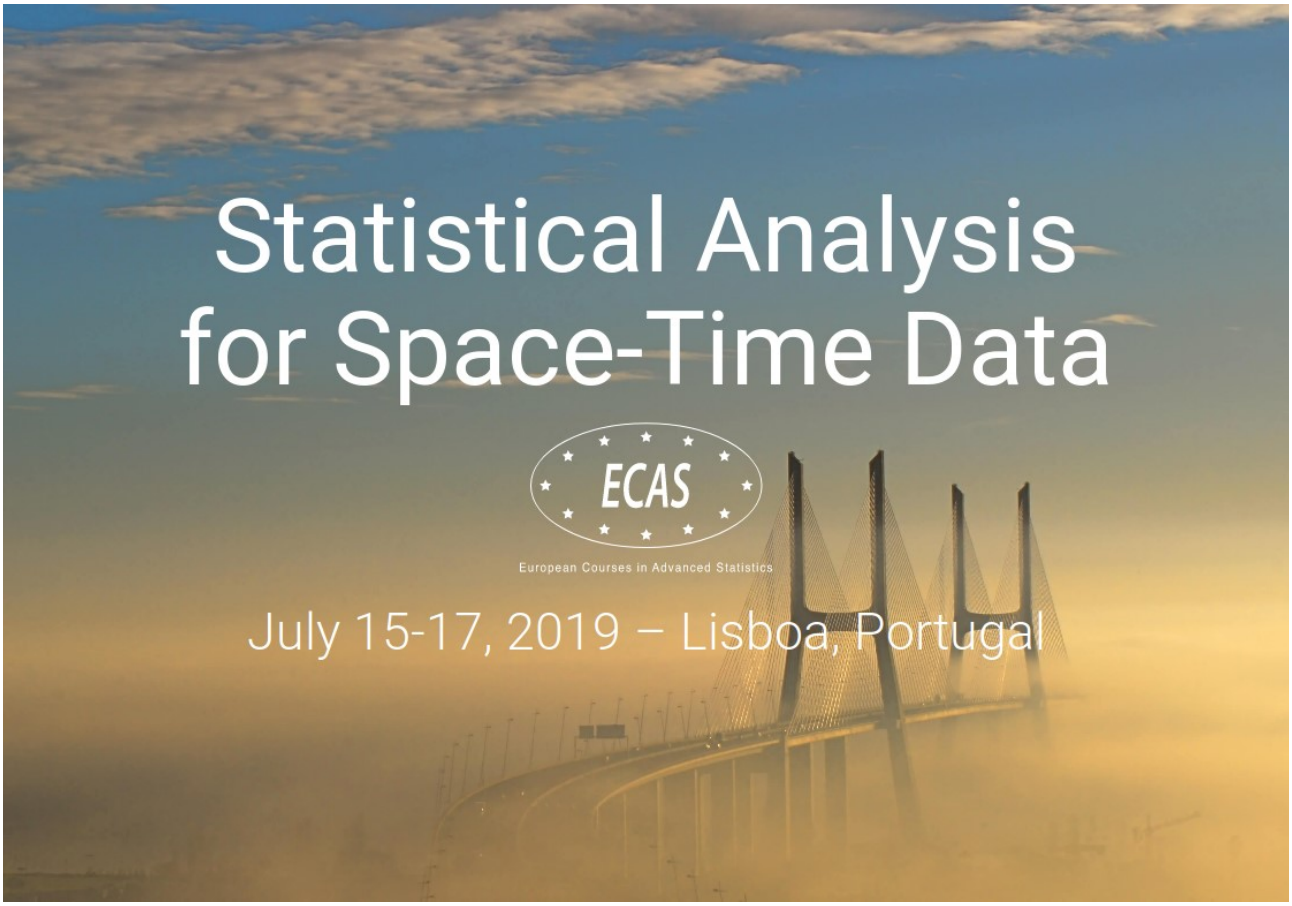


Programme and Abstract Book

Sociedade Portuguesa de Estatística

SEIO

Welcome

Dear participant, welcome!

The *European Courses in Advanced Statistics on Statistical Analysis for Space-Time Data* (ECAS2019) is organized by the Portuguese Statistical Society (SPE) and the Spanish Society of Statistics and Operational Research (SEIO) and is hosted by Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal, on 15-17 July 2019.

The ECAS courses are intended to achieve postgraduate training in special areas of statistics for (especially 1st year) PhD students, researchers, teachers at universities, and professionals interested in the application of new statistical methods.

Due to the proliferation of data sets that are both spatially and temporally indexed, spatio-temporal modelling has received an increasing attention in the last few years. Space-Time data are usually related to applied areas, such as environmental and health sciences, and their analyses focus namely on:

- reading, visualizing, and analysing spatial data, where observations can be identified with geographical locations;
- generating hypotheses for further ecological or epidemiological study;
- producing a smoothed map, identifying spatial and temporal trends;
- formulating policy decisions related to certain disease over space and time.

Invited renowned experts in the field present their methodological advancements with a heavy emphasis on applications in four different courses.

Adrian Baddeley (Curtin University) & Ege Rubak (Aalborg University) lecture *Spatial Point Patterns: Methodology and Applications with R.*
Patrick Brown (St. Michael's Hospital and University of Toronto) lectures *Statistical Models and Inference for Spatio-Temporal Areal Data.*
Liliane Bel (AgroParisTech) lectures *New Trends in Spatio-Temporal Geostatistics.*
Håvard Rue & Haakon Bakka (King Abdullah University of Science and Technology) lecture *Spatial and Spatio-Temporal Models Using the SPDE-Approach.*

A poster session with 26 contributed papers complements the scientific program.

We value all the contributes that justify the scientific merit of the meeting!

The Organizing Committee also thanks the colleagues of the Scientific Committee, the support of the institutions, namely the partners and sponsors listed in the last page, that made this event possible and, of course, all the participants without whom this event would be void!

This document is available on the website `https://ecas2019.math.tecnico.ulisboa.pt/`

Thank you and enjoy your stay!

Lisbon, July 2019

The Organizing Committee

ORGANIZING COMMITTEE

Isabel Natário
Departamento de Matemática da Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa, Portugal
Paulo Soares
Departamento de Matemática do Instituto Superior Técnico
Universidade de Lisboa, Portugal
Soraia Pereira
Departamento de Estatística e Investigação Operacional da Faculdade de Ciências
Universidade de Lisboa, Portugal
Tomás Goicoa Mangado
Departmento de Estadística, Informática y Matemáticas
Universidad Pública de Navarra, Spain
Anabel Forte
Departamento de Estadística e Investigación Operativa
Universitat de València, Spain
Giovani Silva
Departamento de Matemática do Instituto Superior Técnico
Universidade de Lisboa, Portugal

SCIENTIFIC COMMITTEE

Giovani Silva
Departamento de Matemática, Instituto Superior Técnico
Universidade de Lisboa, Portugal
Raquel Menezes
Departamento de Matemática e Aplicações
Universidade do Minho, Portugal
Eduarda Silva
Faculdade de Economia do Porto
Universidade do Porto, Portugal
María Dolores Ugarte
Departmento de Estadística, Informática y Matemáticas
Universidad Pública de Navarra, Spain
Rubén Fernández Casal
Departamento de Matemáticas
Universidade da Coruña, Spain
Ricardo Cao
Departamento de Matemáticas
Universidade da Corunha, Spain

General Information

Venue

The event is hosted by the Faculdade de Ciências da Universidade de Lisboa and it is held in the Amphitheater of the FCiências.ID – Associação para a Investigação e Desenvolvimento de Ciências, located in Campo Grande, C1 building, 3rd floor. 1749-016 Lisbon. The coffee breaks and lunches are also there.
The poster session is held in the lobby of C6 building during the second coffee break of the first day, that exceptionally happens there.

Wireless: `Eduroam`

Posters: The posters should be posted in the corresponding numerical position according to the numeration in this book, in the lobby of C6 building during the lunch time of the first day. On the end of the poster session they ought to be collected.

