



SIS | 2022

51st Scientific Meeting
of the Italian Statistical Society

Caserta, 22-24 June

V: Università
degli Studi
della Campania
Luigi Vanvitelli

SIS
Società
Italiana di
Statistica



www.unicampania.it



Book of the Short Papers

**Editors: Antonio Balzanella, Matilde Bini,
Carlo Cavicchia, Rosanna Verde**



1222-2022
800
ANNI



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO
DI SCIENZE
STATISTICHE



Matilde Bini (Chair of the Program Committee) - *Università Europea di Roma*

Rosanna Verde (Chair of the Local Organizing Committee) - *Università della Campania “Luigi Vanvitelli”*

PROGRAM COMMITTEE

Matilde Bini (Chair), Giovanna Boccuzzo, Antonio Canale, Maurizio Carpita, Carlo Cavicchia, Claudio Conversano, Fabio Crescenzi, Domenico De Stefano, Lara Fontanella, Ornella Giambalvo, Gabriella Grassia - Università degli Studi di Napoli Federico II, Tiziana Laureti, Caterina Liberati, Lucio Masserini, Cira Perna, Pier Francesco Perri, Elena Pirani, Gennaro Punzo, Emanuele Raffinetti, Matteo Ruggiero, Salvatore Strozza, Rosanna Verde, Donatella Vicari.

LOCAL ORGANIZING COMMITTEE

Rosanna Verde (Chair), Antonio Balzanella, Ida Camminatiello, Lelio Campanile, Stefania Capecchi, Andrea Diana, Michele Gallo, Giuseppe Giordano, Ferdinando Grillo, Mauro Iacono, Antonio Irpino, Rosaria Lombardo, Michele Mastroianni, Fabrizio Maturo, Fiammetta Marulli, Paolo Mazzocchi, Marco Menale, Giuseppe Pandolfi, Antonella Rocca, Elvira Romano, Biagio Simonetti.

ORGANIZERS OF SPECIALIZED, SOLICITED, AND GUEST SESSIONS

Arianna Agosto, Raffaele Argiento, Massimo Aria, Rossella Berni, Rosalia Castellano, Marta Catalano, Paola Cerchiello, Francesco Maria Chelli, Enrico Ciavolino, Pier Luigi Conti, Lisa Crosato, Marusca De Castris, Giovanni De Luca, Enrico Di Bella, Daniele Durante, Maria Rosaria Ferrante, Francesca Fortuna, Giuseppe Gabrielli, Stefania Galimberti, Francesca Giambona, Francesca Greselin, Elena Grimaccia, Raffaele Guetto, Rosalba Ignaccolo, Giovanna Jona Lasinio, Eugenio Lippiello, Rosaria Lombardo, Marica Manisera, Daniela Marella, Michelangelo Misuraca, Alessia Naccarato, Alessio Pollice, Giancarlo Ragozini, Giuseppe Luca Romagnoli, Alessandra Righi, Cecilia Tomassini, Arjuna Tuzzi, Simone Vantini, Agnese Vitali, Giorgia Zaccaria.

ADDITIONAL COLLABORATORS TO THE REVIEWING ACTIVITIES

Ilaria Lucrezia Amerise, Ilaria Benedetti, Andrea Bucci, Annalisa Busetta, Francesca Condino, Anthony Cosari, Paolo Carmelo Cozzucoli, Simone Di Zio, Paolo Giudici, Antonio Irpino, Fabrizio Maturo, Elvira Romano, Annalina Sarra, Alessandro Spelta, Manuela Stranges, Pasquale Valentini, Giorgia Zaccaria.

Copyright © 2022

PUBLISHED BY PEARSON

WWW.PEARSON.COM

ISBN 9788891932310

Contents

Preface	XXVII
1 Plenary Sessions	1
Causal inference in air pollution epidemiology <i>Francesca Dominici</i>	2
Clustering of Attribute Data and Network <i>Anuška Ferligoj</i>	11
Bayesian approaches for capturing the heterogeneity of neuroimaging experiments <i>Francesco Denti, Laura D'Angelo and Michele Guindani</i>	17
2 Specialized Sessions	30
Advances in Bayesian nonparametric methodology	31
Repulsive mixture models for high-dimensional data <i>Lorenzo Ghilotti, Mario Beraha and Alessandra Guglielmi</i>	32
Bayesian nonparametric mixtures of directed acyclic graph models <i>Federico Castelletti and Guido Consonni</i>	37
Bayesian Clustering of Brain Regions via Extended Stochastic Block Models <i>Sirio Legramanti, Tommaso Rigon and Daniele Durante</i>	45
Data Science skills for next generation statisticians	52
Cluster based oversampling for imbalanced learning <i>Gioia Di Credico and Nicola Torelli</i>	53
Estimating the effect of remote teaching for university students through generalised linear mixed models <i>Silvia Bacci, Bruno Bertaccini, Simone Del Sarto, Leonardo Grilli and Carla Rampichini</i>	65
Perceived stress across EU countries: does working from home impact? <i>Stefania Capecchi, Francesca Di Iorio and Nunzia Nappo</i>	71

Investigating effects of air pollution on health: a challenge for statisticians	77
Investigating effect of air pollution on health via Spatial-Resolution Varying Coefficient Models <i>Garritt L. Page and Massimo Ventrucchi</i>	78
A statistical framework for evaluating health effect of PM sources <i>Monica Pirani, Georges Bucyibaruta, Gary Fuller, David Green, Anja Tremper, Christina Mitsakou and Marta Blangiardo</i>	84
Adjusting for unmeasured spatial confounding through shrinkage methods <i>Pasquale Valentini, Alexandra M. Schmidt, Carlo Zaccardi and Luigi Ippoliti</i>	91
Explainable Artificial Intelligence methods	98
Multidimensional Time Series Analysis via Bayesian Matrix Auto Regression <i>Alessandro Celani and Paolo Pagnottoni</i>	99
Advances in Classification and Data Analysis	109
Optimizing time slots in scientific meetings: a Latent Dirichlet allocation approach <i>Luca Frigau</i>	110
Clustering artists based on the energy distributions of their songs on Spotify via the Common Atoms Model <i>Francesco Denti, Federico Camerlenghi, Michele Guindani and Antonietta Mira</i>	121
Hidden markov models for four-way data <i>Salvatore D. Tomarchio, Antonio Punzo and Antonello Maruotti</i>	127
Family demography	133
Does family of origin make the difference in occupational outcomes? <i>Annalisa Busetta, Elena Fabrizi, Isabella Sulis and Giancarlo Ragozini</i>	134
Is there a cultural driver pushing Italian low fertility? <i>Francesca Luppi, Alessandro Rosina and Maria Rita Testa</i>	144
Unpaid family work and the subjective well-being of Italian women during lockdown <i>Marina Zannella, Erica Aloé, Marcella Corsi and Alessandra de Rose</i>	155
New Frontiers in the theory of composite indicators	164
Methodological PLS-PM Framework for Model Based Composite Indicators <i>Rosanna Cataldo</i>	165
Open issues in composite indicators construction <i>Leonardo Salvatore Alaimo</i>	176
The posetic approach to the construction of socio-economic indicators: open issues and research opportunities <i>Marco Fattore</i>	186

