includes 208 abstracts of keynote, invited and contributing authors presenting at the conference with talks and posters. Two open calls for special issues connected to climate change and environment will be published in "Statistics & Probability Letters" and "Environmetrics" journals.

On bahalf of the local organising committee and of the scientific committee, we would like to thank all the speakers and we extend a very warm welcome to everyone attending the TIES-GRASPA 2017 conference.

Michela Cameletti and Francesco Finazzi

Contents

Keynote sessions

High-Dimensional Bayesian Geostatistics

S. Banerjee[(a)]

[(a)]University of California, Los Angeles

With the growing capabilities of Geographic Information Systems (GIS) and related software, statisticians today routinely encounter spatial data containing observations from a massive number of locations and time points. Important areas of application include environmental exposure assessment and construction of risk maps based upon massive amounts of spatiotemporal data. Spatiotemporal process models have been, and continue to be, widely deployed by researchers to better understand the complex nature of spatial and temporal variability. However, fitting hierarchical spatiotemporal models is computationally onerous with complexity increasing in cubic order for the number of spatial locations and temporal points. Massively scalable Gaussian process models, such as the Nearest-Neighbor Gaussian Process (NNGP), that can be estimated using algorithms requiring floating point operations (flops) and storage linear in the number of spatiotemporal points. The focus will be on a variety of modeling and computational strategies to implement massively scalable Gaussian process models and conduct Bayesian inference in settings involving massive amounts of spatial data.

Modeling Temporally Evolving and Spatially Globally Dependent Data

E. Porcu[(a)]

[(a)]Newcastle University, Newcastle. Technical University Federico Santa Maria, Valparaìso

The last decades have seen an unprecedented increase in the availability of data sets that are inherently global and temporally evolving, from remotely sensed networks to climate model ensembles. This paper provides a view of statistical modeling techniques for space-time processes, where space is the sphere representing our planet. In particular, we make a distintion between (a) second order-based, and (b) practical approaches to model temporally evolving global processes. The former are based on the specification of a class of space-time covariance functions, with space being the two-dimensional sphere. The latter are based on explicit description of the dynamics of the space-time process, i.e., by specifying its evolution as a function of its past

history with added spatially dependent noise. We especially focus on approach (a), where the literature has been sparse. We provide new models of space-time covariance functions for random fields defined on spheres cross time. Practical approaches, (b), are also discussed, with special emphasis on models built directly on the sphere, without projecting the spherical coordinate on the plane. We present a case study focused on the analysis of air pollution from the 2015 wildfires in Equatorial Asia, an event which was classified as the year's worst environmental disaster. The paper finishes with a list of the main theoretical and applied research problems in the area, where we expect the statistical community to engage over the next decade.

Future Climate Change Projections Through Statistical Analysis of Multi-model Ensembles: Challenges and Opportunities

C. Tebaldi[(a)]

[(a)]National Center for Atmospheric Research, Boulder

I will sketch the reasons for the necessity of utilizing multiple models when addressing future climate change projections and their uncertainty, but also the challenging nature of multi-model ensembles as far as their statistical characteristics are concerned: why analysis of their output provides both difficult hurdles to overcome, when going beyond descriptive statistics, but also opportunities for innovative statistical modeling, algorithms and data exploration. After reviewing the main sources of uncertainties affecting climate model output, and the odd statistical characteristics of the data samples they generate when considered as an ensemble, I will present a few distinct areas of research where I think the injection of fresh perspectives would have the chance of creating a big impact. Among those, model tuning; pattern scaling and emulators; and, perhaps the crux of any analysis of future projection, the problem of relating model skill for current climate simulation to model reliability in the future part of their simulation (i.e., the search for emergent constraints).

Invited sessions

Spatiotemporal Interpolation of Rainfall Variability in Blue Nile Basin for the Period 1998–2016

Y.M. Abdelfattah[a], A.H. El-Shaarawi[b], H. Abou-Ali[c]

[a]The British University in Egypt, El Sherouk, [b]Canadian National Water Research Institute, [c]Cairo University, Cairo

Egypt has reached a state where the quantity of water available is imposing limits on its national economic development. The Blue Nile feeds most of the water that arrives to Egypt. The Blue Nile is exclusively fed by precipitation in the Ethiopian highlands, which is concentrated in a 4-month rainy season, June to September. This study describes the main patterns of rainfall distribution in the Blue Nile Basin using Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) product 3B42 version 7 and characterizes the associated processes using spatiotemporal modeling. The rainfall magnitude is measured using satellite data TRMM 3B43-in monthly time which spans from year 1998 to 2016 at spatial resolution of $0.25^\circ \times 0.25^\circ$ latitude/longitude. Rainfall pattern across the Blue Nile Basin is affected with both climate variability and change. Any change in the behavior of this crucial hydro-climatic variable is spread irregularly throughout the region; rather, it has its own unique localized pattern, both spatially and temporally. Therefore, to explore the intrinsic spatial patterns of rainfall, not only should the spatial and temporal variability be analyzed separately, but the joint spatio-temporal variability should also be explored by applying spatio-temporal semivariogram models and interpolation methods. The implications from the research have a broader significance for Egypt. As it highlights the distribution of rainfall that passes through the Grand Ethiopian Renaissance Dam (GERD) which is planned to be built on the Blue Nile at the Ethiopian- Sudanese border.

Spatio-temporal Log-Gaussian Cox Processes On Earthquake Events

G. Adelfio[(a)], M. Siino[(a)], J. Mateu[(b)]
[(a)]University of Palermo, Palermo, [(b)]University Jaume I, Castellon

In this paper we aim at studying some extensions of complex space-time models, useful for the description of earthquake data. In particular we want to focus on the Log-Gaussian Cox Process (LGCP) model estimation approach, with some results on global informal diagnostics. Indeed, in our opinion the use of Cox processes that are natural models for point process phenomena that are environmentally driven could be a new approach for the description of seismic events. These models can be useful in estimating the intensity surface of a spatio-temporal point process, in constructing spatially continuous maps of earthquake risk from spatially discrete data, and in real-time seismic activity surveillance. Moreover, covariate information varying in space-time can be considered into the LGCP model, providing complex models useful for a proper description of seismic events. LGCP is a Cox process with a stochastic intensity function, depending on a Gaussian process. This construction has some advantages, related to the multivariate Normal distribution features, since the moment properties of the intensity function are inherited by the Cox process. In particular, both estimation and diagnostics, can deal with some higherorder properties, expressed for instance by the intensity and the pair correlation function of the LGCP.

Risk Estimates in Disease Mapping When Discontinuities Are Present

A. Adin[(a)], D. Lee[(b)], T. Goicoa[(a)], M. Ugarte[(a)]
[(a)]Public University of Navarre, Pamplona, [(b)]University of Glasgow, Glasgow

Smoothing risks and detecting cluster structures are two contradictory goals in disease mapping. The main reason for using statistical models in disease mapping is to borrow strength from spatial (and temporal) neighbors to reduce the high variability inherent to classical risk estimators, such as the standardized mortality ratio (SMR). On the contrary, the goal of cluster detection models is to identify groups of areal units that exhibit substantially different high/low risks compared with their neighbors. In this work, a two-stage approach is proposed for both clustering detection and smoothing risks in the presence of local discontinuities. In the first stage, an agglomerative hierarchical clustering algorithm is applied to ancillary data to provide an initial set of cluster conguration candidates, and in the second stage, two-level spatial and spatio-temporal models are applied to each conguration selecting the one that minimizes several information criteria. The superiority of the proposed approach with regard to previous proposals is shown by simulation. The procedure is also illustrated using real data.

A Latent Gaussian Approach for Modeling Geological Sequences: Model, Inference and Conditional Simulation

D. Allard[(a)], C. Gaetan[(b)], P. Fabbri[(c)]

[(a)]INRA, [(b)]Università Ca'Foscari, Venezia, [(c)]Università degli studi di Padova, Padova

The prediction of the soil contamination in an environmental risk area is a problem of relevant public interest. Very often, a proper analysis requires a preliminary geological study. Indeed, understanding the distributions of permeable and impermeable layers under the risk area is necessary in order to estimate the risk that a pollutant could reach a sensible location, like the aquifers. The data available in this sort of studies come from cores drilled at several locations. The data are geological sequences summarized by a vector of thicknesses of different sediments or categories. A 3D reconstruction of the soil characteristics starting from this data requires to specify a spatial model for categorical data. In this talk, we present a new spatial model based on the assumption that there exists a sequence of categories (parent sequence), which is compatible with all observed sequences in the sense that each observed sequence can be obtained from the parent sequence by deleting one or more categories of the parent sequence. The parent sequence can be extracted from the empirical data or defined according the prior knowledge of a scientist about the geological deposition. The thicknesses of each category are modeled as truncated Gaussian random variables that are spatially correlated. This choice allows the possibility that the parent sequence is not completely observed. We will therefore cast the estimation procedure of the model parameters within a Bayesian framework using an MCMC algorithm for sampling all possible sequences. We will illustrate the relative merits of our modeling procedure by means of simulated and real examples. In particular, we will show how we can simulate a 3D reconstruction conditionally on the observed data.

The Use of Spatial Information in Entropy Measures

L. Altieri[(a)], D. Cocchi[(a)], G. Roli[(a)]

[(a)]University of Bologna, Bologna

The concept of entropy, firstly introduced in information theory, rapidly became popular in many applied sciences via Shannon's formula to measure the degree of heterogeneity among observations. A rather recent research field aims at accounting for space in entropy measures, as a generalization when the spatial location of occurrences ought to be accounted for. The main limit of these developments is that all indices are computed conditional on a single chosen distance and do not cover the whole spatial configuration of the phenomenon under study. Moreover, most of them do not satisfy the desirable additivity property between local and global spatial measures. This work follows and extends the route for including spatial components in entropy measures. Starting from the probabilistic properties of Shannon's entropy for categorical variables, it investigates the characteristics of the quantities known as residual entropy and mutual information, when space is included as a

second dimension. This way, the proposal of entropy measures based on univariate distributions is extended to the consideration of bivariate distributions, in a setting where the probabilistic meaning of all components is well defined. As a direct consequence, substantial innovations are achieved. Firstly, Shannon's entropy of a variable may be decomposed into one term, spatial mutual information, accounting for the role of space in determining the variable outcome, and another term, spatial global residual entropy, summarizing the remaining information carried by the variable itself. Secondly, the two terms both satisfy the additivity property, being sums of partial entropies measuring what happens at different distance classes. The superiority of the proposed indices is assessed both via their theoretical properties and via a thorough comparative study. Moreover, they are used for measuring the spatial entropy of a case study consisting on a marked point pattern on rainforest tree species.

Earthquakes and Human Displacement in Italy. A Comparison Between L'Aquila and the Emilia Romagna Cases

E. Ambrosetti[(a)], E.R. Petrillo[(a)]

[(a)]Sapienza Università di Roma, Roma

This paper analyses the environmentally-induced migration and displacement resulting from two earthquakes: Abruzzo (2009) and Emilia Romagna (2012). After a general critical overview of the social science literature on this topic, the main changes in the two migration systems are analysed looking at the different roots and trajectories of the forced human displacement that followed the two earthquakes. Moving from the fact that similar events may occurring in different contexts, may have different outcomes according to the specific vulnerability experienced by the territory, we assess the pre disaster context and recovery period with the aim to offer a comparative analysis of the challenges related to post earthquake demographic movements and post-disaster resettlement. The overall goal is to understand how the two migration systems have been influenced by the pre-existent vulnerabilities and pre-quake social and institutional backgrounds before and after the hazard. Relying on ISTAT data on the internal migration in Italy, we finally offer a general model of how environmental disaster might affect displacement and suggest the main challenges related to the post-disaster governance.

E-mobility Infrastructure for Supporting Private Mobility in Italy

G. Azzone[(a)], M. Arena[(a)], M. Brenna[(a)], A. Calissano[(a)], F. Foiadelli[(a)], M. Longo[(a)], A. Menafoglio[(a)], P. Secchi[(a)], S. Vantini[(a)]
[(a)]Politecnico di Milano, Milano

Today, the issue of e-mobility is in the foreground of the debate concerning sustainable mobility. Major car manufacturers have launched new models of electric cars at competitive prices and consumers are demonstrating increasing curiosity concerning the economic and practical convenience of this new technology. However, for ensuring the diffusion of e-vehicles among private consumers, a crucial factor is represented by the existence of an adequate infrastructure that allows the users the possibility of recharging their cars easily, overcoming the so called "range anxiety". The centrality of the role of the recharging infrastructure has been clearly recognized by the EU community, that has defined a shared framework for the construction of recharging stations for alternative fuels in Europe by the 2020. Still, for implementing these programmatic guidelines, each single country has to face the problem of determining, in operational terms, the number, type and location where the infrastructure has to be build. The speech will deal specifically with this problem, reporting on the results of a project aimed at designing the recharging infrastructure in Italy, in order to ensure the possibility of using e-vehicles across the whole country. In particular, a parametrical model has been adopted in order to deal with the uncertainty that is inherent to different design variables, such as the number of e-vehicles that will be driven in the 2020, the autonomy range and the recharging time that will be allowed by technological development, possible changes in consumer behaviour. Then, the alternative scenarios have been evaluated considering economic environmental and energy effects associated to different choices.

Plant Species Distribution Modeling for Opportunistic Occurrences, a Case Study On Pl@ntNet Data

C. Botella[(a)], A. Joly[(a)], P. Monestiez[(a)], P. Bonnet[(b)], F. Munoz[(c)]
[(a)]INRA, [(b)]CIRAD - AMAP, [(c)]Université de Monptellier 2, Montpellier

Today, citizen sciences programs are quickly emerging and developing across the world, especially thanks to web technologies facilitating data collection, sharing and revision, or species identification, with large communities of contributors. A diversity of contribution schemes exists amongst those programs. While some have well defined data collection protocols oriented toward scientific analysis, others gather opportunistic species occurrences from a large number of contributors. Pl@ntNet is a good example of such a system, in which a smartphone application is open to everyone and helps non-expert users identify a wide spectrum of plant species from pictures in a semi-automated way. The system is generating a flow of around 10000 geolocated species occurrences per day around the world at this time. Such a growing database could contain useful information to get a better estimate of

spatial distribution and dynamics of invasive alien species. However, occurrences being opportunistic, they are not only distributed according to species abundance, but also influenced by the observational process, which is even more problematic since we don't have true absence data. We will present results in this context, in which we model occurrences report with an inhomogeneous Poisson point process depending on observation intensity and species abundance, where abundance is a function of environmental variables. We study two different procedures to estimate species abundance up to a proportionality constant and without bias. The first is a reference method that uses all occurrences of a species group to approximate the relative observation intensity. Then, we propose another procedure that estimates simultaneously relative abundance and observation intensity. We evaluate their precision by comparing relative abundance estimates from Pl@ntNet data to the ones obtained from expert data.

Inlabru: Bayesian Modeling and Point Process Analysis

F.E. Bachl[(a)]

[(a)]University of Edinburgh, Edinburgh

Recent methodological advances have brought point process inference using the Integrated Nested Laplace Approximation (INLA) to the domain of distance sampling, a widely used method of estimating wildlife population abundance. Albeit widely applicable, a key component for more realistic and sophisticated models has been neglected so far. In many cases the observational thinning of the underlying point processes depends on covariates and marks that require a network of spatiotemporally varying models. This introduces non-linearities to the linear predictor which are not covered by the current INLA methodology. Moreover, using the INLA R package for spatial modeling in general and point processes in particular involves error-prone and time consuming technicalities that can be avoided from the user's perspective. As a remedy we introduce inlabru, a package that provides a convenient interface to INLA's spatial modeling functionality and an automated approximation scheme for non-linear predictors. We illustrate inlabru's capabilities using distance sampling data from a series of shipboard line transect surveys of striped dolphins in the eastern tropical Pacific (ETP). This species gathers in schools with a large range of sizes, which significantly influences the likelihood of a group being seen by the observers on a research vessel. The respective group size distribution depends on space and connects the distribution of groups to a distribution of animals we are ultimately interested in. Our inference framework successfully separates detectability, group size and spatial group intensity within a joint Bayesian network formulation.

Going Around the Barriers - Simple Models for Complex Ecological Behaviour

H.C. Bakka[(a)]

[(a)]Norwegian University of Science and Technology, Trondheim

Spatial ecology is dependent on our ability to construct spatial abundance maps, and model habitat preference. To perform these tasks, we need a toolbox of statistical models that are appropriate for modeling species distributions. Spatial models often describe complex relationships, and the real-world implication of a specific statistical assumption, or property, can be difficult to understand. To develop our intuition and understanding, we need to visualize the assumptions inherent in our models. To improve the general methodology we need to develop models with a variety of realistic assumptions. One tool that we need is to allow for spatial barriers, instead of just assuming stationarity. For example, when modeling fish near the shore, the spatial dependencies should go around islands and peninsulas rather than across them. Nature is full of similar barriers. Human-made changes may result in artificial (semi-permeable) barriers. E.g. the impact of a major traffic/shipping lane may, in some cases, be better modelled as a barrier to correlation than as local changes in abundance. In this talk I will present our ongoing work on understanding real-world implications of modeling assumptions, a new, realistic and easy-to-use model for the barrier problem, an application of this model, and the detection (or implication) of human-made barriers.

On Modelling Rain Drop Size Distributions for Hydrological Applications and Remote Sensing

L. Baldini[(a)], E. Adirosi[(a)], E. Volpi[(b)], F. Lombardo[(b)], N. Roberto[(a)], M. Montopoli[(a)]

[(a)]Consiglio Nazionale delle Ricerche Istituto di Scienze dell'Atmosfera e del Clima, Roma, [(b)]Università degli Studi Roma Tre, Roma

The choice of an appropriate analytical expression to model rain Drop Size Distribution (DSD) is fundamental to develop remote sensing methods for estimating precipitation. DSD represents the number of raindrops normalized to the raindrop diameter interval and the volume of air, and it can be estimated either by sampling drops falling through a device measuring area (e.g. instruments such as impact, laser, and 2D video disdrometers) or by measuring Doppler spectra (examples are known as Micro Rain Radar, POSS, Pludix). Previous studies have suggested a rather wide range of theoretical distributions to describe DSDs, with a clear preference for the gamma distribution. In this work, we want to contribute to establish how often and how well these distributions are able to fit experimental data. To this aim, using the maximum likelihood method, we fit models based on lognormal, gamma and Weibull distributions to more than 42.000 1-minute measured drop spectra from instruments (such as 2D video disdrometers) deployed in field campaigns of NASA Ground Validation program of the Global Precipitation

Measurement (GPM) mission. The Kolmogorov–Smirnov (KS) test is applied to investigate the goodness of fit. A model selection technique evaluates the relative quality of each DSD model. The model based on the gamma distribution presents the lowest KS rejection rate. On the other hand, Weibull is the most frequently rejected distribution. Considering for each minute the models that pass the KS test, the light-tailed distributions seem to be more adequate to model the variability exhibited by measured DSDs. However, it was also found that distributions fitting empirical DSDs could be heavy-tailed in a significant number of cases: this causes uncertainty in estimating statistical moments and bulk hydro-meteorological variables. Implications of the results found are discussed from the practical perspective of weather radar remote sensing of hydro-meteorological variables.

Statistical Post-processing of Ensemble Forecasts for Precipitation Accumulation

S. Baran[(a)]

[(a)]University of Debrecen, Debrecen

Recently, all major weather prediction centres provide forecast ensembles of different weather quantities, which are obtained from multiple runs of numerical weather prediction models with various initial conditions and model parametrizations. However, ensemble forecasts are often underdispersive and may also be biased, so that some post-processing is needed to account for these deficiencies. Probably the most popular modern post-processing techniques are the ensemble model output statistics (EMOS) and the Bayesian model averaging (BMA), which provide estimates of the density of the predictable weather quantity. We present an EMOS method for calibrating ensemble forecasts of precipitation accumulation, where the predictive distribution follows a censored and shifted gamma (CSG) law with parameters depending on the ensemble members. The CSG EMOS model is tested on ensemble forecasts of 24-h precipitation accumulation of the eight-member University of Washington mesoscale ensemble and on the 11-member ensemble produced by the operational Limited Area Model Ensemble Prediction System of the Hungarian Meteorological Service. The predictive performance of the new EMOS approach is compared with the fit of the raw ensemble, the generalized extreme value (GEV) distribution-based EMOS model and the gamma BMA method. According to the results, the proposed CSG EMOS model slightly outperforms the GEV EMOS approach in terms of calibration of probabilistic and accuracy of point forecasts and shows significantly better predictive skill that the raw ensemble and the BMA model.

Verification of Severe Thunderstorms Forecast in Lombardia

S. Barberis[(a)]

[(a)]Università Milano Bicocca, Milano

ARPA Lombardia provides weather data and forecasts to support Civil Protection to prevent and mitigate natural hazards. This work focuses on the forecasts verification of severe thunderstorms. Traditional verifications are carried out comparing forecast and observed rainfall accumulation. This study introduces a new method based on the Severity Storm Index (SSI) that summarizes all the phenomena related to storms (large hail, lightning, heavy showers) and their damage potential. An algorithm has been developed to identify and to classify storms on the basis of the Storm Severity Index: each radar detected storm cell is schematized as an ellipse and its SSI value is computed (ranging from 1 to 5). Civil Protection operates on the basis of severe storms forecasts, i.e. those with the highest potential to make a widespread damage, with a 24-hours leading time. Lombardia is divided into eight "alert areas" and a three levels probability of occurrence forecast (unlikely, likely, and very likely) of severe storms is issued for each of them. A basic method for verifying probabilistic forecasts is the Brier score. However, in this application the main interest of the user is in the correctness of the alerts issued. This arise the problem of defining what the event of interest is and how to identify its occurrence. The method developed, based on the so called Area Impact Severity Index (AISI), is discussed and some verification results are presented.

Weather Generators By Means of Truncated and Transformed Gaussian Random Fields

A. Baxevani[(a)]

[(a)]University of Cyprus, Latsia

There are certain common characteristics among a variety of environmental processes, such as precipitation, wind and wind power generation among others. These include periods of time or areas with no activity - which results to distributions that have a mass at zero, and also tails that usually are heavier than those of the Gaussian distribution. We propose a common approach with the use of a latent transformed and truncated Gaussian field. The truncation accommodates the probability masses at zero and the anamorphosis transformation produces the proper marginal distributions. The spatio-temporal dependence is captured by means of dynamic anisotropic covariance models. The methodology has been applied to produce both long-term and short terms spatio-temporal prediction for precipitation data from Sweden and wind power production from a portfolio of wind farms situated in Denmark.

Classification Techniques for the Retrieval of Microphysical Information From Dual-polarization Weather Radars

R. Bechini[(a)], R. Cremonini[(a)]
[(a)]Arpa Piemonte

Weather radars with dual-polarization capabilities provide a large amount of information from the observation of clouds and precipitation using orthogonal polarized signals. The transmission and reception of horizontally and vertically aligned pulses allows the calculation of several meteorological moments from the full covariance matrix. The typical set of radar variables includes reflectivity, differential reflectivity, correlation coefficient, differential phase shift, and Doppler velocity. The range of values assumed by each variable depends on the type of scatterer, whether it is a meteorological target and which type of hydrometeor (rain, snow, hail, ...) is dominant within the radar resolution volume. Since the introduction of dual-polarization techniques in weather radar systems during the seventies, efforts have been devoted to exploit this additional information for the identification of specific precipitation types. In particular, the early studies focused on the discrimination between rain and hail. More in general, the need of a synthesis in terms of precipitation composition from the available radar observations has led to the development of hydrometeor classification algorithms. In this presentation we'll revisit the progress of hydrometeor classification, considering the main techniques that have been developed so far, and focusing in particular on the most widely used algorithms based on fuzzy logic. A discussion of more recent methods including cluster analysis and segmentation techniques will also be given.

Stochastic Rainfall Generation From a Censored Latent Gaussian Field: Micro-scale Parametrization

L. Benoit[(a)], D. Allard[(b)], G. Mariethoz[(a)]
[(a)]University of Lausanne, Lausanne, [(b)]INRA

The characterization of rainfall spatial and temporal variability is a longstanding challenge. Since the physical processes responsible for rainfall generation are still only partially understood, stochastic methods have been widely used to characterize and model rainfall. These models have continuously evolved in the last decades to fit instrumental improvements in resolution and accuracy. Here we propose to study the structure of the rain at micro scale, i.e. over areas of 1-10km^2, and with a fine resolution, i.e. 100m $\times$ 100m $\times$ 30sec. The focus is on characterizing and reproducing spatial rainfall patterns and their dynamics at the scale of interest. To this end, we adopt a stochastic rainfall model based on a censored space-time latent Gaussian field. Such statistical model allows to reproduce the main features of rainfall at subkilometric scale, namely the advection and morphing properties of rain cells over time, the intermittency and the skewed distribution of rainfall, and the decrease of the rain rate near the rain cell boundaries. Inferring the model parameters is non-trivial given the available observation datasets, i.e. high resolution

rain rate time series collected by a network of pluviometers. In this presentation we investigate different methods for the inference of the parameters of the stochastic rainfall model, and we compare their efficiency for micro-scale parametrization of space-time rain fields.

A Segmentation-based Approach for Non-stationary Spatial Modeling of Environmental Data

V.J. Berrocal[(a)]

[(a)]University of Michigan, Ann Arbor

A key component of statistical models for environmental applications is the spatial covariance function, which is traditionally assumed to belong to a parametric class of stationary models whose parameters are estimated using observed data. While convenient, the assumption of stationarity is rarely realistic; as a result, there is a rich literature on alternative strategies that allow the covariance function to vary across space. We consider a Bayesian nonstationary spatial model constructed through the convolution of locally stationary spatial processes. Regions of local stationarity underlying our model are determined via segmentation of the geographic space into homogeneous regions based on the within-region distribution of spatially-varying covariate information. Via Bayesian model averaging, we account for uncertainty in the segmentation process, as well as the situation where multiple spatially-varying covariates are relevant to describing the nonstationary behavior of the process of interest. In this way, spatially-referenced covariates in our model are allowed to drive both the first and second moments of the spatial process. One advantage of our modeling framework over other covariate-driven nonstationary spatial models is that our approach does not require the covariates to be observed at the same location as either the observations or prediction locations. We illustrate our methodology through an application in soil science.

Low-memory Filtering for High-dimensional Satellite Data Time Series

A. Bibov[(a)]

[(a)]Lappeenranta University of Technology, Lappeenranta

Data filtering is process of noise reduction from raw data. One among the widely used filtering approaches is the Kalman filter, which employs analytical description of the process determining the physics behind the data and uses it to correct the noisy measurements. Unfortunately, when either dimension of the data or the number of state parameters of the governing model is large, the classical Kalman filter formulae cannot longer be directly used because of increased compute and memory costs. In the present work we estimate the amount of chlorophyll in the Finnish Gulf due to algae blooming using the image data extracted from satellite spectral observations. The task is complicated by cloudiness, which results in "gaps" in the

observation data. This estimation problem is naturally high-dimensional due to resolution of the satellite images and therefore cannot be solved by classical Kalman filter and smoother algorithms. In order to overcome this problem we employ so-called parallel Kalman filtering, which is essentially a low-memory approximation of the known fixed-lag Kalman smoother. At the current stage of this ongoing work we successfully implement this low-memory approach and manage to obtain smooth estimates regardless of cloud coverage in the raw data.

Assessment of Solar Radiation Components in Brazil Using the BRL Diffuse Radiation Model

J. Boland[(a)]

[(a)]University of South Australia, Adelaide

Quality data regarding direct and diffuse solar irradiation is crucial for the proper design and simulation of solar systems. This information, however, is not available for the entire Brazilian territory. Hourly measurements of global irradiation for more than seven hundred stations over the territory are available. Several mathematical models have been developed over the past few decades aiming to deliver estimations of solar irradiation components when only measurement of global irradiation is available. In order to provide reliable estimates of diffuse and direct radiation in Brazil, the recently presented Boland–Ridley–Laurent (BRL) model is adjusted to the particular features of Brazilian climate data, developing adjusted BRL models on minute and hourly bases. The model is adjusted using global, diffuse and direct solar irradiation measurements at nine stations, which are maintained by INPE in the frame of the SONDA project. The methodology for processing and analysing the quality of the data sets and the procedures to build the adjusted BRL model is thoroughly described. The error indicators show that the adjusted BRL model performs better or similarly to the original one, for both diffuse and DNI estimates calculated for each analysed Brazilian station. For instance, the original BRL model diffuse fraction estimates have MeAPE errors ranging from 16% to 51%, while the adjusted BRL model gives errors from 9% to 26%. Regarding the comparison between the minute and hourly adjusted models, it can be concluded that both performed similarly, indicating that the logistic behaviour of the original BRL model is well suited to make estimates in sub-hourly data sets. Based on the results, the proposed adjusted model can be used to provide reliable estimates of the distribution of direct and diffuse irradiation, and therefore, can help to properly design and reduce the risks associated to solar energy systems.

