

Book of Short Papers SIS 2018

***Editors:* Antonino Abbruzzo - Eugenio Brentari**

Marcello Chiodi - Davide Piacentino



Copyright © 2018

PUBLISHED BY PEARSON

WWW.PEARSON.COM

First printing, November 2018, ISBN-9788891910233

Contents

| | | |
|------------|---|-----------|
| 1 | Preface | 17 |
| 2 | Plenary Sessions | 19 |
| 2.1 | A new paradigm for rating data models. <i>Domenico Piccolo</i> | 19 |
| 2.2 | Statistical challenges and opportunities in modelling coupled behaviour-disease dynamics of vaccine refusal. <i>Plenary/Chris T. Bauch</i> | 32 |
| 3 | Specialized Sessions | 45 |
| 3.1 | 3.1 - Bayesian Nonparametric Learning | 45 |
| 3.1.1 | Bayesian nonparametric covariate driven clustering. <i>Raffaele Argiento, Ilaria Bianchini, Alessandra Guglielmi and Ettore Lanzarone</i> | 46 |
| 3.1.2 | A Comparative overview of Bayesian nonparametric estimation of the size of a population. <i>Luca Tardella and Danilo Alunni Fegatelli</i> | 56 |
| 3.1.3 | Logit stick-breaking priors for partially exchangeable count data. <i>Tommaso Rigon</i> | 64 |
| 3.2 | BDsports - Statistics in Sports | 72 |
| 3.2.1 | A paired comparison model for the analysis of on-field variables in football matches. <i>Gunther Schaubberger and Andreas Groll</i> | 72 |
| 3.2.2 | Are the shots predictive for the football results?. <i>Leonardo Egidi, Francesco Paoli, Nicola Torelli</i> | 81 |
| 3.2.3 | Zero-inflated ordinal data models with application to sport (in)activity. <i>Maria Iannario and Rosaria Simone</i> | 89 |
| 3.3 | Being young and becoming adult in the Third Millennium: definition issues and processes analysis | 97 |
| 3.3.1 | Do Social Media Data predict changes in young adults' employment status? Evidence from Italy. <i>Andrea Bonanomi and Emiliano Sironi</i> | 97 |

| | | |
|-------------|--|------------|
| 3.3.2 | Parenthood: an advanced step in the transition to adulthood. <i>Cinzia Castagnaro, Antonella Guarneri and Eleonora Meli</i> | 106 |
| 3.4 | Economic Statistics and Big Data | 114 |
| 3.4.1 | Improvements in Italian CPI/HICP deriving from the use of scanner data. <i>Alessandro Brunetti, Stefania Fatello, Federico Polidoro, Antonella Simone</i> | 114 |
| 3.4.2 | Big data and spatial price comparisons of consumer prices. <i>Tiziana Laureti and Federico Polidoro</i> | 123 |
| 3.5 | Financial Time Series Analysis | 131 |
| 3.5.1 | Dynamic component models for forecasting trading volumes. <i>Antonio Naimoli and Giuseppe Storti</i> | 131 |
| 3.5.2 | Conditional Quantile-Located VaR. <i>Giovanni Bonaccolto, Massimiliano Caporin and Sandra Paterlini</i> | 140 |
| 3.6 | Forensic Statistics | 146 |
| 3.6.1 | Cause of effects: an important evaluation in Forensic Science. <i>Fabio Corradi and Monica Musio</i> | 146 |
| 3.6.2 | Evaluation and reporting of scientific evidence: the impact of partial probability assignments. <i>Silvia Bozza, Alex Biedermann, Franco Taroni</i> | 155 |
| 3.7 | Missing Data Handling in Complex Models | 161 |
| 3.7.1 | Dependence and sensitivity in regression models for longitudinal responses subject to dropout. <i>Marco Alfo' and Maria Francesca Marino</i> | 161 |
| 3.7.2 | Multilevel analysis of student ratings with missing level-two covariates: a comparison of imputation techniques. <i>Maria Francesca Marino e Carla Rampichini</i> | 170 |
| 3.7.3 | Multilevel Multiple Imputation in presence of interactions, non-linearities and random slopes. <i>Matteo Quartagno and James R. Carpenter</i> | 175 |
| 3.8 | Monitoring Education Systems. Insights from Large Scale Assessment Surveys | 183 |
| 3.8.1 | Educational Achievement of Immigrant Students. A Cross-National Comparison Over-Time Using OECD-PISA Data. <i>Mariano Porcu</i> | 183 |
| 3.9 | New Perspectives in Time Series Analysis | 192 |
| 3.9.1 | Generalized periodic autoregressive models for trend and seasonality varying time series. <i>Francesco Battaglia and Domenico Cucina and Manuel Rizzo</i> | 192 |
| 3.10 | Recent Advances in Model-based Clustering | 201 |
| 3.10.1 | Flexible clustering methods for high-dimensional data sets. <i>Cristina Tortora and Paul D. McNicholas</i> | 201 |
| 3.10.2 | A Comparison of Model-Based and Fuzzy Clustering Methods. <i>Marco Alfo', Maria Brigida Ferraro, Paolo Giordani, Luca Scrucca, and Alessio Serafini</i> | 208 |
| 3.10.3 | Covariate measurement error in generalized linear models for longitudinal data: a latent Markov approach. <i>Roberto Di Mari, Antonio Punzo, and Antonello Maruotti</i> | 216 |
| 3.11 | Statistical Modelling | 224 |
| 3.11.1 | A regularized estimation approach for the three-parameter logistic model. <i>Michela Battauz and Ruggero Bellio</i> | 224 |
| 3.11.2 | Statistical modelling and GAMLSS. <i>Mikis D. Stasinopoulos and Robert A. Rigby and Fernanda De Bastiani</i> | 233 |
| 3.12 | Young Contributions to Statistical Learning | 239 |
| 3.12.1 | Introducing spatio-temporal dependence in clustering: from a parametric to a nonparametric approach . <i>Clara Grazian, Gianluca Mastrantonio and Enrico Bibbona</i> | 239 |

