

IV Jornadas Científicas de Estudiantes de la SEB

(Sociedad Española de Biometría)

Albacete, 5 y 6 de septiembre de 2019

Escuela Politécnica Superior de Albacete
(Edificio Infante Don Juan Manuel)



IV Jornadas Científicas de Estudiantes de la SEB

Agrupación Politécnica de Albacete

Campus de Albacete de la Universidad de Castilla-La Mancha
Albacete, España
5 y 6 de septiembre de 2019
<http://www.biometricsociety.net/iv-jseb/>

Welcome!

Welcome to IV JJSEB19! This event is already the fourth JJSEB congress. After successful congress in Valencia, Barcelona and Bilbao, JJSEB19 is held at the Albacete campus of the University of Castilla-La Mancha (UCLM). Edition after edition, the JJSEB has become in a meeting in which the role of young researchers in scientific spaces is promoted strongly.

Organizing this kind of event has become in one of the main activities of the Sociedad Española de Biometría (SEB), which trusts in its young members the organization of it. JJSEB promotes scientific dissemination at early stages of the scientific career and also offers students a relaxed atmosphere to present their work. Thanks to the economic support of the SEB and the BIOSTATNET network, JJSEB19 is also able to offer several stipends for scholars to help economically.

We would like to thank the Local Committe at the UCLM for the excellent local organization. We would also like to thank Virgilio Gómez-Rubio, the instructor of the specific training course as well as the Super-Women of the round table, María Durbán, Irantzu Barrio, Anabel Forte and Marta Bofill for contributing in the creation of a great program. Finally, a big thank to all of you for coming, for your support, and for helping us to make JJSEB19 a success. We wish us all a great conference! Enjoy JJSEB19 and Albacete!

The Organizing Committee.

Scientific and Local Committee

Scientific Committee

- Amaia Iparragirre (UPV/EHU)
- Blanca Sarzo Carles (UV)
- Elena Lázaro Hervás (IVIA)
- Francisco Palmí Perales (UCLM)
- Guillermo Villacampa (VHIO)
- Irene García-Camacha (UCLM)
- Joaquín Martínez Minaya (UV)
- Leire Cidores Martínez (AZTI/BCAM)
- Lore Zumeta Olaskoaga (BCAM)
- Sergio Pozuelo Campos (UCLM)
- Yovaninna Alarcón Soto (UPC)

Local Committee

- Virgilio Gómez Rubio (UCLM)
- Francisco Parreño Torres (UCLM)
- José Luis Giménez Espinosa (UCLM)
- María Teresa Alonso Martínez (UCLM)
- Francisco Palmí Perales (UCLM)

Contents

Welcome!	iii
Training Course	1
Speakers	3
Development of statistical methodologies to work with ELISpot assays (<i>Yovaninna Alarcón Soto, Klaus Langohr and Guadalupe Gómez Melis</i>)	4
Is arsenic and cadmium exposure causally associated with renal damage? A Mendelian Randomization study (<i>Maria Grau-Perez, Jose D. Bermudez Edo, Jose L. Gomez-Ariza, Zoraida Soriano-Gil, J.Antonio Casasnovas, Josep Redon and Maria Tellez-Plaza</i>)	5
Modelos multinivel aplicados a la comparación de características y resultados de pacientes con diagnóstico de cáncer colorrectal según su vía de detección (<i>Moreno, S., Comas, M., Aguirre, U. and Crespo, C.</i>)	6
A statistical model for analysing the effects of treatments on multivariate time series of microbiota data (<i>I. Creus-Martí, A. Moya and F.J. Santonja</i>)	7
Big problems, integrated solutions (<i>Blanca Sarzo, Ruth King, David Conesa and Jonas Hentati-Sundberg</i>)	9
Mixed models for exploring sardine fishing in the Mediterranean Sea (<i>Gabriel Calvo, Rubén Amorós, Carmen Armero, Maria Grazia Pennino and Luigi Spezia</i>)	10
From data-limited to data-rich: the evolution of sardine in the Bay of Biscay (<i>Leire Cidores, Leire Ibainarriaga, Sonia Sanchez, Andrés Uriarte and Dae-Jin Lee</i>)	10
Variable selection for genome wide association analysis in plant breeding (<i>Iker Oyanguren Monferrer and Anabel Forte</i>)	11
Prognostic factors in the survival of women with breast cancer through a Bayesian competitive risk model (<i>Sandra Ramírez, Carmen Armero and Virgilio Gómez Rubio</i>)	12
Evaluación basada en la modificación de la historia natural de la efectividad de un programa poblacional de atención a la insuficiencia cardíaca (<i>Ania Gorostiza, Arantzazu Arrospide, Igor Larrañaga, Aitziber Barandiarán, Javier Mar and David Hernandez</i>)	13
Statistical Modeling in sports injury prevention (<i>Lore Zumeta-Olaskoaga and Dae-Jin Lee</i>)	14

Pensar antes de actuar: introducción al diseño óptimo de experimentos (<i>Sergio Pozuelo Campos and Irene García-Camacha Gutiérrez</i>)	15
Three-level hierarchical modelling of longitudinal data: Application to the study of worldwide obesity prevalence (<i>Diana Marcela Pérez- Valencia and María Xosé Rodríguez-Álvarez</i>)	16
INLAMSM R-package: Multivariate spatial models using R-INLA. (<i>Fran- cisco Palmí Perales, Virgilio Gómez Rubio and M. A. Martínez Beneito</i>)	17
CompARE: a web app to study Composite Endpoints (<i>Marta Bofill Roig, Jordi Cortés Martínez and Guadalupe Gómez Melis</i>)	18
Fitting Bayesian Dirichlet regression using the integrated nested Laplace approximation (<i>Joaquín Martínez-Minaya, Finn Lindgren, Antonio López-Quílez, Daniel Simpson and David Conesa</i>)	19
Spatial conditional overdispersion models: Modelling mothers' postnatal screening period (<i>Mabel Morales Otero and Vicente Núñez-Antón</i>) . .	20
Design of an alternative delimiting strategy for the <i>Xylella fastidiosa</i> infec- tion in Alicante: a development of a sequential adaptative algorithm for optimising survey and sampling. (<i>E. Lázaro, M. Sesé, A. López-Quílez, D. Conesa, A. Ferrer, V. Dalmau and A. Vicent</i>)	21
Assessing landscape fire hazard by multitemporal automatic classification of Landsat time series using Google Earth Engine in West-Central Spain (<i>N. Quinteros, O. Viedma, I.R. Urbieto and J.M Moreno</i>)	22
Evaluación de la linealidad en inmunoensayos (<i>Laura Santulario Verdú and Susana Pérez Álvarez</i>)	23
Benefits and drawbacks of using stratification factors in clinical trials. When and why should we use it? (<i>Guillermo Villacampa and Rodrigo Dienstmann</i>)	24
Robustness of design for accelerated failure time models with Type I censor- ing (<i>Irene García-Camacha Gutiérrez, María Jesús Rivas-López and Raúl Martín-Martín</i>)	24
Efficiency Analysis for the Misspecification Problem of Response Probability Distribution in Optimal Experimental Design (<i>Sergio Pozuelo Campos, Mariano Amo-Salas and Víctor Casero-Alonso</i>)	25
The smart potential of European regions (<i>Leticia Belmonte Cortés, Víctor Raúl López Ruiz and José Luis Alfaro Navarro</i>)	26
Aplicación de la Teoría de Respuesta al ítem para armonizar diferentes escalas de depresión en el proyecto ATHLOS (<i>Laia Egea-Cortés, Ivet Bayés-Marín, Albert Sánchez-Niubò and David Hernandez</i>)	27
Caracterización del dimorfismo sexual a nivel craneal mediante análisis discriminantes (<i>L.J. Sánchez-Martínez, L.A. González-De la rosa and A. Sánchez-Jiménez</i>)	28
From high-dimensional to functional data: application in SNPs data (<i>Harold Antonio Hernández Roig, Rosa Elvira Lillo Rodríguez and María del Carmen Aguilera Morillo</i>)	29
Overview and new proposals in the development of prediction models with complex survey data (<i>Amaia Iparragirre, Irantzu Barrio and Inmaculada Arostegui</i>)	30

Estudio comparativo de métodos de estimación de los parámetros del modelo de regresión logística con datos de encuestas con diseño complejo (<i>Ane Villanueva, Amaia Iparragirre and Irantzu Barrio</i>)	31
Targeted lipidomic profiling reveals a lipid mediator signature specific of acute-on-chronic liver failure (<i>Arantxa Urdangarin-Iztueta, Cristina López-Vicario, Antonio Checa, José Alcaraz-Quiles, Paolo Caraceni, Alex Amorós, Marco Pavesi, Ferran Aguilar, Richard Moreau, Craig E. Wheelock, Vicente Arroyo, Joan Clària, Núria Planell and David Gómez-Cabrero</i>)	32
Early-life urban exposome and cognitive and psychomotor function in children in 5 European birth cohorts: applying an exposome-wide association study approach (<i>Lucia Alonso, Claire Philippat, Gunn Marit Aasvang, Mark Mon-Williams, Leda Chatzi, Martine Vrijheid, Mark Nieuwenhuijsen, Jordi Sunyer and Mònica Guxens</i>)	33
Improving the transparency of randomised trial manuscripts: results from a randomised trial to enhance author adherence to CONSORT guidelines (<i>David Blanco, Sara Schroter, Adrian Aldcroft, David Moher, Isabelle Boutron, Jamie J Kirkham and Erik Cobo</i>)	37
Evaluation of the prognosis of HPV negative but p16 positive oropharyngeal cancer patients in an international collaborative study (EPIC-OPC) (<i>Sorina Sferle, Sara Tous, Laia Alemany and Guadalupe Gómez Melis</i>)	38
Bayesian inference on microbiota compositional data time-series (<i>J. López and F.J. Santonja</i>)	40
Mediation analysis and Network Mendelian Randomization: methods for the estimation of the causal effect (<i>Coscia C, Molina-Montes E, Muriel A, Pérez T and Malats N</i>)	41
RAAPS: A Shiny App for Risk Assessment around Pollution Sources (<i>Virgilio Gómez Rubio and José Luis Gutiérrez Espinosa</i>)	42
Spatial modelling of <i>Xylella fastidiosa</i> distribution: two case studies in the Mediterranean Basin (<i>M. Cendoya, J. Martínez-Minaya, D. Conesa, A. López-Quílez and A. Vicent</i>)	43
A procedure for Satellite image calibration using UAV imagery for agriculture applications (<i>Amal Chakhar, David Hernandez-Lopez and Miguel A. Moreno</i>)	44
Application of laser scanner systems in precision viticulture (<i>Ana del-Campo-Sánchez, Miguel Moreno, Rocío Ballesteros and David Hernandez</i>) . .	45
Water stress and yield assessment for vineyard orchard by using multispectral and RGB very high resolution aerial images (<i>P. López-García, Moreno, M.A., J.F. Ortega and Ballesteros, R.</i>)	47
Time schedule	49
Location	55
List of participants	57
Index of authors	61

Training Course

Title: Introduction to spatial mapping with R

Instructor: Virgilio Gómez-Rubio (Universidad de Castilla-La Mancha)

Abstract:

This short course will focus on recent developments for spatial mapping and data analysis in R. Currently, R provides a myriad of R packages to handle spatial data, which includes data input/output, visualization and spatial data analysis. The course will provide an introduction to several packages for handling spatial data and making (interactive) maps. Next, the analysis of lattice and geostatistical data will be discussed using relevant examples. All course materials are available from https://github.com/becarioprecario/curso_IVJJSEB.

References:

- Bivand et al. (2013). Applied Spatial Data Analysis with R. Springer, New York.
Lovelace et al. (2019). Geocomputation with R. CRC Press, Boca Raton (FL).

Speakers

Session 1: Medical studies I
September 5th, 09:30 to 10:30

Chair: Lore Zumeta Olaskoaga

Sept 5th
09:30

**Development of statistical methodologies to work with
ELISpot assays**

Yovaninna Alarcón Soto¹, Klaus Langohr¹ and Guadalupe Gómez Melis¹

¹Universitat Politècnica de Catalunya, Barcelona, Spain

The main goal of this work is to provide statistical methodologies to analyze Enzyme-Linked Immunosorbent Spot (ELISpot) assays data. The ELISpot assay is an immunoassay that measures the frequency of cytokine-secreting cells at the single-cell level. It is carried out in a 96-well plate, and an automated ELISpot reader is used for analysis. The assay is therefore easy to perform and allows rapid analysis of a large number of samples. While ELISpot assays allows one to directly visualize and count extremely low frequencies of cytokine secreting T cells amongst millions of non-secreting bystander cells, the interpretation of ELISpot data can become ambiguous when (a) spot numbers in antigen-containing wells are low, (b) spot counts in negative control wells are elevated, and particularly (c) when both of the above occur simultaneously. Thus, the primary task, before any statistical analysis, must be the optimization of the basic assay parameters. This includes low background signal in the negative control wells, and the maximal number of antigen-induced spots in test wells, i.e., the signal to noise ratio is maximized (Dittrich and Lehmann, 2012). Our main variables of interest are spot count and spot size. Spot count corresponds to cytokine-producing cell numbers. Due to diffusion properties, a true spot has a densely coloured center which fades toward the edges; the size or color intensity of the spots is determined by the amount of cytokine released. Spot size refers to relative amounts of cytokine produced per cell.

Although ELISpot assays are widely used in the context of the HIV study, the relationship between spot size or spot count and other variables has not been studied until now. For this reason we are interested in studying if there is a relationship between spot size or spot count, and other different clinical parameters, using different statistical techniques, such as balanced/unbalanced one-way/two way ANOVA (Sahai and Ojeda, 2004) or Generalized Linear Models (GLM). Moreover, as the immunospot reader creates an Excel sheet for every ELISpot assay, we have built a routine in R to read and sort the data and we are planning to create an R package for the statistical analysis of ELISpot assays data.

Keywords: ANOVA, ELISpot assays, Generalized Linear Models, HIV studies.

References

Dittrich, M. and Lehmann P. V. (2012). Statistical analysis of elispot assays. *Hand- book of ELISPOT: Methods and Protocols*, 173–183.

Sahai, H. and Ojeda, M. M. (2004). Analysis of Variance for Random Models, Volume 2: Unbalanced Data: Theory, Methods, Applications, and Data Analysis (Vol. 2). *Springer Science and Business Media*.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*, 88(421), 9–25.

Is arsenic and cadmium exposure causally associated with renal damage? A Mendelian Randomization study

Sept 5th
09:45

Maria Grau-Perez¹, Jose D. Bermudez Edo², Jose L. Gomez-Ariza³, Zoraida Soriano-Gil⁴, J.Antonio Casasnovas⁴, Josep Redon¹ and Maria Tellez-Plaza⁵

¹Area of Cardiometabolic and Renal Risk, Institute for Biomedical Research INCLIVA, Valencia, Spain

²Department of Statistics and Operational Research, University of Valencia, Valencia, Spain

³University of Huelva, Huelva, Spain

⁴Fundación Instituto de Investigación Sanitaria de Aragón (IIS Aragón), Zaragoza, Spain

⁵Department of Chronic Diseases Epidemiology, National Center for Epidemiology, National Institutes for Health Carlos III, Madrid, Spain

Introduction: Traditional epidemiologic studies are limited to evaluate correlational associations due to confounding and other biases linked to observational data. Mendelian randomization (MR) is the use of genetic variation to assess the causal effect of modifiable exposures on health outcomes using observational data. Using a MR approach, we investigated the causal relationship of arsenic (As) and cadmium (Cd) exposure with renal damage among 1323 participants of the Aragon Workers Health Study (AWHS) from Spain.

Methods: Exposure to As and Cd was evaluated by its concentrations in urine. Renal damage was assessed as the urine albumin levels. We first conducted a genome-wide association analysis to identify single nucleotide polymorphisms (SNPs) associated with increased levels of urine metals in an independent database. Identified SNPs were further genotyped in the AWHS for this analysis. We validated 3 and 2 SNPs associated with genetically elevated urine As and Cd levels, respectively. We evaluated the correlational association of As and Cd urine levels with urinary albumin using traditional linear regression models and the causal association using a 2-stage least squares MR method.

Results: In traditional adjusted linear regression models, the geometric mean ratio (95% confidence interval) of urinary albumin by an interquartile range increase in As and Cd levels was 1.03 (0.97, 1.10) and 1.06 (1.01, 1.11), respectively. The F-statistic of the SNPs for As and Cd was 14.8 and 8.7, respectively, suggesting that the instrumental variables for Cd might be weak. The MR analysis suggested a significant causal association of As and Cd exposure with increased urinary albumin. The GMR (95% CI) of urine albumin per metal-increasing allele was 1.58 (1.15, 2.16) and 1.59 (1.18, 2.14) for As and Cd, respectively.

Conclusion: Mendelian randomization is an interesting tool to disentangle causal associations where randomized controlled trials are not feasible. More research is needed to confirm these findings.