Certificates: All the due certificates will be sent by email on the week after the event.

Programme

Monday, 15 July 2019	
9:00 to 9:30	`Reception`
9:30 to 11:00	*Spatial Point Patterns: Methodology and Applications with R* Adrian Baddeley & Ege Rubak
11:00 to 11:30	`Coffee Break`
11:30 to 13:00	*Spatial Point Patterns: Methodology and Applications with R* Adrian Baddeley & Ege Rubak
13:00 to 14:00	`Lunch`
14:00 to 15:00	*Spatial Point Patterns: Methodology and Applications with R* Adrian Baddeley & Ege Rubak
15:00 to 17:30	`Poster Session & Coffee Break`
Tuesday, 16 July 2019	
9:30 to 11:30	*New Trends in Spatio-Temporal Geostatistics* Liliane Bel
11:30 to 12:00	`Coffee Break`
12:00 to 13:00	*Statistical Models and Inference for Spatio-Temporal Areal Data* Patrick Brown
13:00 to 14:00	`Lunch`
14:00 to 15:30	*Statistical Models and Inference for Spatio-Temporal Areal Data* Patrick Brown
15:30 to 16:00	`Coffee Break`
16:00 to 17:30	*Statistical Models and Inference for Spatio-Temporal Areal Data* Patrick Brown
Wednesday, 17 July 2019	
9:30 to 11:30	*New Trends in Spatio-Temporal Geostatistics* Liliane Bel
11:30 to 12:00	`Coffee Break`
12:00 to 13:00	*Spatial and Spatio-Temporal Models Using the SPDE-Approach* Håvard Rue & Haakon Bakka
13:00 to 14:00	`Lunch`
14:00 to 15:30	*Spatial and Spatio-Temporal Models Using the SPDE-Approach* Håvard Rue & Haakon Bakka
15:30 to 16:00	`Coffee Break`
16:00 to 17:30	*Spatial and Spatio-Temporal Models Using the SPDE-Approach* Håvard Rue & Haakon Bakka
17:30	`Farewell`

Courses

Spatial Point Patterns: Methodology and Applications with R

Adrian Baddeley & Ege Rubak

This course is a practical introduction to statistical methodology for analysing spatial point pattern data. Presentations of the basic principles are followed immediately by hands-on exercises with real data using the R language. Topics include exploratory data analysis, nonparametric estimation of intensity, parametric models (Poisson, Cox, Neyman-Scott, Gibbs), model-fitting and simulation, Monte Carlo tests and model diagnostics. The course gives an in-depth introduction to spatstat, an R package for analysing spatial point patterns. Our recently-published textbook can be consulted for more technical details. Familiarity with basic statistical concepts, and with the R language for statistical computation, is assumed.

New Trends in Spatio-Temporal Geostatistics

Liliane Bel

Spatial and spatio-temporal prediction techniques need the modelling of covariance function. We will investigate in this course methods for modelling and estimating the covariance function in some special frameworks : no separable spatio-temporal covariance, covariance on the sphere, estimating the covariance function with massive datasets, non stationary processes. Examples on real datasets will be studied, and R packages devoted to covariance estimation, spatio-temporal prediction and simulation will be presented.

Statistical Models and Inference for Spatio-Temporal Areal Data

Patrick Brown

This course will focus on the most common type of areal data, which are counts of incident cases of a health outcome in areas or regions over time. The focus will be on practical aspects of using these models, as well as interpretation and presentation of results. Computer code using R and INLA will be explained and examples will be provided for participants to run. Markov random field (or conditional autoregressive) models, where risk or intensity in each region depends on risk in neighbouring regions, are the basis of the standard and methods and tools for area-level problems. Various forms of spatial and spatio-temporal Markov random field models will be presented, and Bayesian inference for these models will be motivated and described. The course will conclude with the problem of changes in the spatial boundaries over time (i.e. changes in census regions), and methodology and software using the local-EM algorithm will be explained.

Spatial and Spatio-Temporal Models Using the SPDE-Approach

Håvard Rue & Haakon Bakka

In these lectures we will discuss how to construct (Gaussian) spatial models, in space and space-time, using stochastic partial differential equations (the so-called SPDE approach), and how to do Bayesian analysis of these models efficiently using R-INLA. The SPDE-approach comes with several benefites. First, it generates models, like the Gaussian fields with the Matern covariance function, that is extremely well suited for practical computations through sparse precision matrices. Secondly, it is the natural framework for constructing non-stationary models, like accounting for barriers, and non-separable models in space-time.

Posters

1. **Borrajo**: *Kernel intensity estimation on networks with spatial covariates* by M. I. Borrajo, C. Comas, J. Mateu.

2. **Boutigny**: *Spatial analysis of heterogeneous precipitation data, application to urban hydrology* by Marie Boutigny, Pierre Ailliot, Aurore Chaubet, Benoit Saussol, Antoine Sinquin.

3. **Costa**: *Statistical analysis of warming in Europe through time series modeling* by Marco Costa, Magda Monteiro.

4. **Eckardt**: *Analysing conditional independence in general multivariate spatial and spatio-temporal data based on partial marked point process characteristics* by Matthias Eckardt.

5. **Fernandez-Casal**: *Nonparametric Geostatistics with the npsp package* by Rubén Fernández-Casal.

6. **Fuentes-Santos**: *Nonparametric spatiotemporal point process analysis of gunfire violence in Rio de Janeiro metropolitan area* by Isabel Fuentes-Santos, Wenceslao Gonz alez-Manteiga, Jorge P. Zubellir.

7. **Ghorbani**: *Functional marked point processes - Unifying spatio-temporal frameworks and analysing spatially dependent functional data* by Cronie, O., Ghorbani, M., Yu, J., Mateu, J.

8. **Goicoa**: *SSTCDapp: a user-friendly web application to fir spatio-temporal models for count data* by Goicoa, T., Adin, A., Ugarte, M.D..

9. **Gorshechnikova**: *Likelihood approximation and prediction for large spatial and spatio-temporal datasets using hierarchical matrices* by Anastasiia Gorshechnikova, Carlo Gaetan.

10. **Guerra**: *Solving a linear fractional SPDE for spatial statistics* by Sílvia Guerra, Fernanda Cipriano, Isabel Natário.

11. **Heudtlass**: *Using over-the-counter (OTC) medication sales and the Moving Epidemic Method (MEM) for flu surveillance in Portugal (2016-2019)* by Peter Heudtlass, Nuno Rodrigues, Rúben Pereira, Zilda Mendes, António T. Rodrigues.

12. **Marques**: *Non-stationary spatial regression for modelling monthly precipitation in Germany* by Isa Marques, Nadja Klein, Thomas Kneib.

13. **Monteiro**: *Modelling informative time points: an evolutionary process approach* by Andreia Monteiro, Raquel Menezes, Maria Eduarda Silva.

14. **Monteiro**: *Recent trends in climate change in European cities* by Magda Monteiro, Marco Costa.

15. **Nadifar**: *A Flexible Bayesian Spatial Model For Over-dispersion and Under-dispersion Count Data* by Mahsa Nadifar, Håvard Rue, Hossein Baghishani, Thomas Kneib, Afshin Fallah.

16. **Navarrete-Alfonzo**: *Spatio-temporal analysis of ecosystem services in the Magdalena Basin, Mexico City* by Andrea Navarrete-Alfonzo, Carlos Díaz-Avalos.

17. **Ng**: *A continuous time spatial network model for analyzing bike sharing systems* by Tin Lok James Ng, Andrew Zammit-Mangion.

18. **Pinto da Silva**: *A dynamic Bayesian network for monitoring urban traffic states* by Pedro Pinto da Silva, Matthew Forshaw, Phil Blythe, Stephen McGough.

19. **Riebl**: *Gaussian process responses in distributional regression* by Hannes Riebl, Nadja Klein, Thomas Kneib.