| | | |
|------------|---|------------|
| 3.12.2 | Bayesian inference for hidden Markov models via duality and approximate filtering distributions. <i>Guillaume Kon Kam King, Omiros Papaspiliopoulos and Matteo Ruggiero</i> | 248 |
| 3.12.3 | K-means seeding via MUS algorithm. <i>Leonardo Egidi, Roberta Pappada', Francesco Pauli, Nicola Torelli</i> | 256 |
| 4 | Sollicited Sessions | 263 |
| 4.1 | Advances in Discrete Latent Variable Modelling | 263 |
| 4.1.1 | A joint model for longitudinal and survival data based on a continuous-time latent Markov model. <i>Alessio Farcomeni and Francesco Bartolucci</i> | 264 |
| 4.1.2 | Modelling the latent class structure of multiple Likert items: a paired comparison approach. <i>Brian Francis</i> | 273 |
| 4.1.3 | Dealing with reciprocity in dynamic stochastic block models. <i>Francesco Bartolucci, Maria Francesca Marino, Silvia Pandolfi</i> | 281 |
| 4.1.4 | Causality patterns of a marketing campaign conducted over time: evidence from the latent Markov model. <i>Fulvia Pennoni, Leo Paas and Francesco Bartolucci</i> | 289 |
| 4.2 | Complex Spatio-temporal Processes and Functional Data | 297 |
| 4.2.1 | Clustering of spatio-temporal data based on marked variograms. <i>Antonio Balzanella and Rosanna Verde</i> | 297 |
| 4.2.2 | Space-time earthquake clustering: nearest-neighbor and stochastic declustering methods in comparison. <i>Elisa Varini, Antonella Peresan, Renata Rotondi, and Stefania Gentili</i> | 304 |
| 4.2.3 | Advanced spatio-temporal point processes for the Sicily seismicity analysis. <i>Marianna Siino and Giada Adelfio</i> | 312 |
| 4.2.4 | Spatial analysis of the Italian seismic network and seismicity. <i>Antonino D'Alessandro, Marianna Siino, Luca Greco and Giada Adelfio</i> | 320 |
| 4.3 | Dimensional Reduction Techniques for Big Data Analysis | 328 |
| 4.3.1 | Clustering Data Streams via Functional Data Analysis: a Comparison between Hierarchical Clustering and K-means Approaches. <i>Fabrizio Maturo, Francesca Fortuna, and Tonio Di Battista</i> | 328 |
| 4.3.2 | Co-clustering algorithms for histogram data. <i>Francisco de A.T. De Carvalho and Antonio Balzanella and Antonio Irpino and Rosanna Verde</i> | 338 |
| 4.3.3 | A network approach to dimensionality reduction in Text Mining. <i>Michelangelo Misuraca, Germana Scepi and Maria Spano</i> | 344 |
| 4.3.4 | Self Organizing Maps for distributional data. <i>Rosanna Verde and Antonio Irpino</i> | 352 |
| 4.4 | Enviromental Processes, Human Activities and their Interactions | 353 |
| 4.4.1 | Estimation of coral growth parameters via Bayesian hierarchical non-linear models. <i>Crescenza Calculli, Barbara Cafarelli and Daniela Cocchi</i> | 353 |
| 4.4.2 | A Hierarchical Bayesian Spatio-Temporal Model to Estimate the Short-term Effects of Air Pollution on Human Health. <i>Fontanella Lara, Ippoliti Luigi and Valentini Pasquale</i> | 361 |
| 4.4.3 | A multilevel hidden Markov model for space-time cylindrical data. <i>Francesco Laguna and Monia Ranalli</i> | 367 |
| 4.4.4 | Estimation of entropy measures for categorical variables with spatial correlation. <i>Linda Altieri, Giulia Roli</i> | 373 |
| 4.5 | Innovations in Census and in Social Surveys | 381 |
| 4.5.1 | A micro-based approach to ensure consistency among administrative sources and to improve population statistics. <i>Gianni Corsetti, Sabrina Prati, Valeria Tomeo, Enrico Tucci</i> | 381 |
| 4.5.2 | Demographic changes, research questions and data needs: issues about migrations. <i>Salvatore Strozza and Giuseppe Gabrielli</i> | 392 |

| | | |
|-------------|--|------------|
| 4.5.3 | Towards more timely census statistics: the new Italian multiannual dissemination programme. <i>Simona Mastroluca and Mariangela Verrascina</i> | 400 |
| 4.6 | Living Conditions and Consumption Expenditure in Time of Crises | 409 |
| 4.6.1 | Household consumption expenditure and material deprivation in Italy during last economic crises. <i>Ilaria Arigoni and Isabella Sicilliani</i> | 409 |
| 4.7 | Network Data Analysis and Mining | 418 |
| 4.7.1 | Support provided by elderly Italian people: a multilevel analysis. <i>Elvira Pelle, Giulia Rivellini and Susanna Zaccarini</i> | 418 |
| 4.7.2 | Data mining and analysis of comorbidity networks from practitioner prescriptions. <i>Giancarlo Ragozini, Giuseppe Giordano, Sergio Pagano, Mario De Santis, Pierpaolo Cavallo</i> | 426 |
| 4.7.3 | Overlapping mixture models for network data (manet) with covariates adjustment. <i>Saverio Ranciati and Giuliano Galimberti and Ernst C. Wit and Veronica Vinciotti</i> | 434 |
| 4.8 | New Challenges in the Measurement of Economic Insecurity, Inequality and Poverty | 440 |
| 4.8.1 | Social protection in mitigating economic insecurity. <i>Alessandra Coli</i> | 440 |
| 4.8.2 | Changes in poverty concentration in U.S. urban areas. <i>Francesco Andreoli and Mauro Mussini</i> | 450 |
| 4.8.3 | Evaluating sustainability through an input-stateoutput framework: the case of the Italian provinces. <i>Achille Lemmi, Laura Neri, Federico M. Pulselli</i> | 458 |
| 4.9 | New Methods and Models for Ordinal Data | 466 |
| 4.9.1 | Weighted and unweighted distances based decision tree for ranking data. <i>Antonella Plaia, Simona Buscemi, Mariangela Sciandra</i> | 466 |
| 4.9.2 | A dissimilarity-based splitting criterion for CUBREMOT. <i>Carmela Cappelli, Rosaria Simone and Francesca Di Iorio</i> | 474 |
| 4.9.3 | Constrained Extended Plackett-Luce model for the analysis of preference rankings. <i>Cristina Mollica and Luca Tardella</i> | 480 |
| 4.9.4 | A prototype for the analysis of time use in Italy. <i>Stefania Capecchi and Manuela Michelini</i> | 487 |
| 4.10 | New Perspectives in Supervised and Unsupervised Classification | 493 |
| 4.10.1 | Robust Updating Classification Rule with applications in Food Authenticity Studies. <i>Andrea Cappozzo, Francesca Greselin and Thomas Brendan Murphy</i> | 493 |
| 4.10.2 | A robust clustering procedure with unknown number of clusters. <i>Francesco Dotto and Alessio Farcomeni</i> | 500 |
| 4.10.3 | Issues in joint dimension reduction and clustering methods. <i>Michel van de Velden, Alfonso Iodice D'Enza and Angelos Markos</i> | 508 |
| 4.11 | New Sources, Data Integration and Measurement Challenges for Estimates on Labour Market Dynamics | 514 |
| 4.11.1 | The development of the Italian Labour register: principles, issues and perspectives . <i>C. Baldi, C. Ceccarelli, S. Gigante, S. Pacini</i> | 514 |
| 4.11.2 | Digging into labour market dynamics: toward a reconciliation of stock and flows short term indicators. <i>F. Rapiti, C. Baldi, D. Ichim, F. Pintaldi, M. E. Pontecorvo, R. Rizzi</i> | 523 |
| 4.11.3 | How effective are the regional policies in Europe? The role of European Funds. <i>Gennaro Punzo, Mariateresa Ciommi, and Gaetano Musella</i> | 531 |
| 4.11.4 | Labour market condition in Italy during and after the financial crises: a segmented regression analysis approach of interrupted time series. <i>Lucio Masserini and Matilde Bini</i> | 539 |