Advances in complex sampling strategies	197
Random forest model-assisted estimation for finite population totals <i>Mehdi Dagdoug, Camelia Goga and David Haziza</i>	198
Design-based consistency of the Horvitz-Thompson estimator in spatial sampling <i>Lorenzo Fattorini</i>	208
The responsive-adaptive survey design approach for planning the permanent census of population and housing <i>Claudia De Vitiis, Stefano Falorsi, Alessio Guandalini, Francesca Inglese, Paolo Righi and Marco D. Terribili</i>	216
Socio-demographic aspects of aging in Italy	228
Socio-economic and spatial stratification of frailty in the older population <i>Margherita Silan</i>	229
Time allocation and wellbeing in later life: the case of Italy <i>Annalisa Donno and Maria Letizia Tanturi</i>	241
The role played by migration and fertility on Italy's demographic aging trends: a provincial-level analysis <i>Thais García-Pereiro and Anna Paterno</i>	250
New challenges in the labour market	260
Detecting changes and evolution in specialized professional figures: an application on the Italian IT & Digital sector <i>Andrea Marletta</i>	261
How did the COVID-19 pandemic affect the genderpay gap in EU countries? <i>Antonella Rocca, Paolo Mazzocchi, Giovanni De Luca, Rosalia Castellano and Claudio Quintano</i>	272
Skill Similarities and Dissimilarities in Online Job Vacancy Data across Italian Regions <i>Adham Kahlawi, Lucia Buzzigoli, Laura Grassini and Cristina Martelli</i>	284
Small area estimation methods with socioeconomic applications	292
Exploring Small Area Estimation techniques to address uncertainty in Spatial Price Indexes <i>Ilaria Benedetti and Federico Crescenzi</i>	293
Small Area Estimation of Relative Inequality Indices using Mixture of Beta <i>Silvia De Nicolò and Silvia Pacei</i>	301
Inference for big data assisted by small area methods: an application to OBEC (on-line based enterprise characteristics) <i>Monica Pratesi, Francesco Schirripa Spagnolo, Gaia Bertarelli, Stefano Marchetti, Monica Scannapieco, Nicola Salvati and Donato Summa</i>	305

Statistical methods and models for Sports Analytics	312
The 'hot shoe' in soccer penalty shootouts <i>Andreas Groll and Marius Otting</i>	313
G-RAPM: revisiting player contributions in regularized adjusted plus-minus models for basketball analytics <i>Luca Grassetti</i>	319
Formative vs Reflective constructs: a CTA-PLS approach on a goalkeepers' performance model <i>Mattia Cefis and Eugenio Brentari</i>	323
Integrating available Data Sources for Official Statistics	329
The Use of Administrative Data for the Estimation of Italian Usually Resident Population <i>Marco Caputi, Giampaolo De Matteis, Gerardo Gallo and Donatella Zindato</i>	330
New frontiers for the analysis of the territorial economic phenomena	339
An empirical tool to classify industries by regional concentration and spatial polarization <i>Diego Giuliani, Maria Michela Dickson, Flavio Santi and Giuseppe Espa</i>	340
Comparing non-compensatory composite indicators: a case study based on SDG for Mediterranean countries <i>Francesca Mariani, Mariateresa Ciommi, Maria Cristina Recchioni, Giuseppe Ricciardo Lamonica and Francesco Maria Chelli</i>	346
Evaluating the determinants of innovation from a spatio-temporal perspective. The GWPR approach <i>Gaetano Musella, Giorgia Riveccio and Emma Bruno</i>	354
Dimension Reduction for complex data	366
Discrimination and clustering via principal components <i>Nikolay Trendafilov and Violetta Simonacci</i>	367
Exploratory graph analysis for configural invariance assessment <i>Sara Fontanella, Alex Cucco and Nicola Pronello</i>	373
Penalized likelihood factor analysis <i>Kei Hirose</i>	379

3 Solicited Sessions	385
Bayesian nonparametric modelling and learning	386
A regularized-entropy estimator to enhance cluster interpretability in Bayesian nonparametrics <i>Beatrice Franzolini and Giovanni Rebaudo</i>	387
Exact confidence sets from credible sets with finite amounts of data <i>Bas J. K. Kleijn</i>	399
Empirical Bayesian analysis of componentwise maxima in multivariate samples <i>Simone A. Padoan and Stefano Rizzelli</i>	411
Processing of textual data in large corpora	420
Predictive performance comparisons of different feature extraction methods in a financial column corpus <i>Andrea Sciandra and Riccardo Ferretti</i>	421
Topics and trends in the End-of-Year addresses of the Presidents of the Italian Republic (1949-2021) <i>Matilde Trevisani and Arjuna Tuzzi</i>	428
Thematic analysis on online education issues during COVID-19 <i>Valerio Basile, Michelangelo Misuraca and Maria Spano</i>	437
What do we learn by applying multiple methods in topic detection? A comparative analysis on a large online dataset about mobility electrification <i>Fabrizio Alboni, Margherita Russo and Pasquale Pavone</i>	446
Businesses in industry: new challenges in sustainability, innovation, performance and competitiveness	454
Multidimensional assessment of Eco-Innovation and its link with Marketing Innovations <i>Ida D'Attoma and Marco Ieva</i>	455
Circular Economy practices in the European SMEs: company-level and country-level drivers <i>Francesca Bassi, José G. Dias and Nunzio Tritto</i>	462
The employment effects of Italian Jobs Act. An ex-post impact evaluation <i>Alessandro Zeli and Leopoldo Nascia</i>	474
Statistics for finance: new models, new data	482
The News-Jumps Relationship in the Cryptocurrency Market <i>Ahmet Faruk Aysan, Massimiliano Caporin, Oguzhan Cepni, and Francesco Poli</i>	483
A weighted quantile approach to Expected Shortfall forecasting <i>Giuseppe Storti and Chao Wang</i>	489

Smooth and abrupt dynamics in financial volatility: the MS-MEM-MIDAS	492
<i>Giampiero M. Gallo, Edoardo Otranto and Luca Scaffidi Domianello</i>	
The tail index and related quantities for volatility models	501
<i>Fabrizio Laurini</i>	
Bayesian inference for complex random structures	507
Bayesian nonparametric modeling of mortality curves via functional Dirichlet processes	508
<i>Emanuele Aliverti and Bruno Scarpa</i>	
Bayesian nonparametric clustering of spatially-referenced spike train data	514
<i>Laura D'Angelo</i>	
Bayesian Analysis of Mortality in Iceland via Locally Adaptive Splines	520
<i>Federico Pavone and Sirio Legramanti</i>	
Advances in clustering	526
A Two-step Latent Class Approach with Measurement Equivalence Testing	527
<i>Zsuzsa Bakk, Roberto Di Mari, Jennifer Oser and Marc Hooghe</i>	
Group-wise penalized estimation schemes in model-based clustering	534
<i>Alessandro Casa, Andrea Cappelletto and Michael Fop</i>	
Extending finite mixtures of latent trait analyzers for bipartite networks	540
<i>Dalila Failli, Maria Francesca Marino and Francesca Martella</i>	
A Fast Majorization-Minimization Algorithm for Convex Clustering	551
<i>Daniel J.W. Touw, Patrick J.F. Groenen and Yoshikazu Terada</i>	
Statistical Methods for Complex Evolutionary Data	558
A FANOVA model with repeated measures for detecting patterns in biomechanical data	559
<i>Ana M. Aguilera, Christian Acal and Manuel Escabias</i>	
Modes of variation for Lorenz curves	565
<i>Enea G. Bongiorno and Aldo Goia</i>	
Analyzing textual data through Word Embedding: experiences in Istat	571
<i>Mauro Bruno, Elena Catanese, Massimo De Cubellis, Fabrizio De Fausti, Francesco Pugliese, Monica Scannapieco and Luca Valentino</i>	
Functional Horvitz-Thompson estimator for convex curves	584
<i>Adelia Evangelista, Francesca Fortuna, Stefano Antonio Gattone and Tonio Di Battista</i>	