Multivariate Type-G Random Fields

D. Bolin[(a)]

[(a)]University of Gothenburg, Gothenburg

A new class of non-Gaussian multivariate random fields is formulated using systems of stochastic partial differential equations (SPDEs) with additive spatial type-G noise. We show how to formulate the system to get solutions with marginal Matérn covariance functions for each dimension and derive a parametrization that allows for separate control of cross-covariance and other dependence between the dimensions. Four different constructions of the type-G noise based on normal-variance mixtures are examined. The different constructions result in random fields with increasing flexibility and we show that the simpler constructions, which are often used in the literature, lacks important properties such as ergodicity. The fields are incorporated in a geostatistical model with measurement errors and covariates, for which a computationally efficient likelihood-based parameter estimation method is derived. Finally, we compare the predictive ability of the different multivariate type-G fields and regular multivariate Gaussian random fields for two geostatistical datasets.

Lag-Adjusted Models for Air Pollution Time Series

W. Burr[(a)]

[(a)]Trent University, Peterborough

In this talk we examine recent developments in statistical models for integration of multiple lags in models for population health risk due to air pollution. Several widely-used models are compared and contrasted with a new statistical algorithm for estimation of lag structure and risk, which we have named "synthetic lag". Our new model uses frequency-domain structural estimation of lag coherence to allow for lag-matching of structural components in air pollution and human health metrics. Models are compared using simulation and two real-world data sets: the NMMAPS database of United States cities (1987-2000), and the AHI database of Canadian cities (1984-2012).

Modelling Fish Fauna Assemblages To Detect Factors Affecting Differences Between Coral and Non Coral Habitats

C. Calculli[(a)], A. Pollice[(a)], G. D'Onghia[(a)], L. Sion[(a)]
[(a)]University of Bari Aldo Moro, Bari

The characterization of biotic and abiotic factors affecting fish assemblages has been widely accepted as an essential tool to assess the stability of habitats, to investigate the ecosystem productivity and the overall environmental change. Monitoring coral habitats, well-known for the species richness and abundance, is crucial for marine biologists, due to the many implications in terms of population distribution and dynamics impacting on ecosystem services. Here, the faunal assemblages of a cold-water coral (CWC) site recently recorded along the Apulian margin (central Mediterranean) is investigated in comparison to those recorded outside the coral area. Data come from experimental longline surveys conducted during the 2014 and 2015 spring-summer seasons. Only 11 hauls were sampled for each of the two areas, along transects with increasing depth. Overall, about 20 species were caught, with 6 mostly abundant. Besides ad hoc multivariate statistical methods, mostly specific to community ecology, we propose a stochastic modelling approach that allows addressing some critical issues of the available data. Indeed, the small sample size together with the high observed species heterogeneity pose some interesting modelling challenges. Then, in order to characterize and compare fish assemblages in the two areas, we discuss the possibility to conveniently model the effects of different kinds of environmental factors such as depth, current and habitat. We provide a unified modeling framework for a multivariate zero-inflated response, explicitly including spatial and seasonal components. Finally, we discuss the advantages attained by the proposed approach in describing fish assemblages and quantifying the uncertainty affecting ecological community data.

Modelling and Decomposing Vital Rates: a Non-parametric Approach

C. Camarda[(a)]
[(a)]Institut national d'études démographiques

Demographic events have characteristic age-specific patterns of occurrence. Finding model schedules to summarize the age-pattern of demographic rates has a long tradition, however, parametric models are predominantly used. Many demographic rates shows complex shape in their overall age-pattern. However such pattern can be attributed to different distinct components. While some of the components can be described by a parametric model, such as the Gompertz hazard for adult mortality, many others cannot. Additional complication arises if data are provided in age groups, which is still the case in many official statistics, historical and cause-specific data. In this talk we propose a model that allows to specify rates across a wide range of ages as the sum of several components, which are modelled on the log scale

and are assumed to be smooth, but do not have to follow a particular parametric form. A penalized composite link model is used to decompose complex trajectories into smooth additive components. Parametric and non-parametric forms can be used. Data can be given in age-groups which can eventually be of variable lengths. Furthermore, monotonicity or shape constraints on the components can be incorporated and the model can cope with two-dimensional settings in which age-patterns change over time.

Modeling Uncertainty for Wind Energy Resources in Saudi Arabia

S. Castruccio[(a)], F. Tagle[(a)], P. Crippa[(a)], M.G. Genton[(b)]
[(a)]Newcastle University, Newcastle, [(b)]King Abdullah University of Science and Technology, Thuwal

Facing increasing domestic energy consumption from population growth and industrialization, Saudi Arabia is aiming to reduce its reliance on fossil-fuels and broaden its energy mix by expanding its investment in renewable energy sources, including wind energy. A preliminary task in the development of wind energy infrastructure is the assessment of the wind energy potential, a key aspect of which is the characterization of its spatio-temporal behavior. In this study we examine the impact of internal climate variability on seasonal wind power density variability using 30 simulations from the Community Earth System Model (CESM) Large Ensemble (LENS) project developed at the National Center for Atmospheric Research. Furthermore, a spatio-temporal model for daily wind speed is proposed, with neighbor-dependent cross-temporal dependence and a multivariate skew-t distribution to capture the spatial patterns of higher order moments. The model can be used to generate synthetic time series over the entire spatial domain that adequately reproduces the internal variability in the LENS dataset, thus allowing for fast and relivable assessment of the uncertainty due to internal variability in wind power density.

Estimating Dispersal Kernels Using Citizen Science Data : the Montagu's Harrier Example

J. Chadoeuf[(a)], A. Millon[(b)], V. Lecoustre[(c)], V. Bretagnolle[(c)]
[(a)]INRA, [(b)]IMBE, Aix-Marseille University, Marseille, [(c)]CEBC, CNRS

To be resilient, bird populations must be large enough so as to exchange genetic material and compensate for local population downfalls. Knowing how animals disperse is thus necessary to value population status. When spatial population distribution is large, its survey becomes difficult and citizen science participation becomes helpful. However, it also leads to a spatially heteogeneous and generally not measurable observation effort therefore impairing data interpretation. We show with the example of a Montagu's harrier population survey at the national scale in France, where young birds were marked at the nest and subsequently reported by

citizen science network (see www.busards.com) how to overcome the presence of two heterogeneities when estimating a dispersal kernel: a known spatial heterogeneity of departure points and an unknown spatial heterogeneity of arrival points.

Local Variable Selection Under a Misspecified Spatial Regression Model

C. Chen[(a)]

[(a)]National Changhua University of Education, Changhua

Variable selection and spatial prediction are important issues in spatial statistics. If spatially varying means exist among different subareas, globally fitting a spatial regression model for the study area may be not suitable. To alleviate deviations from model assumptions, we propose a local variable selection criterion to locally select variables for each subarea, where the global spatial dependence of observations is considered and the local characteristics of each subarea are also identified. It results in a composite spatial predictor which provides a more accurate spatial prediction and the corresponding prediction variance is also smaller. Statistical inferences of the proposed method are justified both theoretically and numerically.

Cautionary Tales On Air Quality Improvement in Beijing

S.X. Chen[(a)], S. Zhang[(b)], B. Guo[(c)]

[(a)]Peking University, Beijing; Iowa State University, Ames, [(b)]Peking University, Beijing, [(c)]Southwestern University of Finance and Economics, Chengdu

There has been an encouraging statistic from Beijing showing an annual 9.9% decline in the concentration of PM2.5 in the city in Year 2016 as compared with that of the previous year. While this statistic offered some relief for the inhabitants of the capital experiencing an episode of severe air pollution a few days into the 2017 New Year, we present several analyzes on Beijing's PM2.5 data of past four years along with meteorological data of past seven years. The analyzes reveal non-significant improvement in the fine particulate matter pollution in 2016 as compared to the previous year, and shed more light on aspects of the city's PM2.5 situation in 2016.

Longitudinal Modelling of Crop Root Physiology As a Breed-specific Spatial Response To Environmental Conditions

G.S. Chiu[(a)], A.P. Wasson[(b)], A.B. Zwart[(b)]
[(a)]Australian National University, Canberra, [(b)]CSIRO

To ensure future food security, a key objective of crop breeding programs is to effectively identify which genetic and physiological characteristics of the plant are associated with high yield and/or resistance to environmental stressors. Regarding physiology, the part of the plant that is above-ground is easily observed and thus commonly emphasized. However, the root system is perceivably more sensitive to soil-related stressors yet notoriously challenging to (a) measure and (b) characterize. For (a), recent imaging technology can evaluate the number of roots at regular depths along soil cores that are sampled from the crop field. This method results in 1-dimensional spatial data on within-core root counts. For (b), we develop a modelling framework that regards the spatial count data as longitudinal in nature, exhibiting a parametric trend that depends on the plant's genotype (or "breed"). Under our framework, we define new measures of heritability — the variability among cores that is due to genetics as opposed to noise. The novelty of our methodology lies in the ability to reflect root architecture as a whole by accounting for within-core root counts collectively. Applied to a field study in Australia, our approach indicates an overall heritability of 0.52-0.71 (95% credible interval), which is substantially higher than previous methods. This suggests that our approach is much more effective in discerning root architecture as captured by soil core data.

Identifying the Best Fishing-suitable Areas Using Bayesian Spatio-temporal Models

D. Conesa[(a)]
[(a)]Universitat de València, València

Methodological approaches for modelling fish species occurrence patterns is a very important field of research. Its findings allow researchers to provide management directives. To do so, it is required the understanding of marine biological processes at a spatial scale, but also quantifying the ecosystemic importance of an area. In this talk, we discuss how the use of Bayesian hierarchical models can be very helpful in this context. The basic model of the presence/absence of species expressed in terms of geographical and environmental characteristics can be improved in many aspects. The fist one is to take into account that fish distribution can vary with time: biological spatial processes not only tend to be spatially correlated but also to evolve in time. Consequently, the inclusion of a temporal structure in the model not only improves the spatial interpolation but also allows us to predict forward in time at different locations. Moreover, this approach can be adapted to be used in situations where sampling is likely to be preferential (as it happens usually in fishery). Another important factor when modelling fish species distribution is the

semi-continuous nature of the response variable. Fish abundance is commonly measured in continuous Catch Per Unit Effort (CPUE) values where zero observations are very common at adverse environmental conditions. We will also discuss how to implement this via hurdle models that could share components. Finally, we will also briefly discuss about other issues like misaligment and how to deal with physical barriers.

A Numerical and Experimental Investigation of Uncertainties in Air Temperature Measurements, in the Framework of MeteoMet Project

G. Coppa[(a)], A. Merlone[(a)], C. Musacchio[(b)], A. Quarello[(c)], L. Massano[(c)], T. Vukicevic[(d)]

[(a)]Istituto Nazionale di Ricerca Metrologica, [(b)]Società Meteorologica Italiana, [(c)]Università di Torino, Torino, [(d)]Zavod za metrologiju Crne Gore, Podgorica

The EURAMET project JRP ENV 58 MeteoMet 2 is aimed at the evaluation of the overall measurement uncertainties for Essential Climate Variables (ECVs). The MeteoMet2 project is structured in 3 work packages, each covering a different area of observation: air, sea and land. Assessing the measurement uncertainties is a step forward with respect to the calibration uncertainties: the latter are confined to the controlled environment of a laboratory, but in order to retrieve the former, ad-hoc in-field experiments have to be devised. In this context, several experiments have been designed, installed, run and managed in order to evaluate measurement uncertainties for ground-based air temperature sensors in a wide range of environmental conditions and for different quantities of influence. Here an investigation on these experiments is presented.

Regularization for Grouped Dirichlet-multinomial Regression in the Analysis of Ecological Networks

C. Crea[(a)]

[(a)]University of Guelph, Guelph

Field studies in community ecology have been expanding to include collection of detailed trait information in order to gain a deeper understanding of the functions and services provided by ecosystems. Several studies suggest that the observed counts (interactions) in mutualistic networks, such as plant-pollinator or seeder-forager networks, are driven or facilitated by linkage rules defined by these species traits. Grouped Dirichlet-multinomial (DM) regression models have been used to analyze such mutlinomial count data using a trait-based approach. These models are flexible in that they can accommodate diverse types of overdispersion in the count data, including scenarios in which dispersion is a function of species traits and/or linkage rules. However, in practice the specific species traits/linkage rules relevant to encouraging interactions or contributing to overdispersion in a given ecosystem

are typically unknown and must be learned from the collected trait data. Further, standard forward/backward variable selection strategies can be time consuming and computationally expensive. In this presentation, we consider the issue of variable selection for grouped DM models in the presence of overdispersion. We develop regularization methods for the grouped DM model in which we optimize a penalized likelihood function using lasso-type penalties. In particular, we derive a proximal gradient algorithm to estimate the model parameters and conduct variable selection simultaneously for the standard lasso and adaptive lasso. Through a simulation study and empirical analyses, we compare the regularized regressions for the grouped DM model and discuss their merits and limitations.

A Directional Look On Fires

R.M. Crujeiras[(a)], J. Ameijeiras-Alonso[(a)], A. Rodríguez-Casal[(a)]
[(a)]Universidade de Santiago de Compostela, Santiago de Compostela

Forest fires have become an important environmental problem in southern european countries such a Spain, Portugal, Italy or Greece, not only as natural hazards but also as a result of human activities.The analysis of forest fires in these areas is crucial for understanding their evolution and behavior, in terms of fire pattern (both spatial and temporal) identification. Apart from considering the spatial nature of processes driving fires occurrence, some previous works have also dealt with a directional components, on the specification of wildfires orientation patterns and on the appearance of fire seasons. This presentation will be mainly focused on this part, providing a review on previous works on this area, which share a directional look on fires.

Impact of Logging On the Resilience of SE Asian Forests: Remote Sensing Two Decades of Forest Change

M. Cutler[(a)], D. Boyd[(b)], G. Foody[(b)], D. Burslem[(c)]
[(a)]University of Dundee, Dundee, [(b)]University of Nottingham, Nottingham, [(c)]University of Aberdeen, Aberdeen

In 1996, 52 forest plots were established in a 15 km^2 area of tropical forest surrounding the Danum Valley Field Centre in NE Borneo, Malaysia. The forest is a mixture of primary and disturbed (by selective logging) lowland Dipterocarp forest. The plots were distributed in coupes that had been selectively logged in different years, as well as in intact tropical forest reserves. In essence, the plots provided a gradient of forest disturbance, from pristine forest to high levels of relatively recent disturbance. In each 0.1ha forest plot biomass, species etc were recorded and formed the basis for validating estimates of biomass and tree species diversity using satellite imagery. This papers presents, for the first time, the results of a recensus of the 52 forest plots some 20 years after the original survey, allowing a longer time interval to assess the resilience of these fragile ecosystems. Initial analysis of the in

situ data suggest that in forest areas that were not logged there has been a slight increase in biomass of approximately 1 Mg ha-1. However, in the selectively logged plots increases in biomass were extremely high in comparison with other reported figures from logged tropical forests (up to 12 Mg ha-1). Critically though, there appears to be significant variation depending upon the allometric equation used to estimate biomass and uncertainty resulting from recent advances in species identification that were unavailable at the time of the first survey. From time-series analysis of medium spatial resolution remotely sensed data we have tracked changes in forest biomass over the 20 year period and suggest that the uncertainties arising from remote sensing-based estimates of forest biomass are less than those arising from the in situ data, thus demonstrating the potential of remote sensing-based approaches for estimating changes in forest properties over time.

Air Pollution and Reproductive Behavior in Rome

A. De Rose[(a)], G. Cesaroni[(b)]
[(a)]Sapienza University of Rome, Roma, [(b)]S.S.R. Lazio

Aim of this paper is to investigate the effects of environmental conditions on human reproductive behavior in a large urban context, that is the metropolitan area of Rome. The biological effect of pollution on fecundity and reproductive health of women and men in the industrialized countries is a relatively recent field of research. The experimental, clinical and epidemiological studies indicate that the production of sperm and ovocytes, the process of fertilization and intrauterine and postnatal development of the offspring may be particularly vulnerable targets. Although a significant impact on population level fertility is still to be proved, the worrying about a possible effect of the worsening environmental conditions on reproductive health is hard to ignore. Some association has been already found at local level between air pollution and general fertility rates, but no attempt has been done so far to single out the effect of the physical environment conditions on reproductive experience at the individual level. With this study, we wish to contribute to the research by performing a survival analysis on the length of the first birth interval of married couples living in Rome and observed between 2001 and 2014 in function of the level of pollution at residence. Our hypothesis is that time to conception increases with increasing level of air pollution and controlling for a number of individual, family and contextual factors. Data come from the Municipal Register databases, we selected 34,712 couples living in Rome at May 2008, married between 2001 and 2008, with women aged 15-44, not pregnant and childless at marriage. We assessed exposure to PM10 at residential address using a dispersion model (1 Km grid) for particulate matter. We followed the couples from their marriage to the birth of the first child, or to December 2014.

Conditional Simulation of Gaussian Spatio-temporal Random Field With Specified Covariance, Based On the SPDE Approach.

N. Desassis[(a)], R. Carrizo[(a)], C. Lantuejoul[(a)]

[(a)]Geosciences Center MINES ParisTech, Fontainebleau

Conditional simulations are realizations of a stochastic model that honor a set of observations. They can be used to predict and quantify the uncertainty associated with complicated functions of a space-time random field. Standard approaches like kriging may not be really appropriate insofar as the set of observation can be very large. In such a case, a Stochastic Partial Differential Equation approach (SPDE) is proposed. The SPDE's considered here depends on a functional parameter, called a symbol. The right hand side member is a white noise. When a stationary solution exists, its covariance function can be numerically computed. By comparing it to the empirical covariance function of the observations, the symbol can be estimated.For some differential operators, the solution can be approximated by a Markovianrandom field defined on a meshing of the spatio-temporal domain. Its precision matrix for all the points of the spatio-temporal meshing is obtained, for instance by using finite elements methods (FEM) or a combination of FEM and time finite differences method. This matrix is sparse and kriging or conditional simulations can be performed by usingthe Cholesky decomposition of the submatrix obtained by deleting the rows and columns corresponding to the data.When the meshing size is very high, some iterative methods have to be used instead. For instance, the kriging can be obtained by using block Gauss-Seidel algorithmor a conjugate gradient algorithm with a suitable preconditioner.To illustate this workflow, an application with large data set is treated. The target variable is the daily mean PM10 collected at monitoring stations across France in 2014.

URBANSCOPE: Functional Biclustering of Milan Mobility Patterns

J. Di Iorio[(a)], S. Vantini[(a)], P. Secchi[(a)], M. Arnaboldi[(a)], P. Ciuccarelli[(a)], M. Brambilla[(a)]

[(a)]Politecnico di Milano, Milano

Urbanscope is a project by Politecnico di Milano which aims at creating a macroscope of Milan able to describe how the city is lived by using big data in order to figure out the digital urban dimensions that human sight cannot perceive. Considering that understanding the digital layer of Milan can help improving the real one, Urbanscope is trying to use multidisciplinary approaches to find out new views of the city. In terms of how people move around Milan, mobile phone data, correctly treated, could lead to interesting discoveries. In our case Biclustering methods, here presented from the general multivariate procedures to first functional attempts, make the difference in order to highlight brand new dynamics in the mobility pattern of the city.

Nonparametric Discrimination for Directional Data

M. Di Marzio[(a)], S. Fensore[(a)], A. Panzera[(b)], C.C. Taylor[(c)]
[(a)]Università G. d'Annunzio Chieti-Pescara, Chieti, [(b)]Universita di Firenze, Firenze,
[(c)]University of Leeds, Leeds

We propose nonparametric methods to address the problem of classification for directional data. We focus on local regression and kernel density estimation when the domain is the d-dimensional unit sphere. We provide asymptotic theory for the proposed methods along with simulation results.

Missing Data in Space-time: Long Gaps Imputation Based On Functional Data Analysis

F. Di Salvo[(a)], A. Plaia[(a)], M. Ruggieri[(a)]
[(a)]University of Palermo, Palermo

High dimensional data with spatio-temporal structures are of great interest in many fields of research, but their exhibited complexity leads to practical issues when formulating statistical models. Functional data analysis through smoothing methods is a proper framework for incorporating space-time structures: extending the basic methodology to the multivariate spatio-temporal setting, we refer to Generalized Additive Models for estimating functional data taking the spatial and temporal dependences into account, and to Functional Principal Component Analysis as a classical dimension reduction technique to cope with the high dimensionality and with the number of estimated effects. Since spatial and temporal dependences integrate information of different types and from different sources, this framework serves as synthesis of information and give important opportunities for data processing and analysis, including extremely effective dimension reduction and estimation of missing values. The idea behind is to work with an estimated variance function, represented in terms of the bases and parameters defined in the estimation process, by mean of which the variability is espressed in terms of the main temporal and spatial effects; the functional principal component analysis provides dimensions reduction, determining the uncorrelated linear combinations of the original variables that account for most of the variability expressed by the variance function. The eigenfunctions, or principal component functions, also represent an orthonormal functions set, which can be used to fill gaps in incomplete data: we explore the performance of imputation procedures based on Functional Data Analysis and Empirical Orthogonal Function approaches when missing values, and mainly long gaps, are present in the original data set. In order to compare and validate the proposed procedures, a simulation plan is carried out and some performance indicators are computed under different missing value patterns and in presence of long gaps.

Global Measures of Entropy and Similarity for Genetic Sequences Data

F. Divino[(a)], P. Monaco[(a)], A. Bucci[(a)], G. Naclerio[(a)]
[(a)]University of Molise, Campobasso

Investigations of microbial communities and their composition connected to changes in environmental conditions are of significant interest to control the environmental status and its evolution. Recently, within microbial ecology, the use of molecular techniques based on extraction and analysis of nucleic acids, proteins and lipids from environmental samples has replaced the traditional approaches based on isolation and cultivation in laboratory, allowing to acquire structural and functional information on microbial communities related to environments of interest. This type of techniques is based on the analysis of specific marker-genes and allows to obtain classification at any taxonomic level: domain, kingdom, phylum, class, order, family, genus, species. Denaturing Gradient Gel Electrophoresis (DGGE) and Next Generation Sequencing (NGS) are examples of molecular techniques to assess the composition of microbial communities and measure the microbial biodiversity. A relevant issue in this type of data is represented by the potential lack of consistence when the biodiversity assessment is computed at different taxonomic levels. In fact, when shifting on the taxonomic hierarchy, from domain to species, the respective compositions, although from the same genetic sequence, can produce different values of entropy and similarity indexes. This is due to the trade off between taxonomic information and statistical information across the taxonomic hierarchy: the biodiversity richness increases while the abundance of recognized microbial individuals decreases. In this work we propose global measures of biodiversity in terms of new measures of entropy and similarity, accounting for the typical features of abundance, richness and evenness, across the taxonomic levels observable, in order to summarize the assessment of the biodiversity in environmental samples of interest when data are from genetic microbial sequences. Simulation examples and results from the analysis of NGS data are presented to show the statistical behavior of the new biodiversity indexes.

Spatio-temporal Modeling for Predicting Wind Speed and Direction on the Coast of Valparaiso

M. Díaz[(a)], O. Nicolis[(a)], O. Cuevas[(a)]
[(a)]University of Valparaiso, Valparaiso

Weather forecasting is one of the most important challenges in the world. Despite evident improvements in the last years due to the use of complex numerical models, predicting the wind speed and direction involves some errors especially in the sea where there are no monitoring stations. The main aim of this research is to predict the wind speed and direction by using a spatio-temporal model when the observed data are not available. For estimating the model, we use the outputs of the Weather Research and Forecasting (WRF) model: the FNL (Final Operational Global Analysis) prepared from the Global Data Assimilation System (GDAS) as the response

variable and the GFS (Global Forecast System) as covariate together with a set of smooth temporal basis functions. Since the FNL predictions should be more accurate than the GFS outputs, the idea is to calibrate the GFS simulations using the FNL data. The model is applied to the WRF data simulated between July 15th and 31st, 2015 every one hour with spatial resolution of 1km x 1km on the coast of Valparaiso, Chile. Indicators and maps are used for evaluating the performance of the model.

Bivariate Negative Binomial Distribution for Modelling Correlated Count Data

A.H. El-Shaarawi[(a)]

[(a)]Cairo University, Cairo and Canadian National Water Research Institute

I will discuss the construction of bivariate discrete distributions and their use in the study of count environmental data. The application is concerned with a study to revise the guidelines for bacterial contamination from agricultural sources. The observed variability required accounting for over-dispersion and trend in multivariate count data. The starting point was the derivation of the bivariate Poisson distribution to its generalization to the bivariate negative binomial distribution. An extension of Fisher's Index of dispersion to the multivariate case is also considered. Keywords: Count Data, Water Quality, Correlations; Regulations

---

Testing Stationarity of Functional Time Series

A. Elayouty[(a,b)], M. Scott[(b)], C. Miller[(b)]

[(a)]Cairo University, Cairo, [(b)]University of Glasgow, Glasgow[(b)]

With the emergence of new technologies in data acquisition and environmental sensors, enabling the continuous recording of measurements during a time interval or intermittently at several discrete time points, functional time series analysis has increasingly become an appropriate tool for analysing such high-dimensional temporally correlated data. Motivated by a highly dynamic hydrological high-frequency time series, we propose a novel smooth Functional Principal Component Analysis (FPCA) technique for such non-stationary functional data based on a frequency domain approach. These novel smooth FPCs adapt to the changes in the variability structure of functional data over time, and help to investigate whether the variability and the auto-covariance structure change over time. To detect significant changes in the covariance and variability of functional time series, a bootstrap inference procedure based on these time-varying frequency-domain FPCs is proposed. This inferential procedure can typically be used to check the assumption of covariance-stationarity needed for most functional time series methods. The construction and performance of these time-varying FPCs in the frequency domain as well as the proposed bootstrapping inference procedure are presented and illustrated in this presentation with the aid of an empirical example concerning the excess partial pressure of carbon dioxide (EpCO2) in the River Dee in Scotland.

Regional Trends in Fire Danger Based On Historical Records

S.R. Esterby[(a)], Z. Hübnerová[(b)], J. Holešovský[(b)], S. Taylor[(c)]

[(a)]University of British Columbia Okanagan, Kelowna, [(b)]Brno University of Technology, Brno, [(c)]Canadian Forest Service, Pacific Forestry Centre, Victoria

Studies have shown that the probability of fire ignition can be estimated from fire danger indices. Day and location within a region account for much of the variability and the predictive indicator can differ among regions. Calculations in the Canadian Fire Weather Index (FWI) system are based on consecutive daily weather measurements accumulated into fuel moisture codes and fire behavior indices. The FWI system guides temporal and spatial allocation of fire management resources in Canada. Fire danger classes, used operationally with informal knowledge and experience, are based wholly or in part on ranges of FWI. Thus, the study of the FWI extremes should provide insight into temporal trends in fire danger, relevant under the scenario of climate change. Rising temperature and changes in precipitation patterns are expected to result in higher fire risk, longer periods of high risk, and altered timing of high risk within the fire season, all dependent upon the climate region. Indicators of risk, soil moisture content, and elements of fire activity from the FWI system for about 150 fire stations in British Columbia with records up to 36 years (1970 to 2005) comprise the data for this study. In modelling FWI extremes, two types of non-stationarity are considered. Non-stationarity over consecutive years is modelled by time-dependent scale and shape parameters of the generalized extreme value distribution. Parameters are estimated by maximum likelihood methods. Non-stationarity over the fire season is taken into account by modelling trends for monthly maxima separately. Mapping of p-values of trend tests provides an initial spatial summary and shows differences between regions. Spatial methods for extremes are used to characterize temporal trends within climate regions and make comparisons among regions. Differences in the number of days of elevated forest fire danger are also summarized.

Monitoring Sendai Framework for Disaster Risk Reduction and SDGs: Indicators To Analyse Interactions Among Human and Natural Systems

A. Ferruzza[(a)], G. Tagliacozzo[(a)]

[(a)]Istat, Rome

The accelerating path of Climate Change (CC) is having consequences in terms of impacts and adaptation and represents one of the challenges facing humanity. Sendai Framework for Disater Risk Reduction has been implemented in 2015 and the Open-ended Intergovernmental Expert Working Group on Indicators and Terminology Relating to Disaster Risk produced a list of indicators to monitor 7 goals. The UNECE Task Force Measuring Extreme Events and Disasters (MEED) has been created to clarify the role of official statistics in providing data related to EED,

identify steps how National Statistical Offices (NSOs) in coordination with national agencies responsible for disaster management can support disaster management and risk reduction in monitoring Sendai Framework Process. Sustainable Development Goals (SDGs) has been determined in September 2015 and the UN OEWG defined more than 200 indicators to monitor the 17 goals and the 169 targets. SDGs include issues on climate change and disaster preparedness and risk reduction. The demand for high quality statistics is increasing to develop suitable indicators for monitoring SDGs and Sendai Framework. NSOs are called for actions aimed to improve the usefulness of existing statistics and to develop new statistics, filling information gaps. Measuring and enhancing the relationship between Human and Natural System is complex but statistical innovations could be addressed to enhance the availability of new relevant indicators necessary for the analysis. This work presents and analyses the interaction among the different frameworks and highlights the need to strengthen collection and use of several data sources by integrating data from surveys with administrative data and new sources such as Big Data. To this aim, cooperation and awareness of all agencies and institutions operating in the National Statistical System are essential to exploit the existing information assets and to develop new statistics coherent and consistent with the statistical international frameworks.

