Fifth Workshop on Bayesian Inference for Latent Gaussian Models with Applications

University of Bath

September 14, 2016

Contents

Invited Speakers

Statistical paleo-climate reconstruction: challenges for the statistics community

14 Sept
2:00pm
4W1.7

Andrew Parnell

University College Dublin

I will present the generic problem of reconstructing aspects of past climate based on proxy data from a statistical perspective. There are now numerous data sources to assist with such reconstructions, which all present their own advantages and issues. My group have developed one particular approach, based on Bayesian inversion of causal models of the climate-proxy relationship.Our approach contrasts with much of traditional paleo-climate reconstruction. I will discuss some of the many statistical challenges that lie ahead in creating reconstructions with uncertainty and point out some opportunities for collaboration.

Spatio-temporal log-Gaussian Cox processes for public health data

14 Sept
2:45pm
4W1.7

Theresa Smith

University of Bath

Health data with high spatial and temporal resolution are becoming more common, but there are several practical and computational challenges to using such data to study the relationships between disease risk and possible predictors. These difficulties include lack of measurements on individual-level covariates/exposures, integrating data measured on difference spatial and temporal units, and computational complexity.In this talk, I outline strategies for jointly estimating systematic (i.e., parametric) trends in disease risk and assessing residual risk with spatio-temporal log-Gaussian Cox processes (LGCPs). LGCPs are a type of inhomogeneous Poisson point process where the log intensity surface is a GP. I will present and compare computational tools for using spatio-temporal LGCPs to investigate the roles of environmental and socio-economic risk-factors in the incidence of Campylobacter in England.

---

15 Sept
9:00am
4W1.7

Penalised complexity priors

Sigrunn Holbek Sorbye
UiT The Arctic University of Norway

Latent Gaussian models represent a flexible class of three-stage hierarchical models and have a wide range of applications. However, prior selection for the hyperparameters is often a practical nuisance, as this can have a large impact on posterior results. Examples include prior choices for precision and correlation parameters, governing the smoothness and/or range of the dependency structure modelled by the latent field. Such priors have often been selected in a rather ad-hoc way and the interpretation can be unclear. The recently developed penalised complexity (PC) priors make life easier as they are constructed from given principles. This makes them easy to understand and communicate. PC priors have already been developed for various cases and seem like a promising route to construct default priors in software like R-INLA. In this talk, I will discuss some new cases and derive PC priors for autoregressive processes, fractional Gaussian noise and log-Gaussian Cox processes. Also, I will discuss a very nice and not well-known feature of these priors for model comparison.

---

15 Sept
9:45am
4W1.7

Latent process models in environmental public health surveillance

Gavin Shaddick
University of Bath

Environmental health surveillance models do not typically link health outcome measures (e.g., morbidity counts) with measures of exposure (e.g., air pollution). Moreover, while different health events may arise from a common underlying exposure they are usually treated as though they are independent. The ability to reliably forecast health outcomes under different exposure scenarios would facilitate real-time public health decision making. We propose modelling an environmental exposure as a latent process within a hierarchical Bayesian framework. We evaluate the approach through simulation, comparing univariate models for each health outcome to a combined approach. We also illustrate the method with a case study from British Columbia, where seasonal wildfires affect air quality. We forecast two daily asthma-related health outcomes (prescription dispensations, physician visits) using measures of fine particulate matter concentrations. Overall, the bivariate approach was superior in performance compared to using separate univariate models. Modelling multiple measures of health impact with a common latent exposure provides a more cohesive and flexible framework than traditional methods and we show that this approach can provide timely information for public health surveillance. We will also show extensions that have been developed for cases where data is available over both space and time, allowing dependencies between administrative areas to be exploited in order to increase precision in areas where data may be sparse.

The (Non-)Center of Attention: Why Hierarchical Models Are Deviously Hard To Fit

15 Sept
2:00pm
4W1.7

Michael Betancourt

University of Warwick

Hierarchical models are a powerful means of pooling information across disparate groups but they are often difficult to accurately fit because of their latent geometry.

In this talk I will first introduce how every probability distribution yields a unique geometry, and how these geometries naturally identify the properties necessary for efficient Bayesian algorithms. I will then discuss how hierarchical models manifest particularly pathological geometries and how these pathologies obstruct accurate fits. Finally, I will examine alternative implementations of hierarchical models that can sometimes avoid the pathological behavior and some diagnostics that identify the optimal implementation in practice.

