VIBASS 1

FIRST VALÈNCIA INTERNATIONAL BAYESIAN ANALYSIS SUMMER SCHOOL

17-21 JULY 2017

FACULTY OF MATHEMATICS

DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH

VAlencia BAyesian Research group

VNIVERSITAT DE VALÈNCIA

BioSS

LOCAL ORGANIZING AND SCIENTIFIC COMMITTEES

Carmen Armero	Chair, Universitat de València
Mark Brewer	Biomathematics and Statistics Scotland, BioSS
María Eugenia Castellanos Nueda	Universidad Rey Juan Carlos
David Conesa	Universitat de València
Anabel Forte	Universitat de València
Virgilio Gómez-Rubio	Universidad de Castilla-La Mancha

WELCOME AND PRESENTATION

Welcome to the 1st VIBASS! The organizing committee would like to thank you for being here.

Valencia is regarded one of the main homes of Bayesian inference thanks to some very well-known Bayesian Statisticians: M.J. Bayarri, J.M Bernardo and J. Ferrandiz, among others. Given the continuing interest in Bayesian reasoning we could not think of a better place to hold a Bayesian summer school... and so... here we are: The Valencia International Bayesian Summer School, intended to offer an opportunity to be introduced to Bayesian Statistics.

As you may be aware, the five days of VIBASS are organised in three parts:

1. The first two days include a basic course on Bayesian learning (12 hours), with conceptual sessions in the morning and practical sessions using basic Bayesian R packages in the afternoon. Participants are advised to bring their laptops for the practical sessions.

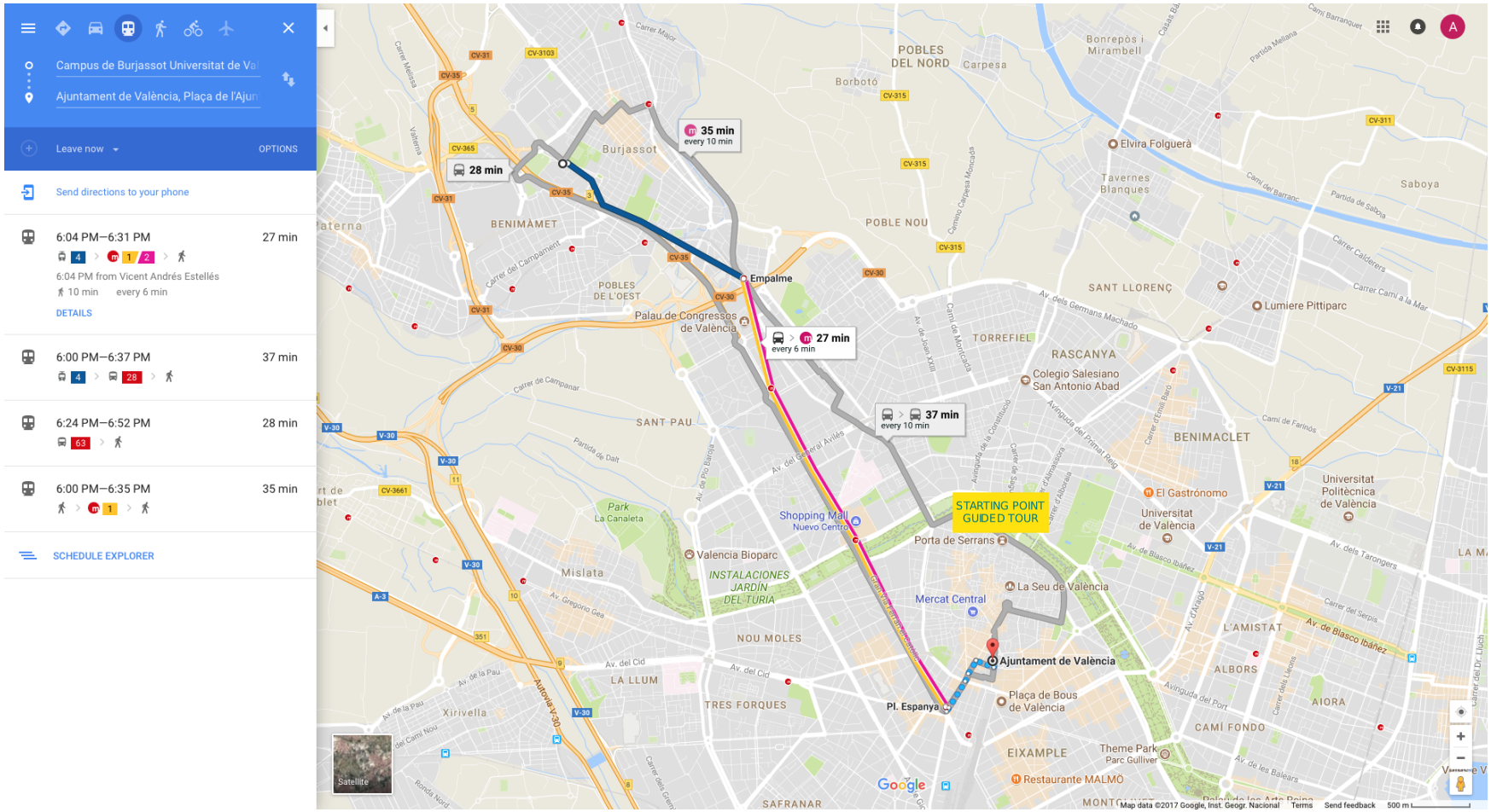
2. The second part in this first edition of VIBASS is dedicated to Stan, a probabilistic programming language for describing data and models for Bayesian inference. This course (12 hours) is provided by Aki Vehtari, Associate Professor in Computational Science at Aalto University (Helsinki), co-author of the third edition of the book Bayesian Data Analysis and member of the Stan development team.