Pairwise Association Between Parts of a Composition

P. Filzmoser[(a)], P. Kynclova[(b)], K. Hron[(c)], C. Reimann[(d)]
[(a)]Vienna University of Technology, [(b)]UNIDO, [(c)]Palacky UNiversity Olomouc, [(d)]Geological Survey of Norway

Correlation analysis belongs to the most widely used tools for analyzing data. Standard correlation analysis is not recommended for compositional data (e.g. concentrations of chemical elements in the soil), because their sample space differs from the standard Euclidean space for which this tool has been developed. The resulting spurious correlations do not reflect the real underlying associations between the parts of the composition. There are several attempts to provide alternative association measures that are meaningful for compositional data. Here we propose an approach where coordinates of a Euclidean space are constructed in a way such that all the relative information of the two parts of interest in symmetrically treated by two coordinates. Any standard correlation measure is then meaningful, and in contrast to other proposal one gets again correlations in the interval [-1,1]. Based on real environmental data, we will demonstrate that already graphical inspection of the two coordinates provides interesting insights into the data: trends, groups, and associations can be revealed, and it can be instructive to link such findings to absolute information (e.g. the raw element concentrations). In the context of geochemistry it turns out that correlations based on these symmetric coordinates serve as a sensitive tool to reveal underlying geochemical processes.

Uncertainty Assessment of Co-located Radiosonde and Remote Sensing Profiles By Harmonization

F. Finazzi[(a)], A. Fassò[(a)], F. Madonna[(b)]

[(a)]University of Bergamo, Bergamo, [(b)]Consiglio Nazionale delle Ricerche

In this work we propose a statistical approach for assessing the co-location uncertainty of radiosonde and remote sensing vertical profiles of essential climate variables. Profiles are co-located in the sense that they are paired on the basis of their mutual distance in space and time. Differences between the profiles are due, on the one hand, to the co-location mismatch, and on the other, to the different vertical resolution of the two measuring instruments. In particular, remote sensing profiles have a much lower resolution than radiosonde profiles. In order to make the profiles comparable, we first harmonize the profiles by smoothing the radiosonde profiles. This is done using smoothing functions that mimic the remote sensing retrieval process. The smoothing functions are characterized by location, scale and shape parameters which are estimated in order to minimize the mean squared error of the difference between radionsonde and remote sensing observations.Considering harmonized co-located profiles, an isotonic regression is implemented in order to assess the co-location uncertainty as a function of both the temporal and the spatial mismatch.

Functional Data Analysis for Energy and Behavioural Analytics

M. Fontana[(a)]

[(a)]Politecnico di Milano, Milano

The energy sector is being quickly revolutionized by the availability of data that are "Big", not only in terms of size, but also in terms of complexity. The widespread diffusion of smart meters, able to provide high frequency data, does allow improvement in demand side management and demand response initiative for both consumers and firms. So far, however, this kind of data has not been yet treated and used at its full potential, given the lack of proper statistical methods able to tackle the size and complexity arising from this type of statistical problems. This will help policy makers by providing them data-driven answers coming from the analysis of complex datasets. In this talk, we will provide an application of Functional on Scalar linear models and Interval-Wise Testing procedures to the analysis of high frequency energy consumption data, coming from a field experiment performed to evaluate the impact of a Demand Response (DR) project developed in the city of Isernia, in southern Italy.

Irradiance Levels and Hourly Temperature: Evidence From the Arctic

K.F. Forbes[a]

[a]The Catholic University of America, Washington

This paper investigates the possible drivers of an upward trend in temperature at Barrow Alaska using a time series model. The model makes use of hourly in situ temperature data, hourly downward total solar irradiance data, hourly upward solar irradiance data, hourly downward longwave irradiance data, and hourly upward longwave irradiance data. The model also employs binary variables for season and the hour of the day. The model is estimated over the period 1 January 2008 through 31 December 2012. There are 25,523 hourly observations in the sample. The model is evaluated using hourly data over the period 1 January 2013 – 31 December 2015. In terms of methodology, a two-step time-series estimation approach is employed. In the first step, the presumption of linearity in terms of the explanatory variables (e.g. the hourly level of downward total solar irradiance) is scrutinized. Based on this analysis, a number of the explanatory variables are entered into the model with nonlinear specifications. The purpose of the second step in the estimation is to obtain parameter estimates that give rise to a white noise error structure that also has the property of asymptotic normality. Step two of the estimation is accomplished by making use of an ARCH/ARMAX model in which the transformed irradiance variables and the seasonality variables from the first step are included as the exogenous inputs. The equation is estimated under the assumption that the standardized residuals correspond to the Student t distribution which accommodates more kurtosis than the Gaussian distribution. The estimation yields a statistically significant relationship between the irradiance levels and hourly temperature. Consistent with the Anthropogenic Global Warming hypothesis, the results indicate that the upward trend in hourly temperature at Barrow Alaska can be accounted for by changes in the levels of longwave irradiance at that same location.

Grouping in the Presence of Artefacts: Model-based Approaches To Avoid Systematic Noise

S.D. Foster[a], N. Hill[b], M. Lyons[c]

[a]The Commonwealth Scientific and Industrial Research Organisation, Hobart, [b]The University of Tasmania, Hobart, [c]University of New South Wales, Sydney

Grouping sites on the basis of their biological content is a frequent goal that has direct scientific and management applications. A useful grouping/clustering analysis identifies the important patterns of variation whilst simultaneously reducing the inherent complexity. Unfortunately, data that are used to find the groupings often have extraneous sources of variation, such as those related to sampling, which are often ignored but should be accounted for when finding the groupings. In ecological studies, this is increasingly common as data sets are now being combined from many smaller survey efforts. We show, through a model-based clustering method, how the

groupings can be obtained, while accounting for these extraneous sources of variation (artefacts). We show, through simulation and application, how the method performs and how it relates to established methods.

Multivariate Conditional Autroregressive Models Within R-INLA

A. Freni Sterrantino[(a)], H. Rue[(b)]

[(a)]Imperial College, London, [(b)]King Abdullah University of Science and Technology, Thuwal

Disease mapping techniques are widely applied in environmental epidemiology to describe the spatial distribution of standardised mortality/morbidity ratios; the most popular model-based approach for disease mapping is framed on a hierarchical Bayesian perspective. Typically the analyses consider one condition (mortality/hospitalisation for a specific cause) and the spatial correlation is accounted for through combinations of a conditional autoregressive structure (CAR) and a standard random effect. There are several programs able to fit this class of models; recently R-INLA is becoming extremely popular as it can easily handle big data in a computationally efficient way. Much research in the last decade was spent in developing sound parametrization on disease mapping for multiple conditions, relying mainly on shared component models or multivariate conditional autoregressive models (MCAR). From a statistical point of view, the Bayesian models are based on the specification of prior distributions that should summarise prior beliefs on the parameters we want to estimate. The main difficulties for the end-user are to understand and interpret correctly the priors and to assign reasonable values to these. Therefore, the MCAR development will be based on the definition of PC-priors for multivariate setting together with the new parametrization. This will allow to (i) correctly specify adjacency matrices also for study regions with island and properly smooth the standardized mortality ratios - this feature is not present in any other software; (ii) allow the user to intuitively select appropriate values for priors to be informative (even weakly) and (iii) provide several possible MCAR specifications that are robustly implemented in R-INLA, also in presence of big data.

Shape Modelling and Clustering of Marine Mammal Dive Profiles

E. Fu[(a)], N. Heckman[(a)]
[(a)]University of British Columbia

Understanding the diving behaviour of marine mammals is important in the study of individual's foraging decisions and efficiencies and species' interaction with the environment. While direct observation of underwater activities is difficult, dive depths are easily recorded via tagging devices, yielding dive profiles, i.e. 2-D plots of depth over time. The profile shapes can then be used to classify dives to infer behavioural states for subsequent analysis. Current approaches to dive type classification mostly rely on multivariate clustering methods applied to selected features of the dive profile. Few approaches model the entire dive profile directly. We directly model the entire dive profile via a functional data approach, viewing the dive profiles as variations of base curves. Even within the same type of dive, the profiles differ in the vertical direction of depth as well as via a lateral warping in shape. To reflect these two modes of variation, we model a dive profile as a composition of an amplitude function and a warping function. Estimation of these two functions is challenging because of the nonlinearity induced by the warping function. We overcome this challenge by using a stochastic approximation version of the EM algorithm. We apply the proposed method to the analysis of elephant seal dive profiles. The fitted warping functions are then used to cluster dive profiles. We compare clustering via warping functions to existing methods which do not take lateral variation directly into account. The warping function approach is found to produce clusters with more homogeneous shape within group.

Brute Force or Hand-waiving: Spatial Interpolation Approaches At Your Service

R. Furrer[(a)]
[(a)]University of Zurich, Zurich

The need for computationally efficient and easy-to-use methodologies to interpolate huge spatial datasets is undisputed. However, classical kriging software in R cannot handle objects larger than 2^{31} elements and thus there is an inherent limit of the size of covariance matrices that can be handled - independent of RAM or time restrictions. On the other hand, ignoring the inherent dependency structures of spatial data results in incorrect uncertainty estimates and may lead to wrong conclusions. In this talk we show how to "free" your software from these limitations such that it is again possible to worry about algorithmic speed. The R package "dotCall64" provides a convenient foreign function interface to call compiled C/C++ and Fortran code. It transparently supports long vectors and features a mechanism to avoid unnecessary copies of function arguments, making it efficient in terms of speed and memory usage. In contrast, we illustrate a fundamentally different "prediction" concept, called gap-filling in our context. That approach is tailored to parallel processing schemes of potentially large computers and relies on spatial or spatio-temporal

neighborhoods but not on the specification of dependency structures. We show that the proposed prediction method is very efficient. However, the computational gain of not specifying the covariance matrix is offset by the enumeration requirements of the algorithms. The procedure is implemented in the R package "gapfill". This is joint work with Florian Gerber.

Measuring the Air Pollution Severity in North China Plain

B. Guo[(a)]

[(a)]Southwestern University of Finance and Economics

In this talk, we focus on the air pollution severity in North China Plain (Beijing, Tianjin and Hebei). To remove the influence of the weather condition, we use the weather adjustment for fair comparison. The adjusted means and quantiles for the six pollutants (PM2.5, PM10, SO2, NO2, O3 and CO) are evaluated to measure the severity.

Climate Adaptation and Risk Management in Insurance

Y.R. Gel[(a)]

[(a)]University of Texas, Dallas

The last few years were characterized by a record number of insurance claims due to severe weather. According to the World Bank report, annual average losses from natural disasters have increased from 50 billion in the 1980s to about 200 billion nowadays. While it is not clear whether the causes of such damages are due to global warming, outdated building standards, non-sustainable city infrastructure and planning, or combination of these factors, the problem becomes increasingly acute and virtually concerns every citizen. Development of efficient and reliable management strategies for climate-induced risk mitigation requires early recognition of vulnerable areas and the extent of the future risk due to environmental factors. Despite the well documented findings on the broad spectrum of weather- and climate-induced risks in the insurance sector, most studies focus pre-dominantly on disaster and catastrophes assessment, and there still exists a relatively limited number of studies, addressing the effect of the so-called "normal" extreme weather on the insurance dynamics. In this talk we discuss utility and limitations of modern statistical and machine learning procedures to quantify and predict weather-related risks in residential insurance. This is a joint work with Vyacheslav Lyubchich, Hali Killbourne and Asim Dey.

Statistics-Based Compression of Global Wind Fields

M.G. Genton[(a)], J. Jeong[(a)], S. Castruccio[(a)], P. Crippa[(a)]
[(a)]King Abdullah University of Science and Technology, Thuwal

Wind has the potential to make a significant contribution to future energy resources; however, the task of locating the sources of this renewable energy on a global scale with climate models, along with the associated uncertainty, is hampered by the storage challenges associated with the extremely large amounts of computer output. Various data compression techniques can be used to mitigate this problem, but traditional algorithms deliver relatively small compression rates by focusing on individual simulations. We propose a statistical model that aims at reproducing the data-generating mechanism of an ensemble of runs by providing a stochastic approximation of global annual wind data and compressing all the scientific information in the estimated statistical parameters. We introduce an evolutionary spectrum approach with spatially varying parameters based on large-scale geographical descriptors such as altitude to better account for different regimes across the Earth's orography. We consider a multi-step conditional likelihood approach to estimate the parameters that explicitly accounts for nonstationary features while also balancing memory storage and distributed computation, and we apply the proposed model to more than 18 million points on yearly global wind speed. The proposed model achieves compression rates that are orders of magnitude higher than those achieved by traditional algorithms on yearly-averaged variables, and once the statistical model is fitted, decompressed runs can be almost instantaneously generated to better assess wind speed uncertainty due to internal variability.

Object-based Classification of Grassland Management Practices From High Resolution Satellite Image Time Series With Gaussian Mean Map Kernels

S. Girard[(a)], M. Lopes[(a)], M. Fauvel[(a)], D. Sheeren[(a)]
[(a)]Inria Grenoble Rhone-Alpes

Grasslands are an important source of biodiversity in farmed landscapes. Agricultural management of grasslands (mowing, grazing...) is essential to maintain their biodiversity. However, an intensive use constitutes a threat to this biodiversity. It is therefore important for conservation ecologists to monitor agricultural practices in each grassland from a year to another. Remote sensing is a useful tool for continuous monitoring of vegetated areas at large extents. This talk deals with the classification of grassland management practices using high resolution satellite image time series. In this work, grasslands are semi-natural elements in fragmented landscapes, they are thus heterogeneous and small elements. Our first contribution is to account for grassland heterogeneity while working at the grassland scale by modeling its pixels distributions by a Gaussian distribution. Our second contribution is to measure the similarity between two grasslands thanks to a Gaussian mean map kernel: the so-called alpha-Gaussian mean kernel. It allows to weight the influence

of the covariance matrix when comparing two grasslands. This kernel is plugged into a Support Vector Machine (SVM) and used for the supervised classification of three management practice types. The dataset is composed of 52 grasslands from south-west France. The satellite data is an intra-annual multispectral time series from Formosat-2. Results are compared to other pixel- and object-based approaches both in terms of classification accuracy and processing time. The proposed modeling showed to be the best compromise between processing speed and classification accuracy. Moreover, it can adapt to classification constraints and it encompasses several similarity measures known in the literature.

Misspecified Identifiability Constraints Produce Wrong PQL and INLA Fits of Spatio-temporal Disease Mapping Models.

T. Goicoa[(a)], A. Adin[(a)], M. Ugarte[(a)], J. Hodges[(b)]
[(a)]Public University of Navarre, [(b)]University of Minnesota

The development of statistical methodology for disease mapping has been enormous in the last few years. New goals and challenges have emerged in parallel with the necessity of looking into the space-time data recorded in modern registries, fostering new flexible models, faster and less demanding fitting strategies, and free software to implement these improvements. The most modern and up to date statistical techniques are now ready for health researchers and policy makers to use in a routine way to produce maps with the temporal evolution of geographical patterns, or to describe the temporal evolution of mortality or incidence risk in a given area. However, this automated form of analyzing mortality data can result in errors and misleading conclusions due to, for example, wrong definition of identifiability constraints which standard software usually sets by default. The goal of this talk is to clarify which constraints should be considered and why they are required when fitting disease mapping spatio-temporal models using integrated nested Laplace approximations (INLA) and penalized quasi-likelihood (PQL).

Modeling Multivariate Complex Spatial Data With INLA

V. Gomez-Rubio[(a)], F. Palmi-Perales[(a)]
[(a)]Universidad de Castilla-La Mancha

The Integrated Nested Laplace Approximation (INLA) provides approximations of the posterior marginal distribution of the parameters of a Bayesian hierarchical model. However, INLA requires the model to be expressed as a latent Gaussian Markov random field. In practice, model fitting with INLA is reduced to the latent models implemented in the R-INLA package. In this work we will describe how to extend the class of models that can be fitted with INLA and R-INLA. In particular, we will consider multivariate spatial data. We will describe how to combine INLA and MCMC for an efficient model estimation. Furthermore, we will show several examples using spatio-temporal cancer mortality in Spain and spatial econometrics models.

Prior Specification in Random Effect Models

F. Greco[(a)]

[(a)]University of Bologna, Bologna

In the context of latent Gaussian models with fixed structure matrices, we discuss prior specification on precision parameters of random effects. We show that several approaches proposed in the literature have the undesirable characteristic that the implied prior on the marginal variance of the random effects depend on the structure matrix in a way that is not controlled by the modeler. The main aim of this talk is to build a coherent strategy for the specification of prior distributions on the precision parameters with the following two main features: (i) when model comprises more than one random effects, the prior is fair with respect to the prior variance allocated to each of them; (ii) dependence of the prior variability on the structure matrix is removed. In this context, we discuss the effect of our proposed prior specification in the Besag York and Mollié model for disease mapping.

Material Deprivation and Social Segregation in the Census Cohort Sudies.

L. Grisotto[(a)], D. Catelan[(a)], A. Biggeri[(a)]

[(a)]University of Florence, Florence

The relationship between material deprivation level and health was largely documented in the epidemiological literature both individual and ecological level.The effect of socio-economic variables on health is complex, and might be mediated by the average level of deprivation of the area in which a person lives. It means that a better off person that lives in a poor area have worst health outcomes than he would have living in a non disadvantaged area. In the sociological literature this was named contextual effect. The heterogeneity in the size of the associations between contextual deprivation and mortality was highlighted across different cities. However, the underlying mechanisms for the effect of context and its heterogeneity remain unclear. The social segregation could explain this heterogeneity. In Sociology, segregation is a "measure of the unevenness of distribution of individual characteristics between organizational units". It is usually used to describe a given macro-unit, like a city or a region or a society. We could be also interested in having segregation measures for each sub-unit of interest. We consider three indices: the isolation index, the dissimilarity index, and the segregation index.The present study aims at analysing the relations between mortality, material deprivation and social segregation in the Tuscan Longitudinal Study, using hierarchical Bayesian models.

Spatial Pattern Clustering for PM2.5 With Application To Anomaly Detection

N. HSU[(a)]

[(a)]National Tsing-Hua University

Air quality surveillance is important in public health and environmental science. The impact of particulate matters on health is particularly concerned in recent years. This work adopts functional data analysis to explore the daily pattern of PM2.5 in Taiwan. Taking into account both spatial and temporal dependence, intraday trajectories at different locations are clustered into groups subject to their common functional characteristics. The connections between cluster membership and covariates are also investigated. This approach achieves two goals. First, by taking the advantages of spatial clustering, more accurate intraday forecast at each location can be obtained. Second, by monitoring the dynamic changes of the functional PCA scores among the global region, we are able to detect anomalies, tracking their patterns and identify the sources of pollution. Taiwan air quality data are used for illustration.

On-line Adaptation of Agent-based Disease Simulation

H. Haario[(a)], V. Shemyakin[(a)]

[(a)]Lappeenranta University of Technology

On-line estimation methods for model parameters of large-scale weather models have been recently introduced. The idea is to avoid the use of costly iterative optimization by just 'monitoring' the performance of perturbed ensemble predictions. Here we adapt the approach to the assimilation of cumulative data of infectious diseases to agent-based models during the outbreak of a disease. We demonstrate the approach both by a synthetic SEIRS case, and with a realistic setting based on a network modeling of the spread of the infection.

Combining a Glacier Dynamics Model With Multiple Surface Data

M. Haran[(a)]

[(a)]Penn State University

The future behavior of the West Antarctic Ice Sheet (WAIS) may have a major impact on future climate. For instance, ice sheet melt may contribute significantly to global sea level rise. Understanding the current state of WAIS is therefore of great interest. WAIS is drained by fast-flowing glaciers which are major contributors to ice loss. Hence, understanding the stability and dynamics of glaciers is critical for predicting the future of the ice sheet. Glacier dynamics are driven by the interplay between the topography, temperature and basal conditions beneath the ice. A glacier dynamics model describes the interactions between these processes. We develop a

hierarchical Bayesian model that integrates multiple ice sheet surface data sets with a glacier dynamics model. Our approach allows us to (i) infer important parameters describing the glacier dynamics, (ii) learn about ice sheet thickness, and (iii) account for errors in the observations and the model. Because we have relatively dense and accurate ice thickness data from the Thwaites Glacier in West Antarctica, we use these data to validate our approach. The long-term goal of this work is to have a general model that may be used to study multiple glaciers in the Antarctic.

Quantifying Uncertainties in the Monthly Mean Zonal Mean, and in Vertical Profiles, of Sparsely Sampled Atmospheric Variables

B. Hassler[(a)], G.E. Bodeker[(a)]

[(a)]Bodeker Scientific, Alexandra

Combining data from different measurement systems to achieve a desired temporal or spatial coverage is complicated. Differences in satellite orbits, measurement techniques, and preferred atmospheric conditions for the measurements prevent a straightforward approach of averaging all available measurements to a realistic and representative mean estimate. Offsets and drifts between measurement systems, uneven or patchy spatial and temporal sampling have to be corrected and therefore influence the uncertainties on the individual measurements. In our approach in creating a monthly mean, zonal mean ozone database, that draws data from several satellite and ground-based measuring systems, we strive to update the uncertainty estimate for each measurement with every single processing step: (1) interpolating the different data sources on a common vertical grid, (2) homogenization of the different data sources to one consistent baseline (utilizing output from a chemical transport model (CTM)), (3) correcting for spatio and temporal measurement biases (also utilizing output of a CTM), and (4) correcting for spatio and temporal autocorrelation between individual measurements. All uncertainties introduced by the data processing are then taken into account when the monthly mean zonal mean values and their uncertainty values are calculated as a last step. Any analysis based on these monthly mean zonal mean values, e.g. ozone trend analyses, can then be provided with realistic uncertainty estimates on the results, and provide a more realistic understanding of the confidence in the results.

Regression Between Parts of Compositional Data Using Symmetric Balances

K. Hron[(a)], P. Filzmoser[(b)], V. Todorov[(c)]

[(a)]Palacky University, [(b)]Vienna University of Technology, [(c)]United Nations Industrial Development Organization

Compositional data can be considered as quantitative descriptions of contributionson a whole, carrying relative information. They occur frequently in environmetrics (e.g., concentrations of elements in a rock), usually represented in units like mg/kg,cps, proportions or percentages. The different parts (components) of a compositional data set cannot be considered independent from each other, since only the ratios between the parts constitute the relevant information to be analysed. Practically, this information can be included in a system of orthonormal logratio coordinates, which allow for interpretation in sense of the original parts. In order to treat both the responseand explanatory components in a symmetric manner, this results in applying so called symmetric balances. The resulting coordinates have the form of logratios of individual parts to a (weighted) "average representative" of the other parts, and thus, they clearly indicate how the respective parts dominate in the composition on average. For estimation of regression coefficients orthogonal regression is appropriate since all compositional parts - also the explanatory variables - are measured with errors. Besides classical (least-squares based) parameter estimation, also robust estimation based on robust principal component analysis is employed. As mostly large-scale data sets are present in environmetrics, the common bootstrap inference is replaced just by an exploratory approach. Theoretical outputs are applied to real data from environmental geochemistry

---

Multivariate Space-time Modelling of Multiple Air Pollutants and Their Health Effects Accounting for Exposure Uncertainty

G. Huang[(a)], D. Lee[(b)], M. Scott[(b)]

[(a)]National Tsing Hua University, Taiwan, [(b)]The University of Glasgow

The long-term health effects of air pollution are often estimated using a spatio-temporal ecological areal unit study, but this design leads to the following statistical challenges: (i) how to estimate spatially representative pollution concentrations for each areal unit; (ii) how to allow for the uncertainty in these estimated concentrations when estimating their health effects; and (iii) how to simultaneously estimate the joint effects of multiple correlated pollutants. This article proposes a novel two-stage Bayesian hierarchical model for addressing these three challenges, with inference based on Markov chain Monte Carlo simulation. The first stage is a multivariate spatio-temporal fusion model for predicting areal level average concentrations of multiple pollutants from both monitored and modelled pollution data. The second stage is a spatio-temporal model for estimating the health impact of multiple correlated pollutants simultaneously, which correctly accounts for the

uncertainty in the estimated pollution concentrations. The novel methodology is motivated by a new study of the impact of both particulate matter and nitrogen dioxide concentrations on respiratory ill health in Scotland between 2007 and 2011, and the results suggest that both pollutants exhibit substantial and independent health effects.

Spatial Prediction of Fine Particulate Matter in Taiwan

H. Huang[(a)]

[(a)]Institute of Statistical Science Academia Sinica

Fine particulate matter (PM2.5) has been a major environmental risk to health in many major cities in the world. In this research, we consider two types of PM2.5 monitoring systems with different measuring accuracies. The first one is composed of high-quality data measured at a small number of environmental monitoring stations. The second one consists of less precise measurements from small air boxes with a wider coverage at a large number of locations. We shall apply a spatial random-effects model to combine the two types of data and obtain the optimal spatial prediction of PM2.5 concentration at any location without restricting to data locations. The proposed method will be applied to analyze PM2.5 data of Taiwan.

Explore the Relations of Pollutant Concentrations With Meteorological and Locational Variables

H. Huang[(a)], M. Xu[(a)]

[(a)]Peking University

To better understand the distribution pattern of fine particulate matters (PM2.5) in Beijing and nearby cities, we apply the hidden dynamic geostatistical (HDG) model discussed in Finazzi and Fassò (2014) and Calculli et al. (2015), and study hourly data collected in year 2016. Our findings show that, factors including meteorology variables, geographic information, other pollutant such as SO2, and their interactions have significant impacts on the level and evolvement of PM2.5. These impacts vary across seasons and regions. In addition, we modify the HDG model by adding calibration components which account for the heterogeneity among reporting of different sites in a certain area. Maximum likelihood estimates are obtained for these calibration components. We specially apply the model to Hebei province and find that some stations have substantially smaller calibration value which indicate under-reporting PM2.5 in these stations.

Spatial-temporal Modeling of Heavy Tailed Data

G. Huerta[(a)]
[(a)]University of New Mexico

In this work we introduce a spatio-temporal process starting with a specific case that allows for Pareto marginal distributions. Dependence in space and time is introduced through the use of latent variables in a hierarchical fashion. We review the construction of the process and study some of its properties. Instead of composite likelihood approaches, we follow a Bayesian approach to estimate the model and show how to obtain posterior inference via MCMC methods. The performance of the process is illustrated with simulations and a pollution dataset of monthly maxima ozone concentration and compared to other alternative models. We also consider generalizations of this model that allow for other marginal distributions and other representations at the latent level.

Methods for Zero-inflated and Autocorrelated Counts With Application To Health Care Performance Analysis

A. Hussein[(a)]
[(a)]University of Windsor

In this talk we review some recent methodologies that have been proposed for temporally correlated and zero-inflated count data. We examine the merits and shortcomings of these approaches by using Monte Carlo simulations. We then use data on the volume of patients at the ED of a hospital to illustrate the practical use of the methods.

Estimation of Abundance From Occurrence Maps

W. Hwang[(a)]
[(a)]National Chung Hsing University

A occurrence map consists of observations from a collection of grids, where for each grid it indicates the presence or absence of a particular species in a geographical area. These data are commonly used in studies of macroecology for understanding the distribution of species and biodiversity estimation. The pattern of spatial occurrence is mainly determined by three factors: abundance, spatial distribution, and the grid size (sampling unit). Challenge to analysis the presence-absence map is to quantify spatial association of species distribution by only using the binary data. Consider the estimation of the population size from a presence-absence map, we show that the effects of ignoring spatial dependnece generally results in a negative bias. To reduce the bias, we develop an approach based on clustered negative binomial models where the size of clusters are unknown. We examine the performance of the estimators in a simulation study and by estimating the abundance of some tree species in a 50-ha rain forest plot.

Data Fusion for Functional Data

R. Ignaccolo[(a)], S. Bande[(b)], M. Franco-Villoria[(a)], A. Giunta[(a)]
[(a)]Università degli Studi di Torino, Torino, [(b)]ARPA Piemonte, Torino

In the air quality framework, chemical transport models (CTMs) simulate air pollutant transport, transformation and diffusion and provide pollutants concentrations on a regular thick grid, whereas observed concentrations are gathered from irregularly spaced sites of monitoring networks. Data fusion methods aim at combining these two kinds of data in order to obtain an accurate air quality assessment. In particular, we are interested in applying a data fusion strategy to nitrogen dioxide, ozone and particular matter concentration in Piemonte, Italy. To this goal, we exploit a functional approach and propose a spatial data fusion strategy based on a functional kriging model with external drift. The proposal will be evaluated and compared to other strategies, such as spatial and spatio-temporal kriging, in terms of prediction perfomance and computational cost.