Explaining Gaussian process models

15 Sept
2:45pm
4W1.7

Aki Vehtari

Aalto University

Gaussian process (GP) models can be used to make flexible multivariate non-linear models with implicit interactions. I discuss some model selection methods which can be used to simplify GP models to make them easier to explain. I also discuss functional shape priors which can improve predictive performance and further improve the explainability of the predictions.

Using Latent Gaussian models for disease mapping

16 Sept
9:00am
4W1.7

Samir Bhatt

University of Oxford

Maps of infectious disease—charting spatial variations in the force of infection, degree of endemicity, and the burden on human health—provide an essential evidence base to support planning towards global health targets.Contemporary disease mapping efforts have embraced statistical modelling approaches to properly acknowledge uncertainties in both the available measurements and their spatial interpolation. The most common such approach is that of LGMs. In my talk I will talk about the applications of LGMs to Malaria and HIV, and discuss approaches to improve predictive performance.

---

16 Sept
4:00pm
4W1.7

Sea Mammal tracks as Gaussian processes: fusing statistical and nonstatistical models

James V. Zidek
University of British Columbia

Many biologging studies deploy biologgers equipped with magnetometers and accelerometers to record animal movements at infra-second frequencies, thus allowing their tracks to be to be reconstructed at high-resolution by dead reckoning (DR). But stochastic factors limit the accuracy of the DR paths. So a conventional (but ad hoc) method was developed, which uses irregularly observed GPS points and simply shifts the DR paths to pass through them. While appealing simple, the conventional method lacks the stochastic foundation that enables quantitative uncertainty statements about the true path to be made. The Bayesian melding (BM) approach provides such a foundation for melding model (the DR path) with data (the GPS measurements). However that approach is computational intensive at the best of times and here the challenges are daunting, due the high dimensional data records. Our implementation of the BM uses a Brownian Bridge process to combine the fine-resolution (but seriously biased) DR path and the sparse (but precise) GPS measurements. But several key approximations and a conditional independence property of the Brownian Bridge process were needed to make it work. A cross-validatory assessment of the method will be described and show that the BM works pretty well, when applied to data obtained from northern fur seals (Callorhinus ursinus) foraging in the Bering Sea. The GPS corrected high-resolution path also revealed that the total distance traveled by the fur seals was much greater than that calculated by simply joining the dots (linear interpolation of the GPS observations)! We also explored use of an integrated Ornstein-Ohlenbeck process (that did not work well), and are developing a generalized Ornstein-Uhlenbeck process that seems promising and these results will also be described.

Contributed Talks

Inference for GLMMs using Laplace approximations: when is it valid?

14 Sept
4:00pm
4W1.7

Helen Ogden

University of Southampton

In latent Gaussian models, including GLMMs, the likelihood function is an integral over the latent variables, and is often very difficult to compute. Several approaches to inference in latent Gaussian models rely on finding a Laplace approximation to the likelihood, and using this approximate likelihood as a proxy for the exact likelihood to do inference on the model parameters. This approach often seems to work well in practice, but there are some examples of models where inference obtained using a Laplace approximation to the likelihood is not close to the inference obtained using the exact likelihood. When should we be worried about the quality of inference obtained using a Laplace approximation to the likelihood? I will discuss recent work on this question in the special case of GLMMs, and consider how the structure of the model should behave to ensure that inference using the Laplace approximation will "tend towards" the inference with the exact likelihood, as the amount of information available about the parameters grows.

Efficient Gaussian process regression to calculate the expected value of partial perfect information in health economics

14 Sept
4:30pm
4W1.7

Gianluca Baio

University College London

The Expected Value of Perfect Partial Information (EVPPI) is a decision-theoretic measure of the "cost" of parametric uncertainty in decision making used principally in health economic decision making. Despite this decision-theoretic grounding, the uptake of EVPPI calculations in practice has been slow. This is in part due to the prohibitive computational time required to estimate the EVPPI via Monte Carlo simulations. However, recent developments have demonstrated that the EVPPI can be estimated by non-parametric regression methods, which have significantly decreased the computation time required to approximate the EVPPI. Under certain circumstances, high-dimensional Gaussian Process regression is suggested, but this can still be prohibitively expensive. Applying fast computation methods developed

in spatial statistics using Integrated Nested Laplace Approximations (INLA) and projecting from a high-dimensional into a low-dimensional input space allows us to decrease the computation time for fitting these high-dimensional Gaussian Processes, often substantially. We demonstrate that the EVPPI calculated using our method for Gaussian Process regression is in line with the standard Gaussian Process regression method and that despite the apparent methodological complexity of this new method, R functions are available in the package BCEA to implement it simply and efficiently.