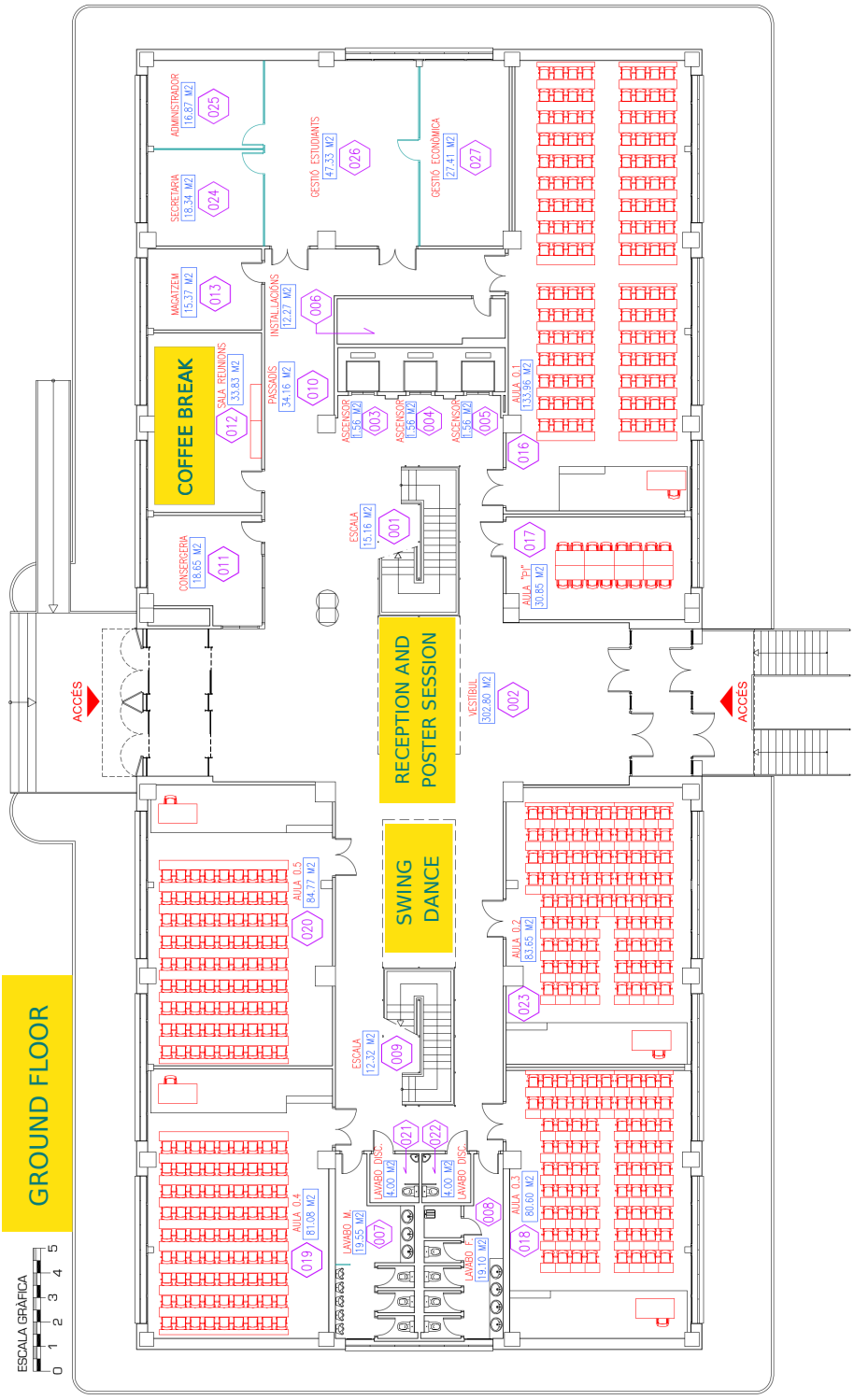
3. The last day, Friday 21st, is devoted to the First VIBASS Workshop. It includes two plenary sessions, with invited speakers researchers Mark Brewer from Biomathematics and Statistics Scotland (BioSS) and Maria Grazia Pennino (Instituto Oceanográfico Español), and a poster session with contributions from many participants of the school.