| | | |
|-------------|---|------------|
| 4.12 | Quantile and Generalized Quantile Methods | 547 |
| 4.12.1 | Multiple quantile regression for risk assessment. <i>Lea Petrella and Valentina Raponi</i> | 547 |
| 4.12.2 | Parametric Modeling of Quantile Regression Coefficient Functions. <i>Paolo Frumento and Matteo Bottai</i> | 550 |
| 4.12.3 | Modelling the effect of Traffic and Meteorology on Air Pollution with Finite Mixtures of M-quantile Regression Models. <i>Simone Del Sarto, Maria Francesca Marino, Maria Giovanna Ranalli and Nicola Salvati</i> | 552 |
| 4.12.4 | Three-level M-quantile model for small area poverty mapping. <i>Stefano Marchetti and Nicola Salvati</i> | 560 |
| 4.13 | Recent Advances on Extreme Value Theory | 560 |
| 4.13.1 | Extremes of high-order IGARCH processes. <i>Fabrizio Laurini</i> | 560 |
| 4.14 | Spatial Economic Data Analysis | 569 |
| 4.14.1 | Spatial heterogeneity in principal component analysis: a study of deprivation index on Italian provinces. <i>Paolo Postiglione, M. Simona Andreano, Roberto Benedetti, Alfredo Cartone</i> | 569 |
| 4.15 | Spatial Functional Data Analysis | 578 |
| 4.15.1 | Object oriented spatial statistics for georeferenced tensor data. <i>Alessandra Menafoglio and Davide Pigoli and Piercesare Secchi</i> | 578 |
| 4.15.2 | A Spatio-Temporal Mixture Model for Urban Crimes. <i>Ferretti Angela, Ippoliti Luigi and Valentini Pasquale</i> | 585 |
| 4.16 | Statistical Methods for Service Quality | 591 |
| 4.16.1 | Cumulative chi-squared statistics for the service quality improvement: new properties and tools for the evaluation. <i>Antonello D'Ambra, Antonio Lucadamo, Pietro Amenta, Luigi D'Ambra</i> | 591 |
| 4.16.2 | A robust multinomial logit model for evaluating judges' performances. <i>Ida Camminatiello and Antonio Lucadamo</i> | 600 |
| 4.16.3 | Complex Contingency Tables and Partitioning of Three-way Association Indices for Assessing Justice CourtWorkload. <i>Rosaria Lombardo, Yoshio Takane and Eric J Beh</i> | 607 |
| 4.16.4 | Finding the best paths in university curricula of graduates to improve academic guidance services. <i>Silvia Bacci and Bruno Bertaccini</i> | 615 |
| 4.17 | Statistical Modelling for Business Intelligence Problems | 623 |
| 4.17.1 | A nonlinear state-space model for the forecasting of field failures. <i>Antonio Pievatolo</i> | 623 |
| 4.17.2 | Does Airbnb affect the real estate market? A spatial dependence analysis. <i>Mariangela Guidolin and Mauro Bernardi</i> | 632 |
| 4.17.3 | Bayesian Quantile Trees for Sales Management. <i>Mauro Bernardi and Paola Stolfi</i> | 640 |
| 4.17.4 | Discrimination in machine learning algorithms. <i>Roberta Pappadá and Francesco Pauli</i> | 648 |
| 4.18 | Statistical models for sports data | 656 |
| 4.18.1 | The study of relationship between financial performance and points achieved by Italian football championship clubs via GEE and diagnostic measures. <i>Anna Crisci, Sarnacchiaro Pasquale e Luigi D'Ambra</i> | 656 |
| 4.18.2 | Exploring the Kaggle European Soccer database with Bayesian Networks: the case of the Italian League Serie A. <i>Maurizio Carpita and Silvia Golia</i> | 665 |
| 4.18.3 | A data-mining approach to the Parkour discipline. <i>Paola Pasca, Enrico Ciavolino and Ryan L. Boyd</i> | 673 |
| 4.18.4 | Players Movements and Team Shooting Performance: a Data Mining approach for Basketball. <i>Rodolfo Metulini</i> | 681 |

| | | |
|-------------|---|------------|
| 4.19 | Supporting Regional Policies through Small Area Statistical Methods | 689 |
| 4.19.1 | Survey-weighted Unit-Level Small Area Estimation. <i>Jan Pablo Burgard and Patricia Dörr</i> | 689 |
| 4.20 | The Second Generation at School | 689 |
| 4.20.1 | Resilient students with migratory background. <i>Anna Di Bartolomeo and Giuseppe Gabrielli</i> | 689 |
| 4.20.2 | Residential Proximity to Attended Schools among Immigrant-Origin Youths in Bologna. <i>Federica Santangelo, Debora Mantovani and Giancarlo Gasperoni</i> | 698 |
| 4.20.3 | From school to ... future: strategies, paths and perspectives of immigrant immediate descendants in Naples . <i>Giustina Orientale Caputo and Giuseppe Gargiulo</i> | 706 |
| 4.21 | Tourism Destinations, Household, Firms | 714 |
| 4.21.1 | The Pricing Behaviour of Firms in the On-line Accommodation Market: Evidence from a Metropolitan City. <i>Andrea Guizzardi and Flavio Maria Emanuele Pons</i> | 714 |
| 4.21.2 | The Migration-Led-Tourism Hypothesis for Italy: A Survey. <i>Carla Massidda, Romano Piras and Ivan Etzo</i> | 724 |
| 4.21.3 | Tourism Statistics: development and potential uses. <i>Fabrizio Antolini</i> | 732 |
| 4.21.4 | Tourism attractiveness in Italy. Some empirical evidence comparing origin-destination domestic tourism flows. <i>Francesca Giambona, Emanuela Dreassi, and Alessandro Magrini</i> | 740 |
| 4.22 | What's Happening in Africa | 748 |
| 4.22.1 | Environmental shocks and internal migration in Tanzania. <i>Maria Francesca Marino, Alessandra Petrucci, and Elena Pirani</i> | 748 |
| 4.22.2 | Determinants and geographical disparities of BMI in African Countries: a measurement error small area approach. <i>Serena Arima and Silvia Polettini</i> | 756 |
| 5 | Contributed Sessions | 765 |
| 5.1 | Advanced Algorithms and Computation | 765 |
| 5.1.1 | Brexit in Italy. <i>Francesca Greco, Livia Celardo, Leonardo Salvatore Alaimo</i> | 766 |
| 5.1.2 | Distance based Depth-Depth classifier for directional data. <i>Giuseppe Pandolfo and Giovanni C. Porzio</i> | 773 |
| 5.1.3 | Approximate Bayesian Computation for Forecasting in Hydrological models. <i>Jonathan Romero-Cuéllar, Antonino Abbruzzo, Giada Adelfio and Félix Francés</i> | 777 |
| 5.1.4 | Customer Churn prediction based on eXtreme Gradient Boosting classifier. <i>Mohammed Hassan Elbedawi Omar and Matteo Borrotti</i> | 783 |
| 5.1.5 | HPC-accelerated Approximate Bayesian Computation for Biological Science. <i>Rita-brata Dutta</i> | 789 |
| 5.1.6 | PC Algorithm for Gaussian Copula Data. <i>Vincenzina Vitale and Paola Vicard</i> | 797 |
| 5.2 | Advances in Clustering Techniques | 803 |
| 5.2.1 | On the choice of an appropriate bandwidth for modal clustering. <i>Alessandro Casa, José E. Chacón and Giovanna Menardi</i> | 803 |
| 5.2.2 | Unsupervised clustering of Italian schools via non-parametric multilevel models. <i>Chiara Masci, Francesca Ieva and Anna Maria Paganoni</i> | 810 |
| 5.2.3 | Chiara Masci, Francesca Ieva and Anna Maria Paganoni. <i>Laura Bocci and Donatella Vicari</i> | 816 |
| 5.2.4 | Robust Reduced k-Means and Factorial k-Means by trimming. <i>Luca Greco and Antonio Lucadamo and Pietro Amenta</i> | 821 |
| 5.2.5 | Dirichlet processes, posterior similarity and graph clustering. <i>Stefano Tonellato</i> | 827 |
| 5.2.6 | Bootstrap ClustGeo with spatial constraints. <i>Veronica Distefano, Valentina Mameli, Fabio Della Marra</i> | 833 |