Coverage Uncertainties in Global Temperature Fields

M. Ilyas[(a)]
[(a)]University College London, London

Temperature measurements are subject to uncertainties. Observational uncertainties come from the non-climatic factors (e.g. station movements, changes in instruments, observing practices, urbanization). Coverage uncertainties arise from the lack of spatial coverage of temperature fields. In other words, it is the uncertainty associated with interpolation estimates. The HadCRUT4 ensemble members present the observational uncertainties in temperatures. These temperature fields are sparse. Therefore, we use the multi-resolution lattice kriging proposed for the full spherical domain. We extend the temperature anomalies over the whole planet and quantify coverage uncertainties in these predicted temperatures. This leads to the generation of a large ensemble of fields sampling coverage uncertainties in HadCRUT4 temperature anomalies. These uncertainties are higher in the past at the global scale. The uncertainties in temperatures are expected to propagate to the classification of ENSO. Therefore, we define the ENSO classification to cope with the uncertainties in temperatures. The assessment of these uncertainties reveals that coverage uncertainties dominate early records. Hence, the impact of these uncertainties is significant at the regional scale as well. We further aim to incorporate parametric uncertainties in the model using Approximate Bayesian Computation (ABC).

Modelling Biodiversity Change From Messy and Biased Data

N. Isaac[a]

[a]Centre for Ecology & Hydrology

We have entered a new geological epoch, the Anthropocene, reflecting the pervasive impact of humans on our planet. One feature the Anthropocene is what ecologists refer to as the biodiversity crisis, or the "Sixth Mass Extinction". Monitoring and understanding biodiversity change is critical in order to enact effective mitigation strategies, but there is a dearth of high quality data for this purpose. Occurrence records, such as those collected by Citizen Science projects, are a rich source of information: the Global Biodiversity Information Facility (GBIF) database now contains over 600 million records. However, occurrence records were not gathered in a systematic manner, leading to numerous biases. I will describe the application of hierarchical Bayesian occupancy-detection models to unstructured occurrence records, and show using computer simulation that the resultant trends are robust to known biases in the data. I will illustrate the use of these models using a suite of examples, including biodiversity indicators and measuring the impact of pesticides on beneficial insects.

Efficiently Modelling Telemetry Data

E.L. Jones[a], H.C. Bakka[b], J.B. Illian[a], T. Photopoulou[c], G. Aarts[d], S. Smout[a]

[a]University of St Andrews, [b]Norwegian University of Science and Technology, [c]Nelson Mandela Metropolitan University, [d]IMARES Wageningen UR

Deploying telemetry tags onto animals has become an increasingly common way of collecting detailed information on animal movement and the availability of covariates in the surrounding environment. These generate a vast amount of information, including how animals move through their environment. Characterising species distribution maps is a fundamental insight of ecology, fulfilling both applied and conservation objectives. However, location information generated through telemetry data are presence-only and species distribution models generated from these data such as those characterising habitat preference or space-use are typically subject to several issues: (1) Where predictions over space and/or time are required, they must be dealt with carefully, and usually pseudo-absences are generated to allow this. The selection of pseudo-absences can in themselves be subjective, and their selection can influence model outputs; (2) spatial resolution of model predictions; (3) temporal correlation; and (4) problems with model convergence due to data limits of the software. To potentially solve some of these issues, we use an Integrated Nested Laplace Approximations – Stochastic Partial Differential Equation (INLA-SPDE) approach to fit spatial point process models using movement data in R-INLA.

Multivariate Point Process Models for Multiple Rain Type Occurrences

M. Jun[(a)]

[(a)]Texas A&M University

We present point process models for dealing with multiple rain type occurrences on a global scale. The model incorporates various atmospheric state variables to account for nonstationarty that rainfall occurrence data exhibit on the globe. Computational methods for inference based on likelihood and computational difficulties due to massive data will be discussed.

Spatio-temporal Bayesian Kriging Models To Predict Residential Air Pollutants Exposure

C. Lagazio[(a)], L. Cecconi[(b)], A. Biggeri[(b)], A. Vicedo Cabrera[(c)], L. Grisotto[(b)], D. Catelan[(b)]

[(a)]University of Genova, Genova, [(b)]University of Florence, Florence, [(c)]Swiss Tropical and Public Health Institute, Basel

A core challenge in epidemiological studies of the impact of air pollution on health is quantifying the exposure of subjects at risk. Self-reported exposure (e.g. to road traffic) was recently criticized as a reliable source of information on air pollution exposure due to potential reporting bias. On the other side, individual monitoring is mostly used in small sized studies due to high costs and difficulties related to monitoring for long periods of time a large number of people. For this reason, in the last years geostatistical methods have become more and more relevant to predict exposure to air pollutants from routine monitoring data. Depending on the type of effect that is investigated, also the time dimension could be relevant. Indeed, while for a long term effect we are only interested in the prediction of an overall spatially referenced exposure, when considering short term effects (0 to 3 days of latency) the time pattern of exposure is of fundamental importance. The simultaneous integration of space and time dimensions in geostatistical models is a complex task, unless separability is assumed. We propose a Bayesian space-time model that explicitly allows several specifications of space-time interaction. The model is a generalization of a class of space-time models used with areal data. The model has been applied to estimate exposure to Sulphur dioxide (SO2) for children in the high risk area of Milazzo – Valle del Mela (Sicily, Italy).

Dynamic Mixtures of Copulas for Cylindrical Time Series

F. Lagona[(a)]
[(a)]University of Roma Tre, Roma

Cylindrical time series are bivariate series of angles and intensities that often arise in environmental research. The statistical analysis of these series has been overlooked due to the special topology of the support on which the measurements are taken (the cylinder), and to the difficulties in modeling the cross-correlations between angular and linear measurements over time. Further complications arise from the skewness and the multimodality of the marginal distributions of the data. A novel mixture model is introduced to accommodate these features. The model is based on a mixture of copula-based cylindrical distributions, whose parameters evolve according to a latent Markov chain. The cylindrical distributions are specified by integrating a von Mises distribution and a Weibull distribution by means of a circular copula. A computationally efficient Expectation-Maximization algorithm is described to estimate the parameters. A parametric bootstrap routine is introduced to compute confidence intervals. These methods are illustrated to segment cylindrical time series of wave heights and directions in the Adriatic sea.

Convergence Analysis of Penalized Least Squares Data Fitting Splines

M. Lai[(a)]
[(a)]University of Georgia Athens

We first review the convergence of penalized least suares data fitting to scattered data in the two dimensional setting using bivariate splines. We shall present the convergence in RMSE and then in the maximum norm error when the number of data locations is not sufficiently many. Then we shall extend the study to the setting of spherical setting.

Modelling Latent Trends From Spatio-temporally Grouped Data Using Composite Link Mixed Models

D. Lee[(a)]
[(a)]Basque Center for Applied Mathematics

Epidemiological data are frequently recorded at coarse spatio-temporal resolutions. The aggregation process is done for several reasons: to protect confidential patients' information, to compare with other datasets at a coarser resolution than the original, or to summarize data in a compact manner. However, we lose detailed patterns that follow the original data, which can be of interest for researchers and public health officials. In this talk we propose the use of the penalized composite link model together with its mixed model representation in order to estimate the underlying

trend behind grouped data at a finer spatio-temporal resolution. This model also allows for the incorporation of fine-scale population into the estimation procedure. We assume the underlying trend is smooth across space and time. We illustrate our proposal with the analysis of data obtained during the largest outbreak of Q-fever in the Netherlands.

A Spatio-temporal Process-convolution Model for Quantifying the Effects of Air Pollution On Respiratory Prescription Rates in Scotland

D. Lee[(a)]

[(a)]University of Glasgow, Glasgow

The rates of respiratory prescriptions vary by GP surgery across Scotland, suggesting there are sizeable health inequalities in respiratory ill health across the country. The aim of this paper is to estimate the magnitude, spatial pattern and drivers of this spatial variation. Monthly data on respiratory prescriptions are available at the GP surgery level, which creates an interesting methodological challenge as these data are not the classical geostatistical, areal unit or point process data types. A novel process-convolution model is proposed, which extends existing methods by being an adaptive smoother via a random weighting scheme and using a tapering function to reduce the computational burden. The results show that particulate air pollution, poverty and ethnicity all drive the health inequalities, while there are additional regional inequalities in rates after covariate adjustment.

Comparison Between Spatio-temporal Random Processes and Application To Climate Model Data

B. Li[(a)]

[(a)]University of Illinois at Urbana-Champaign

Comparing two spatio-temporal processes are often a desirable exercise. For example, assessments of the difference between various climate models may involve the comparisons of the synthetic climate random fields generated as simulations from each model. We develop rigorous methods to compare two spatio-temporal random processes both in terms of moments and in terms of temporal trend, using the functional data analysis approach. A highlight of our method is that we can compare the trend surfaces between two random processes, which are motivated by evaluating the skill of synthetic climate from climate models in terms of capturing the pronounced upward trend of real-observational data. We perform simulations to evaluate our methods and then apply the methods to compare different climate models as well as to evaluate the synthetic temperature fields from model simulations, with respect to observed temperature fields.

Simultaneous Modelling and Estimation of Climate and Weather

F. Lindgren[(a)]

[(a)]The University of Edinburgh

The EUSTACE project will give publicly available daily estimates of surface air temperature since 1850 across the globe for the first time by combining surface and satellite data using novel statistical techniques. Of particular importance is to obtain realistic uncertainty estimates, due to both observation uncertainty and lack of spatio-temporal coverage. To this end, a spatio-temporal multiscale statistical Gaussian random field model is constructed, with a hierarchy of spatio-temporal dependence structures, ranging from weather on a daily timescale to climate on a multidecadal timescale. Connections between SPDEs and Markov random fields are used to obtain sparse matrices for the practical computations. The extreme size of the problem necessitates the use of iterative solvers, which requires using the multiscale structure of the model to design an effective preconditioner.

On Spline-based Approach To Spatial Linear Regression for Geostatistical Data

G. Ludwig[(a)], J. Zhu[(b)], P. Reyes[(c)], C. Chen[(d)], S. Conley[(b)]

[(a)]University of Campinas, Campinas, [(b)]University of Wisconsin-Madison, Madison, [(c)]Kansas State University, [(d)]National Changhua University of Education, Changhua

For spatial linear regression, the traditional approach is to use a parametric linear mixed-effects model such that spatial dependence is captured as a spatial random effect that is a Gaussian process with mean zero and a parametric covariance function. Spline surfaces can be used as an alternative approach to capture spatial variability, giving rise to a semiparametric method that does not require the specification of a parametric covariance structure. The spline component in such a semiparametric method, however, impacts the estimation of the regression coefficients. In this paper, we investigate such an impact in spatial linear regression with spline-based spatial effects. Statistical properties of the regression coefficient estimators are established under the model assumptions of the traditional spatial linear regression. We also develop a method to choose the tuning parameter for the smoothing splines that is tailored toward drawing inference about the regression coefficients. Further, we examine the empirical properties of the regression coefficient estimators under different scenarios of spatial confounding via a simulation study. A data example in precision agricultural research regarding soybean yield in relation to field conditions is presented for illustration.

A Simple Spatio-temporal Model With Principle Splines

L. Ippoliti[(a)], L. Fontanella[(a)], R. Ignaccolo[(b)], P. Valentini[(a)]
[(a)]Università di Chieti-Pescara "G. D'Annunzio", Pescara, [(b)]Università di Torino, Torino

Spatio-temporal modelling has largely been developed through applications in geostatistics, hydrology, meteorology, environment monitoring, health data, functional MRI and facial analysis. Motivated by these applications, the field has adopted various modelling strategies, essentially depending on the underlying objective of the analysis, the scale and type of data. This work is concerned with the specification of a simple three-level hierarchical spatio-temporal model which warrants consideration when small data sets are available and spatial predictions are required. Especially under Gaussian assumptions, the model has several computational advantages and its general form appears to be useful to deal with different types of spatial data. The model has an intuitive appeal and enjoys several advantages. For example, it is well-suited for incorporating the foregoing knowledge at various levels of the modelling, is easy to interpret and facilitates model fitting. Furthermore, within the Bayesian paradigm, it enables exact inference and proper uncertainty assessment within the given specification. The choice of model priors with particular reference to issues that arise in the estimation of the model parameters will be discussed. In practice, assumptions are needed to reduce their total number, and appropriate choices of these priors allow us to regularize our inference. It will be shown that restrictions on the number of parameters can be obtained by imposing patterns of zeroes as "hard constraints" of the model parameter matrix. However, we also discuss the use of "soft constraints", which shrink the estimates in certain directions without forcing them. By following this approach, which is in between posing no restrictions and forcing restrictions, some structures may be favoured probabilistically instead of being imposed. Results from the analysis of real data have shown that our model can provide accurate predictions and is particularly useful when reliable estimates of the parameters of a covariance function is difficult.

Spatio-temporal Clustering of Water Quality Trends

V. Lyubchich[(a)]
[(a)]University of Maryland Center for Environmental Science

Modern climate data sets, including paleoreconstructions, long-term weather monitoring records, and remote sensing data, contain a wealth of space-time information that leads to a variety of challenges related to data storage, management, and analysis. This has sparked an interest in dynamic space-time clustering algorithms that are particularly suitable for the analysis of large data streams. The trend-based clustering algorithm TRUST allows segmentation of space-time processes in real time, but requires the user to set multiple tuning parameters, and this step is usually performed in a subjective manner. Here we propose data-driven automatic approaches to simultaneously select the tuning parameters. We focus on the two most important parameters of the TRUST algorithm, which define short-term closeness of observations across locations and long-term persistence of such closeness within an analyzed

time window. We demonstrate the performance of the enhanced clustering procedures using simulated time series, and illustrate their applicability using long-term water quality records in Chesapeake Bay.

Homogenizing Radiosonde Temperature and Humidity Profile in the BARON for C3S Service

F. Madonna[(a)]

[(a)]Consiglio Nazionale delle Ricerche

Copernicus is a European Union Programme aimed at developing European information services based on satellite Earth Observation and in situ (non-space) data. Among the Copernicus Services, the Climate Change Service (C3S), still under development phase, will combine observations of the climate system with the latest science to develop authoritative, quality-assured information about the past, current and future states of the climate in Europe and worldwide. In the frame The Copernicus Service(s) the Access to Observations for Baseline And Reference Observation Networks will be established by the Service C3S_311a_Lot3, previously named C3S-BARON (Baseline And Reference Observation Networks). The C3S_311A_lot3 is focussed upon access to and redistribution of harmonized data products from atmospheric in-situ observations networks measuring a subset of ECVs which includes surface temperature, atmospheric temperature and humidity (vertical profiles), ozone (column and profiling concentration), wind profiles (from radiosoundings), CO, CO_2 and CH_4 (column concentrations), and water vapour content (columnar from GPS/GNSS only). For these ECVs demonstrable Baseline and Reference quality measurement networks are assured. The successful implementation of the proposal will allow the development of consistent quality control algorithms for in-situ climate data arising from Baseline and Reference networks at various time scales (hourly, daily, monthly, annually). Methods will be developed and implemented to detect and adjust for inhomogeneities due to issues such as instrumentation changes, calibration drifts or observing station relocations and to quantify uncertainty in a consistent and metrologically rigorous manner. The presented work has the aim to share the first results towards the harmonization of the time series of atmospheric temperature and humidity measured by global radiosounding networks and to discuss the methods which will be applied for the release of the harmonized dataset and the related ancillary products to make available to the C3S users.

Multivariate State-space Models for Skewed Environmental Data

R. Ma[(a)], X. Duan[(b)]

[(a)]University of New Brunswick, Fredericton, [(b)]Chuxiong Normal University, Chuxiong

Multivariate time series data frequently arise in environmental and ecological studies. Such data are often skewed with complex cross covariance structures. To have flexibility in accommodating skewness and covariance within the framework of generalized linear mixed models for multivariate time series data, we incorporate serially correlated distribution-free multivariate random effects into Tweedie family of distributions. This Tweedie family has clear ecological interpretation of variance-to-mean relationship characterized by Taylor's power law. Our method is illustrated with application to environmental data.

On Nonseparable Spatiotemporal Covariance Models, With Applications To Bayesian Disease Mapping

Y.C. MacNab[(a)]

[(a)]University of British Columbia, Vancouver

Recently, important advancements have been made in the formulation and implementation of nonseparable multivariate Gaussian Markov random field models for analysis of multivariate lattice data and for Bayesian disease mapping in particular. In this presentation, I show that these advancements open ways for generalizations of nonseparable spatiotemporal covariance models for spatiotemporal data on lattice. Formulating separable spatiotemporal covariance models has been the mainstream approach to spatiotemporal disease mapping. I briefly review key models in recent literature and propose several nonseparable alternatives. The proposals are illustrated for their potentials to facilitate spatiotemporal smoothing for improved risk predictions, in terms of overall goodness-of-fit and increased statistical strength and efficiency.

Spatio-temporal Kriging for Air Quality Forecasting

L. Malherbe[(a)], M. Beauchamp[(a,b)], C. De Fouquet[(b)], M. Valsania[(c)], F. Meleux[(a)], A. Ung[(a)]

[(a)]INERIS, [(b)]Mines ParisTech, Paris, [(c)]University of Turin, Turin

The national PREV'AIR system (www2.prevair.org) delivers daily analyses and forecasts of different atmospheric pollutant concentrations over Europe and France. Analysed maps of the previous day situation (D-1) are produced by combining in-situ measurement data with output data from the chemistry-transport model CHIMERE. To perform this combination, kriging with external drift is applied using a moving neighbourhood. Forecast maps for the current and next two days (D+0, D+1, D+2) are computed according to the same methodology, except that the observation data are replaced by statistical forecasts at the monitoring sites obtained by station specific multilinear regression models. A constraint in that approach – denoted as statistical adaptation - lies in the availability of one to several years of training and validation data in order to build robust and efficient regression models. A way to get round this issue is to consider the analysis and adjustment of the forecast as a unique procedure in a spatio-temporal framework. In this study a spatio-temporal external drift kriging using a spatio-temporal moving neighbourhood has been developed and evaluated over one year on both the European and national scales. In that approach the prediction of D+0 concentrations is similar to an analysis where the available data in the neighbourhood are past observations until D-1 and CHIMERE outputs until D+0. The spatio-temporal covariance model and the results of the evaluation will be presented and commented. Those results will be compared to the performance of the current statistical adaptation. Possible areas for improvement and development will be discussed.

Using Extreme Value Statistics To Forecast Groundwater Floods and Droughts

B. Marchant[(a)], E. Eastoe[(b)], J. Wadsworth[(b)], J. Bloomfield[(a)]

[(a)]British Geological Survey, [(b)]University of Lancaster, Lancaster

Groundwater floods and droughts impact agricultural livelihoods, households, local and national economies and ecosystems. In the UK, previous extreme groundwater events have led to billions of pounds of damage to infrastructure and crops in addition to health impacts and loss of life. Planners and policy makers are increasingly seeking probabilistic forecasts of the characteristics of these extreme events in a changing climate in addition to relatively short-term forecasts of their likelihood a few months ahead. We discuss the various models and methodologies used to predict groundwater floods and droughts. Many studies have considered the temporal variation of groundwater levels within a single borehole. Both empirical and process-based models have been used to represent this variation, to predict the probability of groundwater levels exceeding specific thresholds and to explore how the characteristics and frequency of the extreme events vary over different climate

projections. However, such models consider the entire distribution of groundwater levels whereas extreme events only correspond to the tails of the distribution. We therefore employ extreme value statistical methods. For locations with long-term (>50 year) records of groundwater levels we find that it is possible to calibrate extreme value models that relate the occurrence and duration of extreme events to weather variables. These models are used to forecast drought/flood characteristics under different climate projections and to forecast the likelihood of a drought a few months ahead. Floods evolve more quickly than droughts and are harder to forecast more than a month ahead. We find that the length of the data record limits the use of these models since the parameters become unidentifiable for records substantially shorter than 50 years.

Two Level Modeling of Relative Risk of Dengue Disease At Small Spatial Scale in Colombia

D. Martinez-Bello[(a)], A. Adin[(b)], A. López-Quílez[(a)], M. Ugarte[(b)]

[(a)]Universitat de València, Valencia, [(b)]Public University of Navarre, Pamplona

Dengue is an arboviral infectious disease affecting between 50 to 100 millions of persons every year. Disease mapping is a set of techniques that permits the visualization and modeling of disease incidence in space and time. The objective of this talk is to study dengue incidence in a medium-sized Colombian city with a high dengue incidence in the period January 2009 to December 2015 using Bayesian disease mapping models with several level of spatial aggregation.The city under study is divided in blocks, sections, sectors, and communes, where every level is nested in the next, according to the guidelines of the Colombian statistical office. For the study, we selected as a first aggregation level, the sectors, and as a second spatial level, the communes, making a total of 94 and 17 primary and secondary aggregation units, respectively. We also used the epidemiological period (aggregation of four epidemiological weeks) as the time unit. We geocoded and allocated dengue cases to each of the primary aggregation level and epidemiological period. Space-time interaction models with first and second-level area effects were fitted to the aggregated data. Second-level models were of two kinds: the first ones included only the effects accounting for sector nested in commune, and the second ones included the sector nested in commune effects plus second-level space-time interaction effects. In addition, we fitted second-level models accounting for the spatial variability only among the first-level areas belonging to the same second-level area. We discuss the public health implications based on the results provided by these models for the surveillance and control of the dengue disease.

Robust Model-based Clustering for Longitudinal Circular Data

A. Maruotti[(a)]

[(a)]Libera Università Maria Ss Assunta, Rome

In this work, we propose two general hidden Markov model-based approaches, namely the noise and the trimming approaches, for clustering longitudinal circular data contaminated by atypical observations that may affect parameter estimates and the recovering of the clustering structure. Hidden Markov models are the state of the art in the analysis of (clustered) longitudinal data, and have been increasingly used in circular data analysis. With respect to the noise approach, we consider hidden Markov fittings with the addition of a further component to capture observations that are not consistent with the projected normal hidden Markov model. This is done by adding a uniform conditional density on the circle or representing the noise component by a fixed constant. The trimming approach introduces a robust estimation procedure based on the choice of a representative trimmed subsample. While the noise approach tries to fit the atypical observations in the model, the trimming approach attempts to discard them completely.

Improve Numerical Weather Forecasts Over Astronomical Sites in Chile Using a Kalman Filter

J.C. Marín[(a)], O. Cuevas[(a)]

[(a)]Universidad de Valparaíso, Valparaíso

Numerical Weather Prediction (NWP) models are useful tools in everyday operations at astronomical observatories in the north of Chile. Global and regional meteorological forecasts of near-surface variables for the next 2-5 days are important for the observatories to have a better observational planning and resources saving. However, due to the complex topography where the observatories are located, large errors are present. A Kalman filter was used in this study to decrease the systematic errors shown by numerical weather forecasts in order to improve its performance over astronomical sites. The Kalman was applied to Global Forecast System (GFS) and Weather Forecast and Research (WRF) forecasts at three observatories in the North of Chile. The results show that the mean bias is largely reduced and the root mean squared error decreases when the filter is applied, supporting its use as a useful tool to improve the weather forecasts in the region.

A Flexible Bayesian Model for Clustering Seasonal Time Series With Linear and Circular Components

G. Mastrantonio[(a)]

[(a)]Politecnico di Torino, Torino

In this work we present a new model for clustering cylindrical time data, i.e. a bivariate time series with one linear and one circular observation. We propose a mixture model where the mixing probabilities are time specific and are assumed to follow a Logistic-Normal distribution. We introduce dependence betweens the vectors of mixing probabilities by means of the Gaussian processes representation of the Logistic-Normal distributions. Then through correlation functions defined over a mixed circular-linear domain, we can evaluate seasonal effects. The seasonal periods are considered as latent variables and then estimated along with the other model parameters. We show how to implement the model in a Bayesian framework. We estimate the model on an animal movement dataset and we compare our results with the ones obtained with an hidden Markov model, showing a great improvement in terms of interpretability and DIC (Deviance Information Criterion.)

Two-factor Experiment Analysis for Replicated Marked Point Patterns

J. Mateu[(a)], J. Gonzalez-Monsalve[(a)], B. Lagos[(b)]

[(a)]Universitat Jaume I, Castellon, [(b)]University of Concepcion, Concepcion

Given a sample of replicated marked spatial point patterns, we are interested in providing a procedure to test if the effects of two factors on a measured pattern descriptor are statistically significant. We consider the mark-weighted K-function as a particular descriptor of each pattern in our sample and develop a set of statistics based on classical analysis of variance statistics and their analogues in functional data analysis. Additionally we consider an application related to the metallurgy flotation bubble data. Several other methods to analyse structural differences between groups of replicated spatio-temporal point patterns will be also presented.

Interpretation of Compositional Data Analysis of Geochemical Data With Case Studies From Environmental Management and the Identification of Potential Human Health Impacts.

J.M. McKinley[(a)]

[(a)]Queens University Belfast, Belfast

Geochemical data sets are used for a wide range of applications from environmental management and the identification of potential health impacts from naturally occurring harmful elements or contaminated land. Such data sets can allow the identification of relationships between naturally occurring target elements (e.g., potentially toxic elements, PTEs) in soil such as arsenic and inform epidemiology studies, or predict environmental exposure to nephrotoxins such as lead. Geochemical sampling and analysis campaigns have been routinely used for the detection of geochemical anomalies for the discovery of mineral deposits. Contaminated land remediation is also strongly dependent on an appropriate assessment of the presence and abundance of PTEs and organic contaminants in soil. Informed use of geochemical data is therefore of critical importance to ensure interpretable and meaningful conclusions from nationally funded geochemical sampling campaigns. Northern Ireland was mapped by the Geological Survey of Northern Ireland (GSNI) during 2004-2006 and provides an available soil geochemistry dataset known as the Tellus Survey. The complex and diverse geology of Northern Ireland provides a representation of typical parent rock types and soils and increases the transferability of results from the Tellus Survey. This study explores the interpretability of results for soil, stream sediments and stream water geochemistry data using a compositional data approach to comply with the relative nature of the geochemical data. Both data-driven and knowledge driven approaches are presented. The data-driven approach is used to reveal patterns and processes that may be hidden or masked by dominant processes. A knowledge driven log-ratio approach is explored for known toxins and PTE elemental associations. Correct and meaningful interpretation of the outputs from compositional data analysis of geochemical data is paramount to enable appropriate interaction with the environment, to understand and mitigate any environmental impact on human health.

Biodiversity, Ecological Status Assessment and Management of Rivers and Lakes: New Solutions To Old Issues?

K. Meissner[(a)]

[(a)]Finnish Environment Institute SYKE

Currently, the Water Framework Directive (WFD) aims to achieve and maintain good status for all surface waters and ground waters and to prevent deterioration and ensure the conservation of high water quality. The process of WFD ecological status assessment of surface waters requires the comparison of metrics derived on plant, algae, fish and macroinvertebrate communities to reference conditions. Currently European surface water status assessments in Europe rely on over 300 different national composition, diversity, tolerance or functional metrics or combinations of these. I will look at some challenges surrounding surface water ecological status assessment and provide some examples. Most measures of diversity overlook the underlying complex patterns of biodiversity and very few studies have analyzed the inherent statistical properties of compositional or other metrics and their impact on the reliability of assessments. Further, past data is rarely explicitly used in assessments of current status. With the advent of new promising DNA techniques, data analytics and the potential in Big environmental data it may well be time for novel approaches to surface water status assessment.

Kriging for Non-stationary Object Data Through Random Domain Decomposition

A. Menafoglio[(a)], G. Gaetani[(a)], P. Secchi[(a)]

[(a)]Politecnico di Milano, Milano

The analysis of complex data distributed over large or highly textured regions poses new challenges for spatial statistics. Although methods to deal with spatial object data have been successfully applied in several environmental studies, they rely upon global models for the spatial dependence of the field, that are rarely usable in the presence of large, textured or convoluted domains, with holes or barriers. We here propose a novel methodology for the spatial prediction of possibly non-stationary object data distributed in such kinds of domains, which cope with the data and domain complexities through a divide-et-impera approach. As a key element of innovation, we propose to perform repeated Random Domain Decompositions (RDDs), each defining a set of homogeneous sub-regions where to perform local object-oriented spatial analyses, under stationarity assumptions, to be then aggregated into a final global analysis. In this broad framework, the complexity of the domain can be taken into account by defining its partitions through a non-Euclidean metric that properly represents the adjacency relationships among the observations over the domain. The method we propose is entirely general, and prone to be used with numerous types of object data (e.g., functional data, density data or manifold data), being grounded upon on the theory of Object Oriented Spatial Statistics. As an insightful illustration of the potential of the methodology, we consider the spatial prediction

of aquatic variables in estuarine systems, that are non-convex and very irregularly shaped regions where the narrow areas of land between adjacent tributaries act as barriers. Here, we focus on the analysis and spatial prediction of density data relevant to the study of dissolved oxygen depletion in the Chesapeake Bay (US).