References

- Davey Smith G, Hemani G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet.*, 23:R89-98.
- Burgess, S. and S.G. Thompson. (2015). Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimation. *Chapman and Hall/CRC.*
-

Sept 5th
10:00

Modelos multínivel aplicados a la comparación de características y resultados de pacientes con diagnóstico de cáncer colorrectal según su vía de detección

Moreno, S.¹, Comas, M.^{1,2}, Aguirre, U.^{2,3} and Crespo, C.⁴

¹Institut Hospital del Mar d'Investigacions Mèdiques - IMIM (Barcelona)

²Red de Investigación en Servicios de Salud en Enfermedades Crónicas – REDISSEC

³Unidad de Investigación, Hospital de Galdakao-Usansolo (Galdakao-Usansolo)

⁴Departament d'Estadística – Universitat de Barcelona (Barcelona)

Preámbulo: El cáncer colorrectal se encuentra actualmente entre los cánceres más comunes tanto en mujeres como en hombres. Aunque se han aplicado muchos enfoques científicos en este campo, dos aspectos significativos requieren un mayor enfoque: el desarrollo de reglas de predicción clínica para predecir eventos adversos después del tratamiento quirúrgico de pacientes con cáncer colorrectal y la utilización de los servicios de salud, donde múltiples factores permanecen sin cuantificar.

Material: En una cohorte multicéntrica prospectiva de pacientes intervenidos por cáncer colorrectal con 5 años de seguimiento, se compararán las características y los resultados clínicos, así como la calidad de vida (mediante la utilización de cuestionarios genéricos y específicos) y el uso de los servicios de salud según su modo de detección.

Métodos: Se ha dividido la muestra en 4 grupos de estudio según su vía de detección: clínica o sintomática, cáncer de intervalo, de riesgo alto y de riesgo medio. Entre los diferentes análisis, inicialmente se han realizado modelos bivariantes para identificar los posibles factores de riesgo que indiquen cuales son las diferencias entre los diferentes grupos de detección. de supervivencia mediante el uso de frailty models y se aplicarán modelos mixtos lineales generalizados para resultados continuos (calidad de vida, uso de servicios de salud) y resultados dicotómicos (recidivas). En los análisis realizados, la variabilidad entre los hospitales se tomará en cuenta a través de modelos jerárquicos. Finalmente se han obtenido los mejores modelos mediante el procedimiento del AIC.

Resultados: Los diferentes grupos presentan diferencias en cuanto a rango de edad, nivel educativo e índice de Charlson ponderado, localización del tumor y

estancia hospitalaria entre otros. El alcoholismo y el tabaquismo se presentan como factores de riesgo para la supervivencia a 5 años así como para el uso de servicios sanitarios y la calidad de vida de los pacientes en el mismo periodo.

Conclusiones: Los individuos con detección clínica y cáncer de intervalo presentan diferentes características respecto a los otros grupos, así como peores resultados clínicos, menores complicaciones y menor calidad de vida durante el tiempo de estudio. Además, la variabilidad entre los hospitales tampoco presenta grandes diferencias a lo largo del tiempo de estudio.

Keywords: cáncer colorrectal, modelos jerárquicos, servicios sanitarios.

References

- Quintana, J. M., González, N., Anton Ladislao, et al.** Colorectal cancer health-services research study protocol: the CCR-CARESS observational prospective cohort project. *BioMed Central Cancer*, 16 (435). 10.1186/s12885-016-2475a.y Unidad de Investigación, Hospital Galdakao-Usansolo. Galdakao, Bizkaia. (2016)
- Baré, M., Mora, L., Torà, N., et al.** CCR-CARESS score for predicting operative mortality in patients with colorectal cancer. *British Journal of Surgery*. 10.1002/bjs.10956 Corporació Sanitària Parc Taulí, Parc Taulí. Sabadell, Spain. (2018)
- Filippi, S.** Hierarchical Models – Statistical Methods. University of Oxford. Oxford, United Kingdom. (2015)

A statistical model for analysing the effects of treatments on multivariate time series of microbiota data

Sept 5th
10:15

I. Creus-Martí^{1,4}, A. Moya^{1,2,3} and F.J. Santonja⁴

¹Institut de Biologia de Sistemes (I2Sysbio). Universitat de València-CSIC

²Fundació per al Foment de la Investigació Sanitària i Biomèdica de la Comunitat Valenciana (FISABIO)

³CIBER en Epidemiología i Salut Pública (CIBERESP)

⁴Departament d'Estadística i Investigació Operativa. Universitat de València

There is an increasing interest on understanding the microbiota dynamics and predicting future microbiota behaviours which has motivated the development of dynamical system models. In particular, being able to predict with a model allows us to analyse the response of the microbiota to perturbations. The data available is the relative abundance of the families of bacteria of the microbiota of VIH patients. In this work, we analyse longitudinal microbiota data taking into account the compositional feature of the microbiota data. We assume that the data is distributed using the Dirichlet distribution and we present an autoregressive model that is able to predict the population dynamics. To make a long story short, we present a model that

enables quantifying the effects of treatments on microbiota time-series from VIH patients.

Keywords: time-series, autoregressive model, microbiota.

References

- Aitchison, J.** (1986). The Statistical Analysis of Compositional Data *Chapman & Hall*, London.
- Campbell G, Mosimann E** (1987). Multivariate methods for proportional shape. *American Statistical Association Proceedings of the Section on Statistical Graphics*, pp 10–17.
- Zheng, T., Chen, R.** (2017). Dirichlet ARMA models for compositional time series. *Journal of Multivariate Analysis*, Elsevier, 88(421), 158, pp.31-46.
-

Session 2: Statistical ecology September 5th, 10:30 to 11:30

Chair: Joaquín Martínez Minaya

Big problems, integrated solutions

Sept 5th
10:30

Blanca Sarzo¹, Ruth King², David Conesa¹ and Jonas Hentati-Sundberg³

¹Department of Statistics and O.R., University of Valencia, Spain

²School of Mathematics, University of Edinburgh, UK

³Department of Aquatic Resources, Swedish University of Agricultural Sciences, Sweden

In Ecology, due to the monitoring effort and the inaccessibility of breeding areas, it is often the case that a fraction of the population is monitored. These monitored areas are assumed to be a random sample of the whole population. However, this assumption is often violated. As a consequence, parameter estimates obtained from the monitored areas may be biased (Sanz et al., 2016). This problem is called partial monitoring.

We consider a particular case related to a large capture-recapture-recovery database obtained from the largest colony of Common guillemot (*Uria aalge*) of the Baltic Sea. We model annual survival and resighting probabilities in relation to the age of the individuals through integrated models (Besbeas et al., 2002). In particular, we establish four age categories: 1 (one year old individuals), 2 (two years old individuals), 3 (three years old individuals) and 4 (individuals from four to ten years old).

In this colony, the partial monitoring is present at breeding ledges, where mainly 3 and 4-10 individuals are resighted. As a result, survival probabilities for those age classes are underestimated while recapture probabilities are overestimated. In this study we compare the results obtained from the standard Cormack-Jolly-Seber model and the suitably parameterised integrated model that incorporates recoveries from all individuals in the colony (and not only those at the monitored sites). The results obtained show how the use of the integrated models solve this partial monitoring problem in this colony, although it can be easily extended to other populations.

References

- Besbeas, P., Freeman, S. N., Morgan, B. J. T. and Catchpole, E. A. (2002). Integrating Mark-Recapture-Recovery and Census Data to Estimate Animal Abundance and Demographic Parameters. *Biometrics*, 58: 540-547.
- Sanz-Aguilar, A., J.M. Igual, D. Oro, M. Genovart and Tavecchia, G. (2016). Estimating recruitment and survival in partially monitored populations. *Journal of Applied Ecology*, 53(1): 73-82.

Sept 5th
10:45

Mixed models for exploring sardine fishing in the Mediterranean Sea

Gabriel Calvo¹, Rubén Amorós², Carmen Armero¹, Maria Grazia Pennino³ and Luigi Spezia⁴

¹Universitat de València, Spain

²University of Edinburgh, UK

³Instituto Español de Oceanografía, Vigo, Spain

⁴Biomathematics and Statistics Scotland, Aberdeen, UK

The European pilchard (*Sardina pilchardus*) is one of the most exploited small pelagic fish species. It plays an important role in the transfer of energy from lower to higher trophic level organisms but some worrying changes have recently been observed on its fishing. Experts assure that the abundance of this fish has decreased. In fact, in 2017 the International Council for the Exploration of the Sea advised to stop fishing them for the next fifteen years everywhere.

In this work, we explore the temporal evolution of sardine fishing in the Mediterranean Sea from 1970 to 2014 by means of a few Bayesian longitudinal models which include three types of random variation, i.e., random effects (in the intercept and slope), serial correlation (an autoregressive term) and measurement error. Model assessment is performed through exploration of the conditional and marginal residuals.

Keywords: Fishery estimation, longitudinal models, random variation, serial correlation.

Sept 5th
11:00

From data-limited to data-rich: the evolution of sardine in the Bay of Biscay

Leire Cidores¹, Leire Ibainarriaga¹, Sonia Sanchez¹, Andrés Uriarte¹ and Dae-Jin Lee²

¹AZTI, Sustainable Fisheries Management

²BCAM, Applied Statistics

In the last years sardine (*Sardina pilchardus*) in the Bay of Biscay has evolved from being considered a data-limited stock to being classified by ICES (International Council for the Exploration of the Sea) in Category 2. This means the stock has analytical assessments and forecasts, but they are only treated qualitatively in relative terms. This raises additional difficulties, such as the calculation of reference points (i.e. maximum sustainable yield) and the definition of harvest control rules in relative terms. In this study we conducted management strategy evaluation for sardine in the Bay of Biscay to compare the performance of absolute and relative harvest control rules. All the analyses were done using FLBEIA, which is an R package for the development of bio-economic impact assessments of fisheries management strategies. The population dynamics were based on the latest ICES stock assessment and included uncertainty in the historical period and the biological parameters. Age structured and aggregated abundance indices from different surveys and age

structured observations from the catches were simulated and a stock assessment catch at age statistical model was explicitly incorporated into the simulation. The results will allow to initiate the discussion of the best management strategy for this stock.

Keywords: simulation study, population dynamics, uncertainty estimation, stock assessment models, management strategy evaluation.

Variable selection for genome wide association analysis in plant breeding

Sept 5th
11:15

Iker Oyanguren Monferrer¹ and Anabel Forte¹

¹University of Valencia

In modern plant breeding comparative experiments, there is often a necessity to accurately quantify the underlying genetic structure of industry driven phenotypic traits. With the availability of low cost high-throughput genotyping technology, this quantification usually involves the determination of quantitative trait loci (QTL) through a high dimensional genome wide association analysis pipeline. Within this pipeline, there may be several methods to reduce the dimension of the genetic marker set to contain the most important markers for further scrutiny. In this work we discuss a two steps genome wide association analysis which makes use of a Bayesian variable selection strategy to determine markers significantly linked to a supposed QTL. The method is illustrated with a large Australian wheat panel.

Keywords: QTL, Bayesian model selection, SUPER method.

References

- Bayarri, M. J., Berger, J. O., Forte, A., and García-Donato, G.** (2012). Criteria for Bayesian model choice with application to variable selection. *The Annals of Statistics*, 40, 1550 – 1577.
- Norman, A., et al.** (2017). Increased genomic prediction accuracy in wheat breeding using a large Australian panel. *Theoretical and applied genetics*, 130, 2543 – 2555.
- Wang, Q., Tian, F., Pan, Y., Buckler, E. S., and Zhang, Z.** (2014). A SUPER powerful method for genome wide association study. *PloS one*, 9(9).
-

Sept 5th
15:30

Session 3: Survival analysis and Optimum experimental design

September 5th, 15:30 to 16:45

Chair: Yovaninna Alarcón Soto

Prognostic factors in the survival of women with breast cancer through a Bayesian competitive risk model

Sandra Ramírez¹, Carmen Armero² and Virgilio Gómez Rubio³

¹Universitat Politècnica de València, España

²Universitat de València, España

³Universidad de Castilla-La Mancha, España

The World Health Organization deals with breast cancer is the most frequent cancer among women, impacting 2.1 million women each year, and also causes the greatest number of cancer deaths among women. Breast cancer rates are higher among women in more developed regions, rates are increasing in nearly every region globally (WHO, 2018) . The data of this study comes from a follow-up of 288 months to 1215 patients diagnosed with breast cancer in the region of Biobio in Chile, where 333 died of breast cancer, 121 due to another cause and 761 were still alive at the time of finalization of the study. The objective of this study is to identify those clinical variables (stage of the disease and age at the time of diagnosis, age at the time of menarche, and menopause, age at first birth, if there is a family history of cancer, etc), associated with the prognosis of the women survival with breast cancer. For this, Bayesian models of Cox's and Cause-specific risks (Klein et al, 2014.,Schmorr et al, 2013) are used that consider the clinical variables of the study as covariates.

Keywords: Survival analysis; Measuring time scale; Censored observations.

References

- World Health Organization** (2018). Health topics. *World Health Organization*.
- Schmoor C, Schumacher M, Finke J and Beyersmann J** (2013). Competing risks and multistate models *Clin Cancer*.
- Klein, P., Houwelingen, H., Ibrahim, J., Scheike, T.** (2014). Handbook of Survival Analysis. *Handbooks of Modern Statistical Methods*.

Evaluación basada en la modificación de la historia natural de la efectividad de un programa poblacional de atención a la insuficiencia cardíaca

Sept 5th
15:45

Ania Gorostiza^{1,2}, Arantzazu Arrospide^{1,2,3}, Igor Larrañaga^{1,3}, Aitziber Barandiarán⁴, Javier Mar⁵ and David Hernandez^{1,2,3}

¹Unidad de Investigación AP-OSIs, Organización Sanitaria Alto Deba, Arrasate-Mondragón, Gipuzkoa.

²Instituto de Investigación Sanitaria Biodonostia, Donostia-San Sebastián.

³Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC), Bilbao.

⁴Organización Sanitaria Goierri-Alto Urola, Zumarraga, Gipuzkoa.

⁵Equipo de Atención Primaria de Arrasate-Mondragón, Organización Sanitaria Alto Deba, Arrasate-Mondragón, Gipuzkoa.

El objetivo de este trabajo fue evaluar la efectividad en tres Organizaciones Sanitarias Integradas (OSIs) de un programa integrado de atención a la insuficiencia cardíaca (IC) desarrollado en el País Vasco.

La historia natural de los pacientes con IC se caracteriza por un proceso dinámico con periodos de compensación y descompensación en el que es habitual que los ingresos hospitalarios sean recurrentes y cada vez más frecuentes. Es por ello que las intervenciones dirigidas a la IC han sido comúnmente evaluadas mediante los ingresos hospitalarios de los pacientes.

Para la evaluación del programa se llevó a cabo un estudio observacional retrospectivo basado en las bases de datos administrativas y clínicas de Osakidetza entre los años 2014 y 2018. Se midió el efecto de la intensidad de la intervención sobre el tiempo hasta ingreso por IC, ajustando por otras variables de interés. Para ello se utilizó el modelo de procesos de conteo de Andersen-Gill, una extensión del modelo de riesgos proporcionales de Cox para analizar no solo los primeros eventos de cada paciente, sino también las recurrencias. Estos modelos permiten la inclusión de variables tiempo dependientes en el análisis, lo cual nos posibilitó medir la intensidad de la intervención antes de cada hospitalización.

Se analizó un total de 6.768 pacientes. La edad (hazard ratio (HR)=1,016; IC95% 1,011-1,022), el índice de Charlson (HR=1,067; IC95% 1,047-1,087) y el número de hospitalizaciones previas (HR=1,632; IC95% 1,557-1,712) fueron factores de riesgo de reingreso. La intensidad de la intervención se asoció a una reducción de las hospitalizaciones. También se observaron diferencias entre OSIs, aunque el análisis de la interacción entre la OSI y la intensidad de la intervención mostró que estas diferencias desaparecen a medida que la intervención se intensifica.

Se concluyó que la efectividad de un programa educativo y de monitorización de atención a la IC depende de su nivel de implementación. Además, la evaluación de las intervenciones para la IC basada en los reingresos debe tener en cuenta toda la historia natural de la IC, para lo cual resultan de utilidad las técnicas de análisis de la supervivencia para eventos repetidos.

Keywords: insuficiencia cardíaca; programa integrado; análisis de la supervivencia; eventos repetidos; modelos Andersen-Gill.