---

15 Sept
11:00am
4W1.7

Spatially adaptive Bayesian estimation for probabilistic temperature forecasts

Annette Moeller
University of Goettingen

Uncertainty in the prediction of future weather is commonly assessed through the use of forecast ensembles that employ a numerical weather prediction model in distinct variants. Statistical postprocessing can correct for biases in the numerical model and improves calibration. We propose a Bayesian version of the standard ensemble model output statistics (EMOS) postprocessing method, in which spatially varying bias coefficients are interpreted as realizations of Gaussian Markov random fields. Our Markovian EMOS (MEMOS) technique utilizes the recently developed stochastic partial differential equation (SPDE) and integrated nested Laplace approximation (INLA) methods for computationally efficient inference. We investigate the performance of the MEMOS approach in a comparative study of 24-hour ahead temperature forecasts over Germany based on the 50-member ensemble of the European Centre for Medium-Range Weather Forecasting (ECMWF).

---

15 Sept
11:30am
4W1.7

Bayesian semiparametric inference of space-time trends in extremes of global temperatures

Thomas Opitz
INRA (French National Institute of Agricultural Research)

Extreme value theory provides models and tools for the statistical inference on extreme events whose analysis is crucial in many applications like finance/insurance or climate science. In this context, I will here propose a flexible latent Gaussian space-time regression approach for data in the so-called Gumbel maximum domain of attraction which attracts distributions such as the Gaussian, the log-Gaussian or the Weibull ones. When the data distribution is in the Gumbel domain of attraction, we can model its tail through the exceedance probability $p = P(X > u)$ associated to a fixed high threshold u and an exponential distribution for the excess $(X - u)_+$ above the threshold. By formulating a latent Gaussian regression model for p and the exponential excess, high-dimensional Bayesian space-time inference becomes feasible through the INLA approach. I will here present an application to modeling of space-time trends in global land and ocean monthly mean temperatures, highlighting the extent and spatial variation of climate change on a global scale.

An Additive Mixed Model for Raman Spectroscopy

15 Sept
12:00pm
4W1.7

Matt Moores
University of Warwick

Raman spectroscopy can be used to identify molecules by the characteristic scattering of light from a laser. Each Raman-active dye label has a unique spectral signature, comprised by the locations and amplitudes of the peaks. The presence of a large, nonuniform background presents a major challenge to analysis of these spectra. We introduce a sequential Monte Carlo (SMC) algorithm to separate the observed spectrum into a series of peaks plus a smoothly-varying baseline, corrupted by additive white noise. The peaks are modelled as Lorentzian or Gaussian functions, while the baseline is estimated using a penalised cubic spline. Our model-based approach accounts for differences in resolution and experimental conditions. By incorporating this representation into a Bayesian functional regression, we can quantify the relationship between molecular concentration and peak intensity, resulting in an improved estimate of the limit of detection. We also calculate the model evidence using SMC to investigate long-range dependence between peaks. These methods have been implemented as an R package, using RcppEigen and OpenMP.

A Bayesian hierarchical model for monthly maxima of instantaneous flow

16 Sept
9:45am
4W1.7

Birgir Hrafnkelsson
University of Iceland

We propose a comprehensive Bayesian hierarchical model for monthly maxima of instantaneous flow in river catchments. The Gumbel distribution is used as the probabilistic model for the observations, which are assumed to come from several catchments. Our suggested latent model is Gaussian and designed for monthly maxima, making better use of the data than the standard approach using annual maxima. At the latent level, linear mixed models are used for both the location and scale parameters of the Gumbel distribution, accounting for seasonal dependence and covariates from the catchments. The specification of prior distributions makes use of penalised complexity (PC) priors, to ensure robust inference for the latent parameters. The main idea behind the PC priors is to shrink toward a base model, thus avoiding overfitting. PC priors also provide a convenient framework for prior elicitation based on simple notions of scale. Prior distributions for regression coefficients are also elicited based on hydrological and meteorological knowledge. Posterior inference was done using the MCMC split sampler, an efficient Gibbs blocking scheme tailored to latent Gaussian models. The proposed model was applied to observed data from eight river catchments in Iceland. A cross-validation study demonstrates good predictive performance.