Children, parents, grandparents: a look on changing relationships	590
Changes in social relationships of Italian older people. Evidence from FSS and SHARE Corona surveys <i>Elvira Pelle, Giulia Rivellini and Susanna Zaccarin</i>	591
Internet use and contacts with children among older Europeans <i>Bruno Arpino</i>	600
A time-based comparative approach to study the changing demography of grandparenthood in Italy <i>***Elisa Cisotto, Eleonora Meli and Giulia Cavrini</i>	607
Carry that weight: Parental separation and children's Body Mass Index from childhood to young adulthood <i>Marco Tosi</i>	616
Living conditions, well-being and poverty	622
Analyzing the impact of COVID-19 pandemic on elderly population well-being <i>Gloria Polinesi, Mariateresa Ciommi and Chiara Gigliarano</i>	623
Exploring sustainable food purchasing behaviour using Italian scanner data <i>Ilaria Benedetti, Alessandro Brunetti, Federico Crescenzi and Luigi Palumbo</i>	629
The evaluation of heat vulnerability in Friuli-Venezia Giulia <i>Laura Pagani, Maria Chiara Zamarrotti and Anja Habus</i>	635
Data Science for Functional and Complex Data	641
A parsimonious approach to representing functional <i>Enea G. Bongiorno and Aldo Goia</i>	642
Mixed-effects high-dimensional multivariate regression via group-lasso regularization <i>Francesca Ieva, Andrea Cappozzo, and Giovanni Fiorito</i>	648
The integration of immigrants in Italy: a multidimensional perspective	654
Albanian, Romanian and Italian women's fertility intentions: a comparative perspective among migrants, stayers and natives <i>Thaís García-Pereiro and Anna Paterno</i>	655
Does self-employment in the origin-country affect self-employment after migration? Evidence from Italy and Spain <i>Floriane Bolazzi and Ivana Fellini</i>	662
The impact of integration on immigrants' health behaviours in Italy <i>Giovanni Minchio, Raffaella Rusciani and Teresa Spadea</i>	675
Migration, gender, and the distribution of paid and unpaid labour. Preliminary perspectives on foreign couples in Italy <i>Rocco Molinari, Agnese Vitali and Ester Gallo</i>	687

Sampling techniques for big data analysis	695
Non-probability samples and big data: how to use them? <i>Pier Luigi Conti</i>	696
Combining Big Data with probability survey data: a comparison of methodologies for estimation from non-probability surveys <i>Maria del Mar Rueda, Ramn Ferri-Garcia and Luis Castro-Martin</i>	707
A Bayesian approach for combining probability and non-probability samples surveys <i>Camilla Salvatore, Silvia Biffignandi, Joseph Sakshaug, Bella Struminskaya and Arkadiusz Wisniowski</i>	717
Big data and Official Statistics: some evidences <i>Paolo Righi, Natalia Golini and Gianpiero Bianchi</i>	723
The analysis of students performance and behaviour based on large databases	735
Students enrolled in STEM discipline in Italy: patterns of retention, dropout and switch <i>Valentina Tocchioni, Carla Galluccio, Maria Francesca Morabito and Alessandra Petrucci</i>	736
The routes of Southern Italy University students: an explorative analysis <i>Gabriele Ruiu and Vincenzo Giuseppe Genova</i>	747
A new bipartite matching approach for record linkage: the case of two big Italian databases <i>Martina Vittorietti, Andrea Priulla, Vincenzo Giuseppe Genova, Giovanni Boscaino and Ornella Giambalvo</i>	754
Statistical Methods for Science Mapping	761
A word embedding strategy to study the thematic evolution of ageing and healthcare expenditure growth literature <i>Milena Lopreite, Michelangelo Misuraca and Michelangelo Puliga</i>	762
An automatic approach for bibliographical co-words networks labelling <i>Manuel J. Cobo and Maria Spano</i>	773
Characterising research areas in the field of AI <i>Alessandra Belfiore, Angelo Salatino and Francesco Osborne</i>	780
Mapping evolutionary paths of a society: the longitudinal analysis of the Italian Economia Aziendale <i>Corrado Cuccurullo, Luca D'Aniello and Michele Pizzo</i>	786
Modelling complex structures in ecological data	793
New insights on the ecology and conservation of Mediterranean sharks through the development of Citizen Science networks and new modeling approaches <i>Stefano Moro, Francesco Ferretti, Francesco Colloca</i>	794

An overdispersed Poisson model for forest fires occurrences in Southern Italian municipalities <i>Crescenza Calculli and Serena Arima</i>	798
Assessment of the impact of anthropic pressures on the Giglio island meadow of <i>Posidonia oceanica</i> <i>Gianluca Mastrantonio, Daniele Ventura, Gianluca Mancini and Giandomenico Ardizzone</i>	804
Accounting for observation processes in spatio-temporal ecological data <i>Janine Illian</i>	811
Statistics and indicators for the recovery and resilience plan	815
The prominence of statistical information for the monitoring and effective implementation of the NRRP <i>Andrea Petrella</i>	816
Big Data Analytics in mobile cellular networks as enabler for innovative statistics to evaluate the effects of Recovery and Resilience Plan actions <i>Andrea Zaramella, Dario Di Sorte, Denis Cappellari and Bruno Zamengo</i>	819
Measuring the digital transition within the PA: proposals comparison <i>Susanna Traversa and Enrico Ivaldi</i>	823
Guest Session - European Network for Business and Industrial Statistics (ENBIS)	828
Interpretability in functional clustering with an application to resistance spot welding process in the automotive industry <i>Christian Capezza, Fabio Centofanti, Antonio Lepore and Biagio Palumbo</i>	829
Statistical process monitoring of thermal images in additive manufacturing: a nonparametric solution for in-situ monitoring <i>Panagiotis Tsiamyrtzis, Marco Luigi Giuseppe Grasso and Bianca Maria Colosimo</i>	835
Guest Session - International Biometric Society (IBS) - Italian region	837
Multiple arrows in the Bayesian quiver: Bayesian learning of partially directed structures from heterogeneous data <i>Luca La Rocca, Federico Castelletti, Stefano Peluso, Francesco Claudio Stingo and Guido Consonni</i>	838

4 Contributed Sessions	844
Applications in Machine Learning	845
A neural network approach to survival analysis with time-dependent covariates for modelling time to cardiovascular diseases in HIV patients <i>Federica Corso, Agostino Lurani Cernuschi, Laura Galli, Chiara Masci, Camilla Muccini, Anna Maria Paganoni and Francesca Ieva</i>	846
Analyzing the Correlation Structure of Financial Markets Using a Quantile Graphical Model <i>Beatrice Foroni, Luca Merlo and Lea Petrella</i>	852
Neural Network for statistical process control of a multiple stream binomial process with an application to HVAC systems in passenger rail vehicles <i>Gianluca Sposito, Antonio Lepore, Biagio Palumbo and Giuseppe Giannini</i>	858
Sparse signal extraction via variational SVM <i>Cristian Castiglione and Mauro Bernardi</i>	864
Bayesian modelling and inference 1	870
Bayesian Inference for the Multinomial Probit Model under Gaussian Prior Distribution <i>Augusto Fasano, Giovanni Rebaudo and Niccolò Anceschi</i>	871
Mapping Indicators on the Unit Interval: the tipsae Shiny App <i>Silvia De Nicolò and Aldo Gardini</i>	877
A Bayesian spatio-temporal model of PM10 pollutant in the Po Valley <i>Matteo Gianella, Alessandra Guglielmi and Giovanni Lonati</i>	883
Construction of a proper prior for a Bayesian envelope model <i>Andrea Mascaretti</i>	889
Hilbert principal component regression for bimodal bounded responses <i>Enea G. Bongiorno, Agnese M. Di Brisco, Aldo Goia, and Sonia Migliorati</i>	895
Methods of causal inference	901
Bayesian causal mediation analysis through linear mixed-effect models <i>Chiara Di Maria, Antonino Abbruzzo and Gianfranco Lovison</i>	902
Bootstrap-aggregated adjustment set selection <i>Lorenzo Giammei</i>	908
Exploiting partial knowledge to evaluate the average causal effect via an ABC perspective <i>Giulia Cereda, Fabio Corradi and Cecilia Viscardi</i>	914