20. **Rodrigues**: *Spatial analysis of bat abundance with adjustment for detection probability using distance sampling methods for acoustic data* by Sandra Rodrigues, P. de Zea Bermudez, K. Feridun Turkman, Helena Coelho, Miguel Mascarenhas.

21. **Rodríguez-Cortés**: *Spatio-temporal classification in point patterns under the presence of clutter* by Marianna Siino, Francisco J. Rodríguez-Cortés, Jorge Mateu, Giada Adelfio.

22. **Silva**: *Spatial estimation methods for the abundance of thornback ray* by Daniela Silva, Raquel Menezes, Ivone Figueiredo.

23. **Simoes**: *Accessing real urgencies in hospitals* by Paula Simões, Isabel Natário, Sérgio Gomes.

24. **Sousa**: *Censored linear regression model with AR(1) error term* by Rodney Sousa, Isabel Pereira, Maria Eduarda Silva.

25. **Zimmermann**: *Regional worker migration - stylized facts for Germany* by Mark Trede, Michael Zimmermann.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Kernel intensity estimation on networks with spatial covariates

M. I. Borrajo [1], C. Comas[2], J. Mateu[3]

[1]Lancaster University
[2]Universitat de Lleida
[3]Universitat Jaume I

Abstract

The analysis of point patterns have been of interest from a statistical point of view and many important advances have been made in the recent decades. However the complexity of the mathematical formalisation of point processes has originated a slower theoretical development in this field. Particularly, the analysis of events that occur on a network of lines, such as a road network, is a current challenge due to the different geometrical complexities providing singular structures.

In this work, we propose a kernel intensity estimator for point processes on networks under a model which establishes that the intensity of the process depends on a known covariate. The consistency of our proposal is proved, and the bandwidth selection problem is addressed. Finally, we analyse a dataset involving 2293 wildlife-vehicle collisions over the road network of Catalunya (Spain) during the period 2000-2006.

Keywords: Point processes; Linear networks; Kernel; Covariates.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Spatial analysis of heterogenous precipitation data, application to urban hydrology

Marie Boutigny[1,2], Pierre Ailliot[1], Aurore Chaubet[2], Benoît Saussol[1], Antoine Sinquin[2]

[1]Laboratoire de Mathématiques de Bretagne Atlantique, UMR 6205, Université de Brest, France
[2]Eau du Ponant, Brest, France

Abstract

In urban areas, where an important part of the sewerage system is combined, waste water dumping can occur during rainy weather. Water collection and storage systems are generally designed using hydrological models which describe the functioning of the sewerage systems. Weather conditions – most of all precipitation – are very important forcing factors for such models.

Gauges and radar are the two main data sources for precipitation. Although rain gauges are usually considered as the most reliable, in our case both radar and gauges can contain measurement errors and cannot be considered as representative of the true rainfall. We aim at developing a merging technique that takes into account the uncertainty of both data sources to rebuild a more reliable rain field which can then be used as input to the hydrological model. To model rainfall we use a hierarchical model, with a latent Gaussian field which supports the spatial structure. Radar and gauge data are considered as noisy observations of that field, in a data assimilation framework. The observation equation, that links the Gaussian domain to the precipitation one, is composed of 1) a censoring at zero to model the dry areas, and 2) a transformation of the positive part (*e.g.* power transform) for the wet areas.

Different numerical results with the hierarchical model on precipitation data in the city of Brest will be presented and compared to classical merging techniques. A spatio-temporal extension of the model, in which one the latent field is modelled using a markovian dynamic including an advection term (raincells motion) estimated with radar images, will also be discussed.

Keywords: Spatial statistics, rainfall, data merging, urban hydrology.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Statistical analysis of warming in Europe through time series modeling

Marco Costa[1,2], Magda Monteiro[1,2]

[1] ESTGA - Águeda School of Technology and Management, University of Aveiro
[2] CIDMA - Center for Research & Development in Mathematics and Applications, University of Aveiro

Abstract

Since the warming phenomenon must be monitoring in a smaller scale, a better understanding of the evolution of temperature series in Europe can contribute to a more efficient monitoring and identification of climate change patterns. In this context, statistical methods are crucial in both data analysis and to determine patterns or tendencies, in order to apply efficient environment monitoring processes.

For this purpose, in this work, we present time series of monthly average temperatures in several European locations which were statistical analyzed and modeled with appropriated statistical models. Data set comprises sixteen long-term time series of monthly mean temperatures, in ^{O}C, from January 1900 until February 2017 (available at the Climate Data Online – CDO).

Data is modeled using linear regression models, state space models associated to the Kalman filter. Additionally, it is performed a clustering procedure of smoother trends levels in order to investigate patterns on the warming in Europe.

The clustering procedure show a five clusters solution which discriminates groups of cities in the mean of the trend level. Moreover, this discrimination is more evident in the period after 1980's where the Cluster of the Southwest Europe has the minors trend levels whereas the Cluster of the cities located in the Eastern Europe has the highest trend levels, reaching annual means of 2.5ºC over the seasonal curve in the period between 1900 and 2017.

Keywords: climate change, monthly data, temperature data, time series analysis, state space modeling, clustering.

Acknowledgements Authors were partially supported through Portuguese funds by the Center for Research and Development in Mathematics and Applications (CIDMA) and the Portuguese Foundation for Science and Technology ("Fundação para a Ciência e a Tecnologia" [FCT]), within project UID/MAT/04106/2019.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Analysing conditional independence in general multivariate spatial and spatio-temporal data based on partial marked point process characteristics

Matthias Eckardt[1]

[1]Department of Computer Science, Humboldt-Universität zu Berlin, Berlin, Germany

Abstract

The analysis of multivariate spatial and spatio-temporal data is an emerging field in statistics which poses substantial challenges for the identification and evaluation of potential structural interrelations among different spatially indexed outcomes over time. Due to the ever-increasing availability and accessibility of huge spatial datasets, computationally efficient analysis techniques that not only account for the inherent complexity of such data but also facilitate a clear interpretation increasingly gain importance. A suitable approach should be able to control for the planar arrangement of the outcome and also allow for the simultaneous analysis of different types of multivariate spatial data. A novel graphical model based on the framework of marked spatial point processes is proposed, allowing for a unified treatment of different types of spatial data. In this graphical model, termed the spatial dependence graph model, the edge set is identified through the partial spectral coherence between two components conditional on all remaining components. Beginning with qualitatively and quantitatively marked spatial point processes, extensions to spatial lattice and hybrid as well as spatio-temporal point, lattice and hybrid processes are described. Further, partial statistics which describe pairwise interrelations conditional on all remaining component patterns under study are outlined. In several applications and comparative analyses, the proposed model turns out to be a promising alternative to well-established spatial statistical methods.

Keywords: Conditional independence; Frequency domain; Mixed multivariate spatial and spatio-temporal point-lattice patterns; Multivariate spatial and spatio-temporal processes; Partial spatial and spatio-temporal statistics

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Nonparametric Geostatistics with the npsp package

Rubén Fernández-Casal

Dep. de Matemáticas, Universidad de A Coruña (Spain)

Abstract

In this work the R package `npsp` (R. Fernández-Casal, 2019, Nonparametric spatial statistics, *https://rubenfcasal.github.io/npsp*) is presented. This package implements nonparametric methods for inference on multidimensional geostatistical processes, avoiding the misspecification problems that may arise when using parametric models. The spatial process can be either stationary or show a non-constant trend. Joint estimation of the trend and the semivariogram can be performed automatically, by using the function `np.fitgeo`, or by a step-by-step approach.