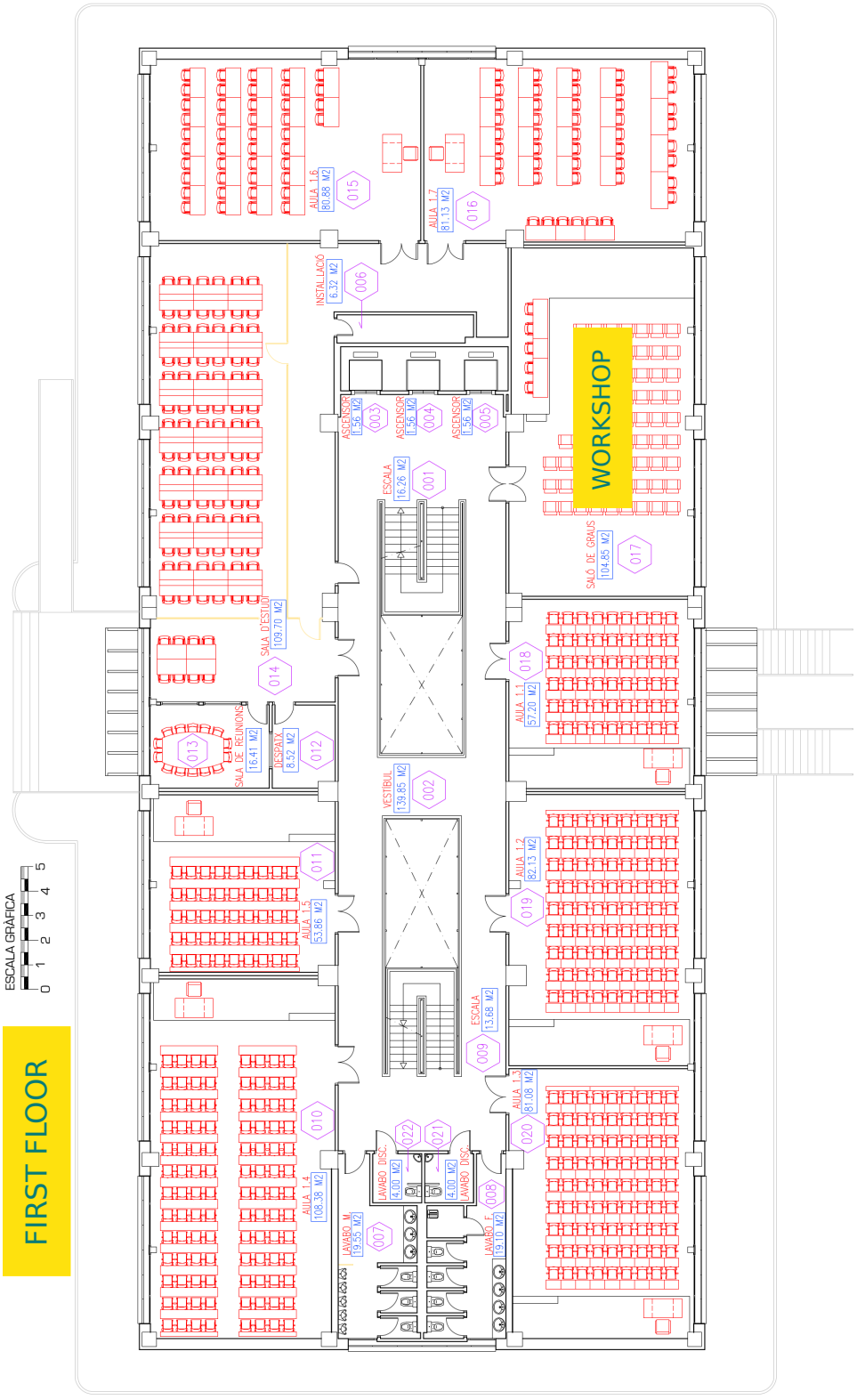
We have also prepared an exciting social program whose main purpose is to get together, to interact and to share knowledge (not only scientific).