References

Andersen PK, Gill RD. Cox's Regression Model for Counting Processes: A Large Sample Study. *Ann Stat.* 1982;10(4):1100-1120. doi:10.1214/aos/1176345976

Oyanguren J, Latorre García PM, Torcal Laguna J, et al. Effectiveness and Factors Determining the Success of Management Programs for Patients With Heart Failure: A Systematic Review and Meta-analysis. *Rev Espanola Cardiol Engl Ed.* 2016;69(10):900-914. doi:10.1016/j.rec.2016.05.012

Sept 5th
16:00

Statistical Modeling in sports injury prevention

Lore Zumeta-Olaskoaga¹ and Dae-Jin Lee¹

¹BCAM - Basque Center for Applied Mathematics

Sports biostatistics is an emergent field in the Sports Science area. Initially proposed by Casals and Finch (2018), the role of the sports biostatistician brings together statistics, epidemiology and sports medicine. Sports science is not exempt from statistical applications and here one of the principal interests relies on the prevention of sports injuries. Statistics can help in gaining insights and understanding why sports injuries occur.

In this talk, a number of important issues regarding the complex nature of sports injury data will be presented. Some statistical aspects when quantifying the injury risk such as measures of injury epidemiology will be shown. Survival analysis will be introduced as a tool to appropriately handle this type of data, identify the main risk factors and model injury occurrence. Also, real-world examples will be used to illustrate these concepts. We will end up looking over the methodological challenges that sports biostatistics currently faces.

Keywords: sports injury, sports biostatistics, survival analysis.

References

Casals M, Finch C. (2018). Sports Biostatistician: a critical member of all sports science and medicine teams for injury. *British Journal of Sports Medicine*, (52) 1457-1461.

Windt J, Ardern C. L, Gabbett T. J, et al. (2018). Getting the most out of intensive longitudinal data: a methodological review of workload–injury studies. *BMJ Open*. 8(10):e022626.

Pensar antes de actuar: introducción al diseño óptimo de experimentos

Sept 5th
16:15

Sergio Pozuelo Campos¹ and Irene García-Camacha Gutiérrez²

¹Departamento de Matemáticas. ETSII Ciudad Real. Universidad de Castilla-La Mancha.

²Departamento de Matemáticas. Escuela de Arquitectura de Toledo. Universidad de Castilla-La Mancha

En la actualidad existe un interés creciente en el estudio del diseño de experimentos, ya que es una parte fundamental del método científico. El análisis de los datos será informativo sólo si los datos por sí mismos lo son (1). Las ciencias experimentales y la ingeniería son campos del conocimiento especialmente interesados en obtener modelos que expliquen adecuadamente los fenómenos objeto de estudio. La obtención de estimadores precisos de los parámetros del modelo es, entre otras, una propiedad deseable para obtener la mejor calidad de la inferencia estadística. Por este motivo, la estrategia de recogida de datos se convierte en un punto crucial para el buen desarrollo del estudio donde factores económicos y limitaciones prácticas entran en juego. El principal objetivo del Diseño Óptimo de Experimentos (DOE) consiste en determinar dónde tomar las observaciones y con qué frecuencia para optimizar algún aspecto del modelo de manera eficiente. En este trabajo se presenta una introducción general a esta teoría, poniendo en valor algunos de los modelos abordados por el grupo Optimum Experimental Design (2) desde esta perspectiva. Algunos de ellos son utilizados para la calibración de instrumental utilizado en radioterapia, para explicar el efecto de una droga en la muerte de células tumorales, para detectar el fenómeno de hormesis o para análisis de supervivencia. Finalmente, se proporcionan diferentes ejemplos prácticos en los que se comprueba cómo un estudio y una planificación previa mejora notablemente la precisión de la inferencia estadística, aprovechando la máxima información recogida de cada unidad experimental realizada.

Keywords: Diseño óptimo de experimentos, inferencia estadística

References

C. Rodríguez Torreblanca, I. M. Ortiz Rodríguez (1999). Diseño óptimo de experimentos para modelos de regresión. *Editorial: Universidad de Almería, España.*

Optimum Experimental Design Group. <http://areaestadistica.uclm.es/oed/>

Session 4: Biostatnet
September 5th, 17:00 to 18:15
Chair: Irene García Camacha

**Three-level hierarchical modelling of longitudinal data:
Application to the study of worldwide obesity prevalence**

Sept 5th
17:00

Diana Marcela Pérez-Valencia¹ and María Xosé Rodríguez-Álvarez^{1,2}

¹BCAM - Basque Center for Applied Mathematics

²IKERBASQUE, Basque Foundation for Science, Spain

This work considers hierarchical or multilevel curve modelling for longitudinal data. Motivated by the need for analyzing the evolution of the difference in the prevalence of obesity between males and females, we propose a three-level nested hierarchical model. In particular, our interest focuses on modelling and estimating the evolution of the difference in the prevalence of obesity from 1975 to 2016, and we have information for 187 countries nested into 10 regions, and regions into the world population (Gender Statistics, The World Bank, <https://datacatalog.worldbank.org/dataset/gender-statistics>). Since important decisions may need to be taken at each level of the hierarchy (countries, regions or worldwide), estimation must be as accurate as possible. Thus, a model which can handle appropriately with hierarchically nested effects is needed. Hierarchical semiparametric and nonparametric regression methods can be found in the literature. Examples include the use of smoothing splines, penalized splines or functional principal component (e.g., Durbán et al., 2005; Fitzmaurice et al., 2011; Brumback and Rice, 1998; Li et al., 2015). In this work, we propose an extension of the spline-based two-level nested hierarchical approach described in Brumback and Rice (1998) and Durbán et al. (2005), among others. We propose the use of penalized splines (P-splines) jointly with second- order differences (Eilers and Marx, 1996) and its representation as a mixed-effect model (Currie and Durban, 2002). This approach implies that each function in the hierarchy is composed by a first order polynomial (modelled as fixed or random, depending on the level) plus a nonlinear effect. Estimation is based on the SOP method presented in Rodríguez-Álvarez et al. (2019). We compare our results with the ones obtained using the methods implemented in the R-package mgcv (Wood, 2011) and using regression with functional data through the R-package refund (Goldsmith et al., 2018).

Keywords: P-splines, hierarchical model, longitudinal data, prevalence of obesity.

References

Brumback, B. A., and Rice, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association*, 93(443), 961-976.

Currie, I. D., and Durban, M. (2002). Flexible smoothing with P-splines: a unified approach. *Statistical Modelling*, 2(4), 333-349.

- Durbán, M., Harezlak, J., Wand, M. P., and Carroll, R. J.** (2005). Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine*, 24(8), 1153–1167.
- Eilers, P. H., and Marx, B. D.** (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11, 89-102.
- Fitzmaurice, G., Laird, N., and Ware, J.** (2011). Applied Longitudinal Analysis, volume Second Edition of Wiley Series in Probability and Statistics *Wiley*.
- Goldsmith, J., et al.** (2018). refund: Regression with Functional Data. R package version 0.1-17. <https://CRAN.R-project.org/package=refund>
- Li, H., et al.** (2015). Methods to assess an exercise intervention trial based on 3-level functional data. *Biostatistics*, 16(4), 754-771.
- Rodríguez- Álvarez, M.X., Durban, M., Lee, D.-J., and Eilers, P.H.C.** (2019). On the estimation of variance parameters in non-standard generalised linear mixed models: Application to penalised smoothing. *Statistics and Computing*, 29, 483-500.
- Wood, S.N.** (2017). Generalized Additive Models: An Introduction with R (2nd edition) *Chapman and Hall/CRC*.
-

INLAMSM R-package: Multivariate spatial models using R-INLA.

Sept 5th
17:15

Francisco Palmí Perales¹, Virgilio Gómez Rubio¹ and M. A. Martínez Beneito²

¹Departamento de Matemáticas, Escuela de ingenieros industriales de Albacete,
Universidad de Castilla-La Mancha

²Área de Desigualdades en Salud, FISABIO

Literature about multivariate spatial models have increased in the last years due to the fact that classical univariate methods are not able to consider the real cohabitation of different individuals or procedures (for instance, different diseases in a disease mapping context). Bearing in mind that multivariate procedures lead to high computation time procedures, different ideas have been developed to speed up fitting multivariate spatial models. On the one hand, some authors have re-parametrized the classical generalizations of the univariate models in order to avoid high computational-time operations within the framework of MCMC algorithms (Botella-Rocamora et. al. 2015). On the other hand, some authors prefer approximate methods, such as the INLA method (Rue et. al. 2009) which is commonly used in order to avoid the high computation time of the MCMC models. This idea is part of the motivation of this work whose main goal is to implement, using R-INLA benefits, multivariate areal models which were proposed for been computed using MCMC algorithms and then re-parametrized in order to be more computationally efficient. Moreover,

comparisons with other existent methodologies have been carried and all the methods implemented have been gathered in an R-package named INLA Multivariate Spatial Models: **INLAMSM**.

Keywords: INLAMSM, Bayesian statistics, R-INLA, Spatial models.

References

- Botella-Rocamora, P., Martinez-Beneito, M.A., Banerjee, S.** (2015). A Unifying Modeling Framework for Highly Multivariate Disease Mapping. *Statistics in Medicine*, , 45, 1548-1559.
- Rue, H., Martino, S., Chopin, N.** (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2), 319-392.

Sept 5th
17:30

CompARE: a web app to study Composite Endpoints

Marta Bofill Roig¹, Jordi Cortés Martínez¹ and Guadalupe Gómez Melis¹

¹Universitat Politècnica de Catalunya, Barcelona, Spain

Context: Composite endpoints are frequently chosen as primary endpoint in randomized controlled trials. Composite endpoints are common in many areas, such as cardiovascular and oncology, where Major Adverse Cardiovascular Events (MACE) and Progression Free Survival (PFS) are the usual primary endpoints. The choice of the primary endpoint, together with the corresponding needed sample size, is critical for the trial success, but deciding which components should be included in the composite endpoint is not straightforward. We have developed an Asymptotic Relative Efficiency (ARE) measure as an aid for this decision. ARE depends on the marginal and joint distributions of the composite components.

Objective: To provide a tool for researchers in studies including several time-to-event or binary outcomes. CompARE, in a quantifiable and friendly way, allows to select the components of a primary composite endpoint and its corresponding sample size and to study the attributes of the composite treatment effect.

Methods: We present the web-app CompARE (<https://cinna.upc.edu/compare>). The user can specify the probabilities of event-occurrence; the expected treatment effect through hazard ratios or odds ratios; and the degree of association between components. As a result, CompARE provides: 1) the ARE values depending on different treatment effects and degrees of association by means of tables and plots; 2) sample size calculations under several scenarios; and 3) some treatment effect measures for the composite endpoint.

Results: We present CompARE for the ZODIAC and TACTICS-TIMI 18 trials with PFS and MACE as primary endpoint, respectively. Through these examples, we illustrate how to decide whether is more efficient to use the composite endpoint

or to remain with one of the components; to compute the sample size under different scenarios and to learn about the impact of the association between components on the composite endpoint. We use CompARE to discuss about the proportionality hazards assumption of the composite endpoint in the ZODIAC trial or the different behavior of relative and absolute effect measures in the TACTICS-TIMI 18 trial.

Conclusions: CompARE is a valuable and helpful tool for the discussion of the primary endpoint to lead the randomized controlled trials regardless whether you have binary or time-to-event components.

Keywords: Software, Clinical Trials, Survival Analysis.

References

- G. Gómez and S. Lagakos (2012). (2013). Statistical considerations when using a composite endpoint for comparing treatment groups. *Statistics in Medicine*.
- M. Bofill and G. Gómez (2018). Selection of composite binary endpoints in clinical trials. *Biometrical Journal*.

Fitting Bayesian Dirichlet regression using the integrated nested Laplace approximation

Sept 5th
17:45

Joaquín Martínez-Minaya¹, Finn Lindgren², Antonio López-Quílez¹, Daniel Simpson³ and David Conesa¹

¹Department of Statistics and OR, University of Valencia

²School of Mathematics, University of Edinburgh

³Department of Statistical Sciences, University of Toronto

Dirichlet regression models can be used to analyze a set of variables lying in a bounded interval that sum up to one exhibiting skewness and heteroscedasticity, without having to transform the data. These data which mainly consist of proportions or percentages of disjoint categories are widely known as compositional data and are common in areas such as ecology, geology, and psychology. Bayesian inference has become a popular tool to deal with complex models such as Dirichlet regression models. The integrated nested Laplace approximation (INLA) is a widely extended methodology that for a large class of models provides higher accuracy for a limited computational budget. However, the implemented R-INLA package can not deal with multivariate likelihoods, such as, in particular, the Dirichlet likelihood. In this work, we propose an expansion of the INLA method for Dirichlet regression.

Keywords: Dirichlet regression, INLA, multivariate likelihood.

References

- Aitchison, J., Egozcue, J. J. (2005). Compositional data analysis: where are we and where should we be heading? *Mathematical Geology*.

Rue, H., Martino, S., Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Sept 5th
18:00

Spatial conditional overdispersion models: Modelling mothers' postnatal screening period

Mabel Morales Otero¹ and Vicente Núñez-Antón¹

¹Departamento de Econometría y Estadística (E.A. III), Universidad del País Vasco
UPV/EHU, Bilbao

Regression models for count data often present overdispersion, a phenomenon that arises when the real variance of the data is larger than the one specified in the model. In addition, when working with georeferenced data, the spatial dependence that may exist among the different locations must be also taken into account in order to produce reliable inference processes from the estimations. In this work we review spatial conditional overdispersion models associated to the binomial distribution, such as the spatial conditional beta binomial and binomial normal models. In these models, spatial association structures are specified in the regression model for the conditional mean. This specification allows researchers to be able to quantify the intensity of the spatial association that may be present in the data. Besides, assuming that the overdispersion is partially generated by the spatial autocorrelation in the data, the model would also be able to capture this phenomenon. We analyze the number of mother who had given birth to their last child between 1999 and 2005 and who underwent a postnatal screening period in different departments in Colombia, and study the behavior of this variable as a function of some speci

c covariates available in this dataset, illustrating the usefulness of the proposed models. Additionally, we also assess the performance of such models when including a conditionally autoregressive (CAR) spatial random effect in the model specification. Models have been fitted with the use of the Markov Chain Monte Carlo (MCMC) algorithms within the context of Bayesian estimation methods.

Keywords: Overdispersion models, Spatial models, Bayesian methods.

References

- Cepeda-Cuervo, E., Córdoba, M. y Núñez-Antón, V.** (2018). Conditional overdispersed models: Application to count area data. *Statistical Methods in Medical Research*, 27(10), 2964–2988.
- Quintero-Sarmiento, A., Cepeda-Cuervo, E. y Núñez-Antón, V.** (2012). Estimating infant mortality in Colombia: Some overdispersion modelling approaches. *Journal of Applied Statistics*, 39(5), 1011-1036.
- De Oliveira, V.** (2012). Bayesian analysis of conditional autoregressive models *Annals of the Institute of Statistical Mathematics*, 64(1), 107–133.
-

Poster session
September 5th, 18:15 to 19:15

Design of an alternative delimiting strategy for the *Xylella fastidiosa* infection in Alicante: a development of a sequential adaptative algorithm for optimising survey and sampling.

Sept 5th
18:15

E. Lázaro¹, M. Sesé², A. López-Quílez², D. Conesa², A. Ferrer³, V. Dalmau³ and A. Vicent¹

¹Centre de Protecció Vegetal i Biotecnologia, Institut Valencià d'Investigacions Agràries, Moncada

²Departament d'Estadística e Investigació Operativa, Universitat de València, Burjassot

³Servei de Sanitat Vegetal, Conselleria d'Agricultura, Medi Ambient, Canvi Climàtic i Desenvolupament Rural, Silla

Xylella fastidiosa (*Xf*) is a regulated quarantine plant pathogen in the European Union (EU). The current legal provisions specify the implementation of delimiting surveys in infested areas to demarcate the geographic extent and implement eradication or containment measures (EU, 2015). North-eastern Alicante Province, Spain, is one of the areas infested by *Xf* in the EU, and so it is currently demarcated and subjected to a delimiting survey activity. Based on the 2018 official surveillance program, in which approximately 100,000 has. were surveyed and 11,000 samples were taken and analysed, this work aimed to improve the efficiency of the current delimiting survey strategy.

A generic delimiting strategy was designed based on a three-phase adaptive approach in which the surveyed area (epidemiological unit) size and sampling intensity were tailored according to the previous phase information (Peyrard et al., 2013). An algorithm was implemented to optimise the number of epidemiological units to be surveyed and the sample size by simulating different random sampling scenarios from the reference data. Strategy evaluation was performed by comparing the delimitation efficacy and incidence estimates between the proposed and the current strategy. Incidence was modelled by means of a Bayesian spatial hierarchical model in which climatic and spatial factors were evaluated using INLA approximation (Lindgren and Rue, 2015). The proposed delimiting strategies achieved to delimit similar extension disease (similar efficacy) with a lower number of samples (better efficiency) than current.

Keywords: spatial sampling, adaptative sampling , simulations-optimization.

References

EU (2015). Commission implementing decision (EU) 2015/789 of 18 May 2015 as regards measures to prevent the introduction into and the spread within the union of *Xylella fastidiosa*. *Oficial Journal of the European Union*, L125: 3653.