| | | |
|------------|--|------------|
| 5.3 | Advances in Statistical Models | 839 |
| 5.3.1 | Regression modeling via latent predictors. <i>Francesca Martella and Donatella Vicari</i> | 839 |
| 5.3.2 | Analysis of dropout in engineering BSc using logistic mixed-effect models. <i>Luca Fontana and Anna Maria Paganoni</i> | 846 |
| 5.3.3 | dgLARS method for relative risk regression models. <i>Luigi Augugliaro and Angelo M. Mineo</i> | 852 |
| 5.3.4 | A Latent Class Conjoint Analysis for analysing graduates profiles. <i>Paolo Mariani, Andrea Marletta, Lucio Masserini and Mariangela Zenga</i> | 858 |
| 5.3.5 | A longitudinal analysis of the degree of accomplishment of anti-corruption measures by Italian municipalities: a latent Markov approach. <i>Simone Del Sarto, Michela Gnaldi, Francesco Bartolucci</i> | 864 |
| 5.3.6 | Modelling the effect of covariates for unbiased estimates in ecological inference methods. <i>Venera Tomaselli, Antonio Forcina and Michela Gnaldi</i> | 870 |
| 5.4 | Advances in Time Series | 876 |
| 5.4.1 | Filtering outliers in time series of electricity prices. <i>Ilaria Lucrezia Amerise</i> . . . | 876 |
| 5.4.2 | Time-varying long-memory processes. <i>Luisa Bisaglia and Matteo Grigoletto</i> | 883 |
| 5.4.3 | Statistical Analysis of Markov Switching DSGE Models. <i>Maddalena Cavicchioli</i> | 889 |
| 5.4.4 | Forecasting energy price volatilities and comovements with fractionally integrated MGARCH models. <i>Malvina Marchese and Francesca Di Iorio</i> | 894 |
| 5.4.5 | Improved bootstrap simultaneous prediction limits. <i>Paolo Vidoni</i> | 900 |
| 5.5 | Data Management | 906 |
| 5.5.1 | Using web scraping techniques to derive co-authorship data: insights from a case study. <i>Domenico De Stefano, Vittorio Fuccella, Maria Prosperina Vitale, Susanna Zaccarin</i> | 906 |
| 5.5.2 | Dealing with Data Evolution and Data Integration: An approach using Rarefaction. <i>Luca Del Core, Eugenio Montini, Clelia Di Serio, Andrea Calabria</i> | 913 |
| 5.5.3 | Monitoring event attendance using a combination of traditional and advanced surveying tools. <i>Mauro Ferrante, Amit Birenboim, Anna Maria Milito, Stefano De Cantis</i> | 919 |
| 5.5.4 | Indefinite Topological Kernels. <i>Tullia Padellini and Pierpaolo Brutti</i> | 925 |
| 5.5.5 | Data Integration in Social Sciences: the earnings intergenerational mobility problem. <i>Veronica Ballerini, Francesco Bloise, Dario Briscolini and Michele Raitano</i> | 931 |
| 5.5.6 | An innovative approach for the GDPR compliance in Big DAta era. <i>M. Giacalone, C. Cusatelli, F. Fanari, V. Santarcangelo, D.C. Sinitó</i> | 937 |
| 5.6 | Developments in Graphical Models | 943 |
| 5.6.1 | An extension of the glasso estimator to multivariate censored data. <i>Antonino Abbruzzo and Luigi Augugliaro and Angelo M. Mineo</i> | 943 |
| 5.6.2 | Bayesian Estimation of Graphical Log-Linear Marginal Models. <i>Claudia Tarantola, Ioannis Ntzoufras and Monia Lupparelli</i> | 950 |
| 5.6.3 | Statistical matching by Bayesian Networks. <i>Daniela Marella and Paola Vicard and Vincenzina Vitale</i> | 956 |
| 5.6.4 | Sparse Nonparametric Dynamic Graphical Models. <i>Fabrizio Poggioni, Mauro Bernardi, Lea Petrella</i> | 962 |
| 5.6.5 | Non-communicable diseases, socio-economic status, lifestyle and well-being in Italy: An additive Bayesian network model. <i>Laura Maniscalco and Domenica Matranga</i> | 968 |
| 5.6.6 | Using Almost-Dynamic Bayesian Networks to Represent Uncertainty in Complex Epidemiological Models: a Proposal. <i>Sabina Marchetti</i> | 974 |

| | | |
|-------------|---|-------------|
| 5.7 | Educational World | 980 |
| 5.7.1 | How to improve the Quality Assurance System of the Universities: a study based on compositional analysis . <i>Bertaccini B., Gallo M., Simonacci V., and Menini T.</i> . | 980 |
| 5.7.2 | Evaluation of students' performance at lower secondary education. An empirical analysis using TIMSS and PISA data.. <i>G. Graziosi, T. Agasisti, K. De Witte and F. Pauli</i> | 985 |
| 5.7.3 | Testing for the Presence of Scale Drift: An Example. <i>Michela Battauz</i> | 991 |
| 5.7.4 | The evaluation of Formative Tutoring at the University of Padova. <i>Renata Clerici, Lorenza Da Re, Anna Giraldo, Silvia Meggiolaro</i> | 996 |
| 5.7.5 | Benefits of the Erasmus mobility experience: a discrete latent variable analysis. <i>Silvia Bacci, Valeria Caviezel and Anna Maria Falzoni</i> | 1001 |
| 5.7.6 | University choice and the attractiveness of the study area. Insights from an analysis based on generalized mixed-effect models. <i>Silvia Columbu, Mariano Porcu and Isabella Sulis</i> | 1007 |
| 5.8 | Environment | 1013 |
| 5.8.1 | The climate funds for energy sustainability: a counterfactual analysis. <i>Alfonso Carfora and Giuseppe Scandurra</i> | 1013 |
| 5.8.2 | Exploratory GIS Analysis via Spatially Weighted Regression Trees. <i>Carmela Iorio, Giuseppe Pandolfo, Michele Staiano, and Roberta Siciliano</i> | 1020 |
| 5.8.3 | A functional regression control chart for profile monitoring. <i>Fabio Centofanti, Antonio Lepore, Alessandra Menafoglio, Biagio Palumbo and Simone Vantini</i> | 1026 |
| 5.8.4 | Understanding pro-environmental travel behaviours in Western Europe. <i>Gennaro Punzo, Rosalia Castellano, and Demetrio Panarello</i> | 1031 |
| 5.9 | Family & Economic issues | 1037 |
| 5.9.1 | Measuring Economic Uncertainty: Longitudinal Evidence Using a Latent Transition Model. <i>Francesca Giambona, Laura Grassini and Daniele Vignoli</i> | 1037 |
| 5.9.2 | Intentions to leave Italy or to stay among foreigners: some determinants of migration projects. <i>Ginevra Di Giorgio, Francesca Dota, Paola Muccitelli and Daniele Spizzichino</i> | 1044 |
| 5.9.3 | Wages differentials in association with individuals, enterprises and territorial characteristics. <i>S. De Santis, C. Freguja, A. Masi, N. Pannuzi, F. G. Truglia</i> | 1050 |
| 5.9.4 | The Transition to Motherhood among British Young Women: Does housing tenure play a role?. <i>Valentina Tocchioni, Ann Berrington, Daniele Vignoli and Agnese Vitali</i> | 1056 |
| 5.10 | Finance & Insurance | 1062 |
| 5.10.1 | Robust statistical methods for credit risk. <i>A. Corbellini, A. Ghiretti, G. Morelli and A. Talignani</i> | 1062 |
| 5.10.2 | Depth-based portfolio selection. <i>Giuseppe Pandolfo, Carmela Iorio and Antonio D'Ambrosio</i> | 1069 |
| 5.10.3 | Estimating large-scale multivariate local level models with application to stochastic volatility. <i>Matteo Pelagatti and Giacomo Sbrana</i> | 1075 |
| 5.11 | Health and Clinical Data | 1081 |
| 5.11.1 | Is retirement bad for health? A matching approach. <i>Elena Pirani, Marina Ballerini, Alessandra Mattei, Gustavo De Santis</i> | 1081 |
| 5.11.2 | The emergency department utilisation among the immigrant population resident in Rome from 2005 to 2015. <i>Eleonora Trappolini, Laura Cacciani, Claudia Marino, Cristina Giudici, Nera Agabiti, Marina Davoli</i> | 1088 |
| 5.11.3 | Multi-State model with nonparametric discrete frailty. <i>Francesca Gasperoni, Francesca Ieva, Anna Maria Paganoni, Chris Jackson and Linda Sharples</i> | 1095 |
| 5.11.4 | A Functional Urn Model for CARA Designs. <i>Giacomo Aleffi, Andrea Ghiglietti, and William F. Rosenberger</i> | 1101 |