The `locpol` function can be used to compute local polynomial kernel estimates of the spatial trend. The bandwidth may be selected by using the generic function `h.cv` which implements different cross-validation approaches, including the corrected generalized cross-validation criterion, so that the spatial dependence is taken into account. In this general setting, the semivariogram can be estimated from the residuals, but their direct use introduces a bias in dependence estimation. The `np.svariso.corr` function implements an iterative algorithm to compute bias-corrected pilot semivariogram estimates. When the trend is constant, nonparametric semivariogram estimates can be obtained with the function `np.svar`. In both cases, the corresponding bandwidth may also be selected with the generic function `h.cv`, which uses the cross-validation relative squared error criterion for pilot variogram estimation by default. To obtain the final variogram, function `fitsvar.sb.iso` may be used to fit a flexible Shapiro-Botha model to the pilot estimates. Currently, only isotropic semivariogram estimation is supported, but the intention is to extend this approach to anisotropy in two components, which could be suitable for modeling spatio-temporal dependency.

The final trend and variogram estimates could be applied to spatial prediction, by using function `np.kriging`, but they also allow making inferences about other characteristics of the process, for instance, by using the bootstrap algorithm implemented in `np.boot`.

Keywords: Local polynomial estimation, Variogram, Kriging, R software.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Nonparametric spatiotemporal point process analysis of gunfire violence in Rio de Janeiro metropolitan area

Isabel Fuentes-Santos[1,2], Wenceslao González-Manteiga[2], Jorge P. Zubellir[3]

[1]Marine Research Institute, CSIC, (Spain)
[2]MODESTYA research group. University of Santiago de Compostela, (Spain)
[3]Instituto Nacional de Matenática Pura e Aplicada (Brazil)

Abstract

The continuous increase of violent crime in Rio de Janeiro metropolitan area (Brazil) motivated the development of the collaborative mobile app *Fogo Cruzado*, which delivers instant alerts every time a user reports a gunfire event. In addition to its contribution to public safety, *Fogo Cruzado* generates a valuable dataset to understand gunfire dynamics and support the development of crime reduction strategies. avoid stray bullets and generates a valuable dataset. This work analyses the 5945 gunfire shootings collected by Fogo Cruzado during 2017. This data set, which comprises the location and time of occurrence of each gunfire event, is a realization of a spatiotemporal point process marked by the number of injured and mortal victims, and by an indicator of police presence.

We have applied nonparametric first and second-order inference tools to analyze spatiotemporal patterns of gunfire in Rio. Kernel intensity estimation with plug-in bandwidth matrix describes the spatial distribution of gunfire and allow us to detect chronic hotsptos. The nonparametric test for comparison of first-order intensities found differences between gunfire events with and without fatalities, whereas the kernel relative risk function detected high mortality risk in areas with low gunfire incidence. Similar results were obtained for gunfires with and without police presence. The log-ratio based first-order separability test rejected the null hypothesis for the different gunfire patterns under study, therefore the spatial distribution of gunfire, fatalities and police presence varied over time. Finally, spatiotemporal inhomogenosu K-tests detected clustering between gunfire events up to 0.5km ad 1 day, thus the occurrence of a gun shooting increases the risk of new events within 0.5 km during the next 24 hours. Fatalities and police interventions have also clustered patterns but with different interaction radius.

The nonseparable first-order distribution and the short-term clustering detected in gunfire events suggest that a self-exciting point process with nonseparable background component shall be a suitable model to predict gunfire hotspots in Rio de Janeiro.

Keywords: First-order intensity, *Fogo cruzado*, gunfire, kernel smoothing, inhomogeneous K-function, separability test

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Functional marked point processes – Unifying spatio-temporal frameworks and analysing spatially dependent functional data

Cronie, O.[1], Ghorbani, M.[1], Yu, J.[1] Mateu, J.[2]

[1]Department of Mathematics and Mathematical Statistics, Umeå University
[2]Department of Mathematics, Universitat Jaume I, Campus Riu Sec, 12071 Castellón, Spain

Abstract

This paper treats functional marked point processes (FMPPs), which are defined as marked point processes where the marks are random elements in some (Polish) function space. Such marks may represent e.g. spatial paths or functions of time. To be able to consider e.g. multivariate FMPPs, we also attach an additionally, Euclidean, mark to each point. We indicate how the FMPP framework quite naturally connects the point process framework with both the functional data analysis framework and the geostatistical framework; in particular we define spatio-temporal geostatistical marking for point processes. We further show that various existing stochastic models fit well into the FMPP framework, in particular marked point processes with real valued marks. To be able to carry out non-parametric statistical analyses for functional marked point patterns, we study characteristics such as product densities and Palm distributions, which are the building blocks for summary statistics such as marked inhomogeneous J-functions and our so-called K-functionals. We finally apply these statistical tools to analyse a few different functional marked point patterns.

Keywords: Càdlàg stochastic process, Correlation functional, Functional marked point process, Intensity functional, Marked inhomogeneous K-functional, Spatio-temporal geostatistical marking.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

SSTCDapp: a user-friendly web application to fit spatio-temporal models for count data

Goicoa, T.[1,2], Adin, A.[1,2], Ugarte, M.D.[1,2]

[1]Department of Statistics, Computer Science and Mathematics, Public University of Navarre.
[2]Institute of Advanced Materials (InaMat), Public University of Navarre.

Abstract

Disease mapping models have become essential in the statistical toolbox of health researchers to discover spatial and spatio-temporal patterns of a disease. In general, complex models are necessary to smooth highly variable crude measures, such as standardized mortality ratios or crude rates. However, in spite of the huge research on spatio temporal models for count data, some applied researchers find them difficult to understand and fit, as there are technical issues that are far from simple. Some theoretical matters include for example a correct identification of the model terms, a key point that requires a deep knowledge of the model structure and mathematical background. Among the fitting problems users might find we can list the choice of priors and hyperpriors, the long computing time required to fit spatio-temporal models using Markov chain Monte Carlo (MCMC) techniques, or simply the selection of appropriate software to fit the models. In this work we present STTCDapp, an interactive web application developed with Shiny to fit a wide range of spatio-temporal models for count data. The application includes some descriptive statistics and classical measures to start with. Then, it offers the possibility of fitting complex hierarchical Bayesian spatio-temporal models in an appealing and easy way. The web application STTCDapp relies on INLA (Integrated nested Laplace approximations) to fit the models and appropriate constraints are automatically and correctly placed to get model identification without the user specification. The web application is available from the web browser at https://emi-sstcdapp.unavarra.es/Login/ and users do not need to install any software in their own computers. All the computations are run in a remote server and the application provides different outputs including summary measures, spatial, temporal, and spatio-temporal patterns, risks or rates estimates and posterior probabilities. The code required to fit the models can be also recovered by the users if needed. An exhaustive user manual together with some examples are available in the web application. Although disease mapping has commonly focused on diseases, and particularly cancer, this web application may also be very useful in other research areas such as gender-based violence, where very recently some research has been conducted to discover spatio-temporal patterns of rapes and dowry deaths in Uttar Pradesh, India.

Keywords: Areal count data, INLA, Remote server, Shiny.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Likelihood approximation and prediction for large spatial and spatio-temporal datasets using hierarchical matrices

Anastasiia Gorshechnikova[1], Carlo Gaetan[2]

[1]University of Padova, Italy
[2]Ca' Foscari University of Venice, Italy

Abstract

Large datasets with irregularly sited locations are difficult to handle for several applications of Gaussian random fields such as maximum likelihood estimation (MLE) and kriging prediction due to a high computational complexity. For relatively large spatial and (or) temporal dimensions the exact computation becomes unfeasible and alternative methods are necessary. Several approaches have been proposed to tackle this problem. Most of them assume a specific form for the covariance function and use different methods to approximate the resulting covariance matrix. The aim is to approximate the covariance functions in a format that facilitates the computation of MLE and kriging prediction with large spatial and spatio-temporal datasets.

For a sufficiently general class of spatial covariance functions, a methodology is developed using hierarchical matrices. This technique involves the partitioning of a matrix into sub-blocks according to specific given conditions with the further approximation of the majority of these sub-blocks by low-rank matrices. The resulting covariance matrix allows for computation of the matrix-vector products and matrix factorisations in a log-linear computational cost followed by an efficient MLE and kriging prediction.