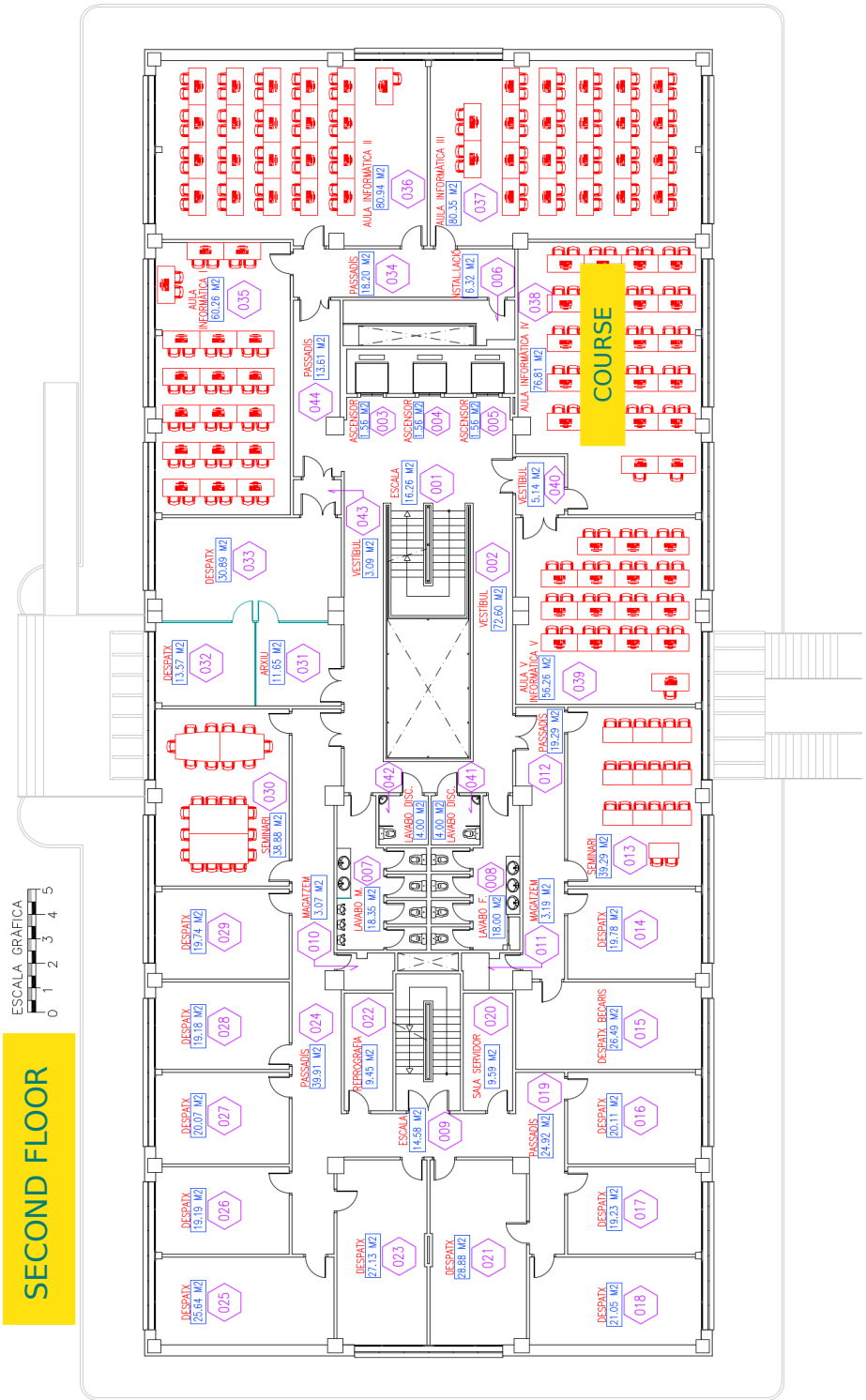
Since this is our first edition (hopefully the first of many more) we encourage you to share your impressions to allow us to improve. But, above everything we want you to learn and enjoy!!!

Welcome to Valencia!


Campus de Burjassot Universitat de Va
Ajuntament de València, Plaça de l'Ajun
Leave now
OPTIONS
Send directions to your phone
6:04 PM–6:31 PM
27 min
6:04 PM from Vicent Andrés Estellés
10 min
every 6 min
DETAILS
6:00 PM–6:37 PM
37 min
6:24 PM–6:52 PM
28 min
6:00 PM–6:35 PM
35 min
SCHEDULE EXPLORER
35 min
every 10 min
28 min
27 min
every 6 min
37 min
every 10 min
STARTING POINT
GUIDED TOUR
Empalme
Burjassot
Palau de Congressos de València
Ajuntament de València
Pl. Espanya
Mercat Central
Porta de Serrans
Shopping Mall
Nuevo Centro
Valencia Bioparc
Park
La Canaleta
Universitat de València
Universitat Politècnica de València
Colegio Salesiano San Antonio Abad
Lumiere Pittiparc
Elvira Folguerà
Plaça de Bous de València
La Seu de València
Theme Park
Parc Gulliver
Restaurante MALMÖ
El Gastrónomo
Satellite
Google
Map data ©2017 Google, Inst. Geogr. Nacional
Terms
Send feedback
500 m


GROUND FLOOR
COFFEE BREAK
RECEPTION AND POSTER SESSION
SWING DANCE
ACCÉS
ACCÉS
ESCALA GRÀFICA
0 1 2 3 4 5
ADMINISTRADOR 16.67 M2 025
SECRETARIA 16.34 M2 024
GESTIÓ ESTUDIANTS 47.33 M2 026
GESTIÓ ECONÒMICA 27.41 M2 027
MAGATZEM 15.37 M2 013
INSTAL.LACIONS 12.27 M2 006
SALA REUNIONS 33.83 M2 012
PASSADÍS 34.16 M2 010
ASCENSOR 5.55 M2 003
ASCENSOR 5.55 M2 004
ASCENSOR 5.55 M2 005
AULA 0.1 133.96 M2 016
CONSERGERIA 16.65 M2 011
ESCALA 15.16 M2 001
AULA "bcn" 30.65 M2 017
VESTÍBUL 302.80 M2 002
AULA 0.5 84.77 M2 020
ESCALA 12.32 M2 009
AULA 0.2 83.65 M2 023
LAVABO DISC 4.00 M2 021
LAVABO DISC 4.00 M2 022
AULA 0.4 81.08 M2 019
LAVABO M 19.55 M2 007
LAVABO F 19.10 M2 008
AULA 0.3 80.60 M2 018