| | | |
|-------------|--|-------------|
| 5.11.5 | Assessment of the INLA approach on gerarchic bayesian models for the spatial disease distribution: a real data application. <i>Paolo Girardi, Emanuela Bovo, Carmen Stocco, Susanna Baracco, Alberto Rosano, Daniele Monetti, Silvia Rizzato, Sara Zamberlan, Enrico Chinellato, Ugo Fedeli, Massimo Rugge</i> | 1107 |
| 5.12 | Medicine | 1113 |
| 5.12.1 | Hidden Markov Models for disease progression. <i>Andrea Martino, Andrea Ghiglietti, Giuseppina Guatteri, Anna Maria Paganoni</i> | 1113 |
| 5.12.2 | A simulation study on the use of response-adaptive randomized designs. <i>Anna Maria Paganoni, Andrea Ghiglietti, Maria Giovanna Scarale, Rosalba Miceli, Francesca Ieva, Luigi Mariani, Cecilia Gavazzi and Valeria Edefonti</i> | 1120 |
| 5.12.3 | The relationship between health care expenditures and time to death: focus on myocardial infarction patients. <i>Luca Grasseti and Laura Rizzi</i> | 1126 |
| 5.12.4 | A multivariate extension of the joint models. <i>Marcella Mazzoleni and Mariangela Zenga</i> | 1132 |
| 5.12.5 | Multipurpose optimal designs for hypothesis testing in normal response trials. <i>Marco Novelli and Maroussa Zagoraiou</i> | 1138 |
| 5.12.6 | Additive Bayesian networks for an epidemiological analysis of swine diseases. <i>Marta Pittavino and Reinhard Furrer</i> | 1144 |
| 5.13 | Population Dynamics | 1150 |
| 5.13.1 | Employment Uncertainty and Fertility: a Meta-Analysis of European Research Findings. <i>Giammarco Alderotti, Daniele Vignoli and Michela Baccini</i> | 1150 |
| 5.13.2 | What Shapes Population Age Structures in the Long Run. <i>Gustavo De Santis and Giambattista Salinari</i> | 1156 |
| 5.13.3 | The impact of economic development on fertility: a complexity approach in a cross-country analysis. <i>NiccolóInnocenti, Daniele Vignoli and Luciana Lazzeretti</i> | 1162 |
| 5.13.4 | A Probabilistic Cohort-Component Model for Population Fore-casting - The Case of Germany. <i>Patrizio Vanella and Philipp Deschermeier</i> | 1167 |
| 5.13.5 | Mortality trends in Sardinia 1992-2015: an ecological study. <i>Vanessa Santos Sanchez, Gabriele Ruiu Marco Breschi, Lucia Pozzi</i> | 1171 |
| 5.14 | Recent Developments in Bayesian Inference | 1177 |
| 5.14.1 | Posterior distributions with non explicit objective priors. <i>Erlis Ruli, Nicola Sartori and Laura Ventura</i> | 1177 |
| 5.14.2 | A predictive measure of the additional loss of a non-optimal action under multiple priors. <i>Fulvio De Santis and Stefania Gubbiotti</i> | 1184 |
| 5.14.3 | Bayesian estimation of number and position of knots in regression splines. <i>Gioia Di Credico, Francesco Pauli and Nicola Torelli</i> | 1190 |
| 5.14.4 | The importance of historical linkages in shaping population density across space. <i>Ilenia Epifani and Rosella Nicolini</i> | 1196 |
| 5.15 | Recent Developments in Sampling | 1202 |
| 5.15.1 | Species richness estimation exploiting purposive lists: A proposal. <i>A. Chiarucci, R.M. Di Biase, L. Fattorini, M. Marcheselli and C. Pisani</i> | 1202 |
| 5.15.2 | Design-based exploitation of big data by a doubly calibrated estimator. <i>Maria Michela Dickson, Giuseppe Espa and Lorenzo Fattorini</i> | 1209 |
| 5.15.3 | Design-based mapping in environmental surveys. <i>L. Fattorini, M. Marcheselli and C. Pisani</i> | 1215 |
| 5.15.4 | Testing for independence in analytic inference. <i>Pier Luigi Conti and Alberto Di Iorio</i> | 1221 |
| 5.15.5 | On the aberrations of two-level Orthogonal Arrays with removed runs. <i>Roberto Fontana and Fabio Rapallo</i> | 1227 |

| | | |
|-------------|---|-------------|
| 5.16 | Recent Developments in Statistical Modelling | 1233 |
| 5.16.1 | Quantile Regression Coefficients Modeling: a Penalized Approach. <i>Gianluca Sottile, Paolo Frumento and Matteo Bottai</i> | 1233 |
| 5.16.2 | Simultaneous calibrated prediction intervals for time series. <i>Giovanni Fonseca, Federica Giummolé and Paolo Vidoni</i> | 1240 |
| 5.16.3 | Reversibility and (non)linearity in time series. <i>Luisa Bisaglia and Margherita Gerolimetto</i> | 1246 |
| 5.16.4 | Heterogeneous effects of subsidies on farms' performance: a spatial quantile regression analysis. <i>Marusca De Castris and Daniele Di Gennaro</i> | 1252 |
| 5.16.5 | On the estimation of high-dimensional regression models with binary covariates. <i>Valentina Mameli, Debora Slanzi and Irene Poli</i> | 1259 |
| 5.17 | Social Indicators | 1265 |
| 5.17.1 | Can a neighbour region influence poverty? A fuzzy and longitudinal approach. <i>Gianni Betti, Federico Crescenzi and Francesca Gagliardi</i> | 1265 |
| 5.17.2 | Weight-based discrimination in the Italian Labor Market: how do ethnicity and gender interact? <i>Giovanni Busetta, Maria Gabriella Campolo, and Demetrio Panarello</i> | 1272 |
| 5.17.3 | The Total Factor Productivity Index as a Ratio of Price Indexes. <i>Lisa Crosato and Biancamaria Zavanella</i> | 1278 |
| 5.17.4 | Monetary poverty indicators at local level: evaluating the impact of different poverty thresholds. <i>Luigi Biggeri, Caterina Giusti and Stefano Marchetti</i> | 1284 |
| 5.17.5 | A gender inequality assessment by means of the Gini index decomposition. <i>Michele Costa</i> | 1290 |
| 5.18 | Socio-Economic Statistics | 1296 |
| 5.18.1 | The NEETs during the economic crisis in Italy, Young NEETs in Italy, Spain and Greece during the economic crisis. <i>Giovanni De Luca, Paolo Mazzocchi, Claudio Quintano, Antonella Rocca</i> | 1296 |
| 5.18.2 | Camel or dromedary? A study of the equilibrium distribution of income in the EU countries. <i>Crosato L., Ferretti C., Ganugi P.</i> | 1303 |
| 5.18.3 | Small Area Estimation of Inequality Measures. <i>Maria Rosaria Ferrante and Silvia Pacei</i> | 1309 |
| 5.18.4 | Testing the Learning-by-Exporting at Micro-Level in light of influence of "Statistical Issues" and Macroeconomic Factors. <i>Maria Rosaria Ferrante and Marzia Freo</i> | 1314 |
| 5.18.5 | The mobility and the job success of the Sicilian graduates <i>Ornella Giambalvo and Antonella Plaia and Sara Binassi</i> | 1320 |
| 5.19 | Statistical Analysis of Energy Markets | 1326 |
| 5.19.1 | Forecasting Value-at-Risk for Model Risk Analysis in Energy Markets. <i>Angelica Gianfreda and Giacomo Scandolo</i> | 1326 |
| 5.19.2 | Prediction interval of electricity prices by robust nonlinear models. <i>Lisa Crosato, Luigi Grossi and Fany Nan</i> | 1333 |
| 5.19.3 | Bias Reduction in a Matching Estimation of Treatment Effect. <i>Maria Gabriella Campolo, Antonino Di Pino and Edoardo Otranto</i> | 1338 |
| 5.20 | Statistical Inference and Testing Procedures | 1344 |
| 5.20.1 | Comparison of exact and approximate simultaneous confidence regions in nonlinear regression models. <i>Claudia Furlan and Cinzia Mortarino</i> | 1344 |
| 5.20.2 | Tail analysis of a distribution by means of an inequality curve. <i>E. Taufer, F. Santi, G. Espa and M. M. Dickson</i> | 1351 |
| 5.20.3 | Nonparametric penalized likelihood for density estimation. <i>Federico Ferraccioli, Laura M. Sangalli and Livio Finos</i> | 1357 |
| 5.20.4 | Rethinking the Kolmogorov-Smirnov Test of Goodness of Fit in a Compositional Way. <i>G.S. Monti, G. Mateu-Figueras, M. I. Ortego, V. Pawlowsky-Glahn and J. J. Egozcue</i> | 1363 |