Intertemporal propensity score matching for casual inference: an application to covid-19 lockdowns and air pollution in Northern Italy	920
<i>Daniele Bondonio and Paolo Chirico</i>	
Methods for Spatio-temporal data	926
Local Spatio-Temporal Log-Gaussian Cox Processes for seismic data analysis	927
<i>Nicoletta D'Angelo, Giada Adelfio, and Jorge Mateu</i>	
Spatial explorative analysis of thyroid cancer in Sicilian volcanic areas	933
<i>Francesca Bitonti and Angelo Mazza</i>	
Using geo-spatial topic modelling to understand the public view of Italian Twitter users: a climate change application	939
<i>Yuri Calleo and Francesco Pilla</i>	
Comparing local structures of spatio-temporal point processes on linear networks	945
<i>Nicoletta D'Angelo, Giada Adelfio, and Jorge Mateu</i>	
DISTATIS-based spatio-temporal clustering approach: an application to business cycles' time series	951
<i>Raffaele Mattera and Germana Scepti</i>	
Developments in composite indicators	957
Bayesian Networks for monitoring the gender gap	958
<i>Flaminia Musella, Lorenzo Giammei, Silvana Romio, Fulvia Mecatti and Paola Vicard</i>	
An Alternative Aggregation Function for the UNDP Human Development Index	964
<i>Manuela Scioni and Paola Annoni</i>	
An ultrametric model for building a composite indicator system to study climate change in European countries	970
<i>Giorgia Zaccaria and Pasquale Samacchiaro</i>	
Functional Weighted Malmquist Productive Index: a proposal for a dynamic composite indicator	975
<i>Annalina Sarra, Eugenia Nissi and Tonio Di Battista</i>	
CFA & PLS-PM for UX-AI Product infused	981
<i>Emma Zavarrone and Rosanna Cataldo</i>	
Fertility, adulthood, and economic uncertainty	987
Uncertainty and fertility intentions: a comparison between the Great Recession and the Covid-19 crisis	988
<i>Chiara Ludovica Comolli</i>	
Interpreting the relationship between life course trajectories and explanatory factors. An example on the transition to adulthood	996
<i>Danilo Bolano, Matthias Studer and Reto Buergin</i>	

The relationship between economic news and fertility: the case of Germany	1002
<i>Maria Francesca Morabito, Raffaele Guetto, Matthias Vollbracht and Daniele Vignoli</i>	
Leaving home among Millennials in Italy: does economic uncertainty matter?	1008
<i>Silvia Meggiolaro and Fausta Ongaro</i>	
Adverse pregnancy outcomes in The United Kingdom following unexpected job loss	1014
<i>Alessandro Di Nallo and Selin Koksai</i>	
Bayesian modelling and inference 2	1020
A Bayesian beta linear model to analyze fuzzy rating responses	1021
<i>Antonio Calcagni, Massimiliano Pastore, Gianmarco Altoe and Livio Finos</i>	
A Mixture Model for Multi-Source Cyber-Vulnerability Assessment	1028
<i>Mario Angelelli, Serena Anima and Christian Catalano</i>	
Hierarchical Bayesian models for analysing fish biomass data	1034
<i>Rita Fici, Antonino Abbruzzo, Luigi Augugliaro and Giacomo Milisenda</i>	
Insights into the derivative-based method for nonlinear mediation models	1040
<i>Claudio Rubino and Chiara Di Maria</i>	
An exploration of Approximate Bayesian Computation (ABC) and dissimilarities	1046
<i>Laura Bondi, Marco Bonetti and Raffaella Piccarreta</i>	
Advances in Categorical and Preference data	1052
On the predictability of a class of ordinal data models	1053
<i>Rosaria Simone and Domenico Piccolo</i>	
Multivariate analysis of binary ordinal data using graphical models	1059
<i>Camilla Caroni, Fabio Alberto Comazzi, Andrea Deretti and Federico Castelletti</i>	
Multinomial Thompson Sampling for adaptive experiments with rating scales	1065
<i>Nina Deliu</i>	
Ranking extraction in nested partially ordered data systems	1071
<i>Marco Fattore, Barbara Cavalletti, Matteo Corsi and Alessandro Avellone</i>	
Towards the definition of distance measures in the preference-approval structures	1077
<i>Alessandro Albano, Mariangela Sciandra and Antonella Plaia</i>	
Covid-19 Assessment and Evaluation 1	1083
Covid-19 impact assessment and inequality decomposition methods	1084
<i>Federico Attili and Michele Costa</i>	

Multiversal methods for model selection: COVID-19 vaccine coverage and relative risk reduction <i>Venera Tomaselli and Giulio Giacomo Cantone</i>	1090
Efficiency and feasibility of two stage sampling designs for estimating SARS-CoV-2 epidemic <i>Pietro Demetrio Falorsi, Vincenzo Nardelli and Giuseppe Arbia</i>	1096
Evaluating the impacts of Covid-19 on the overall Italian death process via Functional Data Analysis <i>Riccardo Scimone, Alessandra Menafoglio, Laura M. Sangalli and Piercesare Secchi</i>	1102
Developing countries, migration and migrants	1107
Domestic violence in Africa: a glance through the DHS survey <i>Micaela Arcaio, Daria Mendola and Anna Maria Parroco</i>	1108
Inequalities in undernutrition among Roma and non-Roma children in Western Balkans: an analysis of the determinants <i>Annalisa Busetta, Valeria Cetorelli and Chiara Puglisi</i>	1114
The manual, communicative and quantitative abilities of native and foreign workers according to their level of education in Italy <i>Camilla Pangallo, Oliviero Casacchia and Corrado Polli</i>	1120
HIV Prevalence in some African Territories: Socio-Economic Drivers <i>Micaela Arcaio, Daria Mendola and Anna Maria Parroco</i>	1126
A longitudinal cross country comparison of migrant integration policies via Mixture of Matrix-Normals <i>Leonardo Salvatore Alaimo, Francesco Amato and Emiliano Seri</i>	1132
Education and job placement	1138
Measuring happiness at work with categorical Principal Component Analysis <i>Ulpiana Kocollari, Maddalena Cavicchioli and Fabio Demaria</i>	1139
Early and accurate: a Machine Learning approach to predict students' final outcome with registry data <i>Lidia Rossi, Marta Cannistrà and Tommaso Agasisti</i>	1146
Students' experience with distance learning during Covid 19 pandemic in Southern Italy <i>Angela Maria D'Ugento and Nunziata Ribecco</i>	1153
Time series methods and Applications	1159
Trend and cycle decomposition in nonlinear time series <i>Maddalena Cavicchioli</i>	1160
Asymptotic properties of the SETAR parameters: a new approach <i>Marcella Niglio and Guy Mèlard</i>	1166
Food prices forecast using post-sampled crowdsourced data with Reg-ARMA model: the case of Nigeria <i>Ilaria Lucrezia Amerise, Gloria Solano Hermosilla, Vincenzo Nardelli and Giuseppe Arbia</i>	1172

Universal change point testing for dependent data <i>Federica Spoto, Alessia Caponera and Pierpaolo Brutti</i>	1178
Change point detection in fruit bioimpedance using a three-way panel model <i>F. Marta L. Di Lascio and Selene Perazzini</i>	1184
Bayesian modelling and inference 3	1190
A dynamic power prior approach to non-inferiority trials for normal means with unknown variance <i>Francesco Mariani, Fulvio De Santis and Stefania Gubbiotti</i>	1191
Bayesian Change-Point Detection for a Brownian Motion with a Total Miss Criterion <i>Bruno Buonaguidi</i>	1197
On the comparison of alternative Bayesian measures of posterior discrepancy <i>Fulvio De Santis and Stefania Gubbiotti</i>	1203
A Bayesian Test for the comparison of two independent populations <i>Mara Manca, Silvia Columbu and Monica Musio</i>	1209
A contribution to the L. J. Savage problem <i>Francesco Bertolino, Silvia Columbu and Mara Manca</i>	1215
Methods for Complex Data	1221
Optimization of delayed rejection adaptive metropolis <i>Daniele Raffo and Antonietta Mira</i>	1222
Dealing with multicollinearity and outliers in multinomial logit model: a simulation study <i>Ida Camminatiello and Antonio Lucadamo</i>	1228
A tool to validate the assumptions on ratios of nearest neighbors' distances: the Consecutive Ratio Paths <i>Francesco Denti and Antonietta Mira</i>	1233
Dimensionality reduction and visualization for interval-valued data via midpoints-ranges principal component analysis <i>Viviana Schisa, Alfonso Iodice D'Enza and Francesco Palumbo</i>	1239
Data-driven design-based mapping of forest resources <i>Sara Franceschi, Rosa Maria Di Biase, Lorenzo Fattorini, Marzia Marcheselli and Caterina Pisani</i>	1245
Environmental data and Climate change	1252
Ensemble model output statistics for temperature forecasts in Veneto <i>Gaetan Carlo, Giummole Federica, Mameli Valentina and Siad Si Mokrane</i>	1253
State of the urban Environment in Italy. A comparative analysis of selected composite indicators <i>Giuseppe Lecardane</i>	1259