Peyrard, N., Sabbadin, R., Spring, D., Brook, B., and Mac Nally, R. (2013).
Model-based adaptive spatial sampling for occurrence map construction. *Statistics and Computing*, 23(1), 29-42.

Lindgren, F., and Rue, H. (2015). Bayesian spatial modelling with R-INLA.
Journal of Statistical Software, 63, 19.

Assessing landscape fire hazard by multitemporal automatic classification of Landsat time series using Google Earth Engine in West-Central Spain

Sept 5th
18:15

N. Quinteros¹, O. Viedma¹, I.R Urbieta¹ and J.M Moreno¹

¹University of Castilla –La Mancha

Annual Land Use and Land Cover (LULC) maps are needed to identify the interaction between landscape changes and wildland fires . Objectives: In this work, we determined fire hazard changes in a representative Mediterranean landscape, through the classification of annual LULC types and fire perimeters, using a dense Landsat Time Series (LTS) during the 1984-2017 period, and MODIS images. Methods: We implemented a semi-automatic process in the Google Earth Engine (GEE) platform to generate annual imagery free of clouds, cloud shadows, and gaps. We compared LandTrendr (LT) and FormaTrend (FT) algorithms that are widely used in LTS analysis, to extract the pixel tendencies and, consequently, assess LULC changes and disturbances such as forest fires. These algorithms allowed us to generate the following change metrics: type, magnitude, direction and duration of change, as well as the pre-change spectral values. Results and conclusions: Our results showed that the FT algorithm was better than the LT algorithm at detecting low-severity changes caused by fires. Likewise, the use of the change metrics' type, magnitude, and direction of change increased the accuracy of the LULC maps by 4% relative to the ones obtained using only spectral and topographic variables. The most significant hazardous LULC change processes observed were: deforestation and degradation (mainly by fires), encroachment (i.e., invasion by shrublands) due to agriculture abandonment and forest fires, and hazardous densification (from open forests and agroforestry areas). Although the total burned area has decreased significantly since 1985, the landscape fire hazard has increased since the second half of the twentieth century. Therefore, it is necessary to implement fire management plans focused on the sustainable use of shrublands and conifer forests; this is so because the stability in these hazardous vegetation types is translated into increasing fuel loads, and thus an elevated landscape fire hazard.

Keywords: FormaTrend; Landsat imagery; land cover changes; LandTrendr; landscape fire hazard; time series analysis; random forest.

Evaluación de la linealidad en inmunoensayos

Sept 5th
18:15

Laura Santulario Verdú¹ and Susana Pérez Álvarez²

¹Universitat Politècnica de Catalunya y Universitat de Barcelona

²Biokit, Dpto. I+D

La validación de los nuevos inmunoensayos que se desarrollan consiste en evaluar si ciertas propiedades de éstos se satisfacen. Una de ellas, y que sólo aplica a los métodos cuantitativos, es la de la **linealidad** de la exactitud del método. La **linealidad** se refiere a la capacidad que tiene el método de reportar resultados que son directamente proporcionales a la concentración de analito en la muestra, es decir, $y = kx$.

Esto hace pensar que el modelo de regresión lineal estimado mediante el método de mínimos cuadrados sea una buena estimación de los datos. Sin embargo, debido a que los datos procedentes de inmunoensayos, por lo general, presentan heterocedasticidad y autocorrelación, se ha decidido analizar si determinados métodos mejoran la estimación lineal (en el sentido de que la estimación presente menor varianza) con respecto al de los mínimos cuadrados. Los métodos propuestos han sido: (1) regresión ponderada, (2) transformación estabilizadora de la varianza y (3) los estimadores de Theil - Sen y Siegel.

La evaluación de las diferentes técnicas estadísticas se ha llevado a cabo mediante la simulación de datos según 4 patrones de la varianza ampliamente observados en datos procedentes de inmunoensayos. Estos patrones son: (1) varianza constante, (2) varianza lineal, (3) coeficiente de variación constante y (4) varianza con forma de gancho. La comparación de estos métodos estadísticos también ha sido evaluada sobre datos reales para verificar que se obtenían resultados similares.

Finalmente, se ha visto que de todos los métodos estudiados el estimador de Theil - Sen es el que presenta menor varianza, sin embargo no hay una gran diferencia con la estimación obtenida mediante el método de mínimos cuadrados, por lo que con esta estimación no se puede garantizar la existencia de mejores resultados en la evaluación de la **linealidad**.

Igualmente, se ha elaborado un protocolo que sirve de guía a los investigadores que han de realizar estos estudios de validación.

Keywords: Heterocedasticidad, regresión lineal, inmunoensayos, rango de linealidad, curva maestra.

References

Referencia1 (1988). Transformation and Weighting in Regression.

Referencia2 (2009). Practical Statistics for the Analytical Scientist.

Referencia3 (1993). A quantitative measure of nonlinearity Clinical Chemistry.

Sept 5th
18:15

Benefits and drawbacks of using stratification factors in clinical trials. When and why should we use it?

Guillermo Villacampa¹ and Rodrigo Dienstmann¹

¹Vall d'Hebron Institute of Oncology (VHIO), Barcelona

The inclusion of stratification factors is increasingly common in clinical trial design, especially in phase III studies. This is explained by two main reasons: i) to prevent imbalance in randomization across known prognostic factors; and ii) to provide a more accurate estimation of treatment effect.

Several techniques exist to control the randomization process according the stratification factors: block randomization and dynamic allocation are the most commonly used. However, some concerns arise when the number of strata increase and, consequently, the risk of imbalance and the probability of empty strata become a real problem. In time-to-event analysis the stratified Cox model (an extension of Cox model) is used to estimate the treatment effect. On one hand, this methodology provides a more accurate point estimation if stratification factors are, in fact, prognostic factors. On the other hand, the estimate standard deviation becomes bigger as the number of strata increase. Under this scenario, several open questions arise in the context of study design. What is the cost of over-stratification? How many prognostic factors to be included as stratification variables? How predictive factors should be treated in study design?

The goal of the current study is to analyse all these questions via simulation under different scenarios. Evaluation of type I/II errors will be studied in detail for each scenario to estimate their consequences for the final treatment effect estimation.

Keywords: Stratification factor, Clinical trial, Randomization, Over-stratification

Sept 5th
18:15

Robustness of design for accelerated failure time models with Type I censoring

Irene García-Camacha Gutiérrez¹, María Jesús Rivas-López² and Raúl Martín-Martín¹

¹Departamento de Matemáticas. Escuela de Arquitectura de Toledo. Universidad de Castilla-La Mancha

²Department of Statistics. University of Salamanca.

Accelerated Failure Time (AFT) models are fairly commonly used in the field of manufacturing, but they are more and more frequent for modeling clinical trial data. These models are defined through the survival function of the time-to-event variable, T. This work deals with construction of robust designs for AFT models with the possibility that the Acceleration Factor (AF) is misspecified when the variance of T is known. In particular, we allow the “true” AF to vary over a neighbourhood of possible functions, $AF(x, \theta) = \exp(\theta'x + g_n(x))$, for some unknown perturbation function g_n . As in Huber (1975), we consider $g_n(x)$ lies in a contaminating space F. Thus, we cannot assess the efficiency of a design through the covariance matrix of the Maximum Likelihood Estimator (MLE). In this case, the estimate is subject

to both “bias error” due to the inadequacy of the model, as well as “variance error” due to sampling. We obtained the asymptotic Mean Squared Error matrix (MSE) of parameter estimates for right censored observations. In order to the asymptotic treatment we will assume that g_n functions are of the order $o(1/\sqrt{n})$. Finally, a typical distribution in AFT, Log-Logistic distribution, was considered to apply the above results.

Keywords: Optimal Experimental Design, Survival Analysis, Robustness, Accelerated Failure Time Models.

References

Huber, J. I. (1975). Robustness and designs. In: J.N. Srivastava, Ed., A Survey of Statistical Design and Linear Models. North Holland, Amsterdam, 287-303.

Efficiency Analysis for the Misspecification Problem of Response Probability Distribution in Optimal Experimental Design

Sept 5th
18:15

Sergio Pozuelo Campos¹, Mariano Amo-Salas¹ and Víctor Casero-Alonso¹

¹Departamento de Matemáticas. ETSII Ciudad Real. Universidad de Castilla-La Mancha.

It is a common assumption in the context of Optimal Experimental Design that the response variable follows a homoscedastic normal distribution. There are, however, other studies that assume different probability distributions based on prior experience or additional information. The main goal of this study is to look at the effect, in terms of efficiency, of misspecification in the probability distribution on optimal design. From the elemental information matrix, which includes information on the probability distribution of the response variable, a generalized Fisher information matrix is obtained. Relevant theoretical results were obtained, for different regression models, comparing heteroscedastic Poisson, gamma and normal distributions. Finally, the analysis was broadened to include a practical case which considers a 4-parameter Hill model, to explain the effect of a pharmaceutical drug on cell development.

Keywords: Elemental Information Matrix; Gamma Distribution; Poisson Distribution; D-optimization; D-efficiency.

Session 5: Multivariate analysis
September 6th, 09:30 to 10:30

Chair: Sergio Pozuelo Campos

Sept 6th
09:30

The smart potential of European regions

Leticia Belmonte Cortés¹, Víctor Raúl López Ruiz¹ and José Luis Alfaro Navarro¹

¹Universidad de Castilla — La Mancha

In an increasingly globalized world, it seems obvious that digital technology has transformed society. Modern society is immersed in a continuous process of knowledge accumulation and new developments. In turn, economic interdependence is growing between areas and less developed regions are being left behind by structural changes in the world economy. This phenomenon has positive effects such as greater development either economic growth, nevertheless at the same time it requires a certain technical level incorporated into the productive tasks and a high-performing human capital applying this new knowledge. Likewise, the development of innovations and improvements in productive activity are part of the market competitive strategy.

The European Union, aware of international influence, considers innovation as the main course of action for smart and sustainable growth. Technological advances have revealed the transcendence of knowledge as the dominant intangible asset. In fact, knowledge creates Intellectual Capital, an asset that can provide a competitive edge. Intellectual Capital, therefore, should be considered an economic investment, just as capital formation or labour.

Normally, productive performance is analysed on a national level; however, what is significant for the whole may not be for parties, and that is why a regional analysis is highly valuable for the diagnosis of economic functioning. The goal is to present a proposal for intellectual valuation at regional level and to know how its effect on productivity can be channelled. The analysis is based on the experience of 271 regions NUTS 2 in Europe by 2015. Through the perspective of intellectual capital, useful elements are provided to observe and understand the behaviour of productive structures in Europe in the face of the digitalisation process and discover singularities and inequalities between regions.

Keywords: human capital, technology, innovation, investigation, productivity.

Aplicación de la Teoría de Respuesta al ítem para armonizar diferentes escalas de depresión en el proyecto ATHLOS

Sept 6th
09:45

Laia Egea-Cortés¹, Ivet Bayés-Marín^{1,2}, Albert Sánchez-Niubò^{1,2} and David Hernandez¹

¹Research, Innovation and Teaching Unit, Parc Sanitari Sant Joan de Déu, Sant Boi de Llobregat, Spain

²Centro de investigación Biomédica en Red de Salud Mental, CIBERSAM, Madrid, Spain

La armonización de datos es una técnica que sirve para lograr o mejorar la comparabilidad de variables y medidas similares recopiladas en estudios diferentes. Armonizar variables como sexo, edad, altura, entre otras, es relativamente sencillo ya que consiste en recodificaciones de categorías o cambio de unidades. En cambio, para variables medidas a través de escalas psicométricas, como por ejemplo depresión, este proceso resulta más complejo dado que para un mismo constructo pueden existir varias escalas con características psicométricas distintas. Por ello, armonizar la variable directamente de la escala propia de cada estudio podría llevar a un sesgo de medida entre los estudios. Un proceso alternativo sería la construcción de una nueva escala con síntomas depresivos armonizados entre los diferentes estudios y la aplicación de modelos de variables latentes.

En la presentación se mostrará la aplicación de este proceso alternativo de armonización usando los datos de las escalas de depresión de 14 estudios longitudinales que se armonizaron en proyecto ATHLOS (Ageing Trajectories of Health: Longitudinal Opportunities and Synergies) con un tamaño de la muestra final de más de 200.000 individuos. En este proceso, primero se seleccionaron 12 ítems que fueron considerados los más comunes e importantes, extraídos de las siguientes escalas de depresión: EURO-D, CIDI, BDI-II, GHQ12 y CES-D versión de 8, 10 y 20 ítems. A continuación se estudió la unidimensionalidad del constructo de depresión utilizando el análisis factorial exploratorio y confirmatorio. Seguidamente, ajustando un modelo de la Teoría de Respuesta al ítem, se calcularon las puntuaciones para cada individuo de manera continua. Por último, se usó la curva ROC para encontrar el punto de corte óptimo para su dicotomización entre individuos con depresión y sin depresión.

Keywords: Armonización, Teoría de Respuesta al ítem, depresión, escalas psicométricas

References

Fortier I., Ratina P., Van den Heuve E.R., et al. Maelstrom Research guidelines for rigorous retrospective data harmonization. En: International Journal of Epidemiology 46.1 (2017), págs. 103-105

Sanchez-Niubo A., Egea-Cortés E., Olaya B., et al. Cohort Profile: The Ageing Trajectories of Health – Longitudinal Opportunities and Synergies (ATHLOS) Project. En: International Journal of Epidemiology dyz077 (2019), págs. 1-11.

Rizopoulos D. ltm: An R Package for Latent Variable Modeling and Item Response Theory Analyses. En Journal of Statistical Software 17.5 (2006), págs. 1-25.

Sept 6th
10:00

Caracterización del dimorfismo sexual a nivel craneal mediante análisis discriminantes

L.J. Sánchez-Martínez¹, L.A. González-De la rosa¹ and A. Sánchez-Jiménez¹

¹Departamento de Biodiversidad, Ecología y Evolución. Facultad de Ciencias Biológicas. Universidad Complutense de Madrid, UCM.

El dimorfismo sexual es el principal patrón de variación biológica entre individuos adultos dentro de las poblaciones humanas actuales. El componente esquelético juega un papel clave en la determinación del sexo en campos como la Antropología o las Ciencias Forenses. No obstante, dentro del conjunto del esqueleto humano, no todos los huesos presentan el mismo grado de dimorfismo, y el cráneo es considerado por detrás de la pelvis, como la estructura más dimórfica. Estudios previos han caracterizado ya algunas de las estructuras craneales que presentan mayor dimorfismo, entre ellas se encuentran los arcos superciliares, las apófisis mastoides o el hueso zigomático. El presente estudio tiene como objetivo caracterizar el dimorfismo sexual a nivel craneométrico. Los datos analizados corresponden a la base de datos de Howells que comprende un total de 2524 cráneos (1368 varones y 1156 mujeres) pertenecientes a 30 poblaciones mundiales diferentes, y de los cuales se tiene información de un total de 46 variables craneométricas cuantitativas pertenecientes tanto al neurocráneo como al esplacnocráneo y consistiendo tanto en medidas de anchura, como de altura o longitud. El análisis de los datos se llevó a cabo mediante el software estadístico R. La metodología empleada para determinar qué variables craneométricas explican mejor este dimorfismo sexual comprendió un Análisis de Componentes Principales (ACP) seguido de un Análisis Discriminante (AD) para optimizar su caracterización. Entre los principales resultados obtenidos el índice craneal, el cual refiere la forma general del cráneo, no mostró un peso relevante a la hora de caracterizar el dimorfismo sexual, siendo éste por tanto más debido al tamaño y no a la forma del cráneo. Además, el ACP indicó que las variables de anchura facial son las que mejor explican la separación de los individuos por sexos. Este fenómeno se puede contrastar y explicar mediante la bibliografía, relacionando una mayor anchura facial en los hombres con posibles efectos de selección sexual. El AD indicó que la variable más dimórfica es la anchura bicigomática. La precisión alcanzada por el modelo a la hora de clasificar los individuos correctamente por sexos fue del 92,24%. Este estudio muestra la utilidad de las variables craneométricas cuantitativas para discriminar el sexo del individuo mediante la construcción de funciones discriminantes.

Keywords: Craneometría, Dimorfismo sexual, Análisis Discriminante.

From high-dimensional to functional data: application in SNPs data

Sept 6th
10:15

Harold Antonio Hernández Roig^{1,2}, Rosa Elvira Lillo Rodríguez^{1,2} and María del Carmen Aguilera Morillo^{1,2}

¹Universidad Carlos III de Madrid

²UC3M-Santander Big Data Institute

Stringing via manifold learning is introduced, a method that allows to treat high-dimensional data as functional data. It takes advantage of the high dimensionality of the data by considering them as discretized observations of some underlying smooth stochastic process. Assuming that data is observed with a scrambled order of its components, stringing starts with a reordering step and then maps the high-dimensional vectors into functions. We extend the earlier version of stringing (Chen, Müller, and Wang, 2011), based on multidimensional scaling, by incorporating manifold learning. Our method is able to recover non-linear relationships between predictors, resulting in a better reordering of the data and a more reliable functional representation. This work reviews the stringing methodology and discusses the techniques related to it. Through simulation studies it is shown that our contribution outperforms the previous approach. We remark the usage of functional principal component analysis through conditional expectation, as a useful tool for functional representation when data is rather noisy. Illustrative examples regarding stringing and functional data representation are provided.