Building substantive beliefs into a prior distribution for the covariance matrix of a spatial multivariate normal distribution

16 Sept
11:00am
4W1.7

Sarah Heaps
University of Newcastle

In many analyses of multi–dimensional data, the dependence structure of the multivariate normal distribution is used to build relationships between variables. We discuss how to construct a prior distribution for the variances and covariances when we wish to convey substantive prior beliefs, with particular emphasis on spatial problems where there is typically no natural ordering of the variables but where, in our prior beliefs, we may well have different degrees of association between different pairs of covariances. Convenient conjugate priors are generally too inflexible and others which address this problem are difficult to interpret. We discuss an approach in which an interpretable prior is constructed and then converted to a flexible structure using a Cholesky decomposition. We apply this to some spatio-temporal data.

Approximate non-linear thinning and probabilistic marks for log Gaussian Cox process models in INLA

16 Sept
11:30am
4W1.7

Fabian Bachl
University of Bath

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 for which spatiotemporally varying models are needed. For instance, when a point represents an aggregate like a group of animals, a link to the distribution of aggregated entities is sought for. This introduces non-linearities to the linear predictor which are not covered by the current INLA methodology. As a remedy we propose approximations within an iterative scheme for INLA capable of solving the initial inference problem. We illustrate the method 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 framework successfully separates detectability and group size as well as animal and group intensity within a joint probabilistic formulation and the iterative INLA approach is used for inference.

Stochastic Neighborhood Structure in Bayesian Spatial Models

16 Sept
12:00pm
4W1.7

Aline Piroutek

Unicamp

Following a Bayesian approach in Disease mapping, we propose a new model where for a fixed neighborhood matrix class, we associate to each matrix of this class an a priori distribution. The goal is to update our knowledge about these unknown matrices after observing the data. The proposed model automatically adapts the neighborhood structure according to the evidence shown in data. We also propose two posterior estimators for matrices in a fixed neighborhood matrix class, allowing an analysis of influence between areas. One of the proposed estimators is simple and intuitive and the other estimator is more robust, taking into account the global influence oh the areas os the map, but it is more time-consuming.

High resolution mapping of poverty indicators using INLA

16 Sept
2:00pm
4W1.7

Jessica Steele

University of Southampton and Flowminder

Poverty is one of the most important determinants of adverse health outcomes globally, a major cause of societal instability, and one of the largest causes of lost human potential. Traditional approaches to estimate local prevalence of poverty, such as small area estimation techniques, rely on census data, which in most low and middle-income countries are unavailable or out-of-date. Alternate measures and data sources are needed to compliment and update poverty estimates between censuses. Here we use hierarchical Bayesian geostatistical models (BGMs) implemented in INLA to estimate fine-resolution poverty indicators using de-identified mobile phone data and remote sensing/GIS data. We find BGMs provide an advantageous approach to accurately predicting poverty metrics with complete modelled uncertainty. Stratifying models into urban and rural components exploits the fine detail offered by dense urban mobile phone networks and improved predictions (highest for wealth index: r2=0.78). Modelling the spatial covariance in the data by defining the neighbourhood structure using the besag model in the INLA function was very important; this resulted in improved measures of fit and lower error. The findings indicate the possibility to estimate and monitor poverty rates at high spatial resolution in countries where census data is out-dated, unavailable or unreliable.

---

16 Sept
2:30pm
4W1.7

Spatio-Temporal Modelling of Criminal Data in Portugal

Conceição Ribeiro
ISE/UALG and CEAUL

The study of the evolution of crime, whether in a temporal level or in a space level, presents a great importance in the definition of measures to improve the welfare of the population. Usually, when you want to analyze the evolution of crime in a given region, it resorts to compare rates of several years. This work aims to extend this analysis and use spatial and temporal models that allow to characterize the trend of crime in the spatial level and in the temporal level. In other words, it intends to understand if over the years and across regions have been changes in crime patterns. To achieve this aim, we use hierarchical Bayesian models and to implement these models we use the INLA methodology through the package R-INLA. These models are applied to data crimes observed in the municipalities of mainland Portugal, from 2011 to 2015.

