BOOK OF ABSTRACTS

Workshop on Bayesian Modeling for Complex Correlated Data

Valencia, May 16th – May 18th, 2023

1 Program Tuesday May 16th

- 10:00 – 10:30 *Registration and coffee welcome*
- 10:30 – 11:40 *Plenary session 1. Chairperson: C. Armero*

  Håvard Rue (KAUST): **INLA-2.0 and preparing the ground for INLA-3.0**.

  In the first part of the talk, I will discuss the "new INLA" was launched in December 2022. The new INLA has another internal parameterisation and use a new take on how to do the "nested" approximations, which boils down to a new take on Variational Inference in the innermost layer. I will discuss how we correct the posterior mean using Variation Inference in some more details.

  In the second part of the the talk, I will discuss various computational strategies for a "massive parallel"-INLA, were we try to prepare the ground for more available cores in the future. This includes dense matrix methods, iterative methods and new sparse solvers with GPU kernels.
- 11:40 – 11:50 *Mini break*
- 11:50 – 13:30 *Invited to contribute PhD student session. Chairperson: R. Amorós*

  Gabriel Calvo (20min): **Applying Bayesian longitudinal models in basketball**

  This presentation will discuss the application of Bayesian longitudinal models in basketball and wheelchair basketball. The first study analyses the hot hand phenomenon in consecutive basketball shots using a Bayesian longitudinal hidden Markov model. The model assumes two possible states, cold or hot, and calculates the probability of success for each throw by considering the corresponding hidden state and the distance to the basket. The model is applied to a real dataset of the Miami Heat team during the 2005-2006 season of the US National Basketball Association.

  In the second study, a Bayesian longitudinal model is used to optimise the lineup of a wheelchair basketball team. Three different performance ratings are considered for each player, and uncertainty from the posterior predictive distribution of the longitudinal model is incorporated into the optimisation problem. The results indicate that the optimal lineup is influenced by factors such as the functional classification and sex of the players. Additionally, the methodology is shown to be applicable to other team sports that cater to individuals with physical disabilities.

  Pablo Escobar (20min): **Studying the effect of the COVID-19 crisis in mental health using Multivariate Spatial Models**

  Suicide is becoming one of the leading causes of death in developed countries, with data showing an increasing trend in the last decades. Moreover, the impact that COVID-19 has had on mental health has been a raising concern since the early days of the pandemic. In

this seminar we will present a PhD project that belongs to a research project named "Social epidemiology of suicide-related 112 calls in covid-19 crisis: A Bayesian spatio-temporal approach", which focus on developing multivariate spatial analysis methods. The aim is to study the interactions between several different categories of emergency calls, spliting the data in 2x2x2 categories: type of caller (victim vs witness callers), gender of the victim (male vs female victims) and period (pre-COVID vs COVID).

Alba Fuster (20min): **Bayesian hierarchical spatio-temporal models for analyzing the cumulative effects of global change and plausible future oceans**

Hierarchical Bayesian spatio-temporal models are used to estimate and predict the distribution of marine species at meso and macroscale. This thesis aims to improve some characteristics of the usual species distribution models (SDMs), such as their interaction with stock assessment models or ecosystem models, and their ability to make long-term predictions at a global scale. The main goal of the thesis is to understand and improve the estimation and prediction of spatio-temporal dynamics of marine species at a regional and global scale under different climate change scenarios. To achieve this, the following specific objectives are developed:

- Apply species distribution models (SDMs) at a global scale and evaluate their use in informing marine ecosystem models (MEMs);
- Develop a validation protocol for analyzing and comparing predicted maps;
- Model complex scenarios with multiple likelihoods.

Jesús Gutiérrez (20min): **Bayesian survival models to study disease progression events in patients with heart failure**

Heart Failure (HF) is a chronic condition which happens when the heart is not able to pump enough blood to supply the patient's tissues. Cardiac Resynchronization Therapy (CRT) is a procedure to implant electrodes in the heart's chambers to make the heart work in a more organized and efficient way. This intervention has shown to have effects on short-term periods of time, but there is scarce data about its long-term benefits.