A real data application on graft-versus-host-disease (GVHD) (Martínez-Laperche et al., 2018) is also addressed. The study deals with patients who had an allogeneic hematopoietic stem-cell transplantation (allo-SCT) in order to treat hematologic malignancies. GVHD is considered the main complication after allo-SCT (affects 30% – 50% of patients) and there are not validated laboratory tests to predict it. Stringing via manifold learning is applied to single-nucleotide polymorphisms (SNPs) data, providing functional genotypes of ill and healthy patients.

Keywords: Functional data analysis, High-dimensional data, Functional principal components analysis, Multidimensional scaling, Manifold learning, Isometric feature mapping, Locally linear embedding; Functional genotype, SNPs data.

References

- Chen, K., Müller, H.-G., and Wang, J. (2011). Stringing High-Dimensional Data for Functional Analysis. *Journal of the American Statistical Association*, 106(493), 275-284.
- Martínez-Laperche, C., Buces, E., Aguilera-Morillo, M.C., ... Buño, I. (2018). A novel predictive approach for GVHD after allogeneic SCT based on clinical variables and cytokine gene polymorphisms. *Blood Advances*, 2(14), 1719–1737.

Sept 6th
10:30

Session 6: Prediction models September 6th, 10:30 to 11:30

Chair: Blanca Sarzo

Overview and new proposals in the development of prediction models with complex survey data

Amaia Iparragirre¹, Irantzu Barrio¹ and Inmaculada Arostegui^{1,2}

¹Departamento de Matemática Aplicada, Estadística e Investigación Operativa,
Universidad del País Vasco (UPV/EHU)

²BCAM - Basque Center for Applied Mathematics

Nowadays, complex survey data are gaining importance between researchers in different fields. The sampling weights are one of the most remarkable differences between simple random sampling data and data collected based on a complex sampling design. Each sampling weight is the number of individuals that one observation represents in the population. However, despite the growing importance of these type of data, we believe that there is still a lack of statistical methodology to develop in this field. The first goal of this work is to make a review of the literature in the context of the development of prediction models with complex survey data.

One of the most discussed topics in this context has been the importance of incorporating the sampling weights in order to obtain accurate estimations of the parameters of prediction models. Specifically, in the context of logistic regression models for dichotomous response variables, the pseudo-likelihood function has been proposed as a modified version of the likelihood function that incorporates the sampling weights. But in addition to the estimation of the model parameters, there are many aspects that should be taken into account in the development of prediction models in order to end up with a good validated model: the selection of the predictors, the functional relation between outcome and predictors, imputation of missing values and evaluation of calibration and discrimination, among others. We believe that the sampling weights should be taken into account in throughout this process, beyond the estimation of the model parameters. In particular, the goals of this work are two fold: a) we focus on the impact of the sampling weights in the imputation process of a relevant predictor variable and, b) we propose a new methodology to impute the missing values of a dichotomous covariate in the context of complex survey data.

Keywords: complex survey data, prediction models, sampling weights.

References

- Binder, D. A.** (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review/Revue Internationale de Statistique*, 279-292.
- Heeringa, S. G., West, B. T., and Berglund, P. A.** (2017). Applied survey data analysis. *Chapman and Hall/CRC*.

Estudio comparativo de métodos de estimación de los parámetros del modelo de regresión logística con datos de encuestas con diseño complejo

Sept 6th
10:45

Ane Villanueva¹, Amaia Iparragirre¹ and Irantzu Barrio¹

¹Departamento de Matemática Aplicada, Estadística e Investigación Operativa,
Universidad del País Vasco (UPV/EHU)

El modelo de regresión logística, es uno de los modelos predictivos utilizados con mayor frecuencia a la hora de analizar el efecto de diversas variables explicativas en una variable dicotómica. Cuando los datos provienen de una muestra aleatoria simple, los parámetros del modelo se estiman generalmente mediante el método de máxima verosimilitud. Sin embargo, cuando se dispone de datos provenientes de una encuesta basada en diseño complejo, cada individuo de la muestra tiene asociado su correspondiente peso muestral, que representa el número de individuos representados en la población. El propósito de este trabajo ha sido analizar el efecto que los pesos muestrales pueden tener -y el alcance del mismo- en la estimación de los parámetros del modelo de regresión logística. Para ello, se han empleado datos simulados basados en los datos de dos encuestas realizadas por el Instituto Vasco de Estadística (Eustat): ESIE (Encuesta sobre la Sociedad de la Información en las Empresas) y PRA (Encuesta de población en relación con la actividad). En ambas encuestas se ha realizado un muestreo estratificado de una etapa, estando los estratos definidos en base a una combinación de 3 variables en el primer caso, y una única variable en el segundo. El estudio de simulación consiste en la obtención de 200 muestras obtenidas a partir de un muestreo estratificado de una etapa. En cada una de estas muestras, se han estimado los parámetros del modelo de regresión logística utilizando dos alternativas: a) máxima verosimilitud, y b) máxima verosimilitud ajustada por los pesos muestrales. Los parámetros obtenidos se han comparado con los parámetros teóricos calculados en la población. Los resultados obtenidos muestran que las estimaciones obtenidas teniendo en cuenta los pesos de muestreo, se aproximan más a los parámetros teóricos. A pesar de que esto ocurre en ambos casos (ESIE y PRA), en el caso de ESIE las diferencias entre los dos métodos de estimación son mayores. Esto se debe a que el diseño del muestreo, y por tanto los estratos considerados en ESIE, son más complejos. Por lo tanto, los resultados del estudio de simulación sugieren que los pesos de muestreo han de ser considerados en el proceso de estimación de un modelo de regresión basado en datos de encuestas con diseño complejo.

Keywords: muestreo complejo, regresión logística, pesos de muestreo.

References

Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review/Revue Internationale de Statistique*, 279-292.

Lumley, T. (2011). Complex surveys: a guide to analysis using R (Vol. 565). *John Wiley & Sons*.

Lumley, T. and Scott, A. (2017). Fitting regression models to survey data. *Statistical Science*, 32(2), 265–278.

Sept 6th
11:00

Targeted lipidomic profiling reveals a lipid mediator signature specific of acute-on-chronic liver failure

Arantxa Urdangarin-Iztueta¹, Cristina López-Vicario², Antonio Checa³, José Alcaraz-Quiles², Paolo Caraceni⁴, Alex Amorós⁵, Marco Pavesi⁵, Ferran Aguilar⁵, Richard Moreau^{5,6}, Craig E. Wheelock³, Vicente Arroyo⁵, Joan Clària^{2,5,7}, Núria Planell¹ and David Gómez-Cabrero¹

¹Translational Bioinformatics Unit, NavarraBiomed, Pamplona.

²Biochemistry and Molecular Genetics Service, Hospital Clínic-IDIBAPS and CIBERehd, Barcelona.

³Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm.

⁴Department of Medical and Surgical Sciences, University of Bologna, Bologna.

⁵European Foundation for the Study of Chronic Liver Failure (EF-Clif) and Grifols Chair, Barcelona.

¹Inserm, U1149, Centre de Recherche sur l’Inflammation (CRI), UMRS1149; Université Paris Diderot-Paris 7, Laboratoire d’Excellence Inflamex, ComUE Sorbonne Paris Cité, Paris.

¹Department of Biomedical Sciences, University of Barcelona, Barcelona.

Individuals with liver failure, severe disruption of the intrahepatic arterial and venous flow, diffuse fibrosis, and portal hypertension are diagnosed with cirrhosis. Traditionally, cirrhosis has been classified as compensated and decompensated.

In patients with cirrhosis or underlying chronic liver disease, Acute-on-chronic liver failure (ACLF) defines a syndrome characterized by acute and severe hepatic abnormalities. However, in contrast to decompensated cirrhosis, ACLF has a high short-term mortality. In 40-50% of the cases it is not possible to identify the cause of ACLF and in the remaining patients, active alcoholism and relapse of chronic viral hepatitis are the most common precipitating factors.

Lipid profiles from the CANONIC study, a prospective observational study including patients with acute decompensated cirrhosis (AD) and acute chronic liver failure (ACLF) were analyzed to assess the added value in risk and prognosis of lipid profiles to the available clinical data.

Non-parametric tests, Kruskal-Wallis test and Wilcoxon test were used to assess the association of the lipids with the disease status and organ failures. Age and gender were regressed out. Two types of logistic regression models to predict 6-month mortality were computed: including only clinical variables and including both lipid profiles and clinical variables. Data was divided in two non-overlapping sets: training set including 80% of the samples and a validation set with the rest. Statistically Equivalent Signatures (SES) algorithm was used to select the “minimal” sets of features with the highest predictive power.

From 59 lipids analyzed, 15 show significant differences between healthy and AD or ACLF patients. Only two lipids show significant differences between AD and

ACLF. This tendency was seen also at follow-up. No significant added value of lipids data to the predictive models was observed.

Keywords: cirrhosis, acute decompensated cirrhosis (AD), Acute-on-chronic liver failure (ACLF), logistic regression, feature selection algorithms.

References

- Schuppan D, Afdhal NH.** Liver cirrhosis. Lancet. 2008;371(9615):838-851. doi:10.1016/S0140-6736(08)60383-9
- Hernaez R, Solà E, Moreau R, Ginès P.** Acute-on-chronic liver failure: an update. Gut. 2017;66(3):541-553. doi:10.1136/gutjnl-2016-312670
- Lagani V, Athineou G, Farcomeni A, Tsagris M, Tsamardinos I.** Feature Selection with the R Package MXM : Discovering Statistically Equivalent Feature Subsets. J Stat Softw. 2017;80(7). doi:10.18637/jss.v080.i07
- Tsamardinos I, Brown LE, Aliferis CF.** The max-min hill-climbing Bayesian network structure learning algorithm. Mach Learn. 2006;65(1):31-78. doi:10.1007/s10994-006-6889-7

Early-life urban exposome and cognitive and psychomotor function in children in 5 European birth cohorts: applying an exposome-wide association study approach

Sept 6th
11:15

Lucia Alonso^{1,2,3}, Claire Philippat⁴, Gunn Marit Aasvang⁵, Mark Mon-Williams⁶, Leda Chatzi⁷, Martine Vrijheid^{1,2,3}, Mark Nieuwenhuijsen^{1,2,3}, Jordi Sunyer^{1,2,3} and Mònica Guxens^{1,2,3,8}

¹ISGlobal, Spain

²Pompeu Fabra University

³Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP)

⁴Inserm, France

⁵Norwegian Institute of Public Health

⁶University of Leeds, UK

⁷University of South California, USA

⁸Erasmus University Medical Centre–Sophia Children’s Hospital, the Netherlands

Introduction. The exposome is a new paradigm that encompasses the totality of human lifelong environmental exposures, including the environmental urban exposures. Several regression-based statistical methods have been developed to assess exposome-health associations. One of the most used is the exposome-wide association study (ExWAS) approach, an exposure-by-exposure method accounting for multiple hypotheses testing.

Objective. We investigated the association between early-life urban exposome and cognitive and psychomotor function in children, applying an ExWAS approach.

Methods. We used data of 11,693 mother-child pairs from 5 European birth cohorts in Spain, France, Greece, UK, and Norway. Information on 30 exposures including air pollution, noise, climate, natural space, and built environment were estimated at each participant's home address during pregnancy. Verbal, non-verbal, and psychomotor function were assessed using validated tests at 3-5 years old. We applied an ExWAS approach to study the association between the exposome and each outcome. We fitted linear regression models independently for each exposure and each outcome. The statistical significance of each association was assessed on the two-sided p-value after correction for multiple testing. Exposures found statistically significant were included in a multiple linear regression model and were retained if their two-sided p-value was below 5%. Models were adjusted for various socio-economic and lifestyle characteristics.

Results. Land use mixture was related to lower verbal function (-0.018 points (95%CI -0.036, -0.001)), while higher population density was related to higher verbal function (0.017 points (95%CI 0.001, 0.033)). Higher humidity levels were associated with lower non-verbal function (-0.081 (95%CI -0.153, -0.044)). Higher distance to green spaces was associated with lower gross motor function (-0.028 (95%CI -0.056, -0.001)) while higher particulate matter exposure was related to higher gross motor function (0.040 (95%CI 0.001, 0.079)). Higher particulate matter exposure was associated with lower fine motor function (-0.045 points (95%CI -0.078, -0.012) and -0.048 points (95%CI -0.093, -0.002), respectively). None of the associations survived correction for multiple testing, thus no multiple linear regression models were performed.

Conclusion. We found that air pollution, natural space, and built environment exposures during pregnancy were associated with small changes in cognitive and psychomotor function in children. We used an ExWAS approach, one of the most used methods developed to assess exposome-health associations. However, alternative methods such as Deletion/Substitution/Addition algorithm, an iterative selection method that selects the most predictive exposures of the outcome by cross-validation, taking into account

Keywords: ExWAS, multiple testing, exposome.

MESA REDONDA: “EL PAPEL DE LAS MUJERES EN LA ESTADÍSTICA”

September 6th, 12:00 to 13:30

Marta Bofill: **Universitat Politècnica de Catalunya (UPC).** Marta Bofill Roig se graduó en Matemáticas en la Universidad de Barcelona (UB) en junio de 2014. Obtuvo el Master de Estadística e Investigación Operativa (MESIO UPC-UB), programa conjunto de la Universidad de Barcelona (UB) y la Universitat Politècnica de Catalunya-BarcelonaTECH (UPC), en junio de 2016. Desde septiembre de 2016, es estudiante de doctorado en el Programa de Doctorado en Estadística e Investigación Operativa de la UPC bajo la supervisión de la Prof. Guadalupe Gómez Melis. Su investigación se centra en el diseño de ensayos clínicos con múltiples variables respuesta.

Irantzu Barrio: **Universidad del País Vasco (UPV/EHU).** Irantzu Barrio es profesora adjunta del Departamento de Matemáticas Aplicadas y Estadística e Investigación Operativa de la Universidad del País Vasco UPV/EHU. Se licenció en Matemáticas en 2006 y obtuvo el Máster en Modelización Matemática, Estadística y Computación en 2010 en la misma universidad. Tras varios años trabajando como estadística en la empresa privada, en el año 2012 empezó a trabajar como docente en el Departamento de Matemática Aplicada, Estadística e Investigación Operativa de la UPV/EHU. En 2015 se doctoró en Matemáticas en la UPV/EHU. Su principal interés investigador se centra en el desarrollo de metodología estadística orientada a la resolución de complejidades derivadas de la investigación científica, principalmente en áreas biomédicas o experimentales. Los temas específicos de interés son el desarrollo y la validación de modelos predictivos. Además, trabaja en el soporte metodológico a investigadores tanto dentro de la facultad de ciencia y tecnología de la UPV/EHU como de unidades de investigación de centros del entorno (centros de investigación y hospitalares) y el Instituto Vasco de Estadística (EUSTAT).

Anabel Forte: **Universitat de Valencia (UV).** Anabel Forte Deltell es profesora Contratada Doctora en el departamento de Estadística e Investigación Operativa de la Universitat de València. Entre sus líneas de investigación destaca la Selección de Modelos Bayesianos, así como el estudio de modelos estadísticos para datos correlados. En estas áreas ha publicado un total de 20 artículos Peer Reviewed además de varios artículos en proceedings y un capítulo de libro gracias a su participación en 8 proyectos competitivos financiados tanto por el gobierno de España como por el de la Generalitat Valenciana. En el ámbito de la docencia y la transferencia del conocimiento, Anabel ha participado en diversos proyectos de innovación docente y de divulgación. Ha impartido cursos en distintas universidades y realizado estancias de docencia en la Universitat Jaume I de Castellón y en la Universidad do Miño en Gimarães (Portugal), esta última financiada con fondos del programa Erasmus+. Ha dirigido dos tesis doctorales además de diversos trabajos fin de grado y máster dentro del grado en Matemáticas y del máster en Bioestadística de la Universitat de València. Además, es vocal de la junta directiva de la Sociedad Española

de Biometría y Secretaría de la Sección OBayes de ISBA (International Society for Bayesian Analysis).