A Functional Data Analysis approach for Climate Model Selection: the case study of Campania Region <i>Veronica Villani, Elvira Romano and Paola Mercogliano</i>	1266
Evolution of scientific literature on climate change: a bibliometric analysis <i>Gianpaolo Zammarchi, Giulia Contu, Maurizio Romano</i>	1273
Energy and material demand of the Italian Regions <i>Flora Fullone, Giulia Iorio, Assunta Lisa Carulli</i>	1279
Health and survivorship	1285
Increasing Inequalities in Mortality by Socioeconomic Position in Italy <i>Chiara Ardito, Nicolás Zengarini, Roberto Leombruni, Angelo d'Errico and Giuseppe Costa</i>	1286
The role of health conditions in the relationship between socio- economic status and well-being: the counterfactual approach in mediation models <i>Sara Manzella and Margherita Silan</i>	1296
Excess economic burden of multimorbidity: a population-based study in Italy <i>Chiara Seghieri, Niccolò Borri, Gaia Bertarelli and Sabina Nuti</i>	1302
Depression-free life expectancy among 50 and older Americans by gender, race/ethnicity and education: the effect of marital disruption <i>Alessandro Feraldi and Cristina Giudici</i>	1308
Disability-free grandparenthood in Italy. Trends and gender differences <i>Margherita Moretti, Elisa Cisotto and Alessandra De Rose</i>	1314
Advances in regression models	1320
Semiparametric M-quantile regression for modelling georeferenced housing price data <i>Riccardo Borgoni, Antonella Carcagni, Alessandra Michelangeli, Nicola Salvati and Francesco Schirripa Spagnolo</i>	1321
Resampling-based inference for high-dimensional regression <i>Anna Vesel, Jelle J. Goeman, Angela Andreella and Livio Finos</i>	1327
Quantile regression coefficient modeling for counts to evaluate the productivity of university students <i>Viviana Carcaiso and Leonardo Grilli</i>	1333
Adaptive smoothing spline using non-convex penalties <i>Daniele Cuntrera and Vito M.R. Muggeo</i>	1339
Conditional tests for generalized linear models <i>Riccardo De Santis, Jelle J. Goeman, Anna Vesely and Livio Finos</i>	1345

Methods and applications in economics and finance	1351
Mixed models for anomaly detection in anti-money laundering aggregate reports <i>Stefano Iezzi and Marianna Siino</i>	1352
On the drivers of Greenwashing risk: evidence from Eurostoxx600 <i>Yana Kostiuk, Costanza Bosone and Paola Cerchiello</i>	1358
Modelling Financial Returns with Finite Mixtures of GED <i>Pierdomenico Dutillo and Stefano Antonio Gattone</i>	1364
Risk Parity strategy for portfolio construction: a kurtosis-based approach <i>Maria Debora Braga, Consuelo Rubina Nava and Maria Grazia Zoia</i>	1370
Fully reconciled probabilistic GDP forecasts from Income and Expenditure sides <i>Tommaso Di Fonzo and Daniele Girolimetto</i>	1376
Latent Class models	1382
Latent thresholds model in classification tasks <i>Giuseppe Mignemi, Andrea Spoto and Antonio Calcagni</i>	1383
Adaptive filters for time-varying correlation parameters <i>Michele Lambardi di San Miniato, Ruggero Bellio, Luca Grassetti and Paolo Vidoni</i>	1389
Bayesian structural learning for Latent Class Model with an application to Record Linkage <i>Davide Di Cecco</i>	1395
Multilevel Latent Class modelling to advise students in self-learning platforms: an application in the context of learning Statistics <i>Roberto Fabbriatore, Zsuzsa Bakk, Roberto Di Mari, Mark de Rooij and Francesco Palumbo</i>	1401
Latent Markov models with associated mixed responses <i>Alfonso Russo and Alessio Farcomeni</i>	1407
Methods for health studies	1413
Beyond the fragility index <i>Piero Quatto and Enrico Ripamonti</i>	1414
Evaluation of the diagnostic-therapeutic paths for schizophrenic patients through state sequences analysis <i>Laura Savaré, Giovanni Corrao and Francesca Ieva</i>	1419
Optimal timing of bone-marrow transplant in myelodysplastic syndromes through multi-state modeling and microsimulation <i>Caterina Gregorio, Marta Spreafico and Francesca Ieva</i>	1425
A fully Bayesian approach for sample size determination of Poisson clinical trials <i>Susanna Gentile and Valeria Sambucini</i>	1431

Compartmental models in epidemiology: Application on Smoking Habits in Tuscany <i>Alessio Lachi, Cecilia Viscardi, Maria Chiara Malevolti, Giulia Carreras and Michela Baccini</i>	1437
Covid-19 Assessment and Evaluation 2	1443
We are in the same storm but not in the same boat: Impact of COVID-19 on UK households <i>Demetrio Panarello and Giorgio Tassinari</i>	1444
A network approach to investigate learning experiences and social support in higher education <i>Ilaria Primerano, Maria Carmela Catone, Giuseppe Giordano, Maria Prosperina Vitale</i>	1450
Physical and cultural activity, internet use and anxiety of Italian university students during the pandemic <i>Giovanni Busetta, Maria Gabriella Campolo and Demetrio Panarello</i>	1456
The digital divide in Italy before and during the pandemic phase <i>Laura Zannella</i>	1462
Covid-19 and financial professional advice <i>Marianna Brunetti and Rocco Ciciretti</i>	1468
Bayesian modelling and inference 4	1472
Bayesian functional mixed effects model for sports data <i>Patric Dolmeta, Raffaele Argiento and Silvia Montagna</i>	1473
Bayesian Optimization with Machine Learning for Big Data Applications in the Cloud <i>Bruno Guindani, Danilo Ardagna and Alessandra Guglielmi</i>	1479
Confidence distributions and fusion inference for intractable likelihoods <i>Elena Bortolato and Laura Ventura</i>	1485
Wasserstein distance and applications to Bayesian nonparametrics <i>Marta Catalano, Hugo Lavenant, Antonio Lijoi and Igor Prunster</i>	1491
Network Analysis and community detection	1497
Community detection in networks: a heuristic version of Girvan Newman algorithm <i>Ilaria Bombelli and Lorenzo Di Rocco</i>	1498
Geographically weighted regression for spatial network data: an application to traffic volumes estimation <i>Andrea Gilardi, Riccardo Borgoni and Jorge Mateu</i>	1504
Asymmetric Spectral Clustering: a comparison between symmetrizations <i>Cinzia Di Nuzzo and Donatella Vicari</i>	1510
Community detection of seismic point processes <i>Valeria Policastro, Nicoletta D'Angelo and Giada Adelfio</i>	1516