Functional Approaches for Sparse Spatiotemporal Satellite Data

C. Miller[(a)], M. Scott[(a)], M. Gong[(a)], C. Wilkie[(a)], R. O'Donnell[(a)]
[(a)]University of Glasgow, Glasgow

Developments in satellite retrieval algorithms continually extend the extraordinary potential of satellite platforms such as the MEdium ReSolution Imaging Spectrometer (MERIS) and the Advanced Along-Track Scanning Radiometer (AATSR) to retrieve information across the Earth at finer spatial resolution. For example, resolution down to 300m for MERIS now enables water quality products (and associated retrieval accuracy) to be produced for lakes, and Sentinel 2A & 2B (launched in June 2015 and March 2017) will enable quantitative retrieval with resolutions down to around 10-60m. Challenges associated with these new environmental data streams are the large volumes of data in space and time, collected as images, and the often large quantities of missing data. Functional data analysis provides an attractive approach to investigating spatiotemporal lake images providing efficient dimensionality reduction, data imputation and data linkage. However, novel developments to functional approaches are required to account for complex uncertainties and sparseness. Specifically, we propose functional Bayesian downscaling for calibrating remotely sensed data using in-situ measurements, and spatiotemporal mixed model functional PCA to reduce dimensionality and provide imputations for missing data. The methods enable satellite data to be bias-corrected using in-situ data which does not have to be matched in space or time. Additionally, spatiotemporal patterns can be estimated from sparse data while efficiently reducing dimensionality using smooth spatial surfaces. Methods will be illustrated using data from the AATSR and MERIS instruments on the European Space Agency satellite platform, which have been used to estimate lake surface water temperature data and ecological properties such as chlorophyll (as an indicator of water quality) for lakes in the projects ARC Lake (http://www.geos.ed.ac.uk/arclake), Diversity II (http://www.diversity2.info/) and GloboLakes.

A Characterization of the Valid Parameter Space of Large Multivariate GMRFs

M. Molinaro[(a)], R. Furrer[(a)]
[(a)]University of Zurich, Zurich

Multivariate Gaussian Markov Random Fields (MGMRFs) are widely used in the statistical analysis of area-based data. The associated precision matrix typically depends on many parameters, in order to model complex spatial dependencies. Since the size of available datasets has been steadily increasing over the years, it is furthermore crucial to develop efficient approaches to make inference computationally feasible. In this study, we are motivated by a Bayesian hierarchical model used to analyze bivariate data coming from several regional climate models. The spatial component is described by seven parameters, which quantify the correlation between temperature and precipitation at different locations on a bidimensional regular lattice. In this context, a crucial challenge that we face in both the frequentist and Bayesian paradigms is to efficiently determine the valid parameter space, namely the domain ensuring (semi)positive-definiteness of the associated precision matrix. Due to the generalized block-Toeplitz structure of the latter matrix, closed-form expressions are in general not available. To overcome this, we provide an asymptotically analytical description by extending a convergence theorem of univariate GRMFs with block-Toeplitz precision to GMRFs with block-circulant precision. A thorough simulation study in R will quantify the rate of convergence of our approach. In addition, we will compare our results to state-of-the-art algorithms for sparse matrices and provide strong evidence of the fact that inference can be performed on very large grids, for which the just mentioned algorithms would fail. We will finally discuss how our approach can be applied to more general cases, namely MGMRFs for multivariate data over a regular lattice with different neighborhood conditions.

International Migration and Family Environment: Evidence From Ireland

I. Mosca[(a)], A. Barrett[(b)]
[(a)]Trinity College Dublin, Dublin, [(b)]Economic and Social Research Institute (ESRI), Dublin

We explore whether older parents of adult children who emigrate experience, in the short term, increases in depressive symptoms and loneliness feelings compared to parents whose children do not migrate. We use data from the first two waves of The Irish Longitudinal Study on Ageing, which is a nationally representative sample of 8500 people aged 50+ living in Ireland. To deal with the endogeneity of migration, we apply fixed-effects estimation models and control for a broad range of life events occurring between the two waves. These include the emigration of a child but also events such as bereavement, onset of disease, retirement and unemployment. We find that depressive symptoms and loneliness feelings increase among the parents of migrant children but that the effect is only present for mothers. As the economic

burden of mental health problems is high, our findings have potentially significant impacts for migrant-sending regions and countries.

Applying Demographic Methods and Concept To the Study of Vulnerability To Global Environmental Change

R. Muttarak[(a)], W. Lutz[(b)]

[(a)]International Institute for Applied Systems Analysis, Laxenburg, [(b)]University of East Anglia, Norwich

How future societies will be different from today's in their capacities to cope with climate change is one of the most significant gaps in our understanding of how dangerous climate change will be for future human well-being. Simply assuming that future societies will be identical to those seen today is misleading because we know that both climate conditions and societies are changing. In this presentation, we contend that the concept of demographic metabolism and the associated methods of multi-dimensional population projections present an effective analytical tool box to forecast important aspects of societal change that directly affect future adaptive capacity. We present an example of how the changing educational composition (as one important demographic characteristic that is quantifiable and forecastable) of the population in the future can influence societies' adaptive capacity. By demonstrating how multi-dimensional population projections have been used to form the human core of the Shared Socioeconomic Pathways (SSP) scenarios, we highlight the relevance of incorporating knowledge and analytical tools from a demography in assessing the likely implications of climate change on future human well-being.

A General Methodological Framework for Identifying Disease Risk Spatial Clusters Based Upon Mixtures of Temporal Trends

G. Napier[(a)]

[(a)]University of Glasgow, Glasgow

We present a novel general Bayesian hierarchical mixture model for clustering areas based on their temporal trends. Our approach is general in that it allows the user to choose the shape of the temporal trends to include in the model, and examples include linear, general monotonic, and changepoint trends. Inference from the model is based on Metropolis coupled Markov chain Monte Carlo (MC3) techniques in order to prevent issues pertaining to multimodality often associated with mixture models. The effectiveness of MC3 is demonstrated in a simulation study, before applying the model to hospital admission rates due to respiratory disease in the city of Glasgow between 2002 and 2011. Software for implementing this model will be made freely available as part of the R package CARBayesST.

Spatio-temporal Mismatch for Aerosol Profiles

I. Negri[(a)], A. Fassò[(a)], L. Mona[(a)], N. Papagiannopoulos[(a)], F. Madonna[(b)], F. Finazzi[(a)]

[(a)]University of Bergamo, Bergamo, [(b)]Consiglio Nazionale delle Ricerche

The high variability both in space and time of tropospheric aerosols is one of the main causes of the high uncertainty related to tropospheric aerosols and their interactions with clouds. Since 2006, CALIOP, the LIDAR onboard CALIPSO specifically designed for aerosol and clouds study, is providing high-resolution vertical profiles of aerosols and clouds on a global scale. How well these CALIOP measurements represent the atmospheric conditions of a surrounding area over a longer time is an important issue to be investigated. An integrated study of CALIPSO and EARLINET correlative measurements opens new possibilities for spatial (both horizontal and vertical) and temporal representativeness investigation of this set of satellite measurements. EARLINET (European Aerosol Research Lidar NETwork) is the first LIDAR network for aerosol studies on continental scale. EARLINET comprises of different instrumental setups, specifics and team expertises, it is a network of different instruments with a wide variety of instrumental specifics, so that there is not a common vertical and temporal sampling overall within the network. The comparison of EARLINET profiles and their CALIPSO counterpart is a straightforward procedure. The main aim of this work is to investigate how the horizontal smoothing impact on the uncertainty term between the satellite and the ground measurement of the aerosol layers. In a first analysis we minimize the RMSE to search for the best horizontal smoothing for CALIOP. A functional regression analysis is also performed to understand how the uncertainty term between the satellite and the ground measurement depends on the horizontal and temporal mismatch error on some other meteorological variables and which of them has a greater impact on this uncertainty.

The Axiomatic Definition of Statistical Depth

A. Nieto-Reyes[(a)], H. Battey[(a)]

[(a)]Universidad de Cantabria, Santander

This work intersects the areas of functional data analysis and statistical depth. Functional data analysis provides an alternative way of studying traditional data objects, recognising that it is sometimes more natural and more fruitful to view a collection of measurements as partially observed realisations of random functions. The main focus of this work is on providing a formal definition of depth for functional data, justified on the basis of several properties. The definition fills an important void in the existing literature. The properties that constitute our definition, not only provide a sophisticated extension of those defining the multivariate depth, recognising topological features such as continuity, contiguity and smoothness, but also implicitly address several common or inherent difficulties associated with functional data.

Assessing Coherence of Global Lake Water Quality Using Functional Data Analysis

R. O'Donnell[(a)]

[(a)]University of Glasgow, Glasgow

Globally, data produced though new sensor technologies and earth observation (EO) instruments on satellites are increasing in terms of both quantity and quality at an unprecedented rate. Earth observation enables us to obtain a richer picture of global environment than we have ever had before and has the potential to change the way in which we monitor our natural resources. While previous work has focused on the study of oceans, recent developments in EO instruments such as MERIS (Medium-Spectral Resolution, Imaging Spectrometer) and AATSR (Advanced Along-Track Scanning Radiometer) from the European Space Agency's Envisat satellite platform have enabled the retrieval of a range of water quality determinants from lakes. This is a crucial step for global assessment of environmental quality as lakes can be considered as sentinels of change due to their sensitivity to climatic conditions. The aim of our work is to investigate temporal and spatial coherence in key features of water temperature and water quality of lakes using EO data, at both a global and within lake scale. We will present results from statistical approaches which have been developed to explore coherence in lake behaviour based on data produced by remote sensing technologies applied to a large set of lakes distributed globally. A functional data analysis approach has been taken and functional clustering methods have been developed to identify groups of lakes which are similar in terms of both trends and seasonal patterns. The methods presented account for the complexities such as missing data (not at random) and measurement uncertainty. Following the identification of coherent groups of lakes, attribution of the temporal patterns which underpin these groups can then be studied.

Extreme Weather Events and Road Accidents in Mexico

S.M. Olivera[(a)], S. Lyubchich[(b)]

[(a)]Universidad Autonoma Metropolitana, Cuajimalpa, [(b)]University of Maryland Center for Environmental Science

The Weather-anomalies associated with car accidents are examined using data extracted from Police Accident Report Forms and hourly weather reports between 1997 and 2016 in Mexico City. The typology of weather-anomalies has been chosen with the understanding that the weather, in special the extreme rain and the high temperature, might be a contributing factor in an accident, although not necessarily the main factor. A road accident is the result of a complex interaction of factors; understanding these factors, that condition car accidents may facilitate the application of traffic procedures during inclement weather and would allow the generation of public policies for the prevention of road accidents.

Mapping Groundwater Arsenic Concentrations in Varanasi, Uttar Pradesh, India

V. Pawlowsky-Glahn[(a)], J. Egozcue[(b)], R.A. Olea[(c)], J.N. Raju[(d)]
[(a)]Universitat de Girona, Girona, [(b)]Universidad Politécnica de Catalunya, Barcelona,
[(c)]United States Geological Survey, Reston, [(d)]Jawaharlal Nehru University, New Delhi

Areas east of Varanasi, Uttar Pradesh, India, are among numerous places within the Ganges River watershed with natural occurrences of groundwater concentrations surpassing World Health Organization (WHO) maximum contaminant levels of 10 parts per billion (ppb) for arsenic in drinking water. This level is approximately equal to 10 microgram/liter (mg/L) arsenic, which are the units reported in the data. Laboratory measurements of 11 ions for waters taken at 95 drinking water wells within a 210 km^2 area show arsenic concentrations in groundwater ranging from 1 to 80 mg/L. Geostatistics and compositional data analysis were applied to this dataset in an effort to clarify spatial conditions related to the occurrence of such elevated levels of arsenic by applying cutting edge mathematical methods. Mapping allowed visualization of the spatial uncertainty associated with the sampling. Resulting probability maps visually portray that groundwater is free of arsenic values exceeding 10 mg/L on the erosional side of a meander of the Ganges River at Varanasi proper. Conversely, two areas show anomalous arsenic concentrations greater than 10 mg/L on the depositional side of the valley underneath and on the right margin of the river, in areas that have recently undergone urban development. The combined use of geostatistics and compositional data analysis is completely general and more mathematically adequate than other hydrochemistry mapping methods being used, and consequently could be applied to a wide range of studies.

Climate Change and Population Movement in Tanziania

A. Petrucci[(a)], E. Pirani[(a)], D. Vignoli[(a)]
[(a)]Università degli Studi di Firenze, Firenze

In recent years, climate change has become a major concern in the international context, and addressing climate challenges has become a priority for a wide range of actors worldwide. Mobility might be expected to increase following extreme events, such as storms, floods, and droughts; as well as in response to gradual changes in temperatures, precipitation, and sea levels. Yet predicting the effects of climate change on population movement remains a challenging task. Clearly, examination of domestic migration-environment processes at the individual-level is critical for informed policy and programs. Using individual-level data from the Tanzania National Panel Survey, this study aims to examine the roles played by droughts or floods, crop diseases, and severe water shortages in inter-district migration in Tanzania following a longitudinal perspective. Given past evidence of the important association between migration and education in Tanzania, the role of education in shaping the impact of climate change inter-district migration will be also scrutinized.

Inferring Accents From Geo-localized Acoustic Speech Recordings I

D. Pigoli[(a)], S. Tavakoli[(a)], J.A. Aston[(a)], J.S. Coleman[(d)]
[(a)]University of Cambridge, Cambridge, [(b)]University of Oxford, Oxford

In this work, we analyse the Audio Edition of the British National Corpus, a large database of geo-localized speech recordings deposited at the British Library. The goal of our project is modelling the variations of speech and accents within England. While dialect variation is of considerable interest in linguistics and other social sciences, traditionally it has been studied using proxies (transcriptions) rather than acoustic recordings directly. We introduce novel statistical techniques to analyse geo-localized speech recordings and to explore the spatial variation of pronunciations continuously over the region of interest, as opposed to traditional isoglosses, which provide a discrete partition of the region. Data of this type require an explicit modelling of the variation in the mean and the covariance structure of the speech process. Our analysis recovered the exceptionality of the dialect features in the Greater London area and the well- known divide between the South-East vs. the North of England. There are however some regions - including the contiguous regions of Northamptonshire, Cambridgeshire, Norfolk and Suffolk and also the somewhat separate location of Hampshire - which show differences from both the South-East and the North. Such a "transitional" or "interface" region is not captured well (and is in fact obscured) by the hard boundaries implied by conventional isogloss maps.

Inference On Forest Attributes and Ecological Diversity of Trees-outside-forests Based On Two-phase Inventories

C. Pisani[(a)], S. Franceschi[(b)]
[(a)]University of Siena, Siena, [(b)]University of Calabria, Rende

Tree-outside-forests (TOF) play important roles in national economies, ecosystems and international efforts for sustainability. Recently, international programmes on global forest resource assessments recognize the TOF role not only in sustainable land management, but also in carbon sequestration, thus allowing for climate changes mitigation. Indeed, in the past, TOF tend to be left out of natural resource assessment, but in 1996 a FAO expert meeting addressed the need of data on TOF and, as a result, the FAO Forestry Department included TOF in the National Forest Monitoring and Assessment Programme. From this perspective, a large-scale investigation was planned in 2008 for the Molise Region, an administrative district of approximately 450000 ha located in Central Italy. The first years were devoted to the collection and storage of 1 m geometric resolution aerial ortho-photos, covering the entire study area and allowing the identification of small groups of trees by visual onscreen analysis. The entire collection of 52796 TOF units scattered throughout the region were subsequently identified. Clearly, the very large population size rendered any complete investigation prohibitive. In 2013, a two-phase sample survey was planned to estimate some attributes of interest and diversity measures, such as

the abundance vector, the basal area, the volume and the carbon storage. In the first phase, the Molise Region was covered by a grid of 4693 hexagons of size 100 ha, and a point was randomly selected within each hexagon. Then, to reduce the sampling effort, a sample of 2347 points/hexagons were selected in a second phase by means of the one per stratum stratified sampling. Finally, in accordance with the point sampling protocol, all of the TOF units intercepted at least by one of these 2347 points were sampled and visited on the ground.

Assessing the Role of the Spatial Scale in the Analysis of Lagoon Biodiversity. A Case-study On the Macrobenthic Fauna of the Po River Delta.

A. Pollice[(a)], É. Marcon[(b)], E. Fano[(c)]

[(a)]Università degli Studi di Bari Aldo Moro, Bari, [(b)]AgroParisTech, Paris, [(c)]Università degli Studi di Ferrara, Ferrara

The analysis of benthic assemblages is a valuable tool to describe the ecological status of transitional water ecosystems, but species are extremely sensitive and respond to both microhabitat and seasonal differences. The identification of changes in the composition of the macrobenthic community in specific microhabitats can then be used as an "early warning" for environmental changes which may affect the economic and ecological importance of lagoons, through their provision of Ecosystem Services. From a conservational point of view, the appropriate definition of the spatial aggregation level of microhabitats or local communities is of crucial importance. The main objective of this work is to assess the role of the spatial scale in the analysis of lagoon biodiversity. First, we analyze the variation in the sample coverage for alternative aggregations of the monitoring stations in three lagoons of the Po River Delta. Then, we analyze the variation of a class of entropy indices by Bayesian mixed effects models, properly accounting for the fixed effects of biotic and abiotic factors and random effects ruled by nested sources of variability corresponding to alternative definitions of local communities. Finally, we address biodiversity partitioning by a generalized diversity measure, namely the Tsallis entropy, and for alternative definitions of the local communities. The main results obtained by the proposed statistical protocol are presented, discussed and framed in the ecological context.

Forecasting of Cases for Climate Sensitive Mosquito-borne Diseases Using Traditional and Online-media Information

L. Ramirez Ramirez[(a)], Y. Gel[(b)], V. Lyubchich[(c)]

[(a)]Centro de Investigacion en Matematicas, Guanajuato, [(b)]University of Texas at Dallas, [(c)]University of Maryland

Emerging and re-emerging climate-sensitive viruses threaten not just a single state or countries, but whole macro-regions and continents. In December 2013 and April 2015 the first cases of chikungunya and zika were reported in the Caribbean and Brazil, respectively. Since then these viruses rapidly spread across the continent attracting a lot of attention from governments and health care professionals. Since data of new diseases in a region is scared, we exploit different source of information such as the originated from surveillance systems, and non-traditional information sources (like online and social media) to propose a forecasting model. In this work we present a forecasting model for chikungunya. This model incorporates information of the number of cases at the beginning of an outbreak and the activity reported by Google Dengue Trend in Mexico. Since this chikungunya virus is transmitted by the same type of mosquito as dengue, we include Google Dengue Trend in Mexico as a proxy for the mosquito population and the mosquito-human interaction in the neighboring countries. The two information sources are incorporated as exogenous covariates of a time series model to predict the epidemic curve.

Characterization of Spatial and Temporal Variability of Exposure To Air Pollution in Epidemiological Studies

A. Ranzi[(a)]

[(a)]Regional Agency for Prevention, Environment and Energy of Emilia-Romagna, Modena

Exposure to chemical and physical agents in the environment can produce a wide range of adverse health consequences. Health effects of air pollution, especially particulate matter (PM), have been widely investigated. Most of the studies rely on few monitors located in urban areas for short-term assessments, and modelling techniques for long-term evaluations, again mainly at city level. The development of models to assess variability of air pollution exposures within cities for assignment to subjects in health studies has been identified as a priority area in the recent years. Geographic information systems (GIS) are being used with increasing frequency in environmental epidemiology studies. Different multidisciplinary approaches and data have been used in this field, ranging from chemical transport model, geostatistics, satellite data. Spatial and temporal variability of air pollution is relevant at different scale; different health outcomes required specific time windows of exposure. As a consequence epidemiological research can benefit from an accurate characterization of both spatial and temporal variability of exposure to air pollution. Advances in the field of exposure assessment of population to air pollution will be reviewed.

Some experiences at national and local level will be presented, together with together with some example on how the quality of the assessment can influence the results of an epidemiological study.

An Integrated Framework for Analyzing Spatially Correlated Functional Data

S. Ray[(a)]

[(a)]University of Glasgow, Glasgow

Datasets observed over space and time have become increasingly important due to its many applications in different fields such as medicine, public health, biological sciences, environmental science and image data. Both spatiotemporal methods and functional data analysis techniques are used to model and analyse these types of data considering the spatial and temporal aspects. In this talk we will present an integral framework for modeling and analysing functional which are spatially correlated. In particular we wish to integrate existing approaches and identify gaps for analyzing a wide variety of spatially correlated functional data and provide the practitioner with objective choices to identify the best method to analyze their data.

Inference Methods for Asymptotically Independent Samples of Extremes

S. Rizzelli[(a)]

[(a)]Bocconi University, Milan

Inferring the dependence among extremes is a crucial aim of multivariate Extreme-Value (EV) analysis. Classical EV theory provides asymptotic models for the tail probability of multivariate distributions that are in the domain of attraction of the so-called multivariate EV distribution. The latter is the limiting distribution for the normalized component-wise maxima. When the copula of the limiting distribution represents independence we call this case "asymptotic independence". In the last few decades, new characterizations of the dependence form in the case of asymptotic independence have emerged, and a new theory on tail probabilities has been established. Such theoretical developments are particularly useful for analyzing data that exhibit a positive dependence which, however, is mitigated at extreme levels. In the first part of the talk, we introduce a dependence function which parallels the classical Pickands function but suits asymptotically independent extremes. We propose: a semi-parametric estimator of it and we discuss its large sample properties; a test for checking if the assumption of asymptotic dependence holds. In the second part of the talk, we focus on the so-called ray dependence (RD) function, which allows to infer the probability of falling in a critical region where at least one variable is extremely large (more than the observed values), under asymptotic independence. We propose for the bivariate case a Bayesian nonparametric inferential approach that allows to simultaneously estimate the RD function and the rate of decay of

the joint tail probability, the so-called coefficient of tail dependence. We discuss the properties and illustrate the performances of the proposed method. Finally, we discuss other functional methods for estimating the probability of observing values in the aforementioned critical regions.

Environment and Causal Inference

J. Robins[(a)]

[(a)]Harvard University, Cambridge

In this talk I review progress made in specific areas of environmental science using formal causal inference methods. I will discuss the healthy worker survivor effect in occupational health, the effect of air pollution on mortality, and the estimation of the number of lives potentially saved by environmental intervention

---

Spatial Clustering of LISA Functions: a Functional Approach

F. Rodríguez-Cortés[(a)], E. Romano[(b)], J. González[(a)], J. Mateu[(a)]

[(a)]University Jaume I, Castellon, [(b)]Universitá della Campania "Luigi Vanvitelli", Caserta

The second-order product density function provides a global measure of the covariance structure by summing up over the contributions from each event observed in the process. However, pair correlation functions can be interpret as their standardised version and they can be used as a tool to discriminate amongst several spatial point pattern structures taking into account their intrinsic inhomogeneity. Individual contributions of the global estimator for the pair correlation function can be considered as a measure of clustering and can be used as an analytic exploratory data tool to examine individual points in a point pattern in terms of how they relate to their neighbouring points. LISA functions can be collected discretely over several distances, either on a fixed or random spatial distance grid and can be considered as a functional entity. In this paper, we propose a functional approach for classifying events with similar local structure by means of LISA functions. The key idea is to work with the estimated pair correlation LISA functions at each event of a point pattern, defining a functional marked point process. We then extend a depth based clustering method to the case of spatially correlated LISA functions, using an appropriated depth-based distance. Performances of the strategy are evaluated through a real and a simulation study.

Recursive Structure of Multivariate Gauss-Markov Random Fields

L. Romagnoli[(a)], L. Ippoliti[(b)], R. Martin[(b)]
[(a)]University of Molise, Campobasso, [(b)]University "G. d'Annunzio", Chieti-Pescara

This work addresses the problem of ML estimation and other inferential issues for multivariate conditional autoregressive CAR models, also known as Gauss-Markov random fields - GMRFs. We assume to work in situations in which the set of observations can be large, can have missing data and can be also corrupted by additive noise. Examples include data from agricultural field trials, important environmental data sources, such as numerical climate or weather model output, and many other gridded data products derived from satellite observations. We discuss a special class of algorithms for dynamic models, based on the Kalman filter, and consider a state space formulation which is related to two equivalent ("backward" or "forward") one-sided representations of a GMRF. These representations are intuitive, and provide a simple and efficient recursive method for calculating the likelihood function under different conditions. A particular attention will be devoted to parametric estimation in presence of missing data. The work concludes with a discussion of simulation studies on parameter estimation, and an application of smoothing and missing data reconstruction procedures to real data.

Functional Depth Function for Seismic Waves Ordering

E. Romano[(a)], R. Verde[(a)], A. Balzanella[(a)]
[(a)]Università degli Studi della Campania "Luigi Vanvitelli", Caserta

This paper introduces a new strategy for ordering misaligned functional data through depth functions. While the literature on depth functions for functional data has become quite extensive in recent years, no proposal addresses this challenge when data are misaligned. Our idea is to introduce a new depth function, based on the established proposals by López-Pintado and Juan Romo (2009), which includes a registration step in the depth measurement. This allows to get a curve ordering, as well as, a median function which account for the different time synchronization in the data. In order to show the effectiveness of our proposal, we will perform some test on seismic waveforms recorded by seismographs. This is motivated by the interest in investigating for a median seismogram able to represent a seismic event in a geographic area keeping into account the time misalignment among the waveforms. We will still show, by means of functional visual instruments, how the method is still able to discover outliers and to highlight the variability among curves.

Covariance Models for Large Spatial Datasets and Their Inference

T. Romary[(a)]
[(a)]Mines ParisTech, Paris

Large spatial datasets are becoming ubiquitous in environmental sciences with the explosion in the amount of data produced by sensors that monitor and measure the Earthsystem. Consequently, methods for the geostatistical analysis of these data have to be developed. Richer datasets lead to more complex modeling but may also prevent from using classical techniques. Indeed, the kriging predictor is not straightforwardly available as it requires the inversion of the covariance matrix of the data. The challenge of handling such datasets is therefore to extract the maximum of information they contain while ensuring the numerical tractability of the associated inference and prediction algorithms.In this work, we will first provide an overview of the different approaches that have been developed in the literature to address this problem. They can be classified into two families, both aiming at making the inversion of the covariance matrix computationally feasible. The tapering approach circumvents the problem by enforcing the sparsity of the covariance matrix, making it invertible in a reasonable computation time. The second available approach assumes, on the contrary, a low rank representation of the covariance function. While both approaches have their drawbacks, we propose a way to combine them. The covariance model is assumed to have the form sparse plus low rank, both terms being possibly non stationary. The choice of the basis functions sustaining the low rank component is data driven and is achieved through a selection procedure. This model expresses as a spatial random effects model and maximum likelihood estimation of the parameters can be conducted through the expectation-maximization algorithm. A particular emphasis will be put on the inference of the covariance matrix of the spatial random effect.

Quantifying Random Uncertainty and Redundancy in Water Vapour Time Series Using Entropy Theory".

M. Rosoldi[(a)], F. Madonna[(a)], G. Pappalardo[(a)]
[(a)]Consiglio Nazionale delle Ricerche

The potential for redundant measurements to reduce uncertainty in atmospheric variables has not been investigated comprehensively for climate observations. Moreover, comparisons among time series of in situ and ground based remote sensing measurements have been performed using several methods, but these are generally based on linear models. In this work, the concepts of entropy (H) and mutual correlation (MC), defined in the frame of the information theory, are applied to the measurements of essential climate variables with the aim of characterizing random uncertainty and redundancy in time series of co-located measurements provided by different surface-based sensors. In particular, time series of integrated water vapor (IWV) and water vapor mixing ratio profiles, obtained at five highly instrumented

GRUAN (GCOS, Global Climate Observing System, Reference Upper-Air Network) stations in the period from 2010-2012, are analyzed in terms of H and MC. Results show that the random uncertainties on the IWV measured with radiosondes, global positioning system, microwave and infrared radiometers, and Raman lidar measurements differed by less than 8%. Comparisons of time series of IWV from ground-based remote sensing instruments with in situ soundings showed that microwave radiometers have the highest redundancy with the IWV time series measured by radiosondes and therefore the highest potential to reduce the random uncertainty of the radiosondes time series. Moreover, the random uncertainty of a time series from one instrument can be reduced by ≈60% by constraining the measurements with those from another instrument. The best reduction of random uncertainty is achieved by conditioning Raman lidar measurements with microwave radiometer measurements. Specific instruments are recommended for atmospheric water vapour measurements at GRUAN sites. The methodology presented in this work for the study of water vapor redundant measurements can be applied to other climate variables.

Some New Developments in the R-INLA Project

H. Rue[(a)]

[(a)]King Abdullah University of Science and Technology, Thuwal

The R-INLA package for doing approximate Bayesian inference for the class of latent Gaussian models has been highly successful and also proven useful for the statistical analysis of environmental data. In this talk, I'll give an overview of new and ongoing developments in the R-INLA project, that we hope users of the package will find useful!

Modeling Spatially Dependent Functional Data By Regression With Differential Regularization

L.M. Sangalli[(a)], M.S. Bernardi[(a)]

[(a)]Politecnico di Milano, Milano

I will present a class of models for the analysis of functional data with complex dependencies, such as spatially dependent curves and time dependent surfaces, possibly observed over complex domains. The models are based on the idea of regression with partial differential regularizations. Among the various modeling features, the proposed method is able to deal with spatial domains featuring peninsulas, islands and other complex geometries. Moreover, it can comply with specific conditions at the boundaries of the problem domain, which is fundamental in many applications to obtain meaningful estimates. The proposed models have the capacity to incorporate problem-specific prior information about the spatial structure of the phenomenon under study and the texture of the domain, formalized in terms of a partial differential equation; this very flexible modeling of space-time variation

allows to naturally account for anisotropy and non-stationarity. Space-varying covariates are accounted for in a semi-parametric framework. The use of advanced numerical analysis techniques makes the models computationally highly efficient. The methodology is illustrated via environmental applications.