María Durban: Universidad Carlos III de Madrid (UC3M). María Durbán es Almeriense, estudió su Licenciatura en Matemáticas en la Universidad de Granada (1993) y decidió emprender sus estudios de postgrado en el Reino Unido donde realizó un Máster en Estadística Matemática en la Universidad de Cambridge (1995), allí es donde experimentó por primera vez lo que realmente era la Estadística Aplicada y esto marcó su carrera desde entonces. A continuación, realizó sus estudios de doctorado en la Heriot-Watt University (Edimburgo, 1999) trabajando en la estimación de tendencias espaciales en enfermedades de plantas. Durante su experiencia postdoctoral en Biomathematics and Statistics Scotland (Edimburgo, 2000) trabajó en temas relacionados con las series temporales multivariantes aplicadas a la predicción de lluvia. Los años de formación en UK fueron para ella determinantes, no solo por la formación recibida, sino por la posibilidad de conocer de cerca a importantes Estadísticas/os que marcaron su carrera y con los que a día de hoy continúa colaborando. En el año 2000 se incorpora como profesora visitante a la Universidad Carlos III de Madrid donde en la actualidad es Catedrática de Universidad. Casi desde el inicio, su área de investigación se ha centrado en los modelos de regresión no-paramétricos, y en particular en los P-splines, en su formulación como modelos mixtos y sobre todo en su aplicación al análisis de datos espacio-temporales en áreas como la Epidemiología, Agricultura y Medicina.

Session 7: Medical studies II September 6th, 15:00 to 16:00

Chair: Guillermo Villacampa

Improving the transparency of randomised trial manuscripts: results from a randomised trial to enhance author adherence to CONSORT guidelines

Sept 6th
15:00

David Blanco^{1,2}, Sara Schroter³, Adrian Aldcroft³, David Moher⁴, Isabelle Boutron⁵, Jamie J Kirkham⁶ and Erik Cobo¹

¹Statistics and Operations Research Department, Universitat Politècnica de Catalunya - Barcelona Tech, Barcelona, Spain.

²INSERM, U1153 Epidemiology and Biostatistics Sorbonne Paris Cité Research Center (CRESS), Methods of Therapeutic Evaluation of Chronic Diseases (METHODS) Team, Paris Descartes University, Sorbonne Paris Cité, Paris F-75014, France.

³The BMJ, London, UK

⁴Centre for Journalology, Ottawa Hospital Research Institute, Ottawa, Canada.

⁵Centre d'Epidémiologie Clinique, Paris Descartes University, Paris, France.

⁶Department of Biostatistics, University of Liverpool, Liverpool, UK.

Improving the transparency and accuracy of randomised controlled trials (RCT) is key to reduce waste in biomedical research. The Consolidated Standards of Reporting Trials (CONSORT) Statement (1) and other reporting guidelines help authors report research methods and findings in a way that they can be understood by a reader, replicated by a researcher, used by a doctor to make a clinical decision, and included in a systematic review (2).

In recent years, biomedical journals have taken different actions to boost the completeness of reporting of published randomised trials. One of the most popular ones consists of requiring authors to submit a populated CONSORT checklist together with their manuscript indicating page numbers corresponding to each item. However, journals usually lack further actions throughout the editorial process to ensure that the corresponding information to each item is actually reported in the manuscript. This has been hypothesized to be one of the reasons why this strategy has not yet achieved optimal results.

In an effort to take full advantage of the submission of populated CONSORT checklists, we are currently performing an RCT (3) in collaboration with BMJ Open, an open-access journal that publishes medical research. Our goal is to evaluate the impact of the following actions on the transparency of the randomised trials published in BMJ Open: 1) assessing during peer review the consistency between the CONSORT checklist submitted by authors and the information that is reported in the manuscript, and 2) writing up a peer review report providing feedback to authors on the inconsistencies found.

In this presentation, we will present the preliminary results of our RCT. Moreover, we will detail how we intervened the editorial process, how authors reacted to the intervention and what challenges we faced. Finally, we will briefly describe other interventions that could be implemented throughout the editorial process in order to improve author adherence to CONSORT and other reporting guidelines.

We hope this presentation will raise awareness on the importance of reporting research transparently and will make attendants reflect on the vital role of biomedical journals taking a step forward to ensure that this happens.

Keywords: Reporting guidelines, transparency, adherence, randomised trial.

References

- Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, et al.** CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* [Internet]. 2010 Aug [cited 2017 Jun 26];63(8):28–55. Available from: <http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L50848715>
- EQUATOR Network.** Available from: <http://www.equator-network.org/>
- Evaluating the Impact of Assessing During Peer Review the CONSORT Checklist Submitted by Authors [Internet]. 2019. Available from: <https://clinicaltrials.gov/ct2/show/NCT03751878>

Sept 6th
15:15

Evaluation of the prognosis of HPV negative but p16 positive oropharyngeal cancer patients in an international collaborative study (EPIC-OPC)

- Sorina Sferle¹, Sara Tous^{2,3}, Laia Alemany^{3,4} and Guadalupe Gómez Melis¹
¹Universitat Politècnica de Catalunya
²Cancer Epidemiology Research Program, Catalan Institute of Oncology (ICO), IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain
³Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain
⁴CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

Background: In 2007, the International Agency of Research on Cancer concluded that there was sufficient evidence for the carcinogenic role of HPV16 in oropharyngeal cancers (OPCs) [1]. However, this cancer etiology is multifactorial and tobacco and alcohol consumption also play an important role. HPV-DNA detection has proven to be insufficient to attribute causality, thus it is crucial to incorporate additional biomarkers to assess the biological and oncogenic potential of HPV, such as p16 expression [2]. In addition, it has been shown that patients with double positivity for HPV-DNA/p16 have a better overall survival compared to any other combination [3]. In spite of this, some scientific societies recommend only performing p16 determination for HPV-relatedness classification. In order to clearly describe the characteristics of HPV-negative/p16-positive patients, since most cohorts had amassed relatively few cases, an international consortium has been created.

Objectives: (1) To describe the distribution of the double HPV-DNA and p16 detection in two retrospective cohorts (Spain and Germany); (2) To identify the

determinants for the combinations of the double detection of HPV-DNA and p16 in OPC patients.

Methodology: Information on sociodemographic and clinical characteristics at diagnosis of primary OPC from patients from two different centers was recruited. Pairwise tests of independence for nominal data were calculated. In order to determine the factors associated to HPV DNA infection and p16 overexpression, multinomial logistic regression analysis was performed using both Frequentist and Bayesian approaches [4, 5].

Results: A total of 1563 patients with a primary OPC were recruited in the two cohorts between 1991 and 2018. These cases were distributed according to their combination of the HPV and p16 detection as follows: double HPV/p16-negative (n=1163), HPV-negative/p16-positive (n=67), HPV-positive/p16-negative (n=71) and double HPV/p16-positive (n=262). There were significant differences between double negative and double positive cases for all the variables; and between double positive and HPV/p16 discordant results for some variables such as recruiting centre, tumour subsite, tumour staging, smoking status, alcohol consumption, period of diagnosis. In tonsil and base of the tongue sites, the risk factors for being double positive were: young cohorts, advanced tumour stage, non-smokers, non-drinkers and diagnosed in recent years. In other oropharyngeal sites, the risk factors to be double positive were to be a non-smoker and a non-drinker patient. The adjusted OR from the Frequentist and the Bayesian, including a random effect on the intercept with center as a grouping variable, approaches were similar, being the magnitudes of the second one slightly larger.

Conclusions: We observed a clear difference in the factors associated to double negative and double positive patients by subsite. HPV-negative/p16-positive patients shared the same determinants with the double negative patients. HPV-positive/p16-negative patients group resembles more to the double negative than to the double positive group confirming that both determinations are necessary to correctly classify patients with OPC. Although we have not observed any difference between the results from the Frequentist and the Bayesian approaches, we propose to use the second one to better characterize this patients.

Keywords: Oropharyngeal Cancer (OPC), Human Papillomavirus (HPV), Odds Ratio (OR), Multinomial Logistic Regression, MCMCglmm

References

IARC Monographs on the Evaluation of Carcinogenic Risks to Humans 2007. Monograph; 89:22376.

Castellsagué X, et al. (2016).HPV Involvement in Head and Neck Cancers: Comprehensive Assessment of Biomarkers in 3680 Patients Journal of the National Cancer Institute; 108(6):djv403.

Mena M, Taberna M, et al. (2018). Double positivity for HPV-DNA/p16INK4a is the biomarker with strongest diagnostic accuracy and prognostic value for

human papillomavirus related oropharyngeal cancer patients. *Oral Oncology*, 78:137-44.

Agresti A, Kateri M. (2011). Categorical data analysis. *Springer*, Heidelberg, Berlin.

Hadfield J D, et al. (2010) MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, 33 (2):1-22.

Bayesian inference on microbiota compositional data time-series

Sept 6th
15:30

J. López¹ and F.J. Santonja¹

¹Departament d'Estadística i Investigació Operativa. Universitat de València

The human microbiota is the collective genomes of the microorganisms that live inside and on the human body. It is composed by many types of microbes as bacteria, archaea, fungi, protists and viruses. These microorganisms resides either in or on of a number of human tissues and biofluids, including the skin, mammary glands, placenta, seminal fluid, uterus, ovarian follicles, lung, saliva, oral mucosa, conjunctiva, biliary and gastrointestinal tracts. Microbiota have been found to be crucial for the health of their host. The human microbiota and their interactions with the host play an important role in basic biological processes and in the development and progression of major human diseases such as infectious diseases, liver diseases, gastrointestinal cancers, metabolic diseases, respiratory diseases, mental or psychological diseases, and autoimmune diseases.

Note that microbiota is inherently dynamics. Thus, longitudinal microbiota analysis can provide rich information of short-and long-term trends of microbial communities. In addition, a common way to express the microbiota data is using relative proportions. As a consequence, compositional data analysis is a valid approach to analyze the microbiota sequencing data.

In this work, we propose a dynamical model in order to analyze trends of microbiota timeseries with a Dirichlet response variable. We will present different modeling proposals and estimation methods.

Keywords: Microbiota, Multivariate Time Series, Dynamic Models, Dirichlet, Bayesian inference, Compositional Data.

Mediation analysis and Network Mendelian Randomization: methods for the estimation of the causal effect

Sept 6th
15:45

Coscia C¹, Molina-Montes E¹, Muriel A², Pérez T³ and Malats N¹

¹Centro Nacional Investigaciones Oncológicas (CNIO) y CIBERONC, Madrid, España

²Unidad de Bioestadística Clínica (UBC), Hospital Universitario Ramón y Cajal, Madrid,
España

³Universidad Complutense de Madrid (UCM), Madrid, España

Introduction: Causal Mediation Analysis (CMA) is a statistical method that allows investigators to consider possible mediator variables in the causal inference chain. The use of genetic variants as Instrumental Variables (IV) in causal inference studies is increasing in the last years. IVs usage in the CMA context is known as Network Mendelian Randomization (MR), but there are few studies describing the appropriate methodology. Classic CMA approaches, such as Potential Outcome Framework (POF), i.e., the counterfactual approach, and Structural Equation Models (SEM), are suitable to implement IVs.

Objective: To compare several CAM approaches including POF, SEMs and Network MR using simulated and real data, within a study on the causal mediation effect between subtypes of T2DM (new-onset NODM and long-standing LSDM) and obesity in pancreatic cancer aetiology.

Methods: Study population (real data): We considered a European multicentre case-control study, the PanGenEU study, with 2,108 PC cases and 1,540 controls. The majority had genotype data. Simulation study: We simulated the outcome variable, as well as the exposure, the mediator and the IVs. The outcome and mediator were considered both as continuous and as binary variables. The exposure was considered in all analyses as a binary variable. We first considered a classic mediation analysis approach using subtypes of T2DM as exposure, obesity as mediator, and vice versa, and pancreatic cancer as outcome using the POF and SEMs. To implement IVs within these methods, we defined specific IVs for the mediator and exposure variables. These IVs were considered as a genetic score of T2DM or obesity-related genetic variants, and included in the fitted final model. In the case of Network MR, we implemented IV methods such as the Wald Ratio (WR) and Multiple-Stage-Least-Squares (MSLS).

Results: We obtained similar results in terms of Odds Ratios (OR) and 95% Confidence Interval (CI) when comparing the classical mediation analysis approaches and their counterparts using IVs. For instance, in POF, the total causal effect of obesity as mediator in the causal pathway between LSDM and PC risk was 1.68 (95%CI: 1.37-2.06) and 1.58 (95%CI: 1.16-2.15), respectively. Regarding Network MR considering the IV methods (WR and MSLS) for binary variables, results were not consistent and need further development.

Conclusions: To obtain the correct causal estimation effect in CMA, it is important to consider the proper methodology. The implementation of IVs in the counterfactual and SEM was possible but requires further development in the context of Network Mendelian Randomization.

**Session 8: Spatial statistics and statistical methods in
agriculture**
September 6th, 16:30 to 17:45
Chair: Elena Lázaro

Sept 6th
16:30

**RAAPS: A Shiny App for Risk Assessment around
Pollution Sources**

Virgilio Gómez Rubio¹ and José Luis Gutiérrez Espinosa¹

¹Escuela de Ingenieros Industriales - Albacete
Universidad Castilla - La Mancha

La importancia de evaluar los riesgos de la polución alrededor de fuentes de emisión de gases es un importante problema espacial estudiado en epidemiología. La metodología usada de casos y controles, dan una versatilidad para el análisis en este tipo de estudios. Como ejemplo, cuando evaluamos el impacto de la polución emitida por industrias, es habitual pensar en la distancia al foco como factor de riesgo en el modelo, así, cuando la localización de los casos y controles son conocidos por el investigador, se pueden proponer un número de modelos que evalúen el impacto de dichas fuentes de polución en la distribución espacial de los casos teniendo en cuenta la distribución de los controles. Sin embargo, este tipo de análisis requieren una manipulación compleja de los datos tal como mapas, visualización espacial, estadísticas espaciales entre otros. También, RAAPS representa una oportunidad y estímulo para quienes generan datos. El personal de centros como hospitales podría trabajar directamente sus datos, optimizando el poco tiempo y recursos que disponen para tareas investigadoras.

Por este motivo, hemos desarrollado una App en el entorno Shiny llamada RAAPS, del inglés (Risk Assessment App around Pollution Sources) con la intención de ayudar a los investigadores que utilizan el método de casos y controles en el campo de la salud. La App se compone de una simple interface donde cargamos los datos, así como la localización de las fuentes de riesgo (en nuestro caso industrias), en una región de estudio fácil de crear y reconocer. De manera similar, un análisis estadístico evalúa el impacto de la localización de las industrias sobre los casos hallados de enfermedad. Esto daría lugar a una importante herramienta a hospitales para la vigilancia epidemiológica y la monitorización de riesgos. El uso de esta App se ilustra con un conjunto de datos con las localizaciones de diferentes tipos de cáncer (pulmón, estómago y riñón) en la ciudad de Alcalá de Henares (Madrid, España).

Se usaron varias industrias cerca de la ciudad para el análisis.

Spatial modelling of *Xylella fastidiosa* distribution: two case studies in the Mediterranean Basin

Sept 6th
16:45

M. Cendoya¹, J. Martínez-Minaya², D. Conesa², A. López-Quílez² and A. Vicent¹

¹Centre de Protecció Vegetal i Biotecnologia, Institut Valencià d'Investigacions Agràries (IVIA), CV-315, Km. 10.7, 46113 Moncada.

²Departament d'Estadística i Investigació Operativa, Universitat De València, Burjassot 46100, Valencia.

Xylella fastidiosa (*Xf*) is a phytopathogenic bacterium with a wide range of host plants, characterized by its genetic diversity. In this study, the effects of climatic and spatial factors on the distribution of *Xf* in two areas of the Mediterranean Basin were analyzed: Apulia (Italy) and Alicante (Spain). The pathogen was confined to the American continent and Taiwan, but in 2013 it was detected in Europe, in Apulia region, where extensive areas were affected by the olive quick decline syndrome, caused by *Xf* subsp. *pauca*. Almond leaf scorch, caused by *Xf* subsp. *multiplex*, was detected in 2017 in Alicante (Spain). Presence/absence data of *Xf* in the official surveys in both regions, Apulia and Alicante, were analyzed using a Bayesian approach through the Integrated Nested Laplace Approximation (INLA) methodology (Rue et al., 2009). Climatic covariates were obtained from the WorldClim database (Fick and Hijmans, 2017). In addition, a categorical variable was included according to Purcell's levels of disease severity based on minimum winter temperature from < 1.1°C to > 4.5°C. These thresholds were defined in North America for Pierce's disease of grapevine, caused by *Xf* subsp. *fastidiosa*, and it is not known if those thresholds can be extrapolated to other *Xf* subspecies or geographic regions. Due to the different surveillance and sampling strategies carried out in the two study areas, the available georeferenced data presented a different structure. In Apulia, data were observed at continuous locations occurring within a defined spatial domain (geostatistical data). Therefore, the spatial effect was included using the Matérn covariance function, approximated as a solution to a Stochastic Partial Differential Equation (SPDE) (Lindgren et al., 2011). In Alicante, the samples were overlapped on a grid, so they were considered lattice data and the spatial effect was included through a conditional autoregressive structure (iCAR). Results showed that in Alicante the *Xf* was detected in all four Purcell's levels, illustrating the climatic adaptability of *Xf* subsp. *multiplex*. In Apulia, with much lower winter temperature variability than Alicante, only the high risk Purcell's class was observed. In both regions, the spatial structure had a strong influence in the models, but not the climatic covariates. In conclusion, disease spread was largely defined by the spatial relationship between geographic locations.