Numerical experiments are performed on simulated spatial and spatio-temporal data to recover the true values of the parameters of the covariance matrix. This method is further applied on a real dataset consisting of measurements of atmospheric carbon dioxide mole fractions. The prediction accuracy and computational time are compared with other methods. As a result, among the methods considered in this study, the approach presented is the most efficient in terms of the root mean-squared prediction error and computational time.

Keywords: computational methods, hierarchical matrices, large datasets, spatio-temporal covariance matrices.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Solving a linear fractional SPDE for Spatial Statistics

Sílvia Guerra[1,2,3], Fernanda Cipriano[1,2], Isabel Natário[1,2]

[1] Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal
[2] Centro de Matemática e Aplicações (CMA), Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal
[3] Nova School of Business & Economics, Portugal

Abstract

The statistical modeling of phenomena whose spatial location is relevant has been increasingly recognized and facilitated by computational advances. Researchers have today access to spatial data on a wide range of subjects, namely ecology, social science, epidemiology, weather, etc. This has provided the study of various spatial models and applications, from which the geostatistical research is of interest here. The main challenge are the computational difficulties associated with modeling the dependence in space, so a novel approach based on stochastic partial differential equation models has been proposed and applied.

In geostatistical models, random variables are represented by a Gaussian field (GF), $x(\mathbf{u})$, $\mathbf{u} \in D \subseteq \mathbb{R}^d$. The counter part of GF is the so called "big N problem", computations scale as $O(N^3)$. A way to address this issue is to work with Gaussian Markov random fields (GMRF) that represent the GF $x(\mathbf{u})$ and use numerical methods for sparse matrices. Lindgren, Rue, and Lindstrom (2011) study a link between GF and GMRF, using the solution of a stochastic partial differential equation (SPDE) corresponding to the Matérn covariance model. The purpose of this poster is to look closely into the derivation of the GF $x(\mathbf{u})$ that solves the linear SPDE

$$(k^2 - \Delta)^{\alpha/2} x(\mathbf{u}) = \mathcal{W}(\mathbf{u}) \tag{1}$$

where $\mathbf{u} \in \mathbb{R}^d$, $\alpha = \nu + d/2$, $k > 0$, and $\nu > 0$ and $\mathcal{W}(\mathbf{u})$ is a spatial Gaussian White Noise with unit variance. The solution is obtain applying Fourier transform methods. We will present the calculations for the expected value and covariance function of the solution, in order to explicitly obtain the Matérn covariance

$$r(\mathbf{u}, \mathbf{v}) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (k\|\mathbf{v} - \mathbf{u}\|)^{\nu} K_{\nu}(k\|\mathbf{v} - \mathbf{u}\|).$$

Keywords: Gaussian fields, Gaussian Markov random fields, Fourier transform, Matérn covariance,stochastic partial differential equation.

Acknowledgments: This work is financed by national funds through FCT - Foundation for Science and Technology - under the projects UID/MAT/00297/2019.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Using over-the-counter (OTC) medication sales and the Moving Epidemic Method (MEM) for flu surveillance in Portugal (2016-2019)

Peter Heudtlass[1], Nuno Rodrigues[2], Rúben Pereira[1], Zilda Mendes[1], António T. Rodrigues[1]

[1)] Centre for Health Evaluation & Research (CEFAR), Portuguese Pharmacy Association (ANF), Lisbon, Portugal
[2)] ACES Oeste Sul, Torres Vedras, Portugal

Abstract

Background: Most patients with influenza-like illness (ILI) probably do not see seek medical attention immediately, if at all. This potentially leads to a lag in detecting the start of flu epidemics and in predicting their peak, when using health care case load data for surveillance. We are trying to improve the timeliness of flu surveillance using geo-referenced daily sales data of over-the-counter (OTC) medication for treatment of early flu symptoms.

Methods: We obtained daily data on ILI cases reported by primary care units in the health zone Oeste Sul, north of Lisbon, Portugal, and daily OTC medication sales from community pharmacies in the same area, covering three flu seasons (2016-2019). Using the first two seasons, a time-warping algorithm and time series correlations, we construct an index (HICORR) of OTC medication whose sales are highly correlated with the ILI case load. We estimate two Moving Epidemic Method (MEM) models, one using HICORR sales and the other one using ILI cases and compare model metrics and the estimated start, peak and intensity of the 2018/2019 flu epidemic using data from the national flu sentinel network as gold standard.

Results: [PRELIMINARY RESULTS] The models using HICORR and ILI perform both well in terms of sensitivity specificity when compared to the gold standard. With HICORR we are detecting the beginning and the peak of the 2018/2019 flu epidemic ahead of the ILI model and the gold standard and obtain a comparable estimate of the intensity.

Conclusions: Data from OTC sales in community pharmacies can complement and strengthen flu surveillance in Portugal, particularly also at the local level. We will pilot flu surveillance based on real-time HICORR sales during the 2019/2020 flu season.

Keywords: Influenza, Surveillance, Community Pharmacies

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Non-Stationary Spatial Regression for Modelling Monthly Precipitation in Germany

Isa Marques[1], Nadja Klein[2], Thomas Kneib[1]

[1] Georg-August-University of Göttingen, Germany
[2] Humboldt-University of Berlin, Germany

Abstract

It is widely accepted that spatial dependencies have to be acknowledged appropriately in data that are spatially aligned. However, most spatial models still assume stationarity. While this assumption considerably facilitates estimation, it is often too restrictive when describing atmospheric phenomena such as precipitation. However, the general applicability of non-stationary models is often hindered as they considerably increase computational complexity and improvements over stationary models are sometimes hard to identify. The stochastic partial differential equation (SPDE) approach to spatial modelling allows for flexible specifications of non-stationary models. Given the German orographic diversity, it makes sense to model precipitation in a non-stationary way. We suggest an SPDE-approach with elevation-based non-stationarity, as well as a more flexible spline-based alternative. Results show that, according to the widely applicable Bayesian information criterion (WAIC), a non-stationary model provides a better fit to the data. Taking German monthly precipitation as a case study, we use a simulation study to explore the ability of the elevation and spline-based models to correctly identify non-stationarity. This ability is analyzed, under a fixed domain, for different hyperpriors, increasing sample size and signal-to-noise ratio. Additional scenarios test different scalings of the spatially varying parameters of the SPDE. The main results show that the stationary data sets are always correctly identified, while identifiability of non-stationary data sets improves with sample size. The elevation-based model performs reasonably well for most sample sizes and it is in general robust to different prior specifications. Nonetheless, it requires some degree of confidence in the hyperprior set, namely in terms of the precision of normally distributed priors. The spline-based model is more data-driven. Finally, some steps are taken towards exact Bayesian inference for the SPDE-approach through Markov Chain Monte Carlo (MCMC). This would allow us to compare the benefits of integrated nested Laplace approximations over MCMC for normally distributed data. Potential extensions of this analysis include alternative ways of modelling non-stationarity and more calibrated priors.

Keywords: Gaussian (Markov) random fields; Monthly precipitation; Non-stationary spatial modelling; Stochastic partial differential equations; Bayesian inference.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Modelling informative time points: an evolutionary process approach

Andreia Monteiro[1], Raquel Menezes[2], Maria Eduarda Silva[3]

[1]University of Minho & CIDMA, Portugal
[2]University of Minho & CBMA, Portugal
[3]Faculty of Economics, University of Porto & CIDMA, Portugal

Abstract

Analysis of experimental data that have been observed at different points in time leads to specific problems in statistical modelling and inference. In traditional time series the main emphasis is on the case when a continuous variable is measured at discrete equispaced time points and there is an extensive body of literature on analysing equally spaced time series data. Nevertheless, unevenly spaced (also called unequally or irregularly spaced) time series data naturally occur in many scientific domains. For example, data related to natural disasters such as earthquakes, floods, or volcanic eruptions which typically occur at irregular time intervals, give rise to irregularly or unevenly spaced time series. A particular situation of irregularly spaced data is that in which the sampling design depends also, for practical constraints, on the observed values. Examples occur in fisheries where the data are observed when the resource is available, in sensoring when sensors keep only some records in order to save memory and in clinical studies, when a worse clinical condition leads to more frequent observations of the patient. In all such situations, there is stochastic dependence between the process under study and the times at which the observations are made, and the observation times are informative on the underlying process. Ignoring this dependence can lead to biased estimates and misleading inferences.