296 patients with heart failure who received CRT between August 2001 and April 2015 have been followed and disease progression events have been registered. These events include the development of congestive heart failure after the therapy, atrial fibrillation, and death from two different causes: cardiovascular and non-cardiovascular.

To study them, we are applying Bayesian survival models. A Bayesian competing risks model for the events cardiovascular and non-cardiovascular deaths have been used, and posterior outputs such as cause-specific hazard function or the overall survival function have been obtained. Some different specifications for the baseline hazard function have been also considered. A multistate survival model for the development of congestive heart failure and atrial fibrillation after CRT has been proposed. For all the models, Bayesian estimation has been performed using MCMC methods with JAGS.

Mario Pereira (20min): **Improving Species Distribution Models using Bayesian feedback**

In ecology we may find scenarios where the same phenomenon (species occurrence, species abundance, etc.) is observed using two different types of samplers. For instance, species data can be collected from scientific surveys with a completely random sample pattern, but also from opportunis- tic sampling (e.g., whale or bird watching fishery commercial vessels), in which observers tends to look for a specific species in areas where they expected to find it.

Species Distribution Models (SDMs) are a widely used tool for analyzing this kind of ecological data. Specifically, we have two geostatistical models available for the above data: an independent model (IM) for the data coming from a complete random sampler and a preferential model (PM) for data from opportunistic sampling.

In this work, we propose a sequential Bayesian procedure to connect these two models through the update of prior distributions. Although, the analysis is restricted to this scheme because of its strong presence in ecology, we also expose the general structure for performing Bayesian feedback between an arbitrary number of models. Implementation of the Bayesian paradigm was done through the integrated nested Laplace approximation (INLA) methodology. Additionally, we have also used the stochastic partial differential equations (SPDE) and finite element method (FEM) approach im- plemented within INLA. This allows us to perform spatial models with high performance and low computational costs. The sequential procedure has been evaluated by simulating several scenarios and comparing the results of sharing information from one model to another using different criteria. Our main results imply that, in general, it is better to share information from the independent (completely random) to the dependent model than the alternative way. However, it depends on different factors such as spatial range or the location resulting from sample distributions.

- 13:30 – 15:30 *Lunch break at Birlibirloque Bar, Calle la Paz, 7*

- 15:30 – 17:10 *Invited session: SPATIAL ANALYSIS. Chairperson: D. Conesa*

  Virgilio Gómez-Rubio (30min): **Hierarchical spatial distribution models**.

  In my talk I will present some results on the use hierarchical Bayesian models for species distribution models of plants in Spain. Spatial models will be fit using the integrated nested Laplace approximation and the results will be compared to other techniques for modeling species distribution. This is joint work with some members of the SABINA research group (UAM, Madrid, Spain).

  Francisco Palmí (20min): **Analysing multivariate spatial data with INLA**

  The joint analysis of different variables is continuously increasing in these day and age. Furthermore, multivariate spatial analysis can be computationally demanding. Therefore, the main objective of this work is to show that R-INLA is a convenient toolbox to analyse different types of multivariate spatial datasets. Some interesting details have been discussed

such as the choice of the prior distribution and the appropriate data structure. Additionally, three different datasets have been analysed in order to illustrate the main objective. The chosen datasets are available in different R packages or can be downloaded from Github. The necessary code to replicate and reproduce the different examples are also available. Finally, it has been shown that R-INLA is a suitable alternative to straightforwardly analyse multivariate spatial datasets.

Sébastien Coube (20min): **A tackle towards the density curse in Gaussian Markov Random Fields sampling: a "blocks and bases" approach**

Gaussian Markov Random field spatial and space-time priors induce a coherence of the latent process that makes naive Gibbs sampling and coordinate descent impracticable. An usual solution is to update the field in spatial blocks. However, when the field is dense in space, we are caught between a rock and a hard place : the blocks must be small enough to be updated in a reasonable time, but they must be big enough to deal with spatial auto-correlation. A very simple scheme based on spatial bases of increasing resolution gives encouraging behaviors with this issue.