Keywords: Hierarchical Bayesian models, INLA, *Xylella fastidiosa*.

References

- Fick, S. E. and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12): 4302–4315.

Lindgren, F., Rue, H., and Lindström (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 73(4): 423–498

Rue, H., Martino, S., and Chopin, N. (2001). Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 71:319-392.

Sept 6th
17:00

A procedure for Satellite image calibration using UAV imagery for agriculture applications

Amal Chakhar¹, David Hernandez-Lopez¹ and Miguel A. Moreno^{1,2}

¹Institute for Regional Development (IDR), University of Castilla-La Mancha, Albacete, Spain.

The launch of Sentinel 2 (A and B) satellites has boosted the development of many applications for instance describing crop status and variability. Sentinel 2 offers ground sample distance (GSD) equal to 10 m and combined constellation revisit time equal to 5 days (ESA, 2015). According to (Padró et al., 2018), the maked efforts in combining different platforms products still insufficient because there is a high risk of uncertainty that these product receive an inadequate radiometric correction that can not allow sensors to perform in coordinated way. The radiometric correction is applied to spaceborne or airborne imagery with the purpose to extract physical earth surface parameters for example reflectance quantities (Richter and Schläpfer, 2016). Based on its many features, such as the flexibility along with the high spectral, temporal and spatial resolution (Herrero-Huerta et al., 2014) imagery obtained from multispectral sensors onboard Unmanned Aerial Vehicles (UAV) is expected to complement field data. In this study, a detailed analysis and comparison of multispectral imagery provided by Sentinel 2 and UAV was performed. The Sentinel 2 and the resampled UAV products was evaluated by considering, first, the intercomparison of NIR and Red reflectance bands, and second, by the intercomparison between the NDVI of the two products at the level of the whole flight zone as well as at different crop level. When downsampling the UAV imagery to the Sentinel 2 pixel size we tested two methods: the nearest neighbor and the average of the UAV pixel values. Likewise, we aimed the elaboration of models with the purpose to understand the relationship between the two different sensor platforms in the study area, and to develop a predictive model for barley, vineyard, almond and olive. Based on multitemporal band-by-band comparison, we concluded that the average resampling method provide better correlation results and lower RMSE than the Nearest Neighbor method. The NDVI fitting models at the whole flight zone suggest that at least 67% (which corresponds to the day with the lowest R2) of the variation in the data can be explained by the linear model. Concerning the predictive models, for barley, we obtained good linear regression model with

strong correlation coefficient varying from 0.86 in 2018.01.19 to 0.921 in 2018.06.13. For almond, vineyard and olive we did not obtain a good linear regression models as the barley model, therefore, we concluded that the herbaceous provide better results than the ligneous.

Keywords: Sentinel 2, UAV, reflectance bands, NDVI, calibration, resampling, downsampling, predictive models.

References

- ESA** (2015). ESA's Optical High-Resolution Mission for GMES Operational Services.
- Herrero-Huerta, M. et al.** (2014). Vicarious radiometric calibration of a multispectral sensor from an aerial trike applied to precision agriculture. *Comput. Electron. Agric.* 108, 28–38.
- Padró, J.C., Muñoz, F.J., Ávila, L.Á., Pesquer, L., Pons, X.** (2018). Radiometric correction of Landsat-8 and Sentinel-2A scenes using drone imagery in synergy with field spectroradiometry. *Remote Sens.* 10. <https://doi.org/10.3390/rs10111687>
- Richter, R., Schläpfer, D.** (2016). ATCOR-2/3 user guide, v.9.0.2 263.

Application of laser scanner systems in precision viticulture

Sept 6th
17:15

Ana del-Campo-Sanchez¹, Miguel Moreno¹, Rocio Ballesteros¹ and David Hernandez¹

¹AgroForestry and Cartographic Precision Research Group, University of Castilla – La Mancha, Spain

Three-dimensional (3D) modelling is a growing remote sensing technique for the agroforestry sector (i.e.[1]). 3D scanning by laser scanner allows to collect most of geometries from agroforestry scenarios using only one geomatic product, termed point clouds. This massive digitization of spatial data involves a wide range of information to be process by powerful computing tools. These tools make possible to extract useful information in the decision making process for management and conservation of agricultural and forestry environments. Scientific community is making much efforts to improve these methodologies and tools as they decrease costs and increase precision and density points.

However, the main challenge of this technology is to close hull points properly to define scanned surfaces. Point clouds captured by laser scanning systems only provide isolated data of the intercepted surfaces by light pulses emitted from the sensor. Therefore, there is no spatial data between one single point and its neighbours. For that reason, defining scanned surfaces relating one point to the next one involves a high computational complexity [2], primarily in complex objects such as the ones in agroforestry systems. Thus, the current challenge consists of moving from a model

expressed as a point cloud to a conventional three-dimensional model that allows us to extract metric characteristics that provide information of interest in the process of analysis or characterization of the modelled object.

Within the framework of precision viticulture, determining the vigour of the vines that composes the vineyard permits take decisions about the inputs to allocate in the plot, determining possible differences in harvest and the probability of diseases and pest affections. Thus, determining the actual volume occupied by each vine could derive a much precise estimation of vigour of the vines.

For these purposes, this research work presents a strategy for determining the volume of individual vines from point clouds collected with a static terrestrial laser scanner (TLS) and an integrated mapping mobile system (MMS), using several algorithms implemented and adjusted for each case (such as convex hull, OctoMap and the L1-medial skeleton method [3–5]), which constitute a problem in the context of biometrics.

Results shows that the two first methods are appropriate when using TLS systems, but they are totally unable for low point cloud density. This problem is solve using the skeleton calculation. These results mean a great advance in the use of laser systems in agroforestry applications, showing the capability of these methods of extracting useful information from dense and sparse point cloud products.

Keywords: Point cloud, Mapping Mobile System, Terrestrial Laser Scanner, Agroforestry Precision.

References

- Keightley, K.E.; Bawden, G.W.** 3D volumetric modeling of grapevine biomass using Tripod LiDAR. *Computers and Electronics in Agriculture* 2010, 74, 305–312.
- Rusu, R.B.; Cousins, S.** 3D is here: Point Cloud Library (PCL). In Proceedings of the 2011 *IEEE International Conference on Robotics and Automation*; 2011; pp. 1–4.
- Papon, J.; Abramov, A.; Schoeler, M.; Wörgötter, F.** Voxel Cloud Connectivity Segmentation - Supervoxels for Point Clouds. In *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*; 2013; pp. 2027–2034.
- Hornung, A.; Wurm, K.M.; Bennewitz, M.; Stachniss, C.; Burgard, W.** OctoMap: an efficient probabilistic 3D mapping framework based on octrees. *Auton Robot* 2013, 34, 189–206.
- Huang, H.; Wu, S.; Cohen-Or, D.; Gong, M.; Zhang, H.; Li, G.; Chen, B.** L1-medial skeleton of point cloud. *ACM Trans. Graph.* 2013, 32, 1.

Water stress and yield assessment for vineyard orchard by using multispectral and RGB very high resolution aerial images

Sept 6th
17:30

P. López-García¹, Moreno, M.A.¹, J.F. Ortega¹ and Ballesteros, R.¹

¹Institute of Regional Development, Campus de Albacete

This study was developed in the framework of precision viticulture and very high resolution remote sensing by using drones. The experiment was carried out along season 2018 in Fuente-Alamo (Albacete), in one experimental plot of 7.5 ha. The studied crop was vineyard (*Vitis vinifera L.*). Six different irrigation treatments were applied in order to study plant behavior regarding water stress and yield. The treatments applied were: 1) rainfed; 2) irrigation; 3) and 4) application of irrigation water including two types of salts; 5) and 6) same salty irrigation treatments only from veraison. 1000 m³/ha were applied to all irrigation treatments. Six flights were made with a UAV with RGB and multispectral cameras on board. The acquired images were photogrammetrically processed to obtain, among other geomatic products, an orthoimage in the RGB range and the multispectral range for each flight event.

These orthoimages were the starting point for calculating some of the main vegetation indexes (VIs) traditionally used for vegetation characterization. Linear relationships between these VIs and accumulated water stress and yield were analyzed. Furthermore, relationships between different multispectral and visible bands were studied using artificial neural networks (ANNs).

The study of water potential did not show significant differences between the different treatments until veraison. There were no differences between treatments with different types of applied salts, neither there were significant differences between the treatments with respect to yield. However, differences were evident between irrigation and rainfed treatments.

Statistical analysis of multispectral VIs and linear relationship with water stress along growing cycle showed Renormalized Difference Vegetation Index (RDVI) as the index with the best adjustment ($R^2 = 0.724$, RMSE = 6.564), while for yield the studied VIs showed inadequate adjustments ($R^2 < 0.5$). In the case of RGB-based VIs, the adjustment for linear relationships and water stress showed the most accurate adjustments for Excess Red Vegetation Index (ExR) ($R^2 = 0.715$, RMSE = 3,299 MPa-days). Adjustment for linear models between yield and RGB VIs was not significant, as it did for multispectral VIs. The use of ANNs, both of water stress and yield, for the two spectral ranges showed a significant improvement of adjustments, being remarkable for ANNs and RGB bands with adjustments of $R^2 = 0.985$, RMSE = 0.313 MPa-days for stress determination and $R^2 = 0.982$, RMSE = 0.216 kg/vine for yield. Thus, networks can be considered the best alternative for the description of these highly non-linear relationships

Keywords: multispectral range, RGB range, UAVs, *Vitis vinifera L.*, water stress, yield

Time schedule

	Miércoles	Jueves	Viernes
8:30 - 9:00		Inscripción y registro	
9:00 - 9:30		Inauguración	
9:30 - 10:30		Sesión 1: Medical studies I	Sesión 5: Multivariate analysis
10:30 - 11:30		Sesión 2: Statistical ecology	Sesión 6: Prediction models
11:30 - 12:00		Coffee break	Coffee break
12:00 - 12:15		Curso R	Mesa redonda: El papel de las mujeres en la estadística
12:15 - 12:30			
12:30 - 12:45			
12:45 - 13:00			
13:00 - 13:15			
13:15 - 13:30			
13:30 - 13:45			
13:45 - 14:00			
14:00 - 14:15			
14:15 - 14:30			
14:30 - 14:45		Comida	Comida
14:45 - 15:00			
15:00 - 15:15			
15:15 - 15:30			
15:30 - 15:45			
15:45 - 16:00		Sesión 3: Survival analysis and Optimum experimental design	Sesión 7: Medical studies II
16:00 - 16:15			
16:15 - 16:30			
16:30 - 16:45			
16:45 - 17:00			
17:00 - 17:15		Sesión 4: Biostatnet	Coffee break
17:15 - 17:30			
17:30 - 17:45			
17:45 - 18:00			
18:00 - 18:15			
18:15 - 19:15		Sesión de póster	Sesión 8: Spatial statistics and statistical methods in agriculture
19:15 - 19:30			
19:30 - 19:45			
19:45 - 20:00			
20:00 - 20:15			
20:15 - 20:30		Actividad social / turística	Clausura
20:30 - 20:45			
20:45 - 21:00			
21:00 - 21:15			
21:15 - 21:30			
21:30 - 21:45			
21:45 - 22:00		Recepción informal	Cena Jornadas
22:00....			

Thursday, September 5th

08:30 Registration

09:00 Welcome

SESSION 1: MEDICAL STUDIES I

Chair: Lore Zumeta Olaskoaga

09:30 “Development of statistical methodologies to work with ELISpot assays”,
Yovaninna Alarcón Soto.

09:45 “Is arsenic and cadmium exposure causally associated with renal damage? A Mendelian Randomization study”, *Maria Grau Perez.*

10:00 “Modelos multínivel aplicados a la comparación de características y resultados de pacientes con diagnóstico de cáncer colorrectal según su vía de detección”,
Sergio Moreno.

10:15 “A statistical model for analysing the effects of treatments on multivariate time series of microbiota data”, *Irene Creus Martí.*

SESSION 2: STATISTICAL ECOLOGY

Chair: Joaquín Martínez Minaya

10:30 “Big problems, integrated solutions”, *Blanca Sarzo.*

10:45 “Mixed models for exploring sardine fishing in the Mediterranean Sea”, *Gabriel Calvo.*

11:00 “From data-limited to data-rich: the evolution of sardine in the Bay of Biscay”,
Leire Cidores.

11:15 “Variable selection for genome wide association analysis in plant breeding”,
Iker Oyanguren.

11:30 Coffee break

12:00 R Course

14:00 Lunch

SESSION 3: SURVIVAL ANALYSIS AND OPTIMUM EXPERIMENTAL DESIGN

Chair: Yovaninna Alarcón Soto

15:30 “Prognostic factors in the survival of women with breast cancer through a Bayesian competitive risk model”, *Sandra Ramírez.*

15:45 “Evaluación basada en la modificación de la historia natural de la efectividad de un programa poblacional de atención a la insuficiencia cardíaca”, *Ania Goroztiza*.

16:00 “Statistical Modeling in sports injury prevention”, *Lore Zumeta Olaskoaga*.

16:15 “Pensar antes de actuar: introducción al diseño óptimo de experimentos”, *Sergio Pozuelo Campus & Irene García Camacha*.

16:45 Coffee break

SESSION 4: BIOSTATNET

Chair: Irene García Camacha

17:00 “Three-level hierarchical modelling of longitudinal data: Application to the study of worldwide obesity prevalence”, *Diana Marcela Pérez Valencia*.

17:15 “INLAMSM R-package: Multivariate spatial models using R-INLA”, *François Palmí Perales*.

17:30 “CompARE: a web app to study composite endpoints”, *Marta Bofill Roig*.

17:45 “Fitting Bayesian Dirichlet regression using the integrated nested Laplace approximation”, *Joaquín Martínez Minaya*.

18:00 “Spatial conditional overdispersion models: Modelling mothers’ postnatal screening period”, *Mabel Morales Otero*.

POSTER SESSION

(18:15-19:15)

1. “Design of an alternative delimiting strategy for the *Xylella fastidiosa* infection in Alicante: a development of a sequential adaptive algorithm for optimising survey and sampling.”, *Elena Lázaro*.
2. “Assessing landscape fire hazard by multitemporal automatic classification of Landsat time series using Google Earth Engine in West-Central Spain”, *Natalia Quinteros*.
3. “Evaluación de la linealidad en inmunoensayos”, *Laura Santulario Verdú*.
4. “Benefits and drawbacks of using stratification factors in clinical trials. When and why should we use it?”, *Guillermo Villacampa*.
5. “Robustness of design for accelerated failure time models with Type I censoring”, *Irene García-Camacha Gutiérrez*
6. “Efficiency Analysis for the Misspecification Problem of Response Probability Distribution in Optimal Experimental Design”, *Sergio Pozuelo Campos*

19:30 Social and tourist activity

22:00 Dinner

Friday, September 6th

SESSION 5: MULTIVARIATE ANALYSIS

Chair: Sergio Pozuelo Campos

- 09:30** “The smart potential of European regions”, *Belmonte Leticia Cortés.*
- 09:45** “Aplicación de la Teoría de Respuesta al ítem para armonizar diferentes escalas de depresión en el proyecto ATHLOS”, *Laia Egea Cortés.*
- 10:00** “Caracterización del dimorfismo sexual a nivel craneal mediante análisis discriminantes”, *Luis Sánchez Martínez.*
- 10:15** “From high-dimensional to functional data: application in SNPs data.”, *Harold Hernández Roig.*

SESSION 6: PREDICTION MODELS

Chair: Blanca Sarzo

- 10:30** “Overview and new proposals in the development of prediction models with complex survey data”, *Amaia Iparraguirre.*
- 10:45** “Estudio comparativo de métodos de estimación de los parámetros del modelo de regresión logística con datos de encuestas con diseño complejo”, *Ane Villanueva.*
- 11:00** “Targeted lipidomic profiling reveals a lipid mediator signature specific of acute-on-chronic liver failure”, *Arantxa Urdangarin Iztueta.*
- 11:15** “Early-life urban exposome and cognitive and psychomotor function in children in 5 European birth cohorts: applying an exposome-wide association study approach”, *Lucía Alonso.*

11:30 Coffee break

MESA REDONDA: “EL PAPEL DE LAS MUJERES EN LA ESTADÍSTICA”

(12:00-13:30)

13:30 Lunch

SESSION 7: MEDICAL STUDIES II

Chair: Guillermo Villacampa

- 15:00** “Improving the transparency of randomised trial manuscripts: results from a randomised trial to enhance author adherence to CONSORT guidelines”, *David Blanco.*

15:15 “Evaluation of the prognosis of HPV negative but p16 positive oropharyngeal cancer patients in an international collaborative study (EPIC-OPC)”, *Sorina Sferle*.