| | | |
|-------------|---|-------------|
| 5.20.5 | Stochastic Dominance for Generalized Parametric Families. <i>Tommaso Lando and Lucio Bertoli-Barsotti</i> | 1369 |
| 5.21 | Statistical Models for Ordinal Data | 1374 |
| 5.21.1 | A comparative study of benchmarking procedures for interrater and intrarater agreement studies. <i>Amalia Vanacore and Maria Sole Pellegrino</i> | 1374 |
| 5.21.2 | Measuring the multiple facets of tolerance using survey data. <i>Caterina Liberati and Riccarda Longaretti and Alessandra Michelangeli</i> | 1381 |
| 5.21.3 | Modified profile likelihood in models for clustered data with missing values. <i>Claudia Di Caterina and Nicola Sartori</i> | 1385 |
| 5.21.4 | Worthiness Based Social Scaling. <i>Giulio D'Epifanio</i> | 1391 |
| 5.21.5 | Direct Individual Differences Scaling for Evaluation of Research Quality. <i>Gallo M., Trendafilov N., and Simonacci V.</i> | 1396 |
| 5.21.6 | A test for variable importance. <i>Rosaria Simone</i> | 1400 |
| 5.22 | Statistical Models New Proposals | 1406 |
| 5.22.1 | Decomposing Large Networks: An Approach Based on the MCA based Community Detection. <i>Carlo Drago</i> | 1406 |
| 5.22.2 | On Bayesian high-dimensional regression with binary predictors: a simulation study. <i>Debora Slanzi, Valentina Mameli and Irene Poli</i> | 1413 |
| 5.22.3 | On the estimation of epidemiological parameters from serological survey data using Bayesian mixture modelling. <i>Emanuele Del Fava, Piero Manfredi, and Ziv Shkedy</i> | 1419 |
| 5.22.4 | An evaluation of KL-optimum designs to discriminate between rival copula models. <i>Laura Deldossi, Silvia Angela Osmetti, Chiara Tommasi</i> | 1425 |
| 5.22.5 | Variational Approximations for Frequentist and Bayesian Inference. <i>Luca Maestrini and Matt P. Wand</i> | 1431 |
| 5.22.6 | Node-specific effects in latent space modelling of multidimensional networks. <i>Silvia D'Angelo and Marco Alfó and Thomas Brendan Murphy</i> | 1437 |
| 5.23 | Statistics for Consumer Research | 1443 |
| 5.23.1 | A panel data analysis of Italian hotels. <i>Antonio Giusti, Laura Grassini, Alessandro Viviani</i> | 1443 |
| 5.23.2 | A Bayesian Mixed Multinomial Logit Model for Partially Microsimulated Data on Labor Supply. <i>Cinzia Carota and Consuelo R. Nava</i> | 1450 |
| 5.23.3 | Comparison between Experience-based Food Insecurity scales. <i>Federica Onori, Sara Viviani and Pierpaolo Brutti</i> | 1456 |
| 5.23.4 | Sovereign co-risk measures in the Euro Area. <i>Giuseppe Arbia, Riccardo Bramante, Silvia Facchinetti, Diego Zappa</i> | 1462 |
| 5.23.5 | Simultaneous unsupervised and supervised classification modeling for clustering, model selection and dimensionality reduction. <i>Mario Fordellone and Maurizio Vichi</i> | 1468 |
| 5.23.6 | Consumers' preference for coffee consumption: a choice experiment including organoleptic characteristics and chemical analysis <i>Rossella Berni, Nedka D. Niki-forova and Patrizia Pinelli</i> | 1475 |
| 5.24 | Statistics for Earthquakes | 1482 |
| 5.24.1 | How robust is the skill score of probabilistic earthquake forecasts? <i>Alessia Caponera and Maximilian J. Werner</i> | 1482 |
| 5.24.2 | Functional linear models for the analysis of similarity of waveforms. <i>Francesca Di Salvo, Renata Rotondi and Giovanni Lanzano</i> | 1489 |
| 5.24.3 | Detection of damage in civil engineering structure by PCA on environmental vibration data. <i>G. Agró, V. Carlisi, R. Mantione</i> | 1495 |