Analysis of Extremes in Climate Model Predictions: Assessment and Comparisons

B. Sanso[(a)], M. Warner[(a)]

[(a)]University of California, Santa Cruz

Evidence from climate model simulations indicates that, with high likelihood, future climates will be subject to an intensification of extreme weather events, like, for example, droughts, hurricanes, dry spells, or episodes of very high temperature. Gauging the probability of occurrence of such events in the future is very important for the assessment of the risks associated with climate change. Such probabilities should be a key component of a sound process of decision making regarding possible adaptation policies. A first step in the quantification of the likelihood of extreme events, as predicted by climate model simulations, is the assessment of the ability of different types of simulations to produce those events, when compared to historical records. In this work we develop hierarchical models to perform threshold based extreme value analysis in multivariate settings to compare the probabilistic behavior of ensembles of global climate model simulations. Our work uses long-term and short-term simulations obtained from the repository of the Coupled Model Intercomparison Project Phase 5 (CMIP5). Together with the assessment of the extremes of the different models, we obtain a unified inference that can be used for the probabilistic quantification of extreme temperature and rainfall events.

The Challenge of Global Water Supply With Time-series Treatment Forecasting

J. Santacruz Cano[(a)]

[(a)]Stock Exchange Studies Institute of Spain

Water supply has become a key element in the actual economic and social debate. Most concretely, water is one of the five main risks that the global economy faces. This study deals with the impact of global water supply over critical economic zones, its relationship with the commodities market (under a "climate change" scenario) towards a time-series treatment forecasting perspective. The dynamics of water and the rest of natural commodities can be studied under the Theory of "Solar Spots" stated by William Stanley Jevons in 1870 that was the first economic interpretation of the Real Business Cycle Theory. Some economic processes follow natural dynamics both in the earth and outside. Water is also a good example of these natural movements, which can be sized by advanced time-series econometric tools. Productivity gains in agriculture and industries, mining and the behavior of non-human provoked climate change, are key factors influenced by water supply.

Joint Point Pattern Modeling for Species Co-occurrence Using Camera Trap Data

E.M. Schliep[(a)], A.E. Gelfand[(b)], D. Taylor-Rodriguez[(c)], J.S. Clark[(b)]
[(a)]University of Missouri, Columbia, [(b)]Duke University, Durham, [(c)]Michigan State University, East Lansing

Camera traps, which activate and record times when an animal crosses in front of the camera, can be used to investigate daily activity patterns of species. Traditional approaches "wrap time" and view the observations from the camera traps as circular data. In circular time, two species can have very similar daily activity patterns and yet never co-occur in a proximate sense. Modeling species activity patterns in linear time preserves this notion of co-occurrence. We propose a multivariate log-Gaussian Cox process to model daily activity patterns of forest mammals. Species-specific intensities are modeled jointly to capture general activity patterns of the species while accounting for possible co-occurrence or competition between the species. Model inference is obtained in a Bayesian framework with an efficient Markov chain Monte Carlo sampling algorithm. Overlap measures using the intensity functions can detect similarities in daily activity patterns. We can infer about the probability of presence of one species in a particular time interval given presence of another species in the same or adjacent interval, addressing the question of proximate co-occurrence.

Spatio-Temporal Multivariate Shared Component Modeling for Cancer Data

V. Schmid[(a)], B. Mahaki[(b)]
[(a)]Ludwig Maximilian University, Munich, [(b)]Isfahan University of Medical Sciences, Isfahan

Joint disease mapping has started to gain attention in the last decade and several methods have been proposed to map two or more diseases jointly. Among the proposals for joint disease mapping, the shared component model has become more popular and several types of this model for different data structures have been introduced. Another recent advance to strengthen inference of disease data has been the extension of purely spatial models to include time and space-time interaction. Such analyses have additional benefits over purely spatial models. However, only a few proposed spatio-temporal models could address analysing multiple diseases jointly. To this end, we present a model which combines the benefits of shared-components with spatio-temporal techniques for multivariate data. In the proposed model, each component is shared by different subsets of diseases, spatial and temporal trends are considered for each component, and the relative weight of these trends for each component for each relevant disease can be estimated. We present an application of the proposed method on incidence rates of seven prevalent cancers in Iran. We show how the model allows to analyse geographical and temporal variation among diseases beyond previous approaches.

Patterns, Missingness and Structure in Spatio-temporal Data

M. Scott[(a)]
[(a)]University of Glasgow, Glasgow

New forms of environmental data, such as those from small, low cost (fixed or mobile) environmental sensors, or from citizen science observations introduce interesting statistical challenges related to lack of structure. The changing nature of much of the data generation processes means that the statistical sampling basis of these new data streams may no longer be consistent with the tradition of statistical sampling (random or otherwise)). The data representativity, quality, and use with fixed site monitoring requires adapted statistical methodology and data fusion models to be developed. From a series of case studies covering air pollution, soil quality, freshwater quality in rivers and global lake attributes including temperature, different statistical approaches to handling such data will be illustrated and some of the challenges explored.

Wildfire Occurrence in Italy: Exploring the Role of Socio-economic, Urban Development and Environmental Factors

L. Secondi[(a)], T. Laureti[(a)], A. Tomao[(a)], L. Mancini[(a)], P. Corona[(a)]
[(a)]University of Tuscia, Viterbo

Wildfires are major disturbances in Mediterranean ecosystems worldwide. Various causes including biophysical and anthropic drivers have been proved to affect the probability of fire occurrence. Furthermore, the vast majority of wildfires are human-induced, particularly in the Countries around the Mediterranean basin. Despite its importance, the influence of human factors on the spatial and temporal patterns of wildfire needs to be better understood. The aim of this study is to explore the drivers of wildfire occurrence and density by disentangling the effect of socio-economic, urban development and environmental variables. A rich longitudinal data set of Italian provinces for the years 2009-2014 is constructed by integrating wildfire occurrence data with information on the socio-economic, demographic, land use and landscape characteristics taken from various official sources. In order to capture socio-economic and demographic changes of each area and to properly consider the heterogeneity of land and urban development within regions, fixed and random effects models are used. Moreover, dynamic panel data models are applied to allow wildfire occurrence to depend on past wildfire events. Land-use and economic intensity, population density and social conditions of inhabitants proved to significantly influence wildfire occurrence. These results enable us to discuss the relationships between wildfire and human development activities as a suitable tool to support the planning of effective fire prevention policies.

Bayesian Estimation of Global Glacial Isostatic Adjustment for Sea Level Rise Re- Evaluation

Z. Sha[(a)], M. Schumacher[(a)], J. Rougier[(a)], J. Bamber[(a)]
[(a)]University of Bristol, Bristol

Glacial isostatic adjustment (GIA) is the viscous response of the solid Earth to past ice loading, primarily during the last Glacial period. It plays a crucial role on changing regional sea level and is, therefore, important for predicting future trends. In general, estimates of the GIA process have been obtained from numerical model simulations (forward models) with assumptions about the Earth structure and ice loading history. These models often lead to significant regional discrepancies due to incomplete information on the true 3-D Earth properties and ice loading history. In this study, we propose an inverse, data-driven approach to provide a GIA estimate that synthesises global model solutions with GPS observations of vertical land motion. GIA is treated as a time invariant Gaussian process on the sphere, with a local Matern covariance function. Following a full Bayesian approach and the principle of stable inference, we use a GIA estimate derived from the forward model ICE-6G (VM5a) as the prior mean, and parameterise the covariance function in terms of a variance and a correlation length. For computational feasibility, we use a sparse matrix representation induced by a Gaussian Markov random field (GMRF) approximation, and a parallel updating algorithm which exploits the conditional independence structure of the GMRF. The result is a new global GIA reconstruction, with a full assessment of uncertainty.

From In-situ To Continuous Modelling of Mosquito Biting Rates

A. Shcherbacheva[(a)], H. Haario[(a)]
[(a)]Lappeenranta University of Technology, Lappeenranta

A discrete agent-based approach enables to study dependences of mosquito-human contact rates and mosquito mortality on various socio-economic and biological factors e.g., coverage with control interventions, typical household size and alterations in mosquito behavior resulting from malaria parasite. The model was initially calibrated with hut-level experimental data and further extended for community-level situation. Model simulations are typically performed at a short time scale of one night, but can be extended to longer periods using continuous mathematical models of malaria transmission. This approach allows quantifying the incidence of malaria for large variety of transmission scenaria.

Rarefaction Techniques and Species Richness in Biodiversity Studies

H. Shimadzu[(a)]

[(a)]Loughborough University, Loughborough

The number of species, namely species richness, is primary account for biodiversity in ecological communities, but also one of the difficult indicators to deal with, despite our familiarity. Difficulties stem from the fact that the number of species we acquire, observed richness, is non-linearly related to other factors, such as the number of individuals, the size of a study area and the completeness of survey. When we compare observed richness of different sites/occasions, it becomes a challenging task since the extent to which each survey can cover regarding space and species may vary among different study sites/occasions. A standardisation method is therefore required for a fair comparison of richness. A common technique called rarefaction has been widely adopted in quantitative ecology; it compares observed richness conditioning upon an equalised number of individuals. This talk discusses an issue that the rarefaction technique may lead us to an unfair comparison of richness, inducing a bias in the observed richness of rarefied samples. In particular, an emphasis is given upon highlighting the bias induction mechanism, illustrating an example data set from biodiversity research.

Multi-resolution Clustering for Dependent Functional Data With Applications To Climate Reconstruction

S. Sjostedt de Luna[(a)], K. Abramowicz[(a)], L. Schelin[(a)], J. Strandberg[(a)]

[(a)]Umea University, Umea

A multi-resolution clustering approach is presented to detect latent groups from observed dependent functional data. Given a lattice of (time) points, a function is observed at each grid point. We assume that the latent unobservable groups vary slowly over time (the lattice). We consider the case when at different time scales (resolutions) different groupings arise, with groups being characterized by distinct frequencies of the observed function-types. We propose and discuss a non-parametric double clustering method, which identifies latent groups at different scales. We present an application of the introduced methodology to the annual seasonal patterns of (varved) lake sediment data, aiming at reconstructing winter climate regimes in northern Sweden at different resolutions during the last six thousand years.

A Dispatched-by-design Architecture for Distribution Systems Based On Exploiting Local Flexibility

F. Sossan[(a)]

[(a)]École Polytechnique Fédérale de Lausanne, Lausanne

Increasing electricity storage capacity is a key factor to increase the proportion of production from stochastic renewable resources, allowing to restore an adequate level of controllability after the displacement of conventional controllable generation, as an alternative to large-scale deployment of fast ramping generating units Electricity can be stored either in two ways: either indirectly, by shifting the consumption of so-called flexible demand, or directly, by implementing storage technologies, like batteries, power to gas and hydro-pumping. An important concern when dealing with distributed storage is the large number of units to control and the fact that they are deployed at the low voltage (LV) level. A comprehensive decision problem accounting for these elements would quickly become intractable due to many decision variables and grid constraints. In this work, we describe the concept of dispatched-by-design operation. It consists in exploiting local flexibility with the objective of dispatching the power flow of certain portion of the network according to a profile, called dispatch plan, established the day before the operation thanks to applying forecasting tools for the local prediction of electricity flows. Such a configuration results in a very tractable formulation because all the complexity of the problem is masked behind the commitment of the operator to track a dispatch plan (which ideally could be submitted to an electricity market) and LV grid constraints are kept local. Experimental results from the EPFL campus are presented, where the dispatchability of a 20 kV medium voltage distribution feeder is achieved by the joint control of a 0.5 MWh lithium battery energy storage system and a smart building equipped with a shiftable electric space heating system.

On the Use and Applicability of the Modified Omori Law

I. Spassiani[(a)], W. Marzocchi[(a)]

[(a)]Istituto Nazionale di Geofisica e Vulcanologia

The Omori-Utsu law, or Modified Omori law (MOL), is often used to describe the temporal decay of the aftershocks in a seismic sequence. It has been proposed in its first version by Omori to describe all the aftershocks following the 1981 Great Nobi earthquake of magnitude 8. Since then, MOL has been embedded in the Epidemic Type Aftershock Sequence (ETAS) models - which are the most reliable earthquake forecasting models in the short-term - to model the temporal correlation of the first generation of aftershocks. We argue that these two applications of MOL are mathematically incoherent. In particular, the distribution of the whole sequence of aftershocks starting from a fixed background event is given by the convolution power series of the first generation shocks' law; interestingly, if the latter is the modified Omori law as in the ETAS models, nothing ensures that the final total temporal rate is again a MOL.Here we analyze the problem in two directions. First

we impose the first generation shocks' law to be a MOL and search for the resulting total distribution that should describe the whole sequence. Second, we impose that total distribution is a MOL, and look for a suitable first generation events' law. Finally, the theoretical calculations are compared with the magnitude distribution of high quality seismic catalogs.

An Anisotropic and Inhomogeneous Hidden Markov Model for the Classification of Water Quality Spatio-temporal Series

L. Spezia[(a)]

[(a)]Biomathematics & Statistics Scotland

A case study about the evaluation of the water quality dynamics in each of the 56 major catchments in Scotland, for a period of 10 years, is presented. Data are obtained by monthly sampling of water contaminants, in order to monitor discharges from the land to the sea. We are interested in the multivariate time series of ammonia, nitrate, and phosphorus. The time series may present issues that make their analysis complex: non-linearity, non-normality, weak dependency, seasonality, and missing values. The goals of this work are the classification of the observations into a small set of homogeneous groups representing ordered categories of pollution, the detection of change-points, and the modelling of data heterogeneity. These aims are pursued by developing a novel spatio-temporal hidden Markov model, whose hierarchical structure was motivated by the data set to study: the observations are displayed on a cylindrical lattice and driven by an anisotropic and inhomogeneous hidden Markov random field. As a result, four hidden states were selected, showing that catchments could be grouped spatially, with a strong relationship with the dominating land use. This work is based on Spezia, Brewer, and Birkel (Environmetrics, 2017).

Prediction of Spatio-Temporal Data: A Comparative Study of Two Kriging Approaches

J. Strandberg[(a)], S. Sjostedt de Luna[(a)], J. Mateu[(b)]

[(a)]Umea Univeristy, Umea, [(b)]Universitat Jaume I, Castellòn de la Plana

Two approaches to predicting spatio-temporal data, by functional kriging (FK) or by spatio-temporal kriging (SPTK), are compared. The performance of the two kriging approaches is evaluated through a simulation study, with data sets covering a variety of different scenarios. For small sample sizes FK typically performs better. For larger sample sizes, SPTK starts to perform better for the stationary spatio-temporal processes, whereas FK continues to work best for the stationary functional random processes. In addition, two applications on real data, one based on salinity curves and the other on Canadian weather data, are also presented.

Adaptive Ensemble Kalman Filters for Online Bayesian State and Parameter Estimation

J. Stroud[(a)]

[(a)]Georgetown University, Washington

This paper proposes new methodology for sequential state and parameter estimation within the ensemble Kalman filter. The method is fully Bayesian and propagates the joint posterior density of states and parameters over time. In order to implement the method we consider two representations of the marginal posterior distribution of the parameters: a grid-based approach and a Gaussian approximation. Contrary to existing algorithms, the new method explicitly accounts for parameter uncertainty and provides a formal way to combine information about the parameters from data at different time periods. The method is illustrated and compared to existing approaches using simulated and real data. This is joint work with Matthias Katzfuss (Texas A&M University) and Christopher Wikle (University of Missouri).

Visualization and Assessment for Properties of Spatio-temporal Covariances

Y. Sun[(a)], H. Huang[(a)]

[(a)]King Abdullah University of Science and Technology, Thuwal

Spatio-temporal covariances are important in modeling geostatistical data and are often used to describe the spatio-temporal variability of the underlying random process. For second-order stationary processes, there exist subclasses of the covariance function by assuming simpler spatio-temporal dependence structure, such as separability and full symmetry. However, it is challenging to visualize and assess such properties from spatio-temporal observations. In this work, we propose a functional data analysis approach by constructing test functions with cross-covariances of each pair of the time series observed from spatial locations. These test functions are functions in time and summarize the property of separability or symmetry for given distances. We then apply to functional boxplot to visualize the functional median and the variability of the test functions, where the extent of departure from zero at all temporal lags indicates the degree of non-separability or asymmetry. We also develop a rank-based nonparametric testing procedure for assessing the significance of the non-separability or asymmetry. The performance of the proposed methods is examined by simulations for various commonly used spatio-temporal covariance models. As applications, we apply our method to different types of real datasets, including weather station data and climate model output.

Inferring Accents From Geo-localized Acoustic Speech Recording II

S. Tavakoli[(a)], D. Pigoli[(a)], J. Aston[(a)], J. Coleman[(b)]

[(a)]University of Cambridge, Cambridge, [(b)]University of Oxford, Oxford

The goal of our project is the modelling the variations of speech and accents within England, using acoustic speech recordings from the Audio Edition of the British National Corpus. We propose a non-isotropic spatial functional mean and covariance regression model, in which sound recordings of a particular vowel of interest - encoded by their Mel-Frequency Cepstral Coefficients (MFCC) - are modelled as a spatial mean plus an error term with varying spatial covariance. Due to the particular form of the recordings available, where multiple recordings are available for a same location, weare able not only to estimate the spatial mean MFCC, but also the spatial covariance field of the error terms, without isotropic assumptions. The mean field is fitted bya local constant smoother, and the covariance fields are fitted using a local constant smoother in the tangent plane. Both are fitted using a variable bandwidth parameter chosen in a data driven way. The output of the estimation procedure can then be visualized using an appropriate dimension reduction, and can also be resynthesized into sounds, which can provide qualitative insight into the proposed model.

Spatio-temporal Models With Space-time Interaction and Their Applications To Air Pollution Data

R. Tsay[(a)]

[(a)]University of Chicago, Chicago

It is of utmost importance to have a clear understanding of the status of air pollution and to provide forecasts and insights about the air quality to the general public and the environmentalists. Previous studies of spatio-temporal models already showed that even short-term exposure to high concentrations of fine atmospheric particulate matters can be hazardous to the health of common people. In this study, we develop a spatio-temporal model with space-time interaction for air pollution data (PM2.5). The proposed model uses a parametric space-time interaction component along with the spatial and temporal components in the mean structure, and introduces a random-effects component specified in the form of zero-mean spatio-temporal processes. For application, we analyze the hourly air pollution data (PM2.5) from Beijing, Chengdu, and Shanghai − three major cities in China and another 15 states from the southern region of Taiwan.

A Bayesian S-FVAR Model To Estimate the Short-term Effects of Air Pollution On Human Health

P. Valentini[(a)], L. Ippoliti[(a)], L. Fontanella[(a)], D. Gamerman[(b)]

[(a)]University G. d'Annunzio Chieti-Pescara, Chieti, [(b)]Universidade Federal do Rio de Janeiro, Rio de Janeiro

There is an extensive and continuing concern over the human health effects of air pollutants. In this work, motivated by a case study on daily hospitalizations for cardiovascular and respiratory diseases in Piemonte and Lombardia regions (Italy), we introduce a Bayesian spatial factor vector autoregressive (S-FVAR) model to study the spatial and temporal association existing between health data and air pollution. The model is developed for handling measurements belonging to the exponential family of distributions and allow the spatial and temporal components to be modeled conditionally independently via a latent factor analysis structure for the (canonical) transformation of the measurements mean function. The proposed model has an intuitive appeal and enjoys several advantages. For example, it describes the spatial-temporal variability of the disease risk and explicitly defines a non-separable spatio-temporal covariance structure of the process. Also, it allows to study how the disease risk at a specific areal unit reacts over time to exogenous impulses from the same or different areal units. Finally, several general structures that make use of different covariate information, can be easily accommodated in the different levels of the hierarchy. The S-FVAR model is developed within a state-space framework and full probabilistic inference for the parameters is facilitated by a Markov chain Monte Carlo (MCMC) scheme for multivariate dynamic systems.

Spatial Hybrid Downscaling of Extreme Precipitation Fields

M. Vrac[(a)], A. Bechler[(b)], L. Bel[(c)]

[(a)]Laboratoire des Sciences du Climat et de l'Environnement - CNRS, [(b)]France Telecom Orange Labs, [(c)]AgroParisTech, INRA

As extreme climate events are rare and often unexpected, society is not well prepared to face them. It is therefore important to evaluate their evolutions in the future. Global Climate Models (GCMs) are generally used to provide future scenarios of precipitation. However, their resolution (200km) is not sufficient to describe efficiently what happens at local scales (a few kilometres or less). Hence, dynamical and statistical downscaling methods have been developed to generate local-scale values of climate variables, based on large-scale information. Nevertheless, both the extreme behaviours and the spatial structures are not always well described by these methods. Moreover, most of the statistical downscaling models cannot simulate values at locations where there is no observed data for calibration. Max-stable processes now emerge as a powerful tool for the statistical modelling of spatial extremes. One recent development in this context is for conditional simulations of spatial fields of maximum values, providing empirical distributions of maximum value at any location in a given region, conditioned by observed values at other locations. We

propose a two-step methodology, called "Spatial Hybrid Downscaling" (SHD), to tackle the problem of spatial downscaling for fields of extreme precipitation. The first step consists in applying a univariate statistical downscaling at given locations. Once this 1d-link is performed, a conditional simulation of max-stable processes is adapted to a flexible process: the extremal t process. This enables us to get conditional distributions and to downscale extreme precipitation values at any point of the region, even in a climate change context.

Spatial Semiparametric Models for Data Over Complex Domain

G. Wang[(a)]

[(a)]College of William & Mary, Williamsburg

In this work, we focus on the variable selection techniques for a class of semiparametric geospatial regression models which allow one to study the effects of covariates in the presence of the spatial information. The smoothing problem in the nonparametric part is tackled by means of bivariate splines over triangulation, which is able to deal efficiently with data distributed over irregularly shaped regions. In addition, we develop a unified procedure for variable selection to identify significant covariates under a double penalization framework, and we show that the penalized estimators enjoy the oracle property. The proposed method can simultaneously identify non-zero spatially distributed covariates and solve the problem of "leakage" across complex domains of the nonparametric part. To estimate the standard deviations of the proposed estimators for the coefficients, a sandwich formula is developed as well. In the end, two simulation examples and a real data analysis are provided to illustrate the proposed methodology.

Planning Environmental Observing Systems of the Future: the Role of Statistics

E.C. Weatherhead[(a)]

[(a)]University of Colorado, Boulder

Historically, statistical expertise is brought in after a good environmental data set has been collected and scientific answers are desired. Often, the results have more uncertainty than was expected or there are retrospective desires for additional data. With enough collaboration between the data analysts and scientists, we can now work together to plan observing systems that will successfully and efficiently allow us to address the most pressing science questions. We can address multiple questions about observing systems including: which locations? With which accuracy and frequency? And what additional observations? These questions need to be addressed for observing systems supporting weather, water, pollution and climate science challenges. Optimization techniques can be employed to assure that robust

scientific results can be achieved. The desire to plan observing systems in an economic manner is expressed by scientists who want the answers and by governmental officials who want to make sure resources are spent wisely. Efforts in this area are beginning in several areas of environmental science; examples will be presented from weather, climate and energy systems. This more careful manner of planning observing systems bring environmental statistics from having a role at the end of data collection to being a partner in addressing environmental questions.

Detection of New Substantial Influences in Spatio-temporal Data

Y. Wu[(a)], X. Sun[(a)]
[(a)]York University, York

In some environmental problems, it is required to find out if there are new substantial influences in an area. We propose a change-point detection approach to tackle such problems. We first introduce a model for modelling the spatio-temporal data, which takes into account effects of different spatial location surroundings, seasonal cyclicities, temporal correlations among observations at the same locations and spatial correlations among observations from different locations. An algorithm is provided for estimating the model parameters. We then present a method for detecting the existence of new substantial influences in an area. Our method is demonstrated through both simulated and real data examples.

Regionalization of Multi-scaled Air Pollutants Based On Functional Principal Component Analysis

M. Xu[(a)], Y. Wang[(a)], Y. Wan[(a)]
[(a)]Peking University, Beijing

Emission patterns of air pollutants may vary across both time and space. They are closely related to local industry and energy consumption structures, but usually not directly observed. To make more effective environment policies, we need to better understand regional emission features. Conventional methods to regionalize air pollutants, such as Empirical Orthogonal Functions (EOF) or Self Organizing Maps (SOM), have limitations in combing information of multi-scaled pollution data. We propose a new method based on Functional Principal Component Analysis (FPCA), in which spatial and temporal correlations are simultaneously considered. The regionalization is realized by a likelihood-based clustering. We apply our method to analyze fine particulates over the North China Plain (NCP) from 2014 to 2016. Our findings indicate that different parts of NCP need different policies to control the air pollution.

A Leslie Matrix Simulation Model for Modelling Wildlife Age-specific Harvest Data

Y. Zhang[(a)]

[(a)]Acadia University, Wolfville

The Leslie matrix model is an age-specific matrix projection model for predicting population growth based on survival and fecundity information. To analyze the wildlife age-specific harvest data, we propose a stochastic Leslie matrix simulation model with harvest in order to generate data of the harvested wildlife population. Under this model, we study the ergodic properties and provide the method for estimating the population growth rate. Finally we apply our simulation model for evaluating two population reconstruction methods.

Matern Space-Time Correlation Functions

T. Zhang[(a)]

[(a)]Purdue University, West Lafayette

Construction of nonseparable and non-fully symmetric space-time correlation functions is essential and difficult in geostatistics. Considering that a fully symmetric space-time covariance function is inappropriate for many spatiotemporal processes, we focus on a way to construct a nonseparable and non-fully symmetric space-time correlation function from any given marginal spatial Matern and margianl temporal Matern correlation functions. Using the relationship between a spatial Matern correlation function and the characteristic function of a multivariate t-distribution, a modified Bochner's representation is proposed and a non-fully symmetric space-time Matern model is constructed. Based on a Taylor expansion, numerical values of the model can be readily obtained. Our numerical results show that the model is useful in accouting for non-fully symmetric dependence in spatiotemporal data analysis.

Integration of Different Sources of Environmental Data in Space and Time

J. Zhu[(a)]

[(a)]University of Wisconsin-Madison, Madison

Technological advances have substantially improved the data collection capacity in occupational exposure assessment. In this talk, we consider spatio-temporal methodology that combines data from two different sources and produces hazard mapping across space and over time in an indoor environment. A new model with continuous index in both space and time is proposed, and a profile likelihood based model fitting procedure is developed that allows fusion of the two types of data. Our methodology is applied to a case study conducted in an engine test facility and dynamic hazard maps are drawn. This is joint work with Tingjin Chu, Kirsten Koehler, Guilherme Ludwig, and Haonan Wang.

Testing Isotropy for Spatial Processes

A.G. Zini[(a)]

[(a)]University of Milano-Bicocca, Milano

Parametric tests for isotropy are proposed, by using of quasi-arithmetic means of spatial covariance functions. Thanks to composite likelihood approach, properties of such tests are checked in increasing domain asymptotic context.

Contributed sessions

Depth-based Methods for Clustering of Functional Data

G. Adelfio[(a)], F. Di Salvo[(a)], G. Sottile[(a)]

[(a)]University of Palermo, Palermo

The problem of detecting clusters is a common issue in the analysis of functional data and some interesting intuitions from approaches relied on depth measures can be considered for construction of basic tools for clustering of curves. Motivated by recent contributions on the problem clustering and alignment of functional data, we also consider the problem of aligning a set of curves when classification procedures are implemented. The variability among curves can be interpreted in terms of two components, phase and amplitude; phase variability, or misalignment, can be eliminated by aligning the curves, according to a similarity index and a warping function. Some approaches address the misalignment as a confounding factor, if it is not suitably taken into account; as opposed to treating phase variability as a nuisance effect, other approaches recognize that both amplitude and phase of curves contain cluster information. The search for suitable transformation of the original data involves the optimization of specific similarities between the warped curves and a natural consequence seems to incorporate the warping step in a clustering approach. Among the similarity indexes considered in the literature on functional data analysis, those defined via statistical depth provide a way to robustly cluster functional data. We implement a procedure exploiting the idea of functional depth, searching both for the set of optimal groups to obtain efficient aligning and clustering curves. We also try to deal with the implications of preprocessing the curves via a warping procedure or alternatively ignoring the misalignment in the further analysis, or explicitly recognizing it as a source of information for clustering. This approach provides an useful tool for analyzing many phenomena and in particular we apply it to seismic curves clustering. Paper supported by the national grant MIUR, PRIN-2015 program, Prot.20157PRZC4.