In this work we consider that the sampling design may depend on entire past history of the process, meaning all the times of the observations as well as the values of these observations. In these situations, the observed time points can be considered informative to the process being studied. In current work, taking into account the natural temporal order underlying available data represented by a time series, then a modelling approach based on evolutionary processes seems a natural choice. We consider maximum likelihood estimation of the model parameters. Numerical studies with simulated and real data sets are performed to illustrate the benefits of this model based approach versus the traditional one for irregularly spaced data.

Keywords: Preferential Sampling; Continuous Time Autoregressive Process; SPDE; Evolutionary Processes.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Recent trends in climate change in European cities

Magda Monteiro[1,2], Marco Costa[1,2]

[1] ESTGA - Águeda School of Technology and Management, University of Aveiro
[2] CIDMA - Center for Research & Development in Mathematics and Applications, University of Aveiro

Abstract

The effects of global warming are already visible all over the world. The increase of droughts, floods, severe storms, and other weather catastrophes is putting vulnerable human and biological populations at risk. However, global warming effects vary around the world, therefore this phenomenon must be monitored at a smaller scale. It is in this context that the statistical analysis of local time series, such as cities, has a special interest in order to monitor temperature rise.

One result of a previous work of the authors is that temperature rise rates in Europe seem to have increased in the last decades when compared with longer periods, such as a century, so it is important to find the change point and use the information from there to get accurate warming rates.

In this work we will study time series of monthly average temperatures in several European locations. Data set comprises fifteen long-term time series of monthly mean temperatures, in ^{o}C, from January 1900 to the March 2019 (available at the Climate Data Online CDO).

In a first step we will find, for each city, the change point time based on a state-space model with a stochastic trend. Two different models will be used in the second step that allow for the analysis of the warming rate in the recent years.

Keywords: Air temperature; Climate change; State-space models; Change-point detection.

Acknowledgements

Authors were partially supported through Portuguese funds by the Center for Research and Development in Mathematics and Applications (CIDMA) and the Portuguese Foundation for Science and Technology (Fundação para a Ciência e a Tecnologia" [FCT]), within project UID/MAT/04106/2019.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

A Flexible Bayesian Spatial Model For Over-dispersion and Under-dispersion Count Data

Mahsa Nadifar[1,3], Håvard Rue[2], Hossein Baghishani[1], Thomas Kneib[3], Afshin Fallah[4]

[1]Shahrood University of Technology
[2] King Abdullah University of Science and Technology
[3]Georg-August-Universität Göttingen
[4]Imam Khomeini International University

Abstract

In recent years, the spatial count data has been a massive concern. Actually, real count data often represent over- or under-dispersion which Poisson regression model is not appreciated. We propose a Bayesian hierarchical modeling approach based on the renewal theory that relates nonexponential waiting times between events and the distribution of the counts, relaxing the assumption of equidispersion at the cost of an additional parameter. Particularly, we extend the methodology for the analysis of spatial count data based on the gamma distribution assumption for waiting times. Setting convenient prior distributions on model parameters is an momentous and reasonable issue in Bayesian statistics that characterize the nature of our uncertainty of parameters. Relying on a recently proposed class of penalized complexity priors motivated from a general set of construction principles, we derive prior structure. The model can be formulated as a latent Gaussian model, and consequently, we can carry out the fast computation by using the integrated nested Laplace approximation method. We investigate the proposed methodology in simulations and a dataset on larynx cancer mortality in Germany, during the years of 1986 to 1990. The analysis of the real dataset and an extensive simulation study show a significant improvement over both Poisson and negative binomial models.

Keywords: Gamma-count distribution; Spatial count data; Bayesian Statistics; Penalized complexity priors; Integrated nested Laplace approximation.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Spatio-temporal analysis of ecosystem services in the Magdalena Basin, Mexico City.

Andrea Navarrete-Alfonzo [1], Carlos Díaz-Avalos [2]

[1] Posgrado en Ciencias de la Sostenibilidad, Universidad Nacional Autónoma de México, Ciudad de México, México.

[2] Department of Probability and Statistics, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Ciudad de México, México.

Abstract

Peri-urban areas maintain a close link with the dynamics of cities, in such a way that patterns and dynamics of urban expansion as well as anthropogenic activities result in changes and even loss of the ecosystem services provided by peri-urban areas. Mexico City has registered intense urban growth since the early 20th, being more notorious at the edge of the urban areas. This trend in city growth affects a preservation zone located mainly in the south of the city, called Suelo de Conservacion (SC). This area has a high environmental value due to its capability to provide environmental services and other goods such as wood to the city. The Magdalena river basin (MRB) is a socio-ecological system that is located within the SC area, and because of its proximity to the urban area, it involves a variety of social, political, economic and environmental processes that influence its dynamics and configuration. The MRB occupies 4% of the SC and contains one of the last non-piped rivers remaining in the city. It provides 50% of the superficial water supply to Mexico City, as well as other important ecosystem services to its inhabitants. The aim of this study is to assess the effect of four factors related to anthropogenic direct change: land-use change, vegetal cover, wastewater discharges, and illegal settlements. The study also aims to assess the effect of biophysical factors, on three ecosystem services generated by the MRB (water supply, carbon storage and recreation areas).

We fitted Generalized Linear Models with spatial component to model changes in several ecosystem services. We constructed scenarios to show the behavior trend of the ecosystem services considered in the study. The modeling allowed to identify the most relevant parameters that affect the provision of ecosystem services, which can serve as a guide for decision makers as well as in the establishment of public policies that ensure the optimal generation of ecosystem services in the basin for the benefit of the inhabitants of the city..

Keywords: ecosystem services, peri-urbanisation, generalized linear regression models, urban sustainability, peri-urban area.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

A continuous time spatial network model for analyzing bike sharing systems

Tin Lok James Ng[1], Andrew Zammit-Mangion [2]

[1]School of Mathematics and Applied Statistics, University of Wollongong
[2]School of Mathematics and Applied Statistics, University of Wollongong

Abstract

Bike-sharing systems allow users to rent bicycles for short trips by accessing a dock within the system to unlock or return bicycles. Such systems, which aim to fill in public transit gaps and to reduce congestion and air pollution, have been growing in number and popularity in major cities across the world. In recent years, bike-sharing data sets have been made available to the public, allowing the analysis of the transportation dynamics, and how it relates to where people live and where they work. For example, the bike-sharing data provided by Capital Bikeshare Washington D.C. includes start date and time of a trip, its duration, starting and ending stations, etc. Such continuous time relational data can be naturally represented using a network structure with time-stamped edges, where the vertices of the network are the stations and the edges are the time-stamped trips between each pair of stations.

Although statistical models have been developed to analyze bike-sharing data, existing methodologies typically require discretization of the time domain, and do not take into account the spatial information of the bike stations. However, such discretization is arbitrary and the lack of consideration of spatial information can lead to inferior predictions. We develop a parsimonious continuous time spatial network model to address these shortcomings. A spatio-temporal process is incorporated in the model to represent how likely a bicycle is unlocked at each station across time, while the interactions between those stations are modeled using a stochastic block structure. The developed model can capture the spatial and temporal variations of bike-sharing activities as well as clustering of the bike stations. Statistical inference of the model is performed using a Bayesian approach to estimate the process and the block structure. Bike-sharing activities between stations and across time can be both visualized and simulated from the estimated model. The developed model is also applicable to other types of continuous time relational data that contain spatio-temporal information.