FIRST FLOOR
ESCALA GRÀFICA
0 1 2 3 4 5
WORKSHOP
AULA 1.6
80.88 M2
015
AULA 1.7
81.13 M2
016
INSTAL.LACIÓ
6.32 M2
006
ASCENSOR
1.56 M2
003
ASCENSOR
1.56 M2
004
ASCENSOR
1.56 M2
005
ESCALA
16.26 M2
001
SALÓ DE GRAUS
104.85 M2
017
SALA D'ESTUDI
109.70 M2
014
AULA 1.1
57.20 M2
018
SALA DE REUNIONS
16.41 M2
013
DESPATX
8.52 M2
012
VESTÍBUL
139.85 M2
002
AULA 1.5
53.86 M2
011
AULA 1.2
82.13 M2
019
ESCALA
13.68 M2
009
AULA 1.3
81.06 M2
020
AULA 1.4
108.38 M2
010
LAVABO DISC.
4.00 M2
022
4.00 M2
021
LAVABO DISC.
LAVABO M
19.55 M2
007
LAVABO F.
19.10 M2
008


SECOND FLOOR
ESCALA GRÀFICA
0 1 2 3 4 5
COURSE
AULA INFORMÀTICA II
80.94 M2
036
AULA INFORMÀTICA III
80.35 M2
037
AULA INFORMÀTICA I
60.26 M2
035
AULA INFORMÀTICA IV
76.81 M2
038
AULA V INFORMÀTICA V
56.26 M2
039
PASSADÍS
18.20 M2
034
PASSADÍS
13.61 M2
044
INSTAL.LACIÓ
6.32 M2
006
ASCENSOR
1.56 M2
003
ASCENSOR
1.56 M2
004
ASCENSOR
1.56 M2
005
ESCALA
16.26 M2
001
VESTÍBUL
5.14 M2
040
VESTÍBUL
3.09 M2
043
VESTÍBUL
72.60 M2
002
DESPATX
30.89 M2
033
DESPATX
13.57 M2
032
ARXIU
11.65 M2
031
SEMINARI
38.88 M2
030
LAVABO DISC.
4.00 M2
042
LAVABO DISC.
4.00 M2
041
PASSADÍS
19.29 M2
012
SEMINARI
39.29 M2
013
MAGATZEM
13.07 M2
010
LAVABO M.
18.35 M2
007
LAVABO F.
18.00 M2
008
MAGATZEM
3.19 M2
011
DESPATX
19.74 M2
029
DESPATX
19.78 M2
014
DESPATX
19.18 M2
028
PASSADÍS
39.91 M2
024
REPROGRAFIA
9.45 M2
022
SALA SERVIDOR
9.59 M2
020
DESPATX BEDARS
26.49 M2
015
DESPATX
20.07 M2
027
DESPATX
20.11 M2
016
ESCALA
14.58 M2
009
PASSADÍS
24.92 M2
019
DESPATX
19.19 M2
026
DESPATX
19.23 M2
017
DESPATX
27.13 M2
023
DESPATX
28.88 M2
021
DESPATX
25.64 M2
025
DESPATX
21.05 M2
018

VIBASS 2017.
EXTENDED PROGRAM FOR THE BAYESIAN INFERENCE COURSE

- **Monday 17th and Tuesday 18th: An Introduction two Bayesian Learning**

  Monday 17th Introduction. All you need is . . . probability. Proportions. Count data. Normal data. Summarising posterior inferences. Estimation and prediction. Nuisance parameters. Joint prior distributions. Joint, conditional and marginal posterior distributions. Hypothesis testing. Bayes factor. Hands-on probability and lacasitos and sweets. Theoretical and computational issues for Beta-Binomial model, Poisson-Gamma model, and Normal-Normal model.

  Tuesday 18th Bayesian statistical modeling. The Basis of Bayesian Generalized linear models. Response variables, covariates, factors (fixed and random). Incorporating space and time in the same way. Hierarchical Bayesian modeling. Hierarchies or levels. Parameters and hyperparameters. Priors and hyperpriors. Introduction to Bayesian computational methods. Laplace approximations, Monte Carlo integration and importance sampling. Markov Chain Monte Carlo: Gibbs sampling and Metropolis Hastings. Convergence, inspection of chains, etc. Software for performing MCMC. Programming your own Metropolis-Hasting algorithm. R Software for inference in Bayesian hierarchical models.