| | | |
|-------------|---|-------------|
| 5.25 | Statistics for Financial Risks | 1501 |
| 5.25.1 | Conditional Value-at-Risk: a comparison between quantile regression and copula functions. <i>Giovanni De Luca and Giorgia Riveccio</i> | 1501 |
| 5.25.2 | Systemic events and diffusion of jumps. <i>Giovanni Bonaccolto, Nancy Zambon and Massimiliano Caporin</i> | 1507 |
| 5.25.3 | Traffic Lights for Systemic Risk Detectio. <i>Massimiliano Caporin, Laura Garcia-Jorcano, Juan-Angel Jiménez-Martin</i> | 1513 |
| 5.25.4 | Bayesian Quantile Regression Treed. <i>Mauro Bernardi and Paola Stolfi</i> | 1520 |
| 5.25.5 | Model Selection in Weighted Stochastic Block models. <i>Roberto Casarin, Michele Costola, Erdem Yenerdag</i> | 1525 |
| 5.26 | Tourism & Cultural Participation | 1529 |
| 5.26.1 | The determinants of tourism destination competitiveness in 2006-2016: a partial least square path modelling approach. <i>Alessandro Magrini, Laura Grassini</i> | 1529 |
| 5.26.2 | Participation in tourism of Italian residents in the years of the economic recession. <i>Chiara Bocci, Laura Grassini, Emilia Rocco</i> | 1536 |
| 5.26.3 | Cultural Participation in the digital Age in Europe: a multilevel cross-national analysis. <i>Laura Bocci and Isabella Mingo</i> | 1542 |
| 5.26.4 | Tourist flows and museum admissions in Italy: an integrated analysis. <i>Lorenzo Cavallo, Francesca Petrei, Maria Teresa Santoro</i> | 1549 |
| 5.26.5 | Posterior Predictive Assessment for Item Response Theory Models: A Proposal Based on the Hellinger Distance. <i>Mariagiulia Matteucci and Stefania Mignani</i> | 1555 |
| 5.27 | Well-being & Quality of Life | 1561 |
| 5.27.1 | Is Structural Equation Modelling Able to Predict Well-being? <i>Daniele Toninelli and Michela Cameletti</i> | 1561 |
| 5.27.2 | The well-being in the Italian urban areas: a local geographic variation analysis. <i>Eugenia Nissi and Annalina Sarra</i> | 1568 |
| 5.27.3 | Comparing Composite Indicators to measure Quality of Life: the Italian "Sole 24 Ore" case. <i>Gianna Agró, Mariantonietta Ruggieri and Erasmo Vassallo</i> | 1574 |
| 5.27.4 | Quality of working life in Italy: findings from Inapp survey. <i>Paolo Emilio Cardone</i> | 1580 |
| 5.27.5 | Well-being indices: what about Italian scenario? <i>Silvia Facchinetti and Elena Siletti</i> | 1587 |
| 5.27.6 | How can we compare rankings that are expected to be similar? An example based on composite well being indicators. <i>Silvia Terzi e Luca Moroni</i> | 1593 |
| 6 | Poster Sessions | 1601 |
| 6.0.1 | A distribution curves comparison approach to analyze the university moving students performance. <i>Giovanni Boscaino, Giada Adelfio, Gianluca Sottile</i> | 1601 |
| 6.0.2 | A Partial Ordering Application in Aggregating Dimensions of Subjective Well-being. <i>Paola Conigliaro</i> | 1608 |
| 6.0.3 | A note on objective Bayes analysis for graphical vector autoregressive models. <i>Lucia Paci and Guido Consonni</i> | 1614 |
| 6.0.4 | Bayesian Population Size Estimation with A Single Sample. <i>Pierfrancesco Alaimo Di Loro and Luca Tardella</i> | 1620 |
| 6.0.5 | Classification of the Aneurisk65 dataset using PCA for partially observed functional data. <i>Marco Stefanucci, Laura Sangalli and Pierpaolo Brutti</i> | 1626 |
| 6.0.6 | Deep Learning to the Test: an Application to Traffic Data Streams. <i>Nina Deliu and Pierpaolo Brutti</i> | 1631 |
| 6.0.7 | Estimating the number of unseen species under heavy tails. <i>Marco Battiston, Federico Camerlenghi, Emanuele Dolera and Stefano Favaro</i> | 1637 |
| 6.0.8 | How to measure cybersecurity risk. <i>Silvia Facchinetti, Paolo Giudici and Silvia Angela Osmetti</i> | 1643 |

| | | |
|--------|---|------|
| 6.0.9 | Implementation of an innovative technique to improve Sauvignon Blanc wine quality. <i>Filippa Bono, Pietro Catanaia and Mariangela Vallone</i> | 1647 |
| 6.0.10 | Investigating the effect of drugs consumption on survival outcome of Heart Failure patients using joint models: a case study based on regional administrative data. <i>Marta Spreafico, Francesca Gasperoni, Francesca Ieva</i> | 1653 |
| 6.0.11 | Mapping the relation between University access test and student's university performance. <i>Vincenzo Giuseppe Genova, Antonella Plaia</i> | 1659 |
| 6.0.12 | Multivariate analysis of marine litter abundance through Bayesian space-time models. <i>C. Calculli, A. Pollice, L. Sion, and P. Maiorano</i> | 1665 |
| 6.0.13 | Power Priors for Bayesian Analysis of Graphical Models of Conditional Independence in Three Way Contingency Tables. <i>Katerina Mantzouni, Claudia Tarantola and Ioannis Ntzoufras</i> | 1669 |
| 6.0.14 | Random Garden: a Supervised Learning Algorithm. <i>Ivan Luciano Danesi, Valeria Danese, Nicolo' Russo and Enrico Tonini</i> | 1675 |
| 6.0.15 | Spatiotemporal Prevision for Emergency Medical System Events in Milan. <i>Andrea Gilardi, Riccardo Borgoni, Andrea Pagliosa, Rodolfo Bonora</i> | 1681 |
| 6.0.16 | Spatial segregation immigrant households in Messina. <i>Angelo Mazza and Massimo Mucciardi</i> | 1687 |
| 6.0.17 | Supervised Learning for Link Prediction in Social Networks. <i>Riccardo Giubilei, Pierpaolo Brutti</i> | 1691 |
| 6.0.18 | Women's empowerment and child mortality: the case of Bangladesh. <i>Chiara Puglisi, Annalisa Busetta</i> | 1697 |

Classification of the Aneurisk65 dataset using PCA for partially observed functional data

Classificazione del dataset Aneurisk65 utilizzando la PCA per dati funzionali parzialmente osservati

Marco Stefanucci, Laura Sangalli and Pierpaolo Brutti

Abstract When functional data are observed over a domain that is subject-specific, most of the techniques for functional data analysis are invalidated. Recently, new methods able to handle this situation were developed and in particular we focus on well-known functional PCA. With the aim of classifying the Aneurisk65 dataset, we apply a few possible methods and we show that carrying out the analysis over the full domain, where at least one of the functional data is observed, may not be the optimal choice. This is also confirmed in a simulation study, where the best interval for classification lies between the common domain and the full domain.

Abstract *Ogniqualvolta dei dati funzionali vengono osservati su un dominio dipendente dal soggetto considerato, non è più possibile utilizzare la gran parte delle tecniche per l'analisi di dati funzionali. Recentemente sono stati sviluppati nuovi metodi in grado di affrontare questa situazione e noi tratteremo la ben nota ACP funzionale. Con lo scopo di classificare il dataset Aneurisk65, abbiamo applicato diversi metodi e mostreremo che eseguire l'analisi sul dominio completo, dove è osservato almeno uno dei dati funzionali, può non essere la scelta ottimale. Ciò è confermato anche da uno studio di simulazione dove il migliore intervallo per la classificazione giace tra il dominio comune e il dominio completo.*

Key words: Functional Data, Partially Observed Data, Classification, Functional PCA

Marco Stefanucci

Sapienza Università di Roma, Piazzale Aldo Moro, 5, 00185 Roma e-mail: marco.stefanucci@uniroma1.it

Laura Sangalli

Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milano e-mail: laura.sangalli@polimi.it

Pierpaolo Brutti

Sapienza Università di Roma, Piazzale Aldo Moro, 5, 00185 Roma e-mail: pierpaolo.brutti@uniroma1.it

In the last 20 years, functional data analysis emerged as one of the fastest growing fields of modern statistics [1]. Functional data can be viewed as realizations of a functional objects, and in general are collected as discrete and noisy observations. The domain where the data are observed is usually common to all the statistical units, while the observation grid may vary across units. When the domain is unit-dependent, standard procedures are invalidated and a practical solution is to restrict the analysis to the intersection of the domains, where all the curves are observed. However, including intervals where the data are only partially observed may result in a more powerful statistical analysis and in the last decade some authors developed methodology handling this particular situation.

The Aneurisk65 dataset is an important example of partially observed functional data. The data consist in the profiles of radius and curvature of the internal carotid artery of 65 subjects suspected to be affected by cerebral aneurysms [2] and are displayed in figure 1. The domain where all the curves are available is highlighted in light-gray. Outside this domain individual observations are progressively lost when moving towards the full domain. In this application it is relevant to investigate whether the morphology of the internal carotid artery influences aneurysms pathogenesis and this can be done considering that the data can be divided in two groups - displayed in orange and blue in the figure - depending on the presence and location of the cerebral aneurysms.