An Explorative analysis of Different Distance Metrics to Compare Unweighted Undirected Networks	1522
<i>Anna Simonetto, Matteo Ventura and Gianni Gilioli</i>	
Gender, attitudes and family ties	1528
Parents of a disabled child in Italy: less healthy but more civically engaged	1529
<i>Nicoletta Balbo and Danilo Bolano</i>	
Searching the nexus between women's empowerment and female genital cutting (FGC)	1535
<i>Patrizia Farina, Liva Ortensi, Thomas Pettinato and Enrico Ripamonti</i>	
Social stratification, gender, and attitudes towards voluntary childlessness in Europe: A double machine learning approach	1539
<i>Danilo Bolano and Francesco C. Billari</i>	
Integrating structuralism and diffusionism to explain the new Italian emigration	1545
<i>Francesca Bitonti</i>	
On the effects of rooted family ties in business networks: The South of Italy in the 19th century	1551
<i>Roberto Rondinelli, Giancarlo Ragozini and Maria Carmela Schisani</i>	
Methods and Applications in Clustering	1557
A semi-supervised clustering method to extract information from the electronic Word Of Mouth	1558
<i>Giulia Contu, Luca Frigau, Maurizio Romano and Marco Ortu</i>	
Spectral approach for clustering three-way data	1564
<i>Cinzia Di Nuzzo and Salvatore Ingrassia</i>	
Double clustering with a matrix-variate regression model: finding groups of athletes and disciplines in decathlon's data	1570
<i>Mattia Stival, Mauro Bernardi, Manuela Cattelan and Petros Dellaportas</i>	
Classification of the population dynamics	1576
<i>Federico Bacchi and Laura Neri</i>	
Locating γ -Ray Sources on the Celestial Sphere via Modal Clustering	1582
<i>Anna Montin, Alessandra R. Brazzale and Giovanna Menardi</i>	
Sampling and Official Statistics	1588
Fisher's Noncentral Hypergeometric Distribution for Population Size Estimation	1589
<i>Veronica Ballerini and Brunero Liseo</i>	
Small area models for skew and kurtotic distributions	1595
<i>Maria Rosaria Ferrante and Lorenzo Mori</i>	

The use of remotely sensed data in sampling designs for forest monitoring	1601
<i>Chiara Bocci, Gherardo Chirici, Giovanni D'Amico, Saverio Francini and Emilia Rocco</i>	
Analyzing different causes of one-inflation in capture recapture models for criminal populations	1607
<i>Davide Di Cecco, Andrea Tancredi and Tiziana Tuoto</i>	
Administrative database and official statistics: an IT and statistical procedure	1613
<i>Caterina Marini and Vittorio Nicolardi</i>	
Spatial modeling and Analyses	1619
Spatial statistics analysis using microdata: an application at agricultural sector	1620
<i>Daniela Fusco, Maria Antonietta Liguori, Valerio Moretti and Francesco Giovanni Truglia</i>	
Bayesian spatial modeling of extreme precipitation	1627
<i>Federica Stolf</i>	
A proposal to adjust local Moran's I for measuring residential segregation	1632
<i>Antonio De Falco and Antonio Irpino</i>	
Accurate directional inference for gaussian graphical models	1637
<i>Claudia Di Caterina, Nancy Reid and Nicola Sartori</i>	
Advances in Classification	1643
Measures of interrater agreement based on the standard deviation	1644
<i>Giuseppe Bove</i>	
A Comparison of accuracy measures for Classification tasks	1650
<i>Amalia Vanacore and Maria Sole Pellegrino</i>	
Iterative Threshold-based Naive Bayes Classifier: an efficient Tb-NB improvement	1656
<i>Maurizio Romano, Gianpaolo Zammarchi and Giulia Contu</i>	
Reprogramming FairGANs with Variational Auto-Encoders: A New Transfer Learning Model	1662
<i>Beatrice Nobile, Gabriele Santin, Bruno Lepri and Pierpaolo Brutti</i>	
Robust statistics	1669
Combinatorial Analysis of Factorial Designs with Ordered Factors	1670
<i>Roberto Fontana and Fabio Rapallo</i>	
Robustifying the Rasch model with the forward search	1676
<i>Anna Comotti and Francesca Greselin</i>	
A novel estimation procedure for robust CP model fitting	1682
<i>Valentin Todorov, Violetta Simonacci, Michele Gallo and Nikolay Trendafilov</i>	

A robust approach for functional ANOVA with application to additive manufacturing <i>Fabio Centofanti, Bianca Maria Colosimo, Marco Luigi Grasso, Alessandra Menafoglio, Biagio Palumbo and Simone Vantini</i>	1688
Modeling unconditional M-quantiles in a regression framework <i>Luca Merlo, Lea Petrella and Nicola Salvati</i>	1692
Model-based clustering	1696
Bayesian mixtures of semi-Markov models <i>Rosario Barone and Andrea Tancredi</i>	1697
Specification of informative priors for capture-recapture finite mixture models <i>Pierfrancesco Alaimo Di Loro, Gianmarco Caruso, Marco Mingione, Giovanna Jona Lasinio and Luca Tardella</i>	1703
Clustering multivariate categorical data: a graphical model-based approach <i>Francesco Rettore, Michele Russo, Luca Zerman and Federico Castelletti</i>	1709
The Gaussian mixture model-based clustering for the comparative analysis of the Healthcare Digitalization Index in the Italian local health authorities <i>Margaret Antonicelli, Michele Rubino and Filomena Maggino</i>	1715
Student performance evaluation	1721
Rasch model versus Rasch Mixture model: strengthens and limits in identifying factors affecting students' performance in mathematics <i>Clelia Cascella</i>	1722
Does taking additional Maths classes improve university performance? <i>Martina Vittorietti, Andrea Priulla and Massimo Attanasio</i>	1728
University dropout and churn in Italy: an analysis over time <i>Barbara Barbieri, Mariano Porcu, Luisa Salaris, Isabella Sullis, Nicola Tedesco and Cristian Usala</i>	1734
The ANOGI for detecting the impact of education and employment on income inequality <i>Elena Fabrizi, Alessio Guandalini and Alessandra Spagnoli</i>	1740
What causes juvenile crime? a case-control study <i>Elena Dalla Chiara and Federico Perali</i>	1747
Methods and Applications in Survival analysis	1753
Recursive partitioning for survival data <i>Ambra Macis</i>	1754
Detecting survival patterns in a digital learning platform <i>Marta Cannistrà, Mara Soncin and Federico Frattini</i>	1760
An extension of proper Bayesian bootstrap ensemble tree models to survival analysis <i>Elena Ballante</i>	1766

Modelling time to university dropout by means of time-dependent frailty COX PH models <i>Mirko Giovio, Paola Mussida and Chiara Masci</i>	1771
Family history in survival and disease development <i>Maria Veronica Vinattieri and Marco Bonetti</i>	1777
Text mining	1783
Topics & metaverse: an explorative analysis <i>Emma Zavarrone, Alessia Forciniti, Emanuele Parisi, Maria Gabriella Grassia</i>	1784
Applying Topic Models to bibliographic search: some results in basketball domain <i>Manlio Migliorati and Eugenio Brentari</i>	1791
Exploiting Text Mining and Network Analysis for future scenarios development: an application on remote working <i>Yuri Calleo, Simone Di Zio and Vanessa Russo</i>	1797
Emotion recognition in Italian political language to predict positionings and crises government <i>Alessia Forciniti and Emma Zavarrone</i>	1803
What does your self-description reveal about you? <i>Riccardo Ricciardi</i>	1809
Variable selection and complete matrix approaches	1815
A Statistical Approach for the Completion of Input-Output Tables <i>Rodolfo Metulini, Giorgio Gnecco, Francesco Biancalani and Massimo Riccaboni</i>	1816
On multivariate records over sequences of random vectors with Marshall-Olkin dependence of components <i>A. Khorrami Chokami and Simone A. Padoan</i>	1822
The joint censored gaussian graphical lasso model <i>Gianluca Sottile, Luigi Augugliaro and Veronica Vinciotti</i>	1829
Variable selection with unbiased estimation: the cdf penalty <i>Daniele Cuntrera, Vito M.R. Muggeo and Luigi Augugliaro</i>	1835
Automatic variable selection for MIDAS regressions: an application <i>Consuelo Rubina Nava, Luigi Riso and Maria Grazia Zoia</i>	1841
Distribution Theory and Estimation	1847
A general framework for unit distributions <i>Francesca Condino, Filippo Domma and Bozidar V. Popovic</i>	1848
Prediction intervals based on multiplicative model combinations <i>Valentina Mameli and Paolo Vidoni</i>	1854
Some advances on pairwise likelihood estimation in ordinal data latent variable models <i>Giuseppe Alfonzetti and Ruggero Bellio</i>	1860