- **Wednesday 19th and Thursday 20th: Bayesian Inference with Stan**

  Wednesday 19th Stan Intro. Quick GLM hands-on with rstanarm, rstan, bayesplot, shinystan. Hamiltonian Monte Carlo + NUTS + step size adaptation. Convergence diagnostics. Effective sample size. Hands-on with rstanarm, rstan, shinystan. Posterior predictive checking. Hands-on with rstanarm, rstan, bayesplot. +Topics based on the feedback

  Thursday 20th Model comparison. Hands-on with rstanarm, rstan, loo. Model averaging, Priors for large p, small n. Hands-on with rstanarm, rstan. Variable selection. Hands-on with rstanarm, projpred +Topics based on the feedback

PROGRAMME

Monday, July 17th. 2017

9:15 - 9:45	**REGISTRATION**
9:45 - 10:00	**WELCOME**
10:00 - 11:30	**LECTURE I** Teacher: **Carmen Armero**.
11:30 - 12:00	**Coffee break**
12:00 - 13:30	**LECTURE II** Teacher: **Carmen Armero**.
13:30 - 15:00	**Lunch**
15:00 - 16:30	**PRACTICAL I** Teachers: **Anabel Forte and Virgilio Gómez-Rubio**.
16:30-17:00	**Orxata break**
17:00- 18:30	**PRACTICAL II** Teachers: **Anabel Forte and Virgilio Gómez-Rubio**.
18:30- 20:00	**SOCIAL PROGRAMME** *Get together Beer* *Location: a bar close to the Venue (we will go together from there)*

Tuesday, July 18th. 2017

10:00 - 11:30	**LECTURE III** Teacher: **David Conesa**.
11:30 - 12:00	**Coffee break**
12:00 - 13:30	**LECTURE IV** Teacher: **David Conesa**.
13:30 - 15:00	**Lunch**
15:00 - 16:30	**PRACTICAL III** Teachers: **Anabel Forte and Virgilio Gómez-Rubio**.
16:30-17:00	**Orxata break**
17:00- 18:30	**PRACTICAL IV** Teachers: **Anabel Forte and Virgilio Gómez-Rubio**.
18:30- 20:00	**SOCIAL PROGRAMME** *Swing Dance* *Location: Hall of the Faculty of Mathematics*

Wednesday, July 19th. 2017

10:00 - 11:30	**STAN SESSION** Teacher: **Aki Vehtari**.
11:30 - 12:00	**Coffee break**
12:00 - 13:30	**STAN SESSION** Teacher: **Aki Vehtari**.
13:30 - 15:00	**Lunch**
15:00 - 16:30	**STAN SESSION** Teacher: **Aki Vehtari**.
16:30-17:00	**Orxata break**
17:00- 18:30	**STAN SESSION** Teacher: **Aki Vehtari**.
19:30- 21:30	**SOCIAL PROGRAMME** *Guided Tour* *Location: Serrano towers, Valencia*

Thursday, July 20th. 2017

10:00 - 11:30	**STAN SESSION** Teacher: **Aki Vehtari**.
11:30 - 12:00	**Coffee break**
12:00 - 13:30	**STAN SESSION** Teacher: **Aki Vehtari**.
13:30 - 15:00	**Lunch**
15:00 - 16:30	**STAN SESSION** Teacher: **Aki Vehtari**.
16:30-17:00	**Orxata break**
17:00- 18:30	**STAN SESSION** Teacher: **Aki Vehtari**.
20:00	**GALA DINNER** *Location: Restaurante Puerta del Mar*

Friday, July 21st. 2017

10:00 - 11:00 **1st VIBASS WORKSHOP**
INVITED SESSION
Mark Brewer.
Bayesian modelling: case studies in ecology and environmental science

11:00 - 11:30 **Coffee break**

11:30 - 12:30 **1st VIBASS WORKSHOP**
INVITED SESSION
Maria Grazia Pennino.
Bayesian Species Distribution Models: an overview

12:30 - 13:30 **1st VIBASS WORKSHOP**
POSTERS

13:30 - 14:00 **VIBASS CLOSING SESSION**

14:00 — **Lunch**

INDEX OF ABSTRACTS

BAYESIAN MODELLING: CASE STUDIES IN ECOLOGY AND ENVIRONMENTAL SCIENCE

Mark Brewer

Bayesian analysis has much to offer scientists working in the ecological and environmental sciences. One of the most important aspects is a completely transparent and explicit handling of uncertainty. We present examples from real applications highlighting how Bayesian methodology provides a straightforward approach to modelling unknowns. These examples include modelling both observation and detection processes in species distribution modelling in ecology, and modelling of unknown source distributions in hydrological analysis of water quality data.

BAYESIAN SPECIES DISTRIBUTION MODELS: AN OVERVIEW

Maria Grazia Pennino

Species Distribution Models (SDMs) are now widely used in ecology for management and conservation purposes across terrestrial, freshwater, and marine realms. The increasing interest in SDMs has drawn the attention of ecologists to spatial models, in particular, to Geostatistical models, which are used to combine observations of species occurrence or abundance with environmental estimates in a finite number of locations in order to predict where (and how much of) a species is likely to be present in unsampled locations. However, nowadays the quantity and the quality of available datasets has substantially increased with respect to the last years, resulting in a higher complexity of the statistical issues that have to be addressed when a SDM is implemented. This complexity has made the inferential and predictive processes challenging to perform. Bayesian statistics has become a good option to deal with these models, due to the easiness in which it handles this complexity by means of the hierarchical models. However, despite the advantages of Bayesian inference, the main challenge still remains in finding an analytic expression for posterior distributions of the parameters and hyperparameters. Several numeric approaches have been proposed such as Markov chain Monte Carlo methods and integrated nested Laplace approximation. This talk presents the most important problems that arise when researchers use SDMs, starting with the different spatial nature of the available data, and then focusing on the different spatial and spatio-temporal structures that can be addressed using Bayesian models.