---

16 Sept
3:00pm
4W1.7

Inequality in Beijing: A Spatial Multilevel Analysis of Perceived Environmental Hazard and Self-rated Health

Guanpeng Dong
ISE/UALG and CEAUL

Social boundaries are potentially important features of the urban landscape. The frontiers between contrasting neighbourhoods are potential sources of conflict and also of enlightenment and interaction. But such boundaries are rarely truly closed there may be a steep contrast with a particular neighbourhood in one direction, but a more gradual blending of neighbourhoods in other directions. This poses some formidable methodological challenges particularly if one seeks to compute statistical inference for the existence of a boundary. This paper presents a method that addresses these issues using an innovative Bayesian spatial statistical model. We demonstrate how the estimated boundaries can be used to compute transparent and easy to understand measures of segregation/diversity based on distance to social boundary. We also look at the question of whether the existence and steepness of social boundaries are indicative of potential social tensions. The approach is illustrated using data on ethnicity and crime of Sheffield, providing evidence that social frontiers are statistically significantly associated with the distributions of different types of crimes in Sheffield.

Contributed Posters

The MCMC split sampler: A block Gibbs sampling scheme for latent Gaussian model

15 Sept
4:00pm
4W Atrium

Oli Palli Geirsson
University of Iceland

Latent Gaussian models (LGMs) form a flexible subclass of Bayesian hierarchical models and have become popular in many areas of statistics and various fields of applications, as LGMs are both practical and readily interpretable. Although LGMs are well suited from a statistical modeling point of view their posterior inference becomes computationally challenging when latent models are desired for more than just the mean structure of a non-Gaussian data density function and when the number of parameters associated with the latent model increase.

We propose a novel computationally efficient Markov chain Monte Carlo (MCMC) scheme which serves to address these computational issue, we refer to as the MCMC split sampler. The sampling scheme is designed to handle LGMs where latent models are imposed on more than just the mean structure of the likelihood; to scale well in terms of computational efficiency when the dimensions of the latent models increase; and to be applicable for any choice of a parametric data density function. The main novelty of the MCMC split sampler lies in how the model parameters of a LGM are split into two blocks, such that one of the blocks exploits the latent Gaussian structure in a natural way and becomes invariant of the data density function.

Bayesian ICA-based source separation of Cosmic Microwave Background by a discrete functional approximation

15 Sept
4:00pm
4W Atrium

Jason Wyse
Trinity College Dublin

In this work we propose an Bayesian ICA-type approach to this separation task with spatially smooth Gaussian priors for the different sources, implemented by a scalable and robust functional approximation, the Integrated Nested Laplace Approximation (INLA). A variant of INLA is proposed in order to address a multi-modality issue which cannot be resolved by the basic approach. The performance of our approach is evaluated on two different datasets.

---

15 Sept
4:00pm
4W Atrium

Spatial Modelling of House Prices in the Dublin Metropolitan Area

James Sweeney

University College Dublin

Assessments of the state of play in the Irish housing market are mainly qualitative at present, based on simple summaries of subsets of the housing population which compare the average value of similar houses in different postcode areas. No model is presented to link the asking price to underlying house price information, or to avail of the latent structure inherent in the data due to its spatially correlated nature. In this poster I will present a proof of concept spatial model for house prices in the greater Dublin area, where the computational efficiency of the SPDE approach of Lindgren et al. (2011) is harnessed to model the spatial latent process. The model appears promising for price prediction given a number of simple property features and provides some interesting results in terms of the factors deemed important in the value of a property.

---

15 Sept
4:00pm
4W Atrium

Spatial regression and spillover effects in cluster randomized trials

Karim Anaya-Izquierdo

University of Bath

We will present a methodology to analyse data from cluster randomized trials taking into account spatial dependence and indirect effects. We use spatial regression models with Gaussian random effects where the corresponding intervention effects have a marginal interpretation. The random effects model spatial dependence using a homoscedastic modification of the intrinsic conditional autoregression (ICAR) model and the indirect effects are modelled using the distance to, and a novel measure of depth within, the intervention arm. We will focus the methodology on count responses and illustrate it using data from a pair-matched CRT against the dengue mosquito vector Aedes aegypti, done in Trujillo, Venezuela.