Functional Data Analysis	1866
A new functional clustering method: the Functional Clustering and Dimension Reduction model <i>Adelia Evangelista and Stefano Antonio Gattone</i>	1867
Nonparametric functional prediction bands: theory with an application to bike sharing mobility demand in the city of Milan <i>Jacopo Diquigiovanni, Matteo Fontana and Simone Vantini</i>	1873
An R package for the statistical process monitoring of functional data <i>Christian Capezza, Fabio Centofanti, Antonio Lepore, Alessandra Menafoglio, Biagio Palumbo and Simone Vantini</i>	1878
Trend filtering for functional regression <i>Federico Ferraccioli, Alessandro Casa and Marco Stefanucci</i>	1884
Conformal prediction for spatio-functional regression models <i>Diana, Romano, Irpino</i>	1890
Tourism and sport studies	1895
Assessing satisfaction of tourists visiting Italian museums: evidence from the eWOM <i>Daria Mendola and Valentina Oddo</i>	1896
COVID-19 pandemic and tourism demand: a comparison between Spain and Italy <i>Caterina Sciortino, Ludovica Venturella and Stefano De Cantis</i>	1902
A compositional analysis of tourism in Europe <i>Francesco Porro</i>	1908
Improving administrative data quality on tourism using Big Data <i>Antonella Bianchino, Armando d'Aniello and Daniela Fusco</i>	1914
Geographical variations of socio-demographic issues	1920
Elderly HCE and health care need: comparing spatially unexplained levels <i>Irene Torrini, Laura Rizzi and Luca Grassetti</i>	1921
Measuring sustainable development at the regional level. The case of Italy <i>Marianna Bartiromo and Enrico Ivaldi</i>	1927
Socio-economic deprivation and COVID-19 infection: a Bayesian spatial modelling approach <i>Antonino Abbruzzo, Andrea Mattaliano, Alessandro Arrigo, Salvatore Scondotto and Mauro Ferrante</i>	1933
Applications in Economics	1939
The measurement of economic security through relative indicators <i>Alessandro Gallo, Silvia Pacei and Maria Rosaria Ferrante</i>	1940

A regional analysis of the efficiency by energy's producers in Italy <i>Gianna Greca, Giuseppe Cinquegrana and Giovanni Fosco</i>	1946
On investigating social and financial aspects of Cardano <i>Stefano Vacca, Marco Ortu, Gianpaolo Zammarchi and Giuseppe Destefanis</i>	1953
Combined permutation test on the effect of age of micro enterprises on the propensity to Circular Economy <i>Stefano Bonnini and Michela Borghesi</i>	1959
Comparison of Two Different Approaches to Measure Economic Access to Food and Insecurity: an Application to Mexican data <i>Stefano Marchetti, Luca Secondi and Adrian Vargas-Lopez</i>	1965
Image analysis and visual methods	1971
Bias correction of the maximum likelihood estimator for Emax model at the interim analysis <i>Caterina May and Chiara Tommasi</i>	1972
Visual and automated methods in digital microscopy to evaluate fungal colonisation on plant roots <i>Ivan Sciascia, Andrea Crosino and Andrea Genre</i>	1977
From satellite images to road pavement type: an object-oriented classification approach <i>Arianna Burzacchi, Matteo Landrò and Simone Vantini</i>	1983
Valid inference for group analysis of functionally aligned fMRI images <i>Angela Andreella, Riccardo De Santis and Livio Finos</i>	1987
Topological persistence for astronomical image segmentation <i>Riccardo Ceccaroni, Pierpaolo Brutti, Marco Castellano, Adriano Fontana and Emiliano Merlin</i>	1993
Statistical assessment and empirical estimation	1999
Confidence regions for optimal sensitivity and specificity of a diagnostic test <i>Gianfranco Adimari, Duc-Khanh To and Monica Chiogna</i>	2000
On the sensitiveness to the memory parameter in the network of tennis <i>Alberto Arcagni, Vincenzo Candila and Rosanna Grassi</i>	2006
Two-part model with measurement error <i>Maria Felice Arezzo, Serena Arima, and Giuseppina Guagnano</i>	2011
Statistical assessment of practical significance <i>Andrea Ongaro, Sonia Migliorati, and Enrico Ripamont</i>	2017
Autoregressive and mixed effects models	2023
Asymptotic Properties of the Nonlinear Least Squares Estimator in HE-HAR Models <i>Emilija Dzuverovic and Edoardo Otranto</i>	2024

A note on testing for threshold non-linearity in presence of heteroskedasticity in time series <i>Simone Giannerini and Greta Goracci</i>	2030
The conditional autoregressive Whart-G model <i>Massimiliano Caporin and Marco Girardi</i>	2036
Semi-parametric generalized linear mixed effects models for binary response for the analysis of heart failure hospitalizations <i>Alessandra Ragni, Chiara Masci, Francesca Ieva and Anna Maria Paganoni</i>	2042
Issues in Data science	2048
etree: Classification and Regression With Structured and Mixed-Type Data in R <i>Riccardo Giubilei, Tullia Padellini and Pierpaolo Brutti</i>	2049
Deep Learning framework for ungrouping coarsely aggregated vital rates <i>Andrea Nigri</i>	2055
Inside the metaverse: analysis of the state of the art and development of a new usage approach based on quality and ethics <i>Vito Santarcangelo, Emilio Massa, Saverio Gianluca Crisafulli, Antonio Ruoto, Angelo Lamacchia, Alessandro D'Alcantara, Alessandro Verderame and Massimiliano Giacalone</i>	2061

A new functional clustering method: the Functional Clustering and Dimension Reduction model

Un nuovo metodo di clustering funzionale: il modello CDR funzionale

Adelia Evangelista and Stefano Antonio Gattone

Abstract In this work a new general procedure to cluster functional observations in a subspace of reduced dimension is proposed. In particular the method simultaneously performs cluster analysis and dimension reduction of functional data. The Functional Clustering and Dimension Reduction (FCDR) is obtained as the combination of two objective functions: the Functional Principal Component Analysis (FPCA) and the Functional Factorial K-Means (FFKM). The advantage of this approach consists in the optimization of a single global objective function which, by means of the selection of a tuning parameter, incorporates several techniques varying from the FPCA to FFKM including intermediate cases of clustering and dimension reduction.

Abstract In questo lavoro viene proposta una nuova procedura generale per raggruppare dati funzionali in un sottospazio di dimensioni ridotte. In particolare, il metodo esegue simultaneamente la cluster analysis e la riduzione della dimensionalità dei dati funzionali. Il metodo Functional Clustering and Dimension Reduction (FCDR) è ottenuto come la combinazione di due funzioni obiettivo: la Functional Principal Component Analysis (FPCA) ed il Functional Factorial K-Means (FFKM). Il vantaggio di questo approccio consiste nell'ottimizzare una sola funzione obiettivo globale che, attraverso la selezione di un parametro di controllo, include diverse tecniche che vanno dalla FPCA al FFKM includendo casi intermedi di raggruppamento e riduzione dei dati.

Key words: Cluster Analysis, Functional Data, Dimension Reduction

Adelia Evangelista

University of Chieti-Pescara, Viale Pindaro, 42, e-mail: adelia.evangelista@unich.it

Stefano Antonio Gattone

University of Chieti-Pescara, Viale Pindaro, 42, e-mail: gattone@unich.it

1 Introduction

Cluster analysis has been well used in multiple areas [5] and it continues to play a significant role for analysts and researchers. Unsupervised classification (or clustering) methods aim to classify a sample of data into homogeneous groups, without having any previous information about the clustering structure. This work deals with the clustering of functional data which received specific consideration in the last decade, as a result of the relevance of the high frequency data collected with modern technologies ([7]). Several clustering techniques for multivariate data have been suitably modified in order to be applied to functional data ([6]). Jacques and Preda ([9]) made an interesting review of the most important works related to functional clustering. They identify three general approaches: dimension-reduction before clustering, model-based clustering and nonparametric distance-based methods. Several approaches have been introduced for clustering univariate and multivariate functional data. See for example ([2][3][10][17][12]).