A SEQUENTIAL STRATEGY FOR BAYESIAN JOINT MODELS OF LONGITUDINAL AND TIME-TO-EVENT DATA: AN APPROACH TO PERSONALISED MEDICINE

D. Alvares, C. Armero, A. Forte and N. Chopin

The statistical analysis of the information generated by medical follow-up is a very important challenge in the field of personalised medicine. As the evolutionary course of a patient's disease progresses, its medical follow-up generates more and more information that should be processed immediately in order to review and update its prognosis and treatment. In this work, our objective focuses on this update process through sequential inference methods for joint models of longitudinal and time-to-event data from a Bayesian perspective. More specifically, we propose the use of sequential Monte Carlo methods for static parameter joint models in order to update the posterior distribution of the parameters, hyperparameters, and random effects with the intention of reducing computational time in each update of the inferential process. Our approach is illustrated by means of a joint model with competing risk events for patients receiving mechanical ventilation in intensive care units.

BAYESIAN STOCK ASSESSMENT MODEL FOR THE SARDINE IN THE BAY OF BISCAY

L. Citores, L. Ibaibarriaga, L. Pawlowski, A. Uriarte and D.J. Lee

Fisheries stock assessment main purpose is to determine the past and current status of a fish population. Bayesian statistical modelling has become an important tool in this area given that it provides a conceptually elegant approach and allows incorporating prior information. This work aims at developing a statistical catch at age Bayesian model for the assessment of the Sardine in the Bay of Biscay, which as most of the fish stocks of commercial interest, is exploited even if their abundance, productivity and sustainability are highly uncertain. It has been constructed using Jags and results have been compared with other non-Bayesian approaches, showing very similar results in terms of stock status. In general, signal to noise ratio in the research surveys make difficult to set the absolute level of the stock abundance. The effect on bias of distinct levels of data uncertainty or fishing mortality have been studied trough simulation.

MARGINALLY SIGNIFICANT RESULTS IN PSYCHOLOGY: REPLICATING AND EXTENDING PRITSCHET ET AL (2016)

A.O. Collentine, C. Hartgerink, and M. Van Assen

We examine the proportion of marginally significant results in 74,489 articles published between 1985 and 2016 in 74 journals belonging to the American Psychological Association (APA). Pritschet et al (2016) previously examined the prevalence of 'marginally significant' results over the years in three subfields of psychology (cognitive, developmental, and social), and concluded that 'marginally significant' results are i) becoming more prevalent over time and ii) are the most prevalent in social psychology. However, their conclusions may be invalidated due to not taking into account differences in the number of reported p-values between fields and over years. We thus adapt their dependent variable and look at the proportion of p-values reported as marginally significant. Beyond replicating Pritschet et al using additional data, we examine the prevalence of marginally significant results in seven additional subfields: Clinical psychology, 'core of psychology', educational psychology, experimental psychology, forensic psychology, health psychology, and organizational psychology.

BAYESIAN MULTIVARIATE POINT PATTERN ANALYSIS FOR SPATIAL EPIDEMIOLOGY

V. Gómez-Rubio, F. Palmí-Perales, G. López-Abente Ortega, R. Ramis-Prieto and P. Fernández-Navarro

In this work, a Bayesian point pattern analysis of differnt diseases is carried out. This analysis is applied to a dataset composed by georeferenced cases of different types of cancer (lung, stomach and kidney) and a group of controls. All of this data is located in Alcalà de Henares, Madrid. Cases included people aged 39 or higher diagnosed with that type of cancer from January of 2012 to June of 2014.

Each disease distribution is considered as a log-Gaussian Cox process (Möller et. al. 1998) and the logarthim of the intensity is modelled as a sum of an intercept and an spatial effect. Using the integrated nested Laplace aproximation (INLA, Rue et. al. 2009) and the stochastic partial differential equation aproximation (SPDE, Lingren et. al. 2011) an estimate of the spatial effects have been computed. Results of each disease have been compared with the control group following the ideas proposed in Diggle et. al. (2007) and Gomez-Rubio et. al. (2015). Future work will be focused on introducing explanatory variables and building a joint modelling of the three diseases and the controls.

BASELINE HAZARD SPECIFICATIONS IN JOINT MODELS OF LONGITUDINAL AND SURVIVAL DATA

E. Lázaro, C. Armero, D. Alvares and M. Rué

The simplest structure of a joint model for longitudinal and survival data considers a linear mixed effects model for the trajectory of the time-dependent predictor and a Cox regression model for the relevant time-to-event. The frequentist paradigm has provided little prominence to the baseline hazard function. In the Bayesian paradigm, however, this function has to be specified.