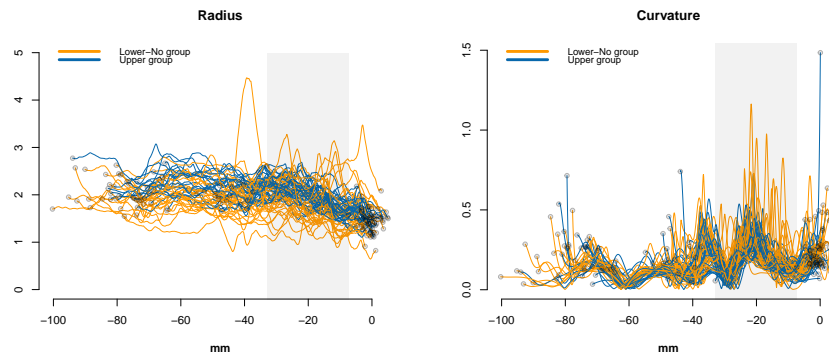


Fig. 1 The Aneurisk dataset.

In [2] a discriminant analysis based on the scores of the principal components of the radius and curvature profiles is presented, restricting the attention to the portion of the domain common across subjects. Our contribution is to improve the discrimination results by considering also portions of the domain where not all data are observed. In doing this, we need a methodology for functional principal components analysis in the case of partially observed data. Here we consider some proposals: the first one [3] is based on a spline reconstruction of the eigenfunctions, the second [4] is a method that estimates the principal component scores using conditional

expectations, the third [5] is a generalization of the power method for extracting eigenvectors of a given matrix and the last one [6] is a fully functional approach in which the missing part of each score is predicted via best linear approximation of its conditional expectation.

We then use these methods to perform a discrimination based on the scores of the first K principal components. We show that carry out the analysis on the largest possible domain may not be the best choice for classification purposes. In fact, we suggest to explore different intervals, ranging from the common domain to the full domain. More specifically, we divide the domain where the data are partially observed in L portions, and we consider a collection of progressively larger domains I_l for $l \in \{0, \dots, L\}$, with $I_{l-1} \subset I_l$, where I_0 is the common domain and I_L is the full domain. We select the optimal number of principal components and the optimal domain extension I_l via cross-validation. For all the methods, the best interval for classification is between the common domain and the full domain and such an interval is the same for all the approaches, see figure 2. In this application, the approach [5] achieves the best result, corresponding to a 9 misclassified subjects, thus outperforming the result in [2]. Note that the application of the methodologies for partially observed data on the full domain does not lead to any improvement in the discrimination.

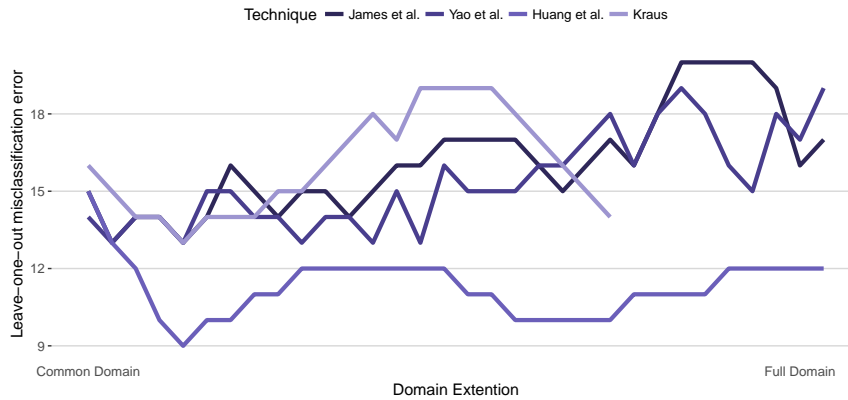


Fig. 2 Performances on the Aneurisk dataset.

In order to understand better the behaviour of the domain extension technique, we carry out simulations. We generate functional data over the interval $I_L = [0, 1]$. We then completely retain the data generated over the interval $I_0 = [1/3, 2/3]$, while we censor them over the intervals $I_{\text{left}} = [0, 1/3]$ and $I_{\text{right}} = [2/3, 1]$, by sampling the starting point of each functional datum uniformly over I_{left} , and its ending point uniformly over I_{right} . The data are generated from a cubic B-splines basis with 16 internal knots, corresponding to a total of 20 bases. We generate two groups of 50 units each with the only difference being the mean function, see figure 3. The 100



Fig. 3 Simulated data.

generated curves are evaluated on a regular grid of 150 points in $[0, 1]$, contaminated with gaussian noise and finally classified over a set of progressively larger domains. This simulation is repeated 50 times.

In order to compare the performances of the four methods, we also apply standard PCA to the fully observed data. The results are presented in figure 4. The figure displays the boxplots of the leave-one-out misclassification error, for various tech-

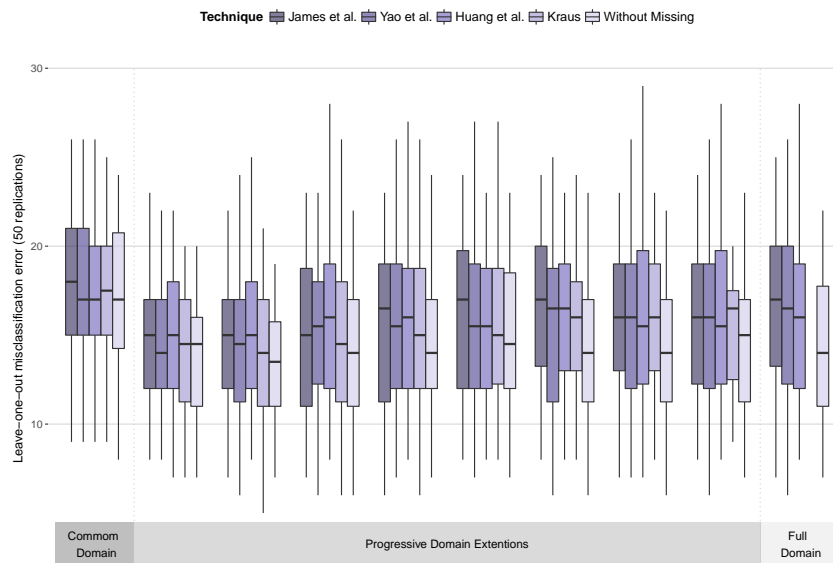


Fig. 4 Simulation results.

niques, for various domain extensions. For all methods, the misclassification error decreases when we start extending the domain with respect to the common domain, but then progressively increases as we approach the full domain. None of the methods outperforms the other. In conclusion, extending the domain with respect to the common domain improves the discrimination between the two groups; on the other hand, larger domain extensions, and in particular the full domain, do not lead to the best discrimination results.

References

1. Ramsay, J. and Silverman, B.W.: *Functional Data Analysis*. Springer (2005)
2. Sangalli, L.M., Secchi, P., Vantini, S. and Veneziani, A.: A Case Study in Exploratory Functional Data Analysis: Geometrical Features of the Internal Carotid Artery. *Journal of the American Statistical Association* **104**, 37-48 (2009)
3. James, G.M. and Hastie, T.J.: Principal component models for sparse functional data. *Biometrika* **87**, 587-602 (2000)
4. Yao, F., Müller, H.G. and Wang, J.L.: Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association* **100**, 577-590 (2005)
5. Huang, J.Z., Shen, H., and Buja, A. : Functional principal components analysis via penalized rank one approximation. *Electronic Journal of Statistics* **2**, 678-695 (2008)
6. Kraus, D.: Components and completion of partially observed functional data. *Journal of the Royal Statistical Society: Series B* **77**, 777-801 (2015)