- 17:10 – 17:40 Orxata break
- 17:40 – 18:30 *Work-group session 1: SDMs book proposal* (X. Barber, A. López-Quílez, D. Conesa and V. Gómez-Rubio)
- 20:30 *Workshop dinner at Entrevins, First floor Calle la Paz, 7*

2 Program Wednesday May 17th

- 9:20 – 10:30 *Plenary session 2. Chairperson: M.A. Martinez-Beneito*

  Håvard Rue (KAUST): **Robustification of LGM's and non-separable models in space-time**.

  In the first part of the talk, I will discuss the issue of assessing the latent Gaussian model (LGM) assumptions, and how to extend LGM to near-LGM's to correct for potential deviations. This has been a long long-standing issue. Our first attempt, more than 10 years ago, lead to the concept of 'Penalized Complexity Priors' as one of the tools required to moving forward. At last, we were able to make progress on this issue, which I will discuss.

  In the second part of the talk I will present the new class(es) of non-separable space-time models and the new package `INLAspacetime` which implement these extensions. Using the SPDE formulation, we can extend the Matern-family to also include time, and then provide a class of physical motivated non-separable space-time models. They have all a unified parameterisation, range in space, time, and marginal variance, and a non-separability parameter. This allows for easy changing models without changing interpretation of the results and prior settings.

- 10:30 – 11:00 *Coffee break*

- 11:00 – 12:10 *Work-group session 2*: **Multiple likelihoods** (X. Barber, D. Conesa, A. Fuster, A. López-Quílez and Håvard Rue)

- 12:10 – 12:20 *Mini break*

- 12:20 – 13:30 *Work-group session 3*: **Multivariate disease mapping** (P. Escobar, A. López-Quílez, M. Martínez-Beneito, F. Palmí and Håvard Rue)

- 13:30 – 15:30 *Lunch break at Birlibirloque Bar, Calle la Paz, 7*

- 20:30 *Dinner at Le Marquis Restaurant - SH Ingles Hotel, Calle Marqués de Dos Aguas, 6*

3 Program Thursday May 18th

- 9:20 – 10:30 *Plenary session 3. Chairperson: A. Forte*

  Håvard Rue (KAUST): **Joint models, no-feedback and graphical models for correlation matrices**

  In this talk I will present our recent work towards joint survival models which is implemented in the package `INLAjoint`. This new interface makes it easy to analyse time-to-event data, like joint competing risk, CoxPH, parametric models. Although its Bayesian, the results also exhibit good frequentist properties.

  Buried withing the joint models formulation, is the question if we should be doing the analysis with no-feedback, to protect for model-miss-specification. This will be discussed.

  Joint models often require correlated "random effects" to some dimension, like 3, 4, 5, ..., k, small, but not that small as the number of (hyper-)parameters is $\mathcal{O}(k^2)$. In order to avoid generic/general "Wishart priors" that might/will not do good, we need a lower-dimensional representation of correlation matrices with graphical/casual interpretation. I will discuss how this can be constructed, and how can construct automatic coherent priors using the Penalized Complexity priors-framework.

- 10:30 – 11:00 *Coffee break*

- 11:00 – 12:10 *Work-group session 4*: **Dirichlet models** (D. Conesa, A. López-Quílez, J. Martínez-Minaya, M. Pereira and Håvard Rue)

- 12:10 – 12:20 *Mini break*

- 12:20 – 13:30 *Work-group session 5*: **Temporal window-dependent covariates in regression models aimed to identify critical windows of exposure** (R. Amorós, C. Iñiguez and Håvard Rue)

- 13:30 – 15:30 *Lunch break at Birlibirloque Bar, Calle la Paz, 7*