Accounting for physical barriers in species distribution modeling with non-stationary spatial random effects

15 Sept
4:00pm
4W Atrium

Haakon Bakka

Norwegian University of Science and Technology

Gaussian latent variable models are commonly used to model spatial data. In these models, the Gaussian random field accounts for the spatial correlation unexplained by covariates, including both unmeasured covariates and spatially correlated noise. A major advantage of these additive models is our ability to separate out the different assumptions and investigate their appropriateness.

In this talk we challenge the assumption of stationarity in cases where there are physical barriers in the study area. In these situations it is important to force the spatial dependencies around the barriers, to make our modelling assumptions more realistic. We develop a non-stationary Gaussian field that treats land as a barrier to spatial correlation. The core idea is to construct a Matérn Gaussian field with range zero on land. This is achieved by defining the field as a continuous solution to a differential equation with spatially varying coefficients, based on the SPDE approach by Lindgren et. al. [2011]. We achieve a computational cost similar to the stationary model.We demonstrate the benefits of the new model with a real application of species distribution modeling (SDM), analysing the reproduction habitats of a commercially important fish species in the Finnish archipelago. SDMs are used to predict the abundance pattern of species and to identify the environmental variables that best describe these patterns, e.g. to aid environmental managers in spatial planning.

There are many important applications where there is a lack of data, and we must put a lot of faith in our modeling assumptions. Then, the appropriateness of the a priori model is very important, but predictive checks may not be able to detect the difference in model fitness. We show that the application herein is an example of this.

Classifying X-Ray Binaries: A Probabilistic Approach

15 Sept
4:00pm
4W Atrium

Giri Gopalan

University of Iceland (currently Harvard Medical School but soon to join Ph.D. there)

(Quoting from ApJ work with Luke Bornn and Saku Vrtilek for which I intend to present a poster on.)

"In X-ray binary star systems consisting of a compact object that accretes material from an orbiting secondary star, there is no straightforward means to decide whether the compact object is a black hole or a neutron star. To assist in this process, we develop a Bayesian statistical model that makes use of the fact that X-ray binary systems appear to cluster based on their compact object type when viewed from a three-dimensional coordinate system derived from X-ray spectral data where the first coordinate is the ratio of counts in the mid- to low-energy band (color 1), the

second coordinate is the ratio of counts in the high- to low-energy band (color 2), and the third coordinate is the sum of counts in all three bands. We use this model to estimate the probabilities of an X-ray binary system containing a black hole, non-pulsing neutron star, or pulsing neutron star. In particular, we utilize a latent variable model in which the latent variables follow a Gaussian process prior distribution, and hence we are able to induce the spatial correlation which we believe exists between systems of the same type. The utility of this approach is demonstrated by the accurate prediction of system types using Rossi X-ray Timing Explorer All Sky Monitor data, but it is not flawless. In particular, non-pulsing neutron systems containing bursters that are close to the boundary demarcating systems containing black holes tend to be classified as black hole systems. As a byproduct of our analyses, we provide the astronomer with the public R code which can be used to predict the compact object type of XRBs given training data." (G.Gopalan, S.D. Vrtilek, L. Bornn ApJ 2015)

To elaborate more on this work and its relevance to latent Gaussian modeling; we used a Monte Carlo algorithm called elliptical slice sampling, due to Murray, Adams, and MacKay (2010 JMLR) to sample from the posterior predictive distribution of compact object type because of the Gaussian process prior on the latent variables. This MCMC algorithm was developed for inference within models employing a latent Gaussian structure and outperforms traditional MCMC schemes such as Gibbs sampling or Metropolis-Hastings in such models. Hence, as part of my poster, I intend to emphasize the critical use of this algorithm and an efficient computational implementation developed with Luke Bornn, available within an R package which relies on Rcpp and RcppEigen.