A New Forecasting Model for High-Dimensional Wind Speed, Wind Direction and Air Pressure Data

D. Ambach[(a)], W. Schmid[(a)]

[(a)]European University Viadrina, Frankfurt

The energy turnaround in Europe increases the importance of renewable energy. One important type of such energy is the wind power, which is mainly affected by the amount of wind speed. Moreover, it is important to locate the best conditions for building a wind farm and to generate much wind power. Therefore, we need to predict the wind direction as well. Moreover, the accurate wind direction could affect the produced wind power by means of the rotor efficiency. Besides, it is possible to move the turbine into the predominant wind direction. Another important trigger is the air pressure, which might be used as a kind of a predictive variable. Hence, it is necessary to predict the variables altogether. We propose a multivariate seasonal time varying threshold autoregressive model with interactions with a threshold seasonal autoregressive conditional heteroscedastic model that is able to predict all variables. This approach includes data mining algorithms to search for periodicity, conditional heteroscedasticity, interactions of different dependent variables and a complex autoregressive structure with non-linear impacts. If the estimation algorithm depicts a significant parameter, it is selected for the modelling and prediction approach, otherwise it is set to 0. The estimation of the model does not use the likelihood method, we apply a high-dimensional shrinkage technique, which allows us to obtain a solution without a distributional assumption for the dependent variables. Our model develops accurate predictions for the short-term. Furthermore, we obtain different new insides of the physical process of wind speed, wind direction and air pressure.

Sustainability Reporting and Social Variables: an Econometric Analysis

M. Antonicelli[(a)]

[(a)]Libera Università Mediterranea Jean Monnet, Casamassima

Sustainability reports are emerging as key documents to 1) signal commitment towards sustainability; 2) manage the dialogue with stakeholders; 3) disclose relevant non-financial information. Although voluntary, sustainability reporting (SR) is spreading rapidly and widely among companies all over the world, as more than two thirds of the S&P 500 publish a non-financial report. Nevertheless, companies face different institutional systems of norms, industry standards and cultural beliefs where they operate, while SR guidelines are defined by international organizations that seek homogeneity and standardization. The friction between the universality of SR guidelines and the difference in national institutional contexts is evidently a major dynamic influencing whether and how companies engage in SR. This research studies the influence of cultural and institutional variables on the publication SR compliant to the Global Reporting Initiative (GRI) G3 and G3.1 Guidelines, with

an A+ Application Level. The cross-country sample includes companies publishing their sustainability reports in the "GRI Sustainability Disclosure Database" from 2005 to 2012. The analysis entails two stages: in the first stage, we observed the Gini index concentration of the reporting activity, in order to give representation of its geographical spreading, then, in the second stage, we performed a multivariate correlation and regression as to indicate the main factors influencing the disclosure of corporate sustainability practices. Our results indicate that institutional, i.e. regulatory initiatives and national culture, are significant in explaining the number of GRI A+ reports published in each country.

A Time-dependent PDE Regularization To Model Functional Data Defined Over Spatio-temporal Domains

E. Arnone[(a)], L. Azzimonti[(b)], L.M. Sangalli[(a)], F. Nobile[(c)]

[(a)]Politecnico di Milano, Milano, [(b)]Università della Svizzera Italiana, Lugano, [(c)]École Polytechnique Fédérale de Lausanne, Lausanne

We propose a new method for the analysis of functional data defined over spatio-temporal domains. These data can be interpreted as time evolving surfaces or spatially dependent curves. The proposed method is based on regression with differential regularization. We are in particular interested to the case when prior knowledge on the phenomenon under study is available. The prior knowledge is described in terms of a time-dependent Partial Differential Equation (PDE) that jointly models the spatial and temporal variation of the phenomenon. We consider various samplings designs, including geo-statistical and areal data. We show that the corresponding estimation problem are well posed and can be discretized in space by means of the Finite Element method, and in time by means of the Finite Difference method. The model can handle data distributed over spatial domains having complex shapes, such as domains with strong concavities and holes. Moreover, various types of boundary conditions can be considered. The proposed method is compared to existing techniques for the analysis of spatio-temporal models, including space-time kriging and methods based on thin plate splines and soap film smoothing.

Sensitivity of Empirical Mode Decomposition and Its Application To Geophysical Data

M.Z. Bahri[(a)], J.J. Sharples[(a)]
[(a)]University of New South Wales, Canberra

Empirical mode decomposition (EMD) is a data-adaptive method that decomposes data (e.g. time series) into a number of constituent components called intrinsic mode functions (IMFs). EMD has an advantage over other methods such as Fourier series and wavelet analysis, as is it more suitable for analysis of non-stationary and non-linear data. EMD has found application in a number of geophysical areas, such as atmospheric science, meteorology, oceanography and climate change studies. In this study, the sensitivity of EMD is examined as a function of interpolation method and then applied to a range of non-stationary data types. The results show that EMD method can be sensitive to the type of non-stationarity of the input data and the interpolation method used in its implementation. This complicates interpretation of EMD output and can lower confidence in the information extracted as IMFs. Given the recent use of EMD in discerning long-term trends, and decadal and multidecadal variations in geophysical time series, it is important to gain a better understanding of the limitations of EMD and to investigate the potential for making the method more robust. This study introduced such a technique, particularly as it applies to extracting information on lower frequency variability and long-term trends. The techniques is illustrated by applying EMD to temperature, sea level, and forest fire danger time series. Moreover, we examine how EMD sensitivity may affect the interpretation of, and confidence in, the resulting long-term trend estimates in those time series.

Modeling Spatiotemporal Precipitation Variability and Glacier Mass Balance in High Mountain Asia

W.F. Christensen[(a)], C.S. Reese[(a)], S. Rupper[(b)], M.F. Christensen[(a)], B. Mabey[(a)], K. Larson[(a)]
[(a)]Brigham Young University, Provo, [(b)]University of Utah, Salt Lake City

With the exception of the earth's polar regions, the High Mountain Asia region (including the Tibetan Plateau) contains more of the world's perennial glaciers than any other. Sometimes called the "third pole" because of its massive storage of ice, High Mountain Asia (HMA) provides water to one-fifth of the world's population. Due to changes in precipitation patterns and temperatures warming faster in HMA than the global average, the region faces increased risk of flooding, crop damage, mudslides, economic instability, and long-term water shortages for the communities down-river. In this talk, we discuss a large, interdisciplinary, multi-institutional research project for characterizing climate change in HMA. We illustrate statistical approaches for extracting consensus estimates of spatiotemporally-correlated climate processes from a suite of climate model outputs and remote-sensing observations, and we discuss the uncertainty quantification needed to inform probability-based decision making.

Assessment of Climate Change By Modeling Long-term Temperature Data Using a Periodic State Space Model

M. Costa[(a)], M. Monteiro[(a)]

[(a)]University of Aveiro, Aveiro

In the last decades, the global warming has been presented on the political and scientific agenda because population have felt its consequences. On the one hand, the global warming is a global phenomenon but it is not reflected equally in every part of the globe or in a country. Thus, the warming phenomenon must be monitored in a smaller scale. This work analyzes long-term time series of monthly mean air temperatures in the three main Portuguese cities: Lisbon, Porto and Coimbra. Statistical time series models were developed in order to accommodate some particular characteristics of this type of time correlated data. In this work, we propose a periodic state space model, associated with an appropriate form of the Kalman filter. This framework allows identifying and quantifying the monthly warming rates taking into account the seasonal behavior and time correlation. Results show that there are different monthly warming rates and the greater annual mean rise in temperature was found in Porto with 2.17ºC whereas in Lisbon and Coimbra were, respectively, 0.62ºC and 0.55ºC per century.

An Improved Stochastic Optimization Method for Estimating Parameters in a Vegetation Model

U. Dahlén[(a)]

[(a)]Lund University, Lund

Processed-based dynamic vegetation models (DVMs) are important component of earth system modelling and important for assessing future climate impacts. The DVMs are often highly complex and non-linear with many tuning parameter. By matching model output to observed CO2 fluxes, we aim to estimate the parameters. Stochastic optimization methods have proved useful for similar optimization problems. These methods use Monte Carlo integration to propagate a probabilistic model that localise regions of high quality solutions. The integration requires multiple evaluation of the DVM. To reduce the number of cost evaluation we propose an improved optimization method that uses previous samples to improve convergence and reduce the number of function evaluations.

Feedback Mechanisms in Integrated Assesment Modeling

K. Dayaratna[(a)]

[(a)]The Heritage Foundation, Washington

Integrated Assessment Models, such as the DICE and FUND models, have been called upon extensively in public policy. Based on Monte Carlo simulation, these models are intended to quantify the economic damages associated with carbon dioxide emissions emitted several hundred years into the future. A neglected component of these models, however, is feedback regarding human response and adaptation. For example, if climate change were to indeed ultimately cause extreme damages decades or centuries into the future, then humans would almost surely attempt to mitigate such damages, which would in turn impact economic growth. We incorporate such a feedback mechanism into the DICE and FUND models that incorporates responses to extreme damages and estimate the associated social cost of carbon. We compare this estimate to those from existing models not incorporating such feedback and offer a variety of avenues for future research.

Clustering of Ocean Waves Using the Hierarchical Spectral Merger Algorithm

C. Euan[(a)]

[(a)]King Abdullah University of Science and Technology, Thuwal

We present a new method for time series clustering which we call the Hierarchical Spectral Merger (HSM) method. This procedure is based on the spectral theory of time series and identifies series that share similar oscillations or waveforms. The extent of similarity between a pair of time series is measured using the total variation distance between their estimated spectral densities. At each step of the algorithm, every time two clusters merge, a new spectral density is estimated using the whole information present in both clusters, which is representative of all the series in the new cluster. The method is implemented in an R package HSMClust. As an application of this method, an algorithm for determining stationary periods for time series of random sea waves is developed, a problem in which changes between stationary sea states is usually slow.

Modelling and Forecasting Air Quality Data With Missing Values Via Multivariate Time Series: Application To Madrid

C. García-Martos[(a)], M. Ramírez Jaén[(b)], M. Sánchez Naranjo[(c)]
[(a)]Universidad Politécnica de Madrid, Madrid, [(b)]Zabala Innovation Consulting, Madrid, [(c)]Escuela Técnica Superior de Ingenieros Industriales, Madrid

In this work we deal with real data obtained from the measurement network in the city of Madrid, where a key issue is the existence of missing data. Several approaches are used to face the problem: seasonal univariate ARIMA models, dimensionality reduction techniques such as the Dynamic Factor Model (DFM) as well as the state-space framework, which is specially useful for dealing with missing data. All the approaches are compared and validated in terms of modelling, interpretation of the estimated parameters as well as forecasting accuracy for many stations and pollutants. This is a crucial task for an issue that affects not only the environment but specially human health in big cities.

Some Meteorological Covariates in Stochastic Fine-sale Modelling of Rainfall

A.P. Garthwaite[(a)]
[(a)]University of Greenwich, London

Using observations from tipping-bucket rain gauges, rainfall at fine-scale time aggregations as low as five-minutes have been modelled with Markov-modulated Poisson processes (MMPP), without the use of disaggregation methods. Where meteorological covariates have been introduced, significant relationships were shown to exist for relative humidity, sea-level air-pressure, and potential temperature, with the volume of rainfall. A class of seasonal models were developed in the univariate case that allowed modelling of several months using a single set of Markov transition parameters, with distinction between months defined by separate Poisson arrival rate parameters. We discuss here the use of a seasonal model that makes no distinction between months in the Poisson arrival rate parameters, but incorporates meteorological data to model fine-scale rainfall distinctly for each calendar month within the winter season. These methods draw on the tractable likelihood function of the Markov modulated Poisson processes which can be optimised numerically through computationally powerful maximum likelihood estimation. The likelihood function is adapted here to consider meteorological information relevant to event arrival rates. Through simulation of rainfall series at a single site in England and analyses of their second order properties, we consider whether a monthly seasonal model is wholly necessary for predicting rainfall when equipped with daily meteorological data relating to air-pressure, temperature and humidity.

Bayesian Spatio-temporal Analysis of Mesothelioma Incidence in the Veneto Region (Italy) Applied To Circumstances of Asbestos Exposure

P. Girardi[(a)], V. Bressan[(a)], E. Merler[(b)]

[(a)]Venetian Mesothelioma Registry, Padova, [(b)]Promotion and Development of Hygiene and Public Health

The association between asbestos exposure and malignant pleural mesothelioma (MPM) is so tight that its spatial distribution follows the historical local use of asbestos. The goal of the study is to analyse the spatio-temporal trend of MPM incidence by circumstance of asbestos exposure in the Veneto Region with a bayesian approach. All incident cases of certain malignant pleural mesotheliomas occurred from year 1988 to 2012 as assessed by the Regional Mesothelioma Register in the residents of the Veneto Region (about 4.5 milion inhabitants) were aggregated at the local health unit level (n=21) of residence. Regional standardized incidence rates were used in order to calculate the expected number of MPM in each health unit. Standardized Incidence Ratio (SIR) and relative 95% Credible Interval (95% CI) by circumstance of asbestos exposures (occupational, non-occupational and undefined) were computed by a BYM (Besag, York and Mollié) spatio-temporal model using an Integrated Nested Laplace Approximation (INLA). The MPMs occurred in the study period (1302 among men, 468 among women) shown an increasing temporal trend. Occupational exposure to asbestos is overall predominant (82.6%), but among women non occupational asbestos exposures are prevailing (35.4%). A percentage of some consistency (9.4% and 23.7% among men and women, respectively) resulted in an undefined asbestos exposure. The spatio-temporal SIR of MPM follows a different pattern by circumstance of exposure and gender. Increased SIRs among subjects with occupational exposures have been observed where asbestos had been used by local industries (municipality of Venice and Padua). Among several health units a growing trend of non-occupational and undefined MPMs is emerging. After a first wave of MPM mainly associated with occupational exposures, MPM occurring in more recent periods are increasingly associated to non-occupational asbestos exposures.

Spatio-temporal Post-processing of Hydrological Ensemble Forecasts

S. Hemri[(a)]

[(a)]Heidelberg Institute for Theoretical Studies, Heidelberg

Hydrologic ensemble forecasts driven by atmospheric ensemble prediction systems need statistical post-processing in order to account for systematic errors in terms of both location and spread. Runoff, and hence also gauge level, is an inherently multivariate process with typical events lasting from hours in case of floods to weeks or even months in case of droughts. This calls for multivariate post-processing techniques that yield well calibrated forecasts in univariate terms and ensure a realistic

spatio-temporal dependence structure at the same time. To this end, the univariate ensemble model output statistics (EMOS) post-processing method is combined with different copula approaches that ensure multivariate calibration throughout the entire forecast horizon and seven gauges along river Rhine. On the one hand we use non-parametric approaches to model the spatio-temporal dependence structure, namely ensemble copula coupling (ECC), which preserves the dependence structure of the raw ensemble, and Schaake shuffle based methods, which preserve the dependence structure of past observations. On the other hand we apply also the parametric Gaussian copula approach (GCA), which estimates the spatio-temporal correlations from training observations.

Of Quantiles and Expectiles: Consistent Scoring Functions, Mixture Representations, and Forecast Rankings

A. Jordan[(a)], W. Ehm[(a)], T. Gneiting[(a,b)], F. Krüger[(c)]

[(a)]Heidelberg Institute for Theoretical Studies, Heidelberg, [(b)]Karlsruhe Institute of Technology, Karlsruhe, [(c)]Heidelberg University, Heidelberg

In the practice of point prediction, it is desirable that forecasters receive a directive in the form of a statistical functional. For example, forecasters might be asked to report the mean or a quantile of their predictive distributions. When evaluating and comparing competing forecasts, it is then critical that the scoring function used for these purposes be consistent for the functional at hand, in the sense that the expected score is minimized when following the directive. We show that any scoring function that is consistent for a quantile or an expectile functional, respectively, can be represented as a mixture of elementary scoring functions that form a linearly parameterized family. Scoring functions for the mean value and probability forecasts of binary events constitute important examples. The elementary scoring functions admit economic interpretations, and the mixture representations allow for simple checks of whether a forecast is preferable to another under any consistent scoring function. Plots of the average scores with respect to the elementary scoring functions, which we call Murphy diagrams, permit detailed comparisons of the relative merits of competing forecasts. I will outline the basic theory and illustrate the use of Murphy diagrams in meteorological and economic examples.

Linear Regression for Interval-valued Data Using Resampling Method

K. Kang[(a)], S. Im[(a)]

[(a)]Hankuk University of Foreign Studies, Seoul

The data used in traditional statistical analysis, such as linear regression, multivariate analysis, are mostly single-valued data that are observed as single value. However, if there is a variation in each observation object such as a person's blood pressure, the degree of air pollution during a day, and the humidity of one month, it cannot be expressed as a single value. In this case, each observation object has a structure and variation internally and hence expressing it as an interval, histogram or table will be more useful and informative. The data of this type is called symbolic dataInterval-valued data, which is one of the symbolic data, is given as an interval in which all observation objects are not single values. So far, many methods for analyzing interval-valued data have been introduced. In this paper, we focus on the method of using linear regression model. In addition, we propose a new method to use truncated normal distribution instead of uniform distribution in the Monte Carlo method. We review several methods for regression analysis for interval-valued data and compare the performance through the various simulation. For real application, we applied our approach to actual data related to the fine dust and we will look at the amount of fine dust that varies with various contaminants in the atmosphere, temperature, and wind speed. We conducted a statistical significance test on the estimated regression coefficients with the advantage that statistical reasoning is possible in the two methods using MCM.

The Correction of Bias in the Bioassessment Based On the Ratio of Observed and Expected Values

S. Kärkkäinen[(a)], H. Hämäläinen[(a)], J. Aroviita[(b)], J. Jyväsjärvi[(c)]

[(a)]University of Jyväskylä, Jyväskylä, [(b)]Finnish Environment Institute, Helsinki, [(c)]University of Oulu, Oulu

In this talk, we consider the ecological status assessment of freshwater bodies based on biological communities and the Reference Condition Approach (RCA). In RCA, the community impacted by human is compared to the communities of reference sites. The decision rule for status assessment is based on the natural variation of selected community metrics on the reference sites. The current rationale is to use as a common threshold a fixed quantile (e.g. 25th) of the ratios between the observed and expected values obtained from reference sites. However, the use of a common quantile may cause severe bias in decision making. We propose a correction method for the bias and compare the novel decision rule to the conventional one in the case of two real data sets, both having reference and impacted sites.

Probabilistic Forecasting and Comparative Model Assessment Based On Markov Chain Monte Carlo Output

S. Lerch[(a,b)], F. Krüger[(c)], T.L. Thorarinsdottir[(d)], T. Gneiting[(a,b)]

[(a)]Karlsruhe Institute of Technology, Karlsruhe, [(b)]Heidelberg Institute for Theoretical Studies, Heidelberg, [(c)]Heidelberg University, Heidelberg, [(d)]Norwegian Computing Center, Oslo

In Bayesian inference, predictive distributions are typically available only through a sample generated via Markov chain Monte Carlo (MCMC) or related algorithms. In this paper, we conduct a systematic analysis of how to make and evaluate probabilistic forecasts from such simulation output. Based on proper scoring rules, we develop a notion of consistency that allows to assess the adequacy of methods for estimating the stationary distribution underlying the simulation output. We then provide asymptotic results that account for the salient features of Bayesian posterior simulators, and derive conditions under which choices from the literature satisfy our notion of consistency. Importantly, these conditions depend on the scoring rule being used, such that the choices of approximation method and scoring rule are intertwined. While the logarithmic rule requires fairly stringent conditions, the continuous ranked probability score (CRPS) yields consistent approximations under minimal assumptions. These results are illustrated in a simulation study and an economic data example. Overall, we find that mixture-of-parameters approximations which exploit the parametric structure of Bayesian models perform particularly well.

Regional Wealth and Environmental Awareness - Using MIMIC Models

S. Lösch[(a)], O. Okhrin[(a)], H. Wiesmeth[(a)]

[(a)]Dresden University of Technology, Dresden

This paper gives an insight into the environmental thinking of the local population in the Russian Federation in terms of the regional economic development. Explaining the latent construct 'environment awareness index', Multiple Indicators-Multiple Causes models are calibrated. This index is derived as a latent variable from various categories of search entries in Yandex, the prominent Russian search engine, during the three years 2014, 2015 and 2016. These indicators are presumably dependent on certain causes, which are also integrated into the models. We find a curvilinear relationship between GRP per capita and environmental awareness. Other possible causes, like industry sector, social status and concentration of emissions seem to play a minor role for environmental consciousness. However, the federal subjects with the highest calculated indices are rather rich and/or lie in the East of Siberia.

Efficient Bayesian Optimisation of Two Penalty Terms in a Spatio-temporal Spline Model With an Application To Groundwater Contamination

M. McLean[(a)], A.W. Bowman[(a)], L. Evers[(a)]
[(a)]University of Glasgow, Glasgow

Data obtained from groundwater monitoring wells has indicated that contaminant concentrations vary more over space than they do over time. Such findings highlight the need for a model that accounts for corresponding differential variability in space and time. To allow for this changeable complexity of the model in different dimensions, a penalised spatio-temporal spline model can be implemented with separate penalty terms for space and time. One drawback to this approach however, when working with spatio-temporal data, is that the optimisation of these two smoothing parameters is computationally expensive. To overcome these computational issues, we have developed an efficient algorithm to determine two optimal smoothing parameters. The model formulation builds on existing methodology which utilises efficient matrix decomposition methods to optimise a single smoothing parameter. Although our model has been developed with groundwater contamination modelling as a primary application, the methodology can be utilised in many spatio-temporal settings.

Bivariate Count Models: a Comparison Between Periodic Bivariate INAR Models and Dynamic Factor Models Applied To Forest Fires

M.S. Monteiro[(a)], I. Pereira[(a)], M. Scotto[(b)]
[(a)]University of Aveiro, Aveiro, [(b)]Instituto Superior Técnico, Lisboa

In the last years the study of count series models has been subject of great interest with the focus centred on univariate models while bivariate, and also multivariate, time series models for counts have less expression in the literature since these models are analytical and computational more demanding. As in univariate case, these models either belong to the class of observation driven model or to the class of parameter driven model and in our work we will apply bivariate count series models of those two classes to bivariate count series of forest fires. The study of forest fires, in its several aspects, is essential to understand the phenomenon and to prevent environmental public catastrophes. In this context the analysis of monthly number of fires along several years is one aspect to have into account in order to better comprehend this thematic. It is from this perspective that we will study the monthly number of forest fires, between 1985 and 2014, in two Portugal neighbouring counties. Due to the periodic nature of the data we will consider periodic bivariate integer valued autoregressive models (PBINAR), which was proposed by the authors in 2015, and dynamic factor models (DF). The performance of these two types of models are analysed in the context of forest fires application.

Spatial Modelling of Mortality Rates With Heteroscedastic Residuals

P. Otto[(a)]

[(a)]European University Viadrina, Frankfurt

Clusters of a high and low observations can be modeled by spatial autoregressive processes, i.e., the observations are assumed to be influenced by their neighbors. The dependence between the observations is defined via the so-called spatial weighting matrix. Moreover, a simple spatial autoregressive process includes an autoregressive parameter. Estimating such kind of process for the mortality caused by lung cancer in all U.S. counties leads to a model having a positive spatial correlation, i.e., the process shows clusters of high and low values. Consequently, there are regions with a higher mortality compared to other regions. This finding is not very surprising. However, we observe that estimated residuals are not homoscedastic, but they rather show clusters of high and low variance, whereas the mean of the residuals is zero. Thus, the error process should be modeled by a spatial ARCH process. We therefore introduce a new spatial model having similar properties, like the well-known ARCH process. Beside environmental covariates, like carbon monoxide, ozone, nitrogen dioxide, sulfate dioxide, and particulate matters, we include economical regressors, namely the personal income per capita, and regressors describing health behavior of people. Interestingly, the particulate matter with an diameter of 2.5 micrometers, and carbon monoxide have no influence on the lung-cancer death rate.

Copula-based Clustering of Concurrent Flood Risks Via Hazard Scenarios

R. Pappada[(a)], F. Durante[(b)], G. Salvadori[(b)], C. De Michele[(c)]

[(a)]University of Trieste, Trieste, [(b)]Università del Salento, Lecce, [(c)]Politecnico di Milano, Milano

Risk assessment of hydrological events is a fundamental task in the framework of environmental sciences, since it may provide valuable strategies to mitigate the impact of extreme (potentially dangerous) events.Copula theory offers a flexible tool to deal with the multivariate nature of hydrological phenomena, which are often characterized by several nonindependent random variables. Here the main interest is in flood risks, which may cause fatalities, displacement of people and damage to the environment, and severely compromise the economic development and activities.The study of the interdependencies among different flood risks across regions is particularly valuable for gaining insight into the presence of homogeneous subregions in terms of hydrological risk. A clustering procedure is proposed, grounded on a copula-based framework for dealing with hazard scenarios recently introduced in literature. Specifically, such scenarios are used to define suitable similarity criteria in terms of the copula expressing the dependence among several random phenomena (e.g., floods) observed at different gauge stations. As a result, the procedure may provide useful information on how (concurrent) flood events, characterized by

various variables, may be interconnected within a given geographical region, thus supplying guidance of utmost importance for assessing the hydrological risks.A case study, involving hydrological data collected in the Po river basin (Northern Italy) is presented, in order to illustrate the practical implementation of the proposed methodology.

Reconstructing Missing Data Sequences of Environmental Data By Means of Spatial Correlation

M. Parrella[(a)], G. Albano[(a)], M. La Rocca[(a)], C. Perna[(a)]
[(a)]Università di Salerno, Salerno

Missing data arise in many statistical analyses, due to faults in data acquisition, and can have a significant effect on the conclusions that can be drawn from the data. In environmental data, the most common approach usually adopted from the Environmental Protection Agencies to handle missing values is by deleting those observations with incomplete information from the study, obtaining a huge underestimation of a lot of indexes usually used for evaluating air quality, for example. In multivariate time series, moreover, it may happen that not only isolated values but also long sequences of some of the time series' components may miss. In such cases, it is quite impossible to reconstruct the missing sequences basing on the serial dependence structure alone. In this work we propose an innovative procedure that aims to reconstruct the missing sequences by exploiting the cross-correlation of the multivariate time series. To this end, we use the technique underlying spatial econometric models. The proposed procedure predicts the missing values in the time series basing on a linear interpolation of the "neighbor" contemporary observations. The weights of the interpolation are given by the strength of the spatial correlation among locations. The procedure has been applied to the pollution data, where the problem of missing sequences is of serious concern, with satisfactory performance.

Assessing Future Flood Risk: the Reliability of Global Projections and the Significance of Future Changes

I. Prosdocimi[(a)], I. Giuntoli[(b)], D. Hannah[(b)]
[(a)]University of Bath, Bath, [(b)]University of Birmingham , Birmingham

This study aims to assess the ability of global models to capture the statistical distribution of high flows for selected catchments in continental United States. As multimodel ensemble experiments become available, changes in flood risk using global models runoff output are estimated to inform mitigation and adaptation strategies. However, model runs suffer from considerable biases and uncertainties and flood risk estimates obtained from their outputs can be largely inconsistent in the control period if compared to those from observed data. These large inconsistencies raise the question of whether the projected changes in flood risk can be fully relied upon and how to better use the outputs of model run to asses the the likely behaviour

of high flows in the future. The statistical properties of peak flows for overlapping record periods from a set of selected streamflow gauges and model runs from corresponding grid-cells fitting are compared. Model runs consist of nine global impact models (GIMs) fed by five global climate models (GCMs) from the ISI-MIP Project. Further, we compare model fits for the future period (2065-2099) for an appraisal of changes in frequencies from the models. Results indicate a large discrepancy between the distribution of observed and modelled data, suggesting that most models fail to reliably simulate high flows. On the other hand, the evidence for the change in the statistical distribution in the future is not clear-cut. Our results call for caution in the use of model projections for the assessment of flood risk changes and highlight the need for a routine evaluation of the model outputs against observations. Finally, the shortcomings of the routine statistical methods used to assess the presence of differences between the statistical properties of model runs for the present and the future time-window are presented and discussed.

Experimental Evaluation of Temperature Uncertainty Components Due To Siting Condition With Respect To WMO Classification

A. Quarello[(a)], R. Ignaccolo[(a)], M. Franco Villoria[(a)], G. Coppa[(b)], A. Merlone[(b)]
[(a)]University of Turin, Turin, [(b)]Istituto Nazionale di Ricerca Metrologica (INRiM), Turin

Automatic weather stations (AWSs) are one of the biggest sources of essential climate variables (ECVs) data, which are necessary to understand the environmental conditions on the surface of the planet. Detection of climate changes based on observed temperature data can be affected by inhomogeneity arising from temporal variation in microscale environments around stations. Indeed, the surface atmospheric air temperature measurements are influenced by the obstacles close the measurement site itself. The purpose of the present study is a contribution to the improvement of the World Meteorological Organization's document "Sustained Performance Classification for Surface Observing Stations on Land", through data analysis of an in-field experiment, within the European project "MeteoMet, Metrology for Meteorology ". In order to improve the WMO/CIMO siting classification for air temperature measurements, three field experiments were carried out, employing a rigorous metrological approach, to evaluate the effect of the presence of different obstacles (roads, trees and buildings) on the uncertainty in air temperature measurement. This work presents a description of the current situation related to the road siting classification for surface observation stations on land in the WMO/CIMO Guide. In accordance with preliminary analysis's results, a semiparametric model for Delta T – the difference between the temperature recorded by a sensor exposed to the effect of the obstacle and the sensor compliant with the current recommended exposure rules - has been developed. The model features also a seasonal factor and smooth terms for the dependence on meteorological variables and their interaction, applied to road siting. Furthermore, the large amount of acquired data (over 13

million fields in 1 million records) gathered with high frequency allows us to exploit a functional data approach.