Keywords: Spatial network, Stochastic Block model, Spatio-temporal process, Bike-sharing systems

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

A Dynamic Bayesian Network for Monitoring Urban Traffic States

Pedro Pinto da Silva[1], Matthew Forshaw[1], Phil Blythe[2,3], Stephen McGough[1]

[1]School of Computing, Newcastle University, United Kingdom
[2]School of Engineering, Newcastle University, United Kingdom
[3]Department for Transport, United Kingdom

Abstract

Tackling traffic congestion in urban cities remains an open challenge. To aleviate congestion, traffic operators and automated traffic control systems (e.g. green-light timer scheduling algorithms) rely extensively on estimates of traffic state derived from real-time traffic data. Whilst large tech companies are successful in crowdsourcing trajectory data, using GPS sensors in mobile phones, in exchange for services that provide real-time traffic information (e.g. Google Traffic), local governments and councils still rely on survey data and a small number of static sensors to create traffic forecasts and inform decision makers in transportation planning. Recently, the increasing use of a new class of sensors, Automatic Number Plate Recognition (ANPR) cameras, provides new insights into the travel patterns of individual vehicles and promises to significantly enhance current systems for traffic management and control.

In this work we propose a Dynamic Bayesian Network (DBN) that learns the state of traffic (free flow, mild congestion, severe congestion) for different routes in the road network, over time, from ANPR data. The structure of the DBN, i.e. the set of routes and their spatial dependencies, is derived from data by analysing individual vehicle trips consisting of several consecutive camera observations. We validate the learned structure by comparing it to the empirical structured used by expert traffic operators with knowledge of Newcastle's road network and ANPR cameras. The observation model consists of average vehicle speeds (one per vehicle per route) which we assume are independent conditional on the corresponding parent hidden node. We test our model on a smaller subset of the network and estimate the parameters using the Expectaction-Maximisation (EM) algorithm. For the expectaction (E) step, often intractable, we analyse the computational complexity of exact inference (forward-backward algorithm) and consider methods for approximate inference (Boyen-Koller, Rao-Blackwellised particle filter).

Empirical results from the learned model will shed light into the capabilities of ANPR networks to provide a clearer picture of urban traffic flow. Ultimately, the proposed model is a step towards a real-time traffic monitoring and forecast system that can be used to combat urban traffic congestion.

Keywords: Urban analytics, Urban Computing, Real-Time Traffic Monitoring, Hidden Markov Models, Dynamic Bayesian Networks

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Gaussian Process Responses in Distributional Regression

Hannes Riebl[1], Nadja Klein[2], Thomas Kneib[3]

[1]Chair of Statistics, University of Göttingen
[2]Chair of Statistics, Humboldt University of Berlin
[3]Chair of Statistics, University of Göttingen

Abstract

We propose a new class of regression models with conditionally independent Gaussian processes (GPs) as responses. Our model is embedded into the framework of structured additive distributional regression models, which have been popularized as Generalized Additive Models for Location, Scale, and Shape (GAMLSS). The distributional regression framework allows us to link all distributional parameters of the responses to covariates via structured additive predictors. In our case, the distributional parameters are the parameters of the mean and the covariance function of the GPs.

For inference and implementation, we follow the `bamlss` R package. Specifically, we use an efficient MCMC sampling scheme for fully Bayesian inference. The proposals in the Metropolis-Hastings algorithm are based on iterative weighted least squares (or equivalently, on Fisher scoring), and the parameters are updated in a backfitting-like nested loop through (1) the distributional parameters and (2) the linear covariate effects and the spline-based smooth terms. The implementation of the likelihood, score, and Fisher information needs to be efficient enough to evaluate them repeatedly during MCMC sampling. The main challenge is to process the large covariance matrices of the GPs at the observed points. Inversion of these matrices should be avoided and quantities related to them should be cached whenever possible.

Finally, we use our model to analyze growth curves of 85 deciduous trees from three different regions in Germany. For each tree, the growing seasons 2012 and 2013 were recorded in a high temporal resolution using electronic circumference dendrometers. Our model uses a scaled Weibull CDF as a mean function and a (slightly adjusted) Matérn covariance function. The mean function has three parameters (limit, shape, and scale), and the covariance function has two parameters (standard deviation and range), each of which we link to covariates like tree species or size. Our model is spatio-temporal in the sense that the GP responses are temporal and spatial smoothers are used in the predictors. We find, among other results, that beeches have longer growing seasons and less "random" fluctuation in their growth than the average tree in our dataset.

Keywords: Generalized additive models for location, scale, and shape; Structured additive predictors; Bayesian inference; Tree growth.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Spatial analysis of bat abundance with adjustment for detection probability using distance sampling methods for acoustic data

Sandra Rodrigues[1,2], P. de Zea Bermudez[1], K. Feridun Turkman[1], Helena Coelho[2], Miguel Mascarenhas[2]

[1]CEAUL - Centro de Estatística e Aplicações, Universidade de Lisboa, Portugal.
[2]Bioinsight, Lda., Portugal.

Abstract

Spatial analysis of bats has become increasingly important in environmental impact assessment studies (EIA), as it allows to determine areas where bats are more abundant and, while avoiding these areas for construction of the project, it is expected that less impact over the species will be induced. However, current spatial analysis with bats have been performed without adjusting for detection bias because the determination of detection functions for acoustic data of bats is not easy. Bats are nocturnal species and most are very small, which makes estimation of distance to the bats very difficult. In this study we aim to apply a spatio-temporal point process model (PPM) developed for distance sampling data while applying an existing protocol that allows to determine distances from acoustic data. Additionally, we intent to compare the results this models with currently used models that do not incorporate adjustments for detection bias. Models will be compared on how results differ in terms of how bat abundance varies in space and its relation to covariates. The dataset comprises opportunistic data gathered in the framework of an EIA study located in western South Africa. Bat locations were obtained from car-based transects using an acoustic bat detector (Echo Meter EM3, Wildlife Acoustics, inc.). The energy of the sound ("power") of calls produced by bats was extracted and used as a proxy for distance of bats to the transect. We used distance-power relations from previous studies to obtain the best estimate of the distances. This method enables to calculate detection functions that will be used in the PPM to scale the intensity function resulting in a thinned process that represents the decrease of detection of bats with distance. An inhomogeneous PPM will be fitted to the data and a variety of meteorological and environmental covariates (e.g. temperature, habitat type, etc) will be examined. It is expected that this model will better estimate bat abundance over the study area, compared to other spatial analysis, because it accounts not only for variations of detection probability over space and time, but also assumes spatial dependence between observations. Future enhancements to this approach should be to use species-specific distance-power relations to account for differences in how calls from different species influence distance estimates.

Keywords: Bat spatial distribution, spatial point pattern, distance sampling, acoustic data.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Spatio-temporal classification in point patterns under the presence of clutter

Marianna Siino[1], Francisco J. Rodríguez-Cortés[2], Jorge Mateu[3], Giada Adelfio[4]

[1]Istituto Nazionale di Geofisica e Vulcanologia, Centro Nazionale Terremoti, Rome, Italy
[2]Escuela de Estadística, Universidad Nacional de Colombia, Medellín, Colombia
[3]Departament of Mathematics, University Jaume I, Castellón, Spain
[4]Dipartimento di Scienze Economiche, Aziendali e Statistiche, Università degli Studi di Palermo, Palermo, Italy

Abstract

We consider the problem of detection of features in the presence of clutter for spatio-temporal point patterns. In the spatial context, it is used the K-th nearest-neighbour distances to classify points between clutter and features. They proposed a mixture of distributions whose parameters were estimated using an EM algorithm. This paper extends this methodology to the spatio-temporal context by considering the properties of the spatio-temporal K-th nearest-neighbour distances. For this purpose we make use of a couple of spatio-temporal distances which are based on the Euclidean and the maximum norms. We show close forms for the probability distributions of such K-th nearest-neighbour distances, and present an intensive simulation study together with an application to earthquakes.