---

15 Sept
4:00pm
4W Atrium

Valid parameter space of multivariate Gaussian Markov random fields

Mattia Molinaro
University of Zurich

Gaussian Markov Random Fields (GMRFs) are commonly employed in several disciplines, for instance image analysis and spatial statistics (e.g., to find a suitable intensity distribution of an image or to analyze areal data). There are several reasons behind this success. First, efficient numerical routines can be used while performing statistical inference. Second, depending on the chosen parametrization, GMRFs can handle complicated correlation patterns in the data. However, even for a model with only a moderate number of parameters, great care must be taken when determining the valid parameter space, namely the set ensuring positive-definiteness of the associated precision matrix.

In this work, we present a theoretical framework for the determination of the valid parameter space for multivariate GMRFs which depend on several parameters and are defined over a regular bidimensional lattice. We consider precision matrices having a generalized block-Toeplitz structure and provide asymptotically closed-form expressions for their spectra. We thus generalize classical results for univariate fields

and fill the gap in the literature for the multivariate case. Additionally, we discuss a thorough simulation study, whose goal is to assess the rate of convergence of our approximation as a function of the lattice size.

Our approach allows to assess the validity of parameters without the need of constructing the precision matrix and computing its Cholesky decomposition. Considering the size of modern datasets,we illustrate the substantial computational efficiency gain of the method with a simulation study.

PC priors for degrees of freedom in P-spline models

15 Sept
4:00pm
4W Atrium

Massimo Ventrucci

University of Bologna

Penalized regression on B-splines, proposed by Eilers and Marx (1996) and denoted as P-splines, is a flexible and stable approach for fitting smooth functional effects in flexible regression models, GAM, STAR etc. The Bayesian approach to P-splines assumes an intrinsic Gaussian Markov random field (IGMRF) prior on the spline coefficients, conditional on a precision hyper-parameter. Typically, a conjugate Gamma prior is assigned to the IGMRF precision, as an attempt of vagueness, but this often results in a very informative prior for the degrees of freedom of the model. We hence consider building priors directly on the degrees of freedom, to "control" complexity of the smooth function to be estimated. Following this idea, we propose Penalized Complexity (PC) priors for the number of effective degrees of freedom in additive P-spline models for Gaussian responses. We discuss two extensions still in progress: implementation of the PC prior for degrees of freedom in Generalized P-splines and in spatial or spatio-temporal smoothing via Bayesian P-splines.

Not long, but wide. Farm structure changes over 3 decades within the Alpine arc: a spatio-temporal analysis

15 Sept
4:00pm
4W Atrium

Oyvind Hoveid

Norwegian Bioeconomic Research Institute

Data for more than 4000 villages at 4 time points in 7 countries within the Alpine Arc are analyzed with regard to possible drivers and out of sample prediction using SPDE-INLA. Co-author: Christian Hoffmann

---

15 Sept
4:00pm
4W Atrium

Random effects binary model to respondent driven-sampling data

Leonardo Bastos
Oswaldo Cruz Foundation

Respondent-driven sampling (RDS) is a procedure to sample hard-to-reach populations. It has been widely used in several countries, especially in the monitoring of HIV/AIDS and other sexually transmitted infections. Hard-to-reach populations have a key role in the dynamics of such epidemics and must inform evidence-based initiatives aiming to curb their spread. However, there is little research on estimating risk factors for health events in these hard-to-reach populations sampled by the RDS approach. In this paper, we present a random effects binary model for RDS data. We propose including random effects to take into account a possible network dependence among the study participants. The proposed model is illustrated in a RDS study for estimating risk factors for HIV and syphilis among men who have sex with another men (MSM) in Campinas, Brazil.

---

15 Sept
4:00pm
4W Atrium

Markov Switching Smooth Transition HYGARCH Model: Stability and Forecasting

Saeid Rezakhah Varnousfaderani
Amirkabir University of Technology

The HYGARCH models are commonly are applied in modeling the long rang dependence in volatility. To provide a more flexible model we propose Markov Switching Smooth Transition HYGARCH (MSST-HYGARCH) model that the volatility in each state is a convex combination of two components, short memory GARCH and long memory FIGARCH but with different weights. We consider the cases where there is a hidden Markov switch between two latent Gaussian models for expressing high volatility and low volatility effects. The necessary and sufficient condition for the asymptotic stability in variance is derived. The competitive performance of the proposed model is shown by applying simulated data and also $S\&P$ indices of some special period.