15:30 “Bayesian inference on microbiota compositional data time-series”, *Jesua López*.

15:45 “Mediation analysis and Network Mendelian Randomization: methods for the estimation of the causal effect”, *Claudia Coscia*.

16:00 Break

SESSION 8: SPATIAL STATISTICS AND STATISTICAL METHODS IN AGRICULTURE
Chair: Elena Lázaro

16:30 “RAAPS: A Shiny App for Risk Assessment around Pollution Sources”, *José Luis Gutiérrez Espinosa*.

16:45 “Spatial modelling of *Xylella fastidiosa* distribution: two case studies in the Mediterranean Basin”, *Martina Cendoya*.

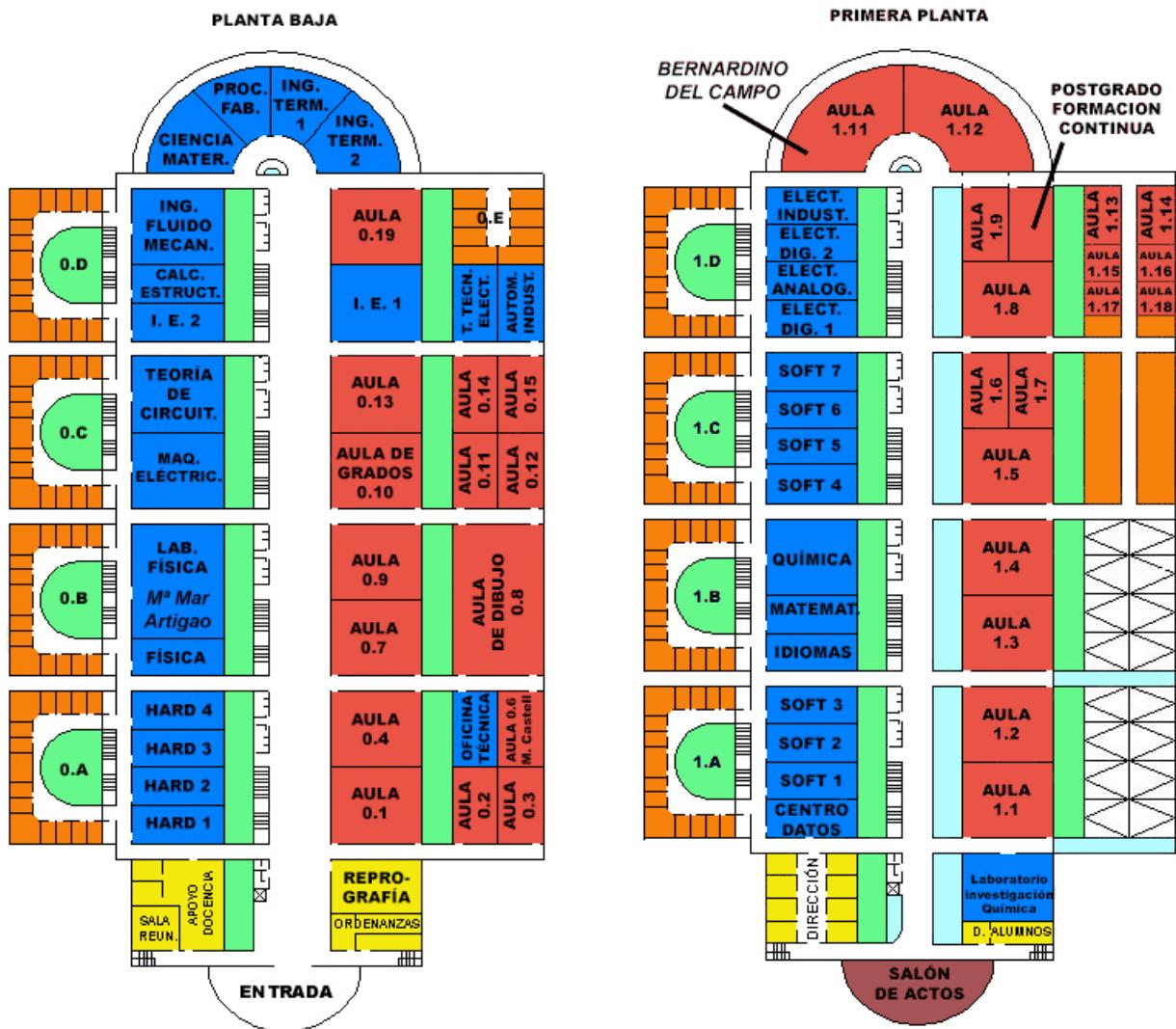
17:00 “A procedure for Satellite image calibration using UAV imagery for agriculture applications”, *Amal Chakhar*.

17:15 “Application of laser scanner systems in precision viticulture”, *Ana del Campo Sánchez*.

17:30 “Water stress and yield assessment for vineyard orchard by using multispectral and RGB very high resolution aerial images”, *Patricia López*.

Location

Most of the congress will be in the auditorium (Salón de actos). The course will be in classroom 0.1 and meals in the cafeteria.



List of participants

In alphabetical order:

Nº	Name	Institution	e-mail
1.	Àlvar Pineda Romero	Universitat Politècnica de Catalunya	alvar.pinedaromero@gmail.com
2.	Amaia Iparragirre	Universidad del País Vasco (UPV/EHU)	amaia.iparragirre@ehu.eus
3.	Amal Chakhar	Universidad de Castilla-La Mancha	amal.chakhar@uclm.es
4.	Ana del Campo Sánchez	PAFyC-Universidad de Castilla-La Mancha	ana.delcampo@uclm.es
5.	Anabel Forte Deltell	Universitat de València	anabel.forte@uv.es
6.	Ane Villanueva Etxebarria	Universidad del País Vasco (UPV/EHU)	anevillanueva1@gmail.com
7.	Àngela Sebastià Bargues	Universidad de Castilla-La Mancha	angela160794@hotmail.com
8.	Ania Gorostiza	Biodonostia	ania.gorostiza@biodonostia.org
9.	Arantxa Urdangarin Iztueta	Navarrabiomed	arantxa.urdangarin.iztueta@navarra.es
10.	Blanca Sarzo Carles	Universitat de València	blanca.sarzo@uv.es
11.	Cecilia Superchi	Universitat Politècnica de Catalunya	cecilia.superchi@upc.edu
12.	Claudia Coscia Requena	Centro Nacional Investigaciones Oncológicas (CNIO)	ccosciar@ext.cnio.es
13.	David Blanco	Universitat Politècnica de Catalunya	david.blanco.tena@upc.edu

Nº	Name	Institution	e-mail
14.	Diana Marcela Pérez Valencia	BCAM - Basque Center for Applied Mathematics	dperez@bcamath.org
15.	Elena Lázaro Hervás	IVIA - Instituto Valenciano de Investigaciones Agrarias	lazaro_ele@gva.es
16.	Francisco Palmí Perales	Universidad de Castilla - La Mancha	francisco.palmi@uclm.es
17.	Gabriel Calvo Bayarri	Universitat de València	g.calvobayarri@gmail.com
18.	Gabriela Viteri	Universidad de Castilla - La Mancha	gabriela.viteri@uclm.es
19.	Guillermo Villacampa Javierre	Vall d'Hebron Institut d'Oncologia (VHIO)	villacampa.u@gmail.com
20.	Harold Antonio Hernández Roig	Universidad Carlos III de Madrid	hahernan@est-econ.uc3m.es
21.	Idoia Eizaguirre Peral	Universidad del País Vasco (UPV/EHU)	idooia24@gmail.com
22.	Iker Oyanguren Monferrer	Universitat de València	ikeroyangurenmon@gmail.com
23.	Irantzu Barrio Beraza	Universidad del País Vasco (UPV/EHU)	irantzu.barrio@ehu.eus
24.	Irene Creus Martí	Universitat de València	icreus@alumni.uv.es
25.	Irene García-Camacha Gutiérrez	Universidad de Castilla-La Mancha	irene.garciacamacha@uclm.es
26.	Jesús López Mañez	Universitat de València	jesualoma@gmail.com
27.	Joaquín Martínez Minaya	Universitat de València	jomarminaya@gmail.com
28.	José Luis Gutiérrez Espinosa	Universidad de Castilla-La Mancha	joseluis.gutierrez@uclm.es
29.	Klaus Langohr	Universitat Politècnica de Catalunya	klaus.langohr@upc.edu

Nº	Name	Institution	e-mail
30.	Laia Egea Cortés	Universitat Politècnica de Catalunya - Fundación Sant Joan de Déu	laia.e@hotmail.com
31.	Laura Santulario Verdú	Universitat Politècnica de Catalunya	laurasantulario88@gmail.com
32.	Leire Citores Martínez	BCAM - Basque Center for Applied Mathematics- AZTI	lcitores@azti.es
33.	Leticia Belmonte-Cortés	Universidad de Castilla-La Mancha	leticia.belmonte@alu.uclm.es
34.	Lidia Herrero Huertas	Centro Nacional de Epidemiología	lidia.herrero.huertas@gmail.com
35.	Lore Zumeta Olaskoaga	BCAM - Basque Center for Applied Mathematics	lzumeta@bcamath.org
36.	Lucia Alonso García	ISGlobal	lucia.alonso.sta@gmail.com
37.	Luis Javier Sánchez Martínez	Universidad Complutense de Madrid	luisja02@ucm.es
38.	Luz Anamar González de la Rosa	Universidad Complutense de Madrid	luzagonz@ucm.es
39.	Mabel Morales Otero	Universidad del País Vasco (UPV/EHU)	mabel.morales@ehu.eus
40.	Maria Durban Reguera	Universidad Carlos III de Madrid	marialuz.durban@uc3m.es
41.	Maria Grau Pérez	Instituto de Investigación Biomedica de Valencia - INCLIVA	maria.grau.perez@gmail.com
42.	Marta Bofill Roig	Universitat Politècnica de Catalunya	marta.bofill.roig@upc.edu
43.	Martina Cendoya Martínez	Instituto Valenciano de Investigaciones Agrarias (IVIA)	cendoya_mar@externos.gva.es
44.	Natalia Quintero	Universidad de Castilla-La Mancha	natalia.quintero@uclm.es
45.	Patricia Lopez García	Universidad de Castilla-La Mancha	patricia.lopez@uclm.es

Nº	Name	Institution	e-mail
46.	Pilar Sanjuán Martín	Universidad Complutense de Madrid	pilsanju@ucm.es
47.	Queralt Miró Catalina	Universitat Internacional de Catalunya	queraltmiro@gmail.com
48.	Sandra Milena Ramirez Buelvas	Universidad Politécnica de Valencia	sanrabue@doctor.upv.es
49.	Sergio Moreno Fornés	CEEISCAT	sermorfornes13@gmail.com
50.	Sergio Pozuelo Campos	Universidad de Castilla-La Mancha	sergio.pozuelo@uclm.es
51.	Sorina Sferle	Universitat Politècnica de Catalunya	sorina.j15@gmail.com
52.	Xavier Barber	Universidad Miguel Hernández de Elche	xbarber@umh.es
53.	Yovaninna Alarcón Soto	Universitat Politècnica de Catalunya	yovaninna.alarcon@upc.edu

Index of authors

- A
Muriel, 41
- A.
Ferrer, 21
López-Quílez, 21, 43
Moya, 7
Sánchez-Jiménez, 28
Vicent, 21, 43
- Adrian
Aldcroft, 37
- Aitziber
Barandiarán, 13
- Albert
Sánchez-Niubò, 27
- Alex
Amorós, 32
- Amaia
Iparragirre, 30, 31
- Amal
Chakhar, 44
- Ana
del-Campo-Sánchez, 45
- Anabel
Forte, 11
- Andrés
Uriarte, 10
- Ane
Villanueva, 31
- Ania
Gorostiza, 13
- Antonio
Checa, 32
López-Quílez, 19
- Arantxa
Urdangarin-Iztueta, 32
- Arantzazu
Arrospide, 13
- Blanca
Sarzo, 9
- C
Coscia, 41
- C.
Crespo, 6
- Carmen
Armero, 10, 12
- Claire
Philippat, 33
- Craig E.
Wheelock, 32
- Cristina
López-Vicario, 32
- D.
Conesa, 21, 43
- Dae-Jin
Lee, 10, 14
- Daniel
Simpson, 19
- David
Blanco, 37
Conesa, 9, 19
Gómez-Cabrero, 32
Hernandez, 13, 27, 45
Hernandez-Lopez, 44
Moher, 37
- Diana Marcela
Pérez-Valencia, 16
- E
Molina-Montes, 41
- E.
Lázaro, 21
- Erik
Cobo, 37

- F.J.
 - Santonja, 7, 40
- Ferran
 - Aguilar, 32
- Finn
 - Lindgren, 19
- Francisco
 - Palmí Perales, 17
- Gabriel
 - Calvo, 10
- Guadalupe
 - Gómez Melis, 4, 18, 38
- Guillermo
 - Villacampa, 24
- Gunn Marit
 - Aasvang, 33
- Harold Antonio
 - Hernández Roig, 29
- I.
 - Creus-Martí, 7
- I.R
 - Urbleta, 22
- Igor
 - Larrañaga, 13
- Iker
 - Oyanguren Monferrer, 11
- Inmaculada
 - Arostegui, 30
- Irantzu
 - Barrio, 30, 31
- Irene
 - García-Camacha Gutiérrez, 15, 24
- Isabelle
 - Boutron, 37
- Ivet
 - Bayés-Marín, 27
- J.
 - López, 40
 - Martínez-Minaya, 43
- J.Antonio
 - Casasnovas, 5
- J.F.
 - Ortega, 47
- J.M
 -
- Moreno, 22
- Jamie J
 - Kirkham, 37
- Javier
 - Mar, 13
- Joan
 - Clària, 32
- Joaquín
 - Martínez-Minaya, 19
- Jonas
 - Hentati-Sundberg, 9
- Jordi
 - Cortés Martínez, 18
 - Sunyer, 33
- José
 - Alcaraz-Quiles, 32
- José Luis
 - Gutiérrez Espinosa, 42
- José Luis Alfaro Navaro, 26
- Jose D.
 - Bermudez Edo, 5
- Jose L.
 - Gomez-Ariza, 5
- Josep
 - Redon, 5
- Klaus
 - Langohr, 4
- L.A.
 - González-De la rosa, 28
- L.J.
 - Sánchez-Martínez, 28
- Laia
 - Alemany, 38
 - Egea-Cortés, 27
- Laura
 - Santulario Verdú, 23
- Leda
 - Chatzi, 33
- Leire
 - Cidores, 10
 - Ibainarriaga, 10
- Leticia Belmonte Cortés, 26
- Lore
 - Zumeta-Olaskoaga, 14
- Lucia

- Alonso, 33
- Luigi
 - Spezia, 10
- M.
 - Cendoya, 43
 - Comas, 6
 - Sesé, 21
- M.A.
 - Martínez Beneito, 17
 - Moreno, 47
- Mònica
 - Guxens, 33
- Mabel
 - Morales Otero, 20
- María del Carmen
 - Aguilera Morillo, 29
- María Jesús
 - Rivas-López, 24
- María Xosé
 - Rodríguez- Álvarez, 16
- Marco
 - Pavesi, 32
- Maria
 - Grau-Perez, 5
 - Tellez-Plaza, 5
- Maria Grazia
 - Pennino, 10
- Mariano
 - Amo-Salas, 25
- Mark
 - Mon-Williams, 33
 - Nieuwenhuijsen, 33
- Marta
 - Bofill Roig, 18
- Martine
 - Vrijheid, 33
- Miguel
 - Moreno, 45
- Miguel A.
 - Moreno, 44
- N
 - Malats, 41
- N.
 - Quinteros, 22
- Núria
- Planell, 32
- O.
 - Viedma, 22
- P.
 - López-García, 47
- Paolo
 - Caraceni, 32
- R.
 - Ballesteros, 47
- Raúl
 - Martín-Martín, 24
- Richard
 - Moreau, 32
- Rocio
 - Ballesteros, 45
- Rodrigo
 - Dienstmann, 24
- Rosa Elvira
 - Lillo Rodríguez, 29
- Rubén
 - Amorós, 10
- Ruth
 - King, 9
- S.
 - Moreno, 6
- Sandra
 - Ramírez, 12
- Sara
 - Schroter, 37
 - Tous, 38
- Sergio
 - Pozuelo Campos, 15, 25
- Sonia
 - Sanchez, 10
- Sorina
 - Sferle, 38
- Susana
 - Pérez Álvarez, 23
- T
 - Pérez, 41
- U.
 - Aguirre, 6

V.	Virgilio
Dalmau, 21	Gómez Rubio, 12, 17, 42
Víctor	Yovaninna
Casero-Alonso, 25	Alarcón-Soto, 4
Víctor Raúl López Ruíz, 26	Zoraida
Vicente	Soriano-Gil, 5
Arroyo, 32	
Núñez-Antón, 20	