Keywords: Clutter; Earthquakes; EM algorithm; Features; Mixtures; Nearest-neighbour distances; Spatio-temporal point patterns.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Spatial estimation methods for the abundance of *thornback ray*

Daniela Silva[1], Raquel Menezes[1], Ivone Figueiredo[3]

[1]Centro de Biologia Molecular e Ambiental, Dep. de Matemática, Universidade do Minho
[2]Instituto Português do Mar e da Atmosfera

Abstract

This study has been partially supported by research project "PREFERENTIAL, PTDC/MAT-STA/28243/2017". The main objective is to compare alternative spatial modeling methods, aiming to estimate the abundance of fish, in particular the *thornback ray* found along the Portuguese coast. The analyzed data concern the total number of specimens and their weight at each fishing haul, for IPMA demersal cruises from 2013 to 2016, in a total of 212 locations. Additionally, for 2015 and 2016, we also have an indicator abundance of *hake* and *small hake* caught at each fishing haul, collected in about half of these locations. The methods of spatial analysis to be considered are: Spatial Kernel smoothing; Kriging with External Drift; and a bayesian approach based on a Zero-Inflated model.

The preliminary analysis indicates that in 75% of locations there was no *thornback ray* and in 11.9% of locations there was no *small hake* and in about 7% there was no *hake* at all. Moreover, based on Knox and the Mantel tests, we conclude that there is no space-time interaction along the four years of study. We will then treat the different surveys as one unique survey to maximize the number of locations with non-zero data.

Whereas the abundance of *thornback ray* may depend on external factors, we investigate the importance of the variable depth of the fishing location and an indicator of *hake* abundance. Thus, these were considered as possible explanatory variables in the Kriging with External Drift method and in the Zero-Inflated model. We may note that for those locations where the covariate number of *hake* specimens per fishing haul of *hake* was missing, then these values were estimated by ordinary kriging. According to the results of Kriging with External Drift, if depth or the amount of *small hake* increase, both the abundance or biomass (species weight per fishing haul) indicator for *thornback ray* are expected to decrease. The results of the Zero-Inflated model allow us to conclude that for each unit increase in the depth variable, the mean amount and weight of *thornback ray* should decrease about 1%. Furthermore, number of specimens of *small hake* seems to explain the amount of *thornback ray*, not being statistically significant when considering the biomass indicator as response variable.

Summarizing, spatial modeling incorporates several interrelated methods that allow informed conclusions to be drawn, which support decision-making.

Keywords: Fishery Applications. Kriging Methods. Zero-Inflated Models.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Accessing Real Urgencies in Hospitals

Paula Simões[1,2,5], Isabel Natário[1,3], Sérgio Gomes[4]

[1]Centro de Matemática e Aplicações (CMA), Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal
[2]Departamento de Estatística e Investigação Operacional, Faculdade de Ciências, Universidade de Lisboa, Portugal
[3]Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal
[4]Direção Geral de Saúde, Portugal
[5]Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Portugal

Abstract

Hospital emergency departments are often misused by non-urgent patients, a major factor regarding hospital costs that ought be mitigated. This study performs a detailed health spatial econometrics analysis of the non-urgent emergency situations (classified by Manchester triage), by area, linking them with the efficient use of the Saude24 (S24) line, in terms of the S24 saving calls. A saving call is a call of a user that had as initial intention to go to an urgency department but that after calling the S24 line changed his/her mind.

The aim of this project is to develop an economic understanding of the advantages of the health line for the health system and to learn about the political and economic factors that influence health policies at global and regional levels. This enables to realize if these management policies helped to reduce the unnecessary per capita costs of the health care, using a savings index and its spatial effectiveness in solving the non-urgent emergency situations.

Given the spatial nature of the data and resorting to INLA methodology, the proportion of non-urgent cases in the Portuguese urgency hospital departments are modeled in an autoregressive way. The spatial structure is accounted for through a set of random effects. The model additionally includes regular covariates and the spatially lagged covariate savings index, related with the S24 saving calls. The response in one area depends not only on the (weighted) values of the response in its neighbourhood and of the considered covariates, but also on the (weighted) values of the covariate savings index measured in each neighbour, by means of a spatial Durbin model.

Keywords: Spatial Econometrics; Bayesian analysis; Autoregressive Models, Saude24 line, Hospital management.

Acknowledgments: This work is financed by national funds through FCT - Foundation for Science and Technology - under the projects UID/MAT/00297/2019.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Censored Linear Regression Model With AR(1) Error Term

Rodney Sousa[1], Isabel Pereira[1], Maria Eduarda Silva[2]

[1] CIDMA & University of Aveiro
[2] CIDMA & University of Porto

Abstract

Linear Regresion model (LR) with autocorrelated error term arises when we deal with time series data. In this case the model parameters can be estimated using Feasible Generalized Least Squares (FGLS) method. Censored data occur in studies where the values of the latent variable are accessible only in a restrictive interval.

For this data, FGLS method result in inconsistent estimates. In particular the bias increases with the proportion of censored values. In this work, we study three alternative methods to estimate LR model with AR(1) error term for censored data; namely, Approximate Maximum Likelihood (MVA), Approximate Bayesian Computation (ABC) and Gibbs with Data Augmentation (GDA). We perform two simulation studies in order to investigate the behaviour of the estimates, considering different censored proportions (5%, 20% and 40%) and different sample sizes (50, 100 and 500). The results show that the GDA method is consistent even when the proportion of censored values is high.

Keywords: Linear Regression, autocorrelation, Autoregressive Processes, Censored Data, Approximate Bayesian Computation, MCMC methods, Data Augmentation.

Statistical Analysis for Space-Time Data
July 15-17, 2019 – Lisboa, Portugal

Regional worker migration – stylized facts for Germany

Mark Trede[1], Michael Zimmermann[2]

[1]Westfälische Wilhelms-Universität Münster, Department of Economics, Am Stadtgraben 9 48143 Münster, mark.trede@uni-muenster.de
[2]Westfälische Wilhelms-Universität Münster, Department of Economics, Am Stadtgraben 9 48143 Münster, michael.zimmermann@wiwi.uni-muenster.de

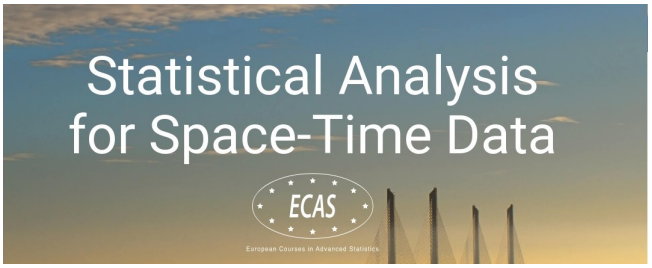
Abstract

In this paper, we study individual workers' biographies at the local level using the administrative German SIAB[1] dataset spanning from 1975 to 2014. In addition we utilize a new dataset about regional price levels. We show that there exists huge regional variation in unemployment rates, nominal as well as real wages, and housing costs. The distinction between urban and rural areas plays a substantial role in determining those economic factors. In particular, we show that the real wage gap between East and West Germany still persists while the unemployment rates tend to converge. Further, we consider monthly worker flows across 326 different regions (roughly equivalent to "Landkreise"). Compared to unemployed workers in prosperous regions, unemployed workers in depressed regions are less likely to a new working place in another region. Instead, most migration happens in the groups of unemployed workers in dense and active regions, rising over time. Employed workers are less willing to move and have procyclical fluctuations in their moving rates. We present the stylized facts about many aspects of the local German labour markets in a systematic way.

Keywords: labour mobility, business cycle fluctuations, regional disparities

[1]Our data set is an employment panel provided by the Institute for Employment Research. It contains a 2% random sample of integrated labour market biographies in Germany from 1975 to 2014.

Sociedade Portuguesa de Estatística



Partners and Sponsors

CEAUL
Centro de Estatística e Aplicações
Universidade de Lisboa

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

CIM
CENTRO INTERNACIONAL DE MATEMÁTICA

FCT
FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Omnidea

PREFERENTIAL: PTDC/MAT-STA/28243/2017