---

15 Sept
4:00pm
4W Atrium

Joint modeling of longitudinal and survival data with INLA

Ines Sousa
Universidade do Minho, Portugal

A Bayesian latent Gaussian spatial model for areal and point observations and predictions

15 Sept
4:00pm
4W Atrium

Ingelin Steinsland & Thea Roksvag
NTNU

This work is motivated of the problem of ungauged basins in hydrology. For most catchments there are no observation of run-off, nor any observations of precipitation within the catchment. Further, it is common to evaluate precipitation interpolation at points only even when the main interest is for an catchment, i.e. an area. The aim of this work is to; 1) Learn about the ungauged basin by using annual precipitation and run-off data for several years together with a model for evaporation. 2) To explore if, how and when evaluation of predictions for a points (within the catchment) and for an area are different. To get information about the catchment we use annual run-off data (areal observations) and/or precipitation data (point observations) in the surrounding area together with a model for evaporation. We set up a spatial Bayesian model for precipitation, with both point and areal observations. This model include both a spatial term for the climatology (constant over years), and a spatial term for year-to-year variation. It is a Bayesian latent Gaussian model, and a stochastic partial differential equation (SPDE) approach to spatial modelling. Inference is done using Laplace approximation within the R-package r-inla.The model is tested both through simulation studies and a case study of Norwegian catchments, and both point and areal observations are valuable. Further, we find that the ranking of observation designs can change when one evaluate for areal predictions instead of point predictions.

Maximum midge abundance and the potential spread of bluetongue virus in the UK

15 Sept
4:00pm
4W Atrium

Adam Butler & Beth Purse & Gustaf Rydevik & Kate Searle
BioSS & CEH & Roslin Institute & CEH

Bluetongue is a disease in ruminants such as cattle and sheep, caused by the bluetongue virus (BTV) and spread from host to host via midges. It can cause symptoms such as fever, lameness, swelling of the mouth and tongue, and even death in severe cases. While it is an endemic disease in countries surrounding the equator, the UK has only experienced one outbreak of bluetongue, in 2007, that had a severe economic impact. There has recently been a number of outbreaks of BTV in France, causing concern for the potential of a new UK introduction. We have used surveillance data of the three key UK midge species that are able to carry BTV (Culicoides obsoletus complex, C. pulicaris, and C. impunctatus) to develop a modelling framework in INLA for predicting the maximum annual abundance at sites across the UK based on environmental covariates (wild and domestic host densities or habitat suitability, meteorological variables, topography and landcover). Based on an established mathematical formula combining midge density, livestock density and temperature, we finally calculate an upper bound for the expected number of

new cases caused by a single infected midge at different locations across the UK, if its introduction were to coincide with a peak in midge density.

People

Scientific Committee

- Janine Illian (University of St Andrews)
- Finn Lindgren (University of Bath)
- Daniel Simpson (University of Bath)
- Andrea Riebler (NTNU, Trondheim)

Organising Committee

- Finn Lindgren (University of Bath)
- Daniel Simpson (University of Bath)

Participants

- Karim Anaya-Izquierdo
- Fabian Bachl
- Gianluca Baio
- Haakon Bakka
- Leonardo Bastos
- Michael Betancourt
- Samir Bhatt
- David Borchers
- Aoibheann Brady
- Francesco Brizzi
- Alessandra Carioli
- Jeremy Colman

- Guanpeng Dong
- Emiko Dupont
- Oli Palli Geirsson
- Giri Gopalan
- Sarah Heaps
- Sigrunn Holbek
- Oyvind Hoveid
- Birgir Hrafnkelsson
- Merrilee Hurn
- Chris Jarvis
- Alex Karagiannis
- Eliud Kibuchi
- Christos Kokaliaris
- Annette Moeller
- Aline Martines Piroutek
- Mattia Molinaro
- Matt Moores
- Kathryn Morrison
- Helen Ogden
- Thomas Opitz
- Andrew Parnell
- Saeid Rezakhah Varnousfaderani
- Conceio Ribeiro
- Thea Roksvag
- Gustaf Rydevik
- Maria Lie Selle
- Gavin Shaddick
- Theresa Smith

- Ines Sousa
- Jessica Steele
- Ingelin Steinsland
- James Sweeney
- Matthew Thomas
- Aki Vehtari
- Massimo Ventrucci
- Jason Wyse
- Yuan Yuan
- James V. Zidek
- Amanuel

Author Index