We analysed the impact of different parametric and non-parametric proposals for the baseline hazard function in the Bayesian framework. We considered a Weibull distribution as a parametric choice, and piecewise constant and B-splines basis functions as non-parametric. Within the non-parametric, we evaluated different prior scenarios that impose different smooth shape conditions.

The proposals were discussed in a real study devoted to assess the relationship between the risk of death or be discharged alive in intensive care units and a longitudinal severity index. Model selection was based on the DIC and the WAIC.

BAYESIAN SURVIVAL MODELS FOR ASSESSING VIRULENCE CHANGES IN FOODBORNE PATHOGENS

E. Lázaro, C. Armero, D. Alvares, M. Sanz-Puig, D. Rodrigo and A. Martínez

Survival times are common outcomes in virulence assays which are usually performed as a part of pathogenicity studies. We implemented a Bayesian Cox model with the aim of assessing virulence changes in a foodborne pathogen as a consequence of different frequencies of application of a validated preservation treatment.

The model was estimated under several baseline hazard functions (h0(t)) that account for the natural course of the infection and are closely connected with the estimation of survival profiles. We specified h0(t) considering a Weibull distribution as a parametrical choice and piecewise exponential and B-splines basis functions as non-parametric options. Different smoothing restrictions were assessed in non-parametric proposals by setting different prior scenarios.

The subsequent posterior distributions were approximated using MCMC methods through the JAGS software. Model selection was evaluated in terms of the Deviance Information Criteria (DIC) and the Log Pseudo-Marginal Likelihood (LPML).

SPATIO-TEMPORAL MODEL FOR THE GENETIC RESISTANCE TO ASH-DIEBACK

F. Muñoz and A. Dowkiw

Ash-dieback is an invasive fungal disease of ash trees characterised by leaf loss and crown dieback in infected trees. First detected in Poland in the early 1990s, it rapidly progressed throughout Europe causing large numbers of deaths and threatening today the survival of the species.

We modelled the evolution of the disease from 2010 to 2014 within a field experiment located in northern France, with 777 trees from 23 progenies originating from three french provenances. The objective was to assess the genetic (co)variation associated with two different symptoms: Crown Dieback (CD) and Collar Lesion (CL), since this has direct implications for breeding and for the management of the disease.

Due to a high proportion of non-symptomatic observations of CL we used a mixture Binomial-Gamma for modelling the probability of infection and its conditional severity separately at the data level. For CD and both components of CL, a latent gaussian field accounted for some specific relevant covariates, the temporal global trend, the genetic effects and a spatio-temporal structure that represented the regionalized relative exposure to pathogen agent.

DEALING WITH MCMC AND INLA APPROACHES IN GAUSSIAN STATE-SPACE MODELS FOR DYNAMIC POPULATIONS

J. Pavani, C. Armero and D. Conesa

The knowledge of the size of the evolution of a given population as well as its growth rate is an important element to plan relevant decisions. Nonetheless, this is not an easy task because during the process of data collection the population size can suffer modifications due to births, deaths and other movements. One of the most common approaches in dynamic population estimation is state-space models. Such modeling is based on two different Markovian processes. The first one describes the underlying (unobserved) population dynamics whilst the second one connects the observation to the state population process. In this work, we focus on the implementation of Gaussian state-space models for dynamic populations through methods based on MCMC and INLA approaches in three different studies, two of them related to wild animal species and the third devoted to estimate the size of the Spanish and the Valencian Community population.

MODELLING JUVENILE SURVIVAL IN COMMON GUILLEMOTS (*Uria aalge*): BAYESIAN CORMACK-JOLLY-SEBER MODELS WITH AGE EFFECTS

B. Sarzo, C. Armero, D. Conesa, J. Hentati-Sundberg and O. Olsson

Juvenile survival, defined as survival from fledging (year zero) to maturity (from 5 years old), is an important life-history parameter in seabirds. Indeed, reliable data during first two years of life are rare because of the long period of unobservability after fledging. One of the most common statistical methods in Ecology that allows to jointly estimate recapture and survival probabilities are the Cormack-Jolly-Seber models. They are formulated in terms of state-space models and they can also be interpreted as Hidden Markov models. In this work we model either annual survival and resighting probabilities in relation to the age of the individuals. We present two models which differ in the number of age classes established. The results show that around 50 % of the birds survive annually during their first year of life, and this probability increases as they become older. The posterior mean of the resighting probability at age one is similar in both models and the lowest one, which shows the lack of detectability at this age.

BAYESIAN HIERARCHICAL MODELLING OF THE OLIVE QUICK DECLINE SYNDROME IN SOUTH-EASTERN ITALY

A. Vicent, J. Martínez-Minaya, A. López-Quílez and D. Conesa

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 plant disease epidemiology. The complexity of these models makes the inferential and predictive processes challenging to perform. Bayesian statistics represents a good alternative, because it is based on the premise that both information and uncertainty can be expressed in terms of probability distributions. Despite the advantages of 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) and integrated nested Laplace approximation (INLA). Here, we present different spatio-temporal analyses using INLA for the geographical spread of the olive quick decline syndrome, a lethal plant disease caused by the bacterium Xylella fastidiosa in south-eastern Italy.