A Hidden Markov Approach To the Analysis of Cylindrical Space-time Series

M. Ranalli[(a)]

[(a)]University of Roma Tre, Roma

Motivated by segmentation issues in marine studies, a new hidden Markov model is proposed for the analysis of cylindrical space-time series, that is, bivariate space-time series of intensities and angles. The model is based on a hierarchical mixture of cylindrical densities, where the parameters of the mixture components vary across space according to a latent Markov field, whereas the parameters of the latent Markov field evolve according to the states of a hidden Markov chain. It allows to segment the data within a finite number of latent classes that vary over time and across space and that represent the conditional distributions of the data under specific environmental conditions, simultaneously accounting for unobserved heterogeneity, spatial and temporal autocorrelation. Further, it parsimoniously accommodates specific features of environmental cylindrical data, such as circular-linear correlation, multimodality and skewness. Due to the numerical intractability of the likelihood function, parameters are estimated by a computationally efficient EM-type algorithm that iteratively alternates the maximization of a weighted composite likelihood function with weights updating. The effectiveness of the proposal is tested in a case study that involves speeds and directions of marine currents in the Gulf of Naples observed over time, where the model was capable to cluster cylindrical data according to a finite number of latent classes varying over time that are associated with specific environmental conditions of the sea.

Spatial Sampling With Spreading the Samples in the Auxiliary Variables' Space

E. Rocco[(a)], B. Bertaccini[(a)]

[(a)]Università degli Studi di Firenze, Firenze

Statistical units selected from a territory are generally spatially correlated which means that two neighbouring statistical units tend to be more similar than two distant statistical units. It is a well-known finding in the literature on sampling from spatial populations that in a situation of this type it is advantageous to select units well spread over the territory. Moreover, when the values taken by some auxiliary variables are known for each unit of the population, the sampling method should be able to take into account both the spatial structure of the population and the available auxiliary information. The aim of this study is the identification of an appropriate sampling technique for the analysis of the seismic vulnerability of the residential buildings in the city of Florence. A census evaluation of the whole town

would be desirable but it is unfeasible since the collection for each building of all data needed to calculate its index of seismic vulnerability is complex and reliable estimates predict that a complete analysis could take about 30,000 man/days. In choosing the sampling strategy, we have considered three elements: 1) the fact that the buildings have a spatial location; 2) the availability for each building of information on its geographical location (i.e. the census tract and the cadastral parcel) and on other variables related to the outcome of interest, such as the year of construction, the type of construction (masonry / concrete) and the number of floors; 3) the need to identify sub-areas with different levels of seismic vulnerability in addition to produce an estimate of the average risk of vulnerability for the whole municipal territory.

Statistical Exploration of Temporal and Spatial Patterns in Measured Pollutant Concentrations

R. Smith[(a)]

[(a)]Centre for Ecology and Hydrology, Edinburgh

Many countries operate atmospheric pollutant monitoring networks providing long-term trends and often annual maps to check for changes in ecosystem drivers and responses to policy initiatives, both nationally and internationally. The UK has monitored rainfall ion chemistry since 1986 and these observational data are used to generate temporal trends and spatial maps of concentrations and deposition, thus forming one element of the emission/deposition budget for the country. With advances in spatio-temporal modelling developed since these networks were set up, there is now potentially much improved estimation of non-linear trends over time, better modelling of the spatial processes that produce the concentration maps, and a significant improvement in the estimation of the uncertainties. This study explores how some of these analyses can help define spatial and temporal patterns and extract additional information from existing data.

Cross-National Comparisons On Pro-Environmental Behaviours – A Multilevel Modelling Approach

V.H. So[(a)], G. Durrant[(a)], P. Smith[(a)]

[(a)]University of Southampton, Southampton

This study investigates the cross-national differences in general pro-environmental behaviours by identifying how individual and country-level factors influence individuals' behaviours using a multilevel modelling framework. Moreover, it also explores how personal environmental attitudes have different degrees of effect on how people behave across nations using random slope model. Analysis is conducted on the 2010 Environmental module of the International Social Survey Programme (ISSP), a cross-national survey that mainly deal with environmental behaviours and attitudes towards environmental related issues. Preliminary result shows country has

a significant effect on people's reported environmental behaviours after controlling for individuals' sociodemographic factors and mode of interview. The findings also suggest that personal values, environmental attitudes and political beliefs have substantial impacts on their daily behaviours. Moreover, the inclusion of the random slope in the final models also provides evidence that the effects of individuals' environmental attitudes on their environmental behaviours vary significantly across countries.

A Stochastic Forecast Model for International Migration in Germany

P. Vanella[(a)], P. Deschermeier[(b)]

[(a)]Gottfried Wilhelm Leibniz Universität, Hannover, [(b)]Cologne Institute for Economic Research. Cologne

International Migration for decades is one of the most discussed topics in Europe. Especially the European Union traditionally is an attractive target region for international migration for many people. The low life standard in most African countries and the agitations in the Arabic countries since 2011, including the emergence of the Islamic State, lead to new records in immigration and the numbers of asylum seekers in Europe. Furthermore, the financial crisis in Europe since 2007 altogether with the enlargements of the European Union by Bulgaria, Romania and Croatia lead to huge migration flows from the southern to the northern European countries. Since data on international migration are very error-prone and are extremely volatile due to the high dependency of economic, social, natural and political environment, estimation of future migration is a very sophisticated undertaking. For future planning of resources and investment strategies as well as appropriate controlling of migration flows, good migratory projections are essential though, which traditionally tend to be inaccurate on the one hand and don't consider uncertainty adequately on the other. This paper proposes an approach for estimation of age- and sex-specific net migration (ASSNM) using principal component time series models for forecasting and taking the above mentioned factors into consideration. We illustrate the modeling approach at the example of net migration in Germany, although the method may be in general be applied for other countries in the European Union as well.

Semi-parametric Models for Nonstationary Environmental Extremes

E. Zanini[(a)]

[(a)]University of Lancaster, Lancaster

Joint work with: E. Eastoe, J. Tawn, P. Jonathan, D. Randell

Realistic and accurate environmental design conditions are fundamental for the construction and maintenance of marine structures. They require an understanding of waves and storm severity, which is often achieved through an extreme value analysis of measured or hind-cast significant wave height. In order to determine as realistic and accurate design criteria as possible, it is fundamental to account for the multiple covariates that characterise the data, particularly wave direction and season. A methodology that accurately models these covariate effects is essential for reliable estimates.To model the nonstationarity in the Generalised Pareto parameters, we focus on two main formulations, namely P-splines and a proposed linear combinations of Gaussian densities. Inference is carried out using MCMC, and we consider both the case where the number of components is fixed and when it's unknown. Reversible jump MCMC are used to estimate models with unknown number of components, and Bayesian Additive Regression Splines (BARS) are used for the spline model. Return value estimates and performance for the different formulations are compared, while potential advantages and computational cost issues are considered both for one-dimensional covariates and for extensions to higher dimensions.Key Words: covariate modelling; environmental; Bayesian; BARS; Reversible Jump MCMC.

Poster session

Multimode: a Package for Multimodality Assessment

J. Ameijeiras-Alonso[(a)], R. Crujeiras[(a)], A. Rodríguez-Casal[(a)]
[(a)]Universidade de Santiago de Compostela, Santiago de Compostela

The multimode package for the statistical software R includes different tools for testing and exploring the number of modes (local maxima) in a real density function. Different testing methods for determining the number of modes are implemented in this package, such as the critical bandwidth and the excess mass. Multimode also includes a function providing the different proposals available in the literature for testing if the number of modes of a sample is equal to a given number (against if it is greater). In addition, once the number of modes is known, this package includes a function estimating the location of the modes and antimodes (local minima) and their density values. Different graphical methods for analyzing, in an exploratory way, the number and location of modes are also provided in the multimode package. The tools available in this package will be illustrated in some examples in applied fields, including climate, geological and environmental problems.

Multispecies Spatial Distribution Modelling in Fisheries: a Prey-predator Example

X. Barber[(a)], D. Conesa[(b)], A. Lopez-Quilez[(b)], I. Paradinas[(b)], M. Pennino[(c)]
[(a)]Universitas Miguel Hernadez, [(b)]Universitat de Valencia, [(c)]INSTITUTO ESPAÑOL DE OCEANOGRAFÍA

The different biotic interactions must be taken into account by implementing a geostatistical approach of multivariate response in the species distribution models. In fisheries distribution species, scientists increasingly recommend multispecies approaches to fisheries management. There are many situations of interest in which this issue is not solved, such as predator-prey and host-parasitoid cases. Here, we present a Bayesian multivariate spatial modelling for the occurrence patterns of cohabitants species: Fish (Hake or Monkfish) and crustaceans/mollusc as the major preys identified. We propose a conditional coregionalized Bayesian linear model that makes use of a reparametrization of the variance, and this reparametrization allows an easier way to elicitate a prior distribution for the variance. This modelling provides a better approach to the prediction problem of these species compared with that which analyzes the distribution of species separately.

Comparing the Performance of Space-time Models in the D-STEM Software

Y. Borisova[(a)], F. Finazzi[(b)], M. Scott[(a)], D. Lee[(a)]
[(a)]University of Glasgow, [(b)]University of Bergamo

Modelling air pollution is vital in order to be able to improve the air quality and thus, improve people's quality of life. In Scotland, the city with the worse air pollution is Glasgow. Using the D-STEM software in MatLab, different space-time models were fitted using the meteorological conditions (wind speed, wind direction, temperature, etc) as covariates and nitrogen dioxide, NO2, concentration as measured by the Automatic Urban Monitoring Network (AURN) sensors as a response. The data were collected from eight different stations around the City of Glasgow. The models' performance was compared using cross-validation techniques.

Non Parametric State-space Model for Missing-data Imputation in Environmental Time Series

T. Chau[(a)], P. Ailliot[(b)], V. Monbet[(a)]
[(a)]IIRMAR, University of Rennes 1, [(b)]LMBA, University of Brest

Missing data are present in many environmental datasets and our work aims at developing a general method for imputing them. State-space models (SSM) have already extensively been used in this framework. The basic idea consists in introducing the true environmental process, which we aim at reconstructing, as a latent process and model the data available at neighbouring sites in space and/or time conditionally to this latent process. A key input of SSMs is a stochastic model which describes the temporal evolution of the environmental process of interest. In many applications, the dynamic is complex and can hardly be described using a tractable parametric model. Here we investigate a data-driven method where the dynamical model is learned using a non-parametric approach and historical observations of the environmental process of interest. From a statistical point of view, we will address various aspects related to SSMs in a non-parametric framework. First we will discuss the estimation of the filtering and smoothing distributions, that is the distribution of the latent space given the observations, using both sequential Monte Carlo approaches and extensions of the usual Kalman recursions in conjunction with local linear regression. Then, a more difficult and original question consists in building a non-parametric estimate of the dynamics which takes into account the measurement errors which are present in historical data. We will propose an EM-like algorithm where the historical data are corrected recursively. The methodology will be illustrated and validated using both simulations and multisite wind data in North West of France.

Spatio-temporal Statistical Analysis of Measured and Simulated Fine-resolution Datasets Generated From the North Wyke Farm Platform

S. Curceac[(a)], P. Atkinson[(b)], A. Milne[(a)], L. Wu[(a)], P.Harris[(a)]
[(a)]Rothamsted Research, [(b)]Lancaster University

The North Wyke Farm Platform (NWFP) is a systems-based experiment with the remit to research grassland livestock productivity and ecosystem responses to alternative management practices. The system consists of 15 hydrologically isolated sub-catchments. Among various measured data sets from the NWFP, this study is focused on baseline 15-minute water run-off variables (chemistry and flow), as well as meteorological and soil moisture data - all available at the 15 sites. For this baseline period (2011 to 2013), all 15 sub-catchments operated under the same permanent pasture management system. The time series data have been statistically explored to investigate their behaviour, seasonal patterns and address any data quality issues (for example where variables contain missing values and outliers). This analysis included correlating the same variable across the 15 sub-catchments to investigate its spatial dependencies, cross-correlating between different variables to examine spatial co-dependencies and an analysis of extremes. Extreme events were considered by using the block maxima and the peaks-over-threshold approaches and appropriate distributions were fitted. The size and the number of the extremal clusters, the autocorrelation function and the extremograms were used to evaluate short-range extremal dependence. The same set of statistical techniques were also used to assess the simulated time-series outputs from mechanistic, process-based (PB) model for the same temporal variables of the NWFP system. This helped to identify factors that are poorly accounted for in the PB model's parameterisation. The PB model can be improved accordingly, or alternatively its errors can also be modelled statistically, to provide a hybridised form. In this paper, we relay the first steps in such research, where we develop the conceptual framework; and also present some initial findings from the pre-processing and exploratory analysis of the measured and simulated data.

A Bayesian Data Integration Model for Air Pollution Concentration

C. Forlani[(a)], M. Blangiardo[(a)], F.B. Piel[(a)]
[(a)]Imperial College London

Air pollution is a major challenge for public health, and the interest in its measurement, modelling and effects on human health, in particular respiratory and cardiovascular diseases, is increasing. Air pollutants act simultaneously, and their concentrations are correlated in space and in time. The objective of this study is building and comparing different models for air pollution concentration using data integration techniques, accounting for spatial misalignment. The environmental data are provided by multiple sources: DEFRA (modelled annual air pollution

concentrations on a 1x1km grid), Met office (modelled daily air pollution concentrations on a 12x12km grid) and King's College London (monitoring daily air pollution concentrations and meteorological confounders). Through a Bayesian hierarchical framework we specify a model to link the data sources, accounting for the change of support problem as well as for the measurement error, in order to predict the concentration at grid level (1x1km). In addition, we evaluate the role of spatial and temporal covariates (e.g. site-type, meteorology). The analysis for model adjustment considers a single pollutant (NO2) in the area of Greater London for the years 2007-2011. This analysis will build the basis for a multi-pollutant model, which takes into account the correlation between the major pollutants (PM10, PM2.5, O3, NO2, NOX) through a temporal effect varying across these; further developments may include the implementation of an epidemiological model to evaluate the link between exposure to air pollution and cardio-respiratory conditions at the small area level, propagating the uncertainty from the exposure estimates.

Penalized Complexity Priors for Varying Coefficient Models

M. Franco Villoria[(a)], M. Ventrucci[(b)], H. Rue[(c)]

[(a)]University of Torino, [(b)]University of Bologna, [(c)]King Abdullah University of Science and Technology

Varying coefficient models (VCMs) can be seen as a general class of models that includes as special cases the generalized linear model, generalized additive models, dynamic generalized linear models or even the more recent functional linear model. In practice, a VCM is useful in the presence of an effect modifier, a variable that changes the effect of a covariate of interest on the response. In a Bayesian hierarchical framework, the varying coefficient can be described by a vector of random effects distributed at prior as a Gaussian Markov Random Field. In this work, we present the use of penalized complexity priors for VCMs, introducing a natural base model for different (temporally and spatially) structured priors. We illustrate the application of these priors on an epidemiological case study.

Predicting Missing Values in Spatio-temporal Satellite Data

F. Gerber[(a)]

[(a)]Institute for Mathematics University of Zurich Winterthurerstrasse 190 CH-8057 Zürich

Time series of remotely sensed optical data often contain data points of low product quality, related to atmospheric contamination or angular configuration for example. After detecting and removing such data points, the resulting data product is sparse and contains missing values. This is problematic for applications and signal processing methods that require temporally continuous data sets. To address this sparsity, we present a new gap filling method, which is designed to scale with the computational resources via parallel computing. We predict each missing value separately

based on data points in the spatio-temporal neighborhood around the missing data point. The prediction of the missing values and the estimation of the corresponding prediction uncertainties are based on sorting algorithms and quantile regression. The gap filling method was applied to a remote sensing based vegetation index from Alaska and tested with realistic scenarios featuring between 20% and 50% missing data. Validation against established methods showed that the proposed method has a good performance in terms of the root mean squared prediction error, which was between 0.041 and 0.060 and lower compared to the others methods for all test scenarios. The method is available in the open-source R package gapfill. We demonstrate its performance using a real data example and show how it can be tailored to specific data sets. The computational workload can be distributed among several computers, rendering the method applicable to large data sets. Due to the flexible software design, users can control and redesign relevant parts with little additional effort. This makes it an interesting tool for gap filling satellite data and for the future development of gap filling methods.

A Predictive Stock Portfolio Selection Strategy for Socially Responsible Investment

S. Jeong[(a)]

[(a)]Hankuk University of Foreign Studies, South Korea

Socially responsible investment (SRI) has now turned up as an effective alternative to traditional investment strategies even for the mainstream investors. The criteria to select socially responsible stocks are classified into environmental, social and governance(ESG), and there are several dimensions under each category by which companies are judged. This work presents a predictive stock portfolio selection tool for SRI, assuming some of those dimensions would form a set of feature variables to predict future controversy issues of a firm. Various machine learning tools are compared with each other for learning a model that predicts the probability of involvement in various ESG-related controversy issues using historical data. And, over a five-year time span, the performance of portfolios selected based on the predicted probabilities are compared with that of the market benchmark.

Additive Bayesian Network Approach Applied To Time Series and Longitudinal Datasets

G. Kratzer[(a)]

[(a)]PhD student at Zurich University

In recent years, Additive Bayesian Networks (ABN) analysis has been successfully used in many fields from sociology to veterinary epidemiology. This approach has shown to be very efficient in embracing the correlated multivariate nature of high dimensional datasets and in producing data driven model instead of expert based models. ABN is a multidimensional regression model analogous to generalised linear modelling but with all variables as potential predictors. The final goal is to construct the Bayesian network that best support the data. When applying ABN to time series dataset it is of high importance to cope with the autocorrelation structure of the variance-covariance matrix as the structural learning process relies on the estimation of the relative quality of the model for the given dataset. This is done by using common model selection score such as AIC or BIC. We adapt the ABN framework such that it can handle time series and longitudinal datasets and then generalize the time series regression for a given set of data. We implement an iterative Cochrane-Orcutt procedure in the fitting algorithm to deal with serially correlated errors and cope with the between- and within- cluster effect in regressing centred responses over centred covariate. We illustrate the use of ABN in a time series longitudinal case with a classical dataset used in time series regression and compare the result with a classical approach.

The 2015 Heat Wave in Central Europe and Meteorological Factors Associated With Major Heat Waves in the Context of EURO-CORDEX RCM Simulations

J. Kysely[(a)], O. Lhotka[(a)], E. Plavcova[(a)]

[(a)]Institute of Atmospheric Physics & Global Change Research Institute, Czech Academy of Sciences

We evaluate the severe 2015 heat wave in Central Europe as to climatological characteristics and meteorological factors related to its severity and persistence. Considering heat waves as spatial events and using the E-OBS dataset, we show that it was the second most severe and longest heat wave in this region since 1950, after the 1994 heat wave. By analysing severity of past heat waves in the E-OBS data, we define 'major heat waves' as those with the extremity index of at least 25°C (the 1994, 2006, and 2015 heat waves meet this criterion) and examine their associated meteorological factors (atmospheric circulation, precipitation, net shortwave radiation, and soil moisture deficits). All major heat waves were associated with atmospheric blocking, precipitation deficit, reduced evaporative fraction, and increased net shortwave radiation. In the following step, we evaluate the simulation of major heat waves in Central Europe and their links to the aforementioned factors in historical runs of EURO-CORDEX regional climate models (RCMs) in the 1970–2016 period. The

magnitude of major heat waves in RCMs is linked to relative contributions of large-scale circulation and land–atmosphere interactions and considerably varies among individual RCMs. In some cases, the seemingly good reproduction of major heat wave's magnitude is erroneously achieved through favourable circulation conditions compensated by a substantial surplus of soil moisture or vice versa. This finding points to different driving mechanisms of major heat waves in some RCMs, which should be taken into account when analysing future projections of these events.

Highly Structured Spatial Models As a Tool for Analyzing the Spread of Diseases and Species Distributions

J. Martínez-Minaya[(a)], A. Vicent[(b)], A. López-Quílez[(a)], F. Picó[(c)], A. Marcer[(d)], D. Conesa[(a)]

[(a)]Universitat de València, Valencia, [(b)]Instituto Valenciano de Investigaciones Agrarias (IVIA), [(c)]Estación Biológica de Doñana, CSIC, [(d)]CREAF and Universitat Autònoma de Barcelona

In the last years, the use of complex statistical models has increased to improve our knowledge on the spread of diseases and the distribution of species, being of great interest in Ecology and Epidemiology. Complexity in these models arises for instance when including the use of beta likelihoods and spatial effects. This complexity makes the inferential and predictive processes challenging to perform. Bayesian statistics represent a good alternative to deal with these models, because it is based on the idea that the information and uncertainty can be expressed in terms of probability distributions. Moreover, this complexity can be readily handled with hierarchical Bayesian models without much difficulty. However, despite the different advantages of the Bayesian inference, the main challenge is to find an analytic expression for posterior distributions of the parameters and hyperparameters. Several numeric approaches have been proposed such as Markov chain Monte Carlo methods (MCMC) or integrated nested Laplace approximation (INLA). Here, we present three different complex real problems which can be approached with hierarchical Bayesian models using INLA. In particular, a beta regression model with random effects to study a persimmon disease caused by the fungus Mycosphaerella nawae in the Comunitat Valenciana region in Spain, and a beta spatial regression to study the spatial distribution of the genetic diversity of the plant Arabidopsis thaliana in the Iberian Peninsula. In addition, we show a preliminary analysis of an emerging plant disease, known as the olive quick decline syndrome and caused by the bacterium Xylella fastidiosa, which is expanding rapidly in the southern region of Apulia in Italy, Corsica, continental France, as well as outbreaks in Balearic Islands in Spain.

Trend Surface Models Estimation With Outliers

A. Meilán Vila[(a)], R. Crujeiras Casais[(b)]

[(a)]Universidade de Santiago de Compostela and Universidade da Coruña, [(b)]Universidade de Santiago de Compostela

Trend surface models try to relate observations of a spatial process which varies continuously with the geographic locations in which those observations are taken, trying to establish patterns of increase or decrease with respect to the coordinates through (usually linear) regression models. In these models, two sources of variability can be distinguished: a regression function or trend which would gather large-scale variability and the error term, which would represent small-scale variability. The main difference between these models and the classical linear regression is that the errors present a dependence structure, which is usually assumed to be intrinsically or second-order stationary. That dependence structure is usually unknown but should be included in the model, through the covariogram (if the process is second-order stationary) or the variogram (if the process is intrinsically stationary). The problem of trend surface estimation can be solved through least squares tools (where pilot estimations of the variogram are used) or maximum likelihood (the estimation of the parameters of the trend and of the dependence are approached jointly, under the assumption of normality). Even so, the trend surface estimation has only been discussed in the case of "simple" statistical processes, under the assumption of normality. However, even under this distributional premise, observed samples may present certain complexities, as the presence of outliers. The aim of this work is to propose a new procedure to perform the trend surface models estimation which protect the conclusions against the influence of outliers. The idea would be to combine the trend surface models estimation using iterative least squares (taking into account the dependence structure of the errors) with the use of pseudo-data (obtained by a previous smoothing procedure) to mitigate the effect of outliers. In order to study the performance of this procedure, a simulation study is carried out.

A Stochastic Model for Generatic Typhoon Trajectory Scenarios Based On A Gaussian Process Model and Its Application

S. Nakano[(a)]

[(a)]The Institute of Statistical Mathematics

A typhoon is an intense tropical cyclone which forms in the north-west region of the Pacific Ocean, and it is one of the major causes of serious natural disasters in the East Asia region. In order to enable us to evaluate the risks of typhoon hazards for each of specific regions, we have developed a data-driven stochastic simulator which can generate various possible typhoon scenarios. In this stochastic simulator, typhoon scenarios are generated from a Gaussian process model of which the parameters were determined based on the data of past typhoon trajectories for more than sixty years. This Gaussian process model provides a probability distribution of possible

spatial patterns (as a function of longitude and latitude) of typhoon properties, and each typhoon scenario can be generated from one realization of spatial patterns of typhoon properties. Actually, the sampling of the realizations of spatial patters can be generated only by calculating the typhoon properties along each typhoon trajectory, and thus we can efficiently generate a large number of typhoon scenarios. We will demonstrate how this stochastic simulator works.

Big Misaligned Space-time Sensor Data On Plant-soil Interactions - Bayesian Semiparametric Regression With Measurement Error

A. Pollice[(a)], G. Jona Lasinio[(b)], R. Rossi[(c)], M. Amato[(d)], T. Kneib[(e)], S. Lang[(f)]
[(a)]Università degli Studi di Bari Aldo Moro, [(b)]Sapienza Università di Roma, [(c)]Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, [(d)]Università della Basilicata, [(e)]Georg August Universität, [(f)]Universität Innsbruck

Alfalfa is one of the most important forage legumes in the world and tools for the rapid assessment of plant-soil variability are crucial for the implementation of tactical and strategic management. We analyze the relationship between Alfalfa NDVI repeated measurements and multi-depth soil electrical resistivity (ER) used as a proxy of soil variability for a seven hectares stand at 210 m a.s.l. in South Italy. Observations of both quantities were made using sensors at a very fine spatial resolution. NDVI data were obtained at 4 sampling occasions with point locations changing over time and the number of samples ranged from 110k to 250k. ER data where obtained in one occasion at three different depths. Modeling these data involves the solving of several issues mostly linked to the spatial and temporal misalignment and the large data size. In this work we present a Bayesian model based approach to all of the above. NDVI and ER data are interpolated (downscaled) on a 2574 points square grid covering the field. To account for the uncertainty on the covariate (ER) we adopt a measurement error type approach expanding on current literature to include replicates, heterogeneous variances and dependent replicates. Semiparametric regression with fixed and random effects allows to analyze the nonlinear relation between NDVI and ER and control many acknowledged sources of process variability. Estimates are obtained using an original BayesX implementation. Comparable alternative model specifications all show ability in detecting the shape of the NDVI-ER relation. Better predictive performances are obtained accounting for the spatial heterogeneity of the measurement error and time-varying residual variability of NDVI.

A Comparative Analysis of Methods for Identification and Characterization of Earthquake Clusters

R. Rotondi[(a)], A. Peresan[(b)], S. Gentili[(b)], E. Varini[(a)]

[(a)]CNR - Istituto di Matematica Applicata e Tecnologie Informatiche, Milano (I),
[(b)]Istituto Nazionale di Oceanografia e Geofisica Sperimentale, Udine (I)

Identification and statistical characterization of seismic clusters may provide useful insights about the features of seismic energy release and their relation to physical properties of the crust within a given region. Moreover, in seismology some studies on hazard assessment require to remove secondary events (aftershocks) from the catalogue, leaving only main shocks; that is called declustering. Seismicity shows clustering in space, highlighted by the concentration of earthquakes along regional fault networks, and clustering in time primarily, but not only, associated with the increase of seismic activity immediately after large earthquakes leading to aftershock sequences.An attempt of combining both these aspects is done through the notion of generalized distance defined on the space-time-size domain, which measures the correlation between earthquakes in terms of Gutenberg-Richter law for the magnitude and of fractal dimension of the epicentres distribution. We aim at comparing declustering techniques, ranging from classical space-time windows methods to methods based on nearest-neighbour distance. A comparative analysis of resulting clusters is carried out for the most relevant earthquakes in North-Eastern Italy, as reported in the bulletins compiled at the National Institute of Oceanography and Experimental Geophysics (INOGS)since 1977. The examination is then extended to selected earthquake sequences associated with a different seismotectonic setting, namely to events that occurred in the region struck by the recent Central Italy destructiveearthquakes, making use of data recorded by the National Institute of Geophysics and Volcanology (INGV). The study shows that the data-driven approach, based on the nearest-neighbor distances, can be satisfactorily applied to decompose the seismic catalog into background seismicity and individual sequences of earthquake clusters, also in areas characterized by moderate seismic activity, where the standard declustering techniques may turn out rather gross approximations.

Data Fusion Model for Large Spatial Datasets

C. Wang[(a)]

[(a)]PhD Candidate in Biostatistics and Epidemiology, Department of Mathematics, University of Zurich

The availability of data has increased dramatically in past years. Multivariate environmental data and highly detailed social-economic data are readily to be analysed to address different research interests. Moreover, the linkage between diverse datasets can be more easily established with the openness trend of database hosts and organizations. We propose a novel data fusion model that combines data from multiple sources. More specifically, the model assumes a continuous latent spatial process, which in turn effects both individual point-referenced observations and aggregated areal data. The point and areal data do not necessarily represent the same

variables or even come from the same distribution. Unlike previous models that become almost infeasible due to large number of data points, we make use of the nearest neighbour Gaussian process (NNGP), which significantly reduce computation time by inducing sparsity in covariance matrix. In addition, utilizing recent No-U-Turn-Sampler (NUTS) in Stan with its interface to R, we implement computationally feasible data fusion model to make inference on the underlying spatial process. We illustrate our fusion model with a simulated dataset and compare the results with widely used geostatistical models.

Author Index