This work is based on the Clustering and Dimension Reduction (CDR) technique proposed by Vichi *et al* [15] which simultaneously identifies the optimal subspaces and the optimal clustering for multivariate data. The idea is that dimension reduction can improve the clustering recovery when the data incorporate unimportant information and the clusters lie in a subspace of reduced dimension ([3]). To this end, the proposed general model called Functional Clustering and Dimension Reduction (FCDR) represents an extension to the functional case of the CDR technique. The work is organized as follows. Section 2 describes Functional Principal Component Analysis (FPCA), Functional Factorial K-Means (FFKM) ([17]), and Functional Reduced K-Means ([8, 16]). The objective function of the FCDR model is presented in Section 3. Some concluding remarks are given in Section 4.

2 Background knowledge: FPCA, FFKM and FRKM

2.1 FPCA

A functional random variable X takes values in a Hilbert space of functions defined on a continuous subset of \mathcal{R} , say T . Let $\{x_i(t)\}_{i=1}^n$ be the set of observed curves where $t \in T$. Without loss of generality assume that data are centered, i.e. the mean function has been subtracted and $\sum_{i=1}^n x_i(t) = 0$, for all $t \in T$. Let $\Phi = \{\phi_1, \phi_2, \dots, \phi_J\}$ be a set of basis functions. X can be expanded in terms of basis functions as

$$X(t) = \sum_{j=1}^J c_{ij} \phi_j(t) = c_i^T \Phi. \quad (1)$$

FPCA searches an optimal representation of the curves into a functional space of reduced dimension. In other words it tries to minimize the reconstruction error of the

Q -dimensional approximation of the data given by the following objective function:

$$\ell_{FPCA}(\Psi) = \sum_{i=1}^n \|x_i(t) - \sum_{q=1}^Q a_{iq} \psi_q(t)\|^2 \quad (2)$$

where $a_{iq} = \int_T x_i(t) \psi_q(t) dt$ are the principal component scores, projections of $x_i(t)$ on the orthonormal eigenfunctions ψ_q with $\int_T \psi_q(t) \psi_r(t) dt = 1$ if $q = r$ and 0 otherwise. The eigenfunctions ψ_q belong to the same linear space spanned by the basis Φ , i.e. $\psi_q = \sum_{j=1}^J b_{jq} \phi_j(t) = b_q^T \Psi$ with $\Psi = \{\psi_1, \psi_2, \dots, \psi_Q\}$.

Let C be the $n \times J$ matrix of basis coefficients with generic component c_{ij} , W the $J \times J$ matrix with generic component given by $w_{js} = \int_T \phi_j(t) \phi_s(t) dt$, and B the $J \times Q$ matrix with generic column given by b_q . The FPCA loss function can be expressed in matrix form as follows:

$$\ell_{FPCA}(B) = \| [C - CWBB^T] \Phi \|^2. \quad (3)$$

2.2 FFKM

Cluster analysis and dimension reduction are often performed in tandem ([1]). A standard approach is to first apply a dimension-reduction technique to obtain a fewer number of components than the number of variables, and subsequently using the extracted components for clustering objects (*tandem clustering*). For example, one may apply standard clustering algorithm on the functional principal component scores a_{iq} [11]. A more efficient way to proceed is to implement the clustering simultaneously with dimension reduction. The main subspace functional clustering are the Functional Reduced K-Means (FRKM) ([8, 16]) and the Functional Factorial K-Means (FFKM) ([17]). FFKM applies to functional data the Factorial K-Means technique ([14]) where a clustering and a continuous factorial model are fitted simultaneously to multivariate data.

Let $U = \{u_{ig}\} (i = 1, \dots, n; g = 1, \dots, G)$ be the binary matrix defining the cluster parameters, where $u_{ig} = 1$ if $x_i(t)$ belongs to the g -th cluster and 0 otherwise. Let n_g be the size of cluster g , the cluster centroids functions $\bar{x}_g(t) = \frac{1}{n_g} \sum_{i=1}^n u_{ig} x_i(t)$ are assumed to lie in a lower dimensional space. FFKM aims at identifying the optimal projection expressing the cluster structure of the data by minimizing the within-cluster deviance of the clusters in the reduced space given by the following objective function:

$$\ell_{FFKM}(U, \Psi) = \sum_{i,g} u_{ig} \left\| \sum_{q=1}^Q \int_T x_i(t) \psi_q(t) dt \psi_q(t) - \sum_{q=1}^Q \int_T \bar{x}_g(t) \psi_q(t) dt \psi_q(t) \right\|^2 \quad (4)$$

subject to the weight functions $\{\psi_q\}_{q=1}^Q$ being orthonormal.

The FFKM loss function can be expressed in matrix form as follows:

$$\ell_{FFKM}(U, B) = \|CWB - H_U CWB\|^2 \quad (5)$$

where $H_U = U(U^T U)^{-1} U^T$ is the projection matrix onto the space spanned by the columns of U .

2.3 FRKM

Functional Reduced K-Means (FRKM) attempts to find an optimal subspace for clustering functional data by maximizing the between cluster deviance in the reduced space. The loss function of FRKM is given by:

$$\ell_{FRKM}(U, \Psi) = \sum_{i=1}^n \sum_{g=1}^G u_{ig} \|x_i(t) - \sum_{q=1}^Q \int_T \bar{x}_g(t) \psi_q(t) dt \psi_q(t)\|^2 \quad (6)$$

subject to the weight functions $\{\psi_q\}_{q=1}^Q$ being orthonormal.

In matrix form the loss function becomes

$$\ell_{FRKM}(U, B) = \| [C - H_U CWB B^T] \Phi \|^2. \quad (7)$$

3 The Functional CDR

The two methodologies described in Section 2 have been shown to perform better than the tandem analysis ([17]). However, one technique maximizes the between cluster deviance (FRKM) and the other minimizes the within cluster deviance (FFKM) in the reduced space. Following Vichi *et al* [15] a more general model can be considered which combines the two methodologies in order to optimize both the within and the between deviance: the clustering and dimension reduction model (CDR). The Functional CDR loss function is developed as the sum of the FPCA loss function given in (3) and the FFKM loss function given in (5):

$$\ell_{FCDR}(U, B) = \alpha \ell_{FPCA}(B) + (1 - \alpha) \ell_{FFKM}(U, B). \quad (8)$$

Hence, the minimization of the loss function (8) is a trade-off amid FPCA and FFKM. Clearly, by setting the value of the constant $\alpha = 0$, (8) becomes the FFKM loss function, vice versa fixing $\alpha = 1$ it is equivalent to minimize the loss function of the FPCA. By varying α between 0 and 1 different weights are given to FPCA and FFKM. Interestingly, it can be shown that with $\alpha = 0.5$ the FCDR is equivalent to the FRKM loss function. Therefore, the clustering solution depends on the chosen value of the tuning parameter α , whose selection becomes crucial.

Vichi *et al* [15] proposed to use the pseudo F index in order to automatically select the value of α ([4]). In this work different criterion to select the value of α are

considered and their performance evaluated through a simulation study. A criterion could be clustering stability proposed by [18]. The idea is to evaluate the clustering algorithm on different samples taken from the same population. If the algorithm is good, it should assign the units to the same cluster from one sample to another. An alternative criterion might be the *gap statistic* suggested by [13] for estimating the number of groups in a data set.

4 Future works

A simulation study to evaluate the performance of the FCDR algorithm together with the criterion for the choice of the tuning parameter α is currently under development.

A possible application of the FCDR method could be the identification of homogeneous areas in an environmental setting. Indeed, the FCDR method seems to be an optimal solution when dealing with meteorological variables where correlation and redundant information are usually found in the data.

References

1. Ben-Hur, A., Guyon, I.: Detecting stable clusters using principal components analysis. In: Brownstein, M. J., Khodursky, A.B. (eds.) *Functional Genomics*, pp.159-182. Human Press (2003)
2. Bongiorno, E.G., Goia, A.: Classification methods for hilbert data based on surrogate density. *Comput Stat Data Anal* **99**(C), 204-222 (2016)
3. Bouveyron, C., Jacques, J.: Model-Based Clustering of Time Series in Group-Specific Functional Subspaces. *Adv Data Anal Classif* **5**, 281–300 (2011)
4. Calinski, T., Harabasz, J.: A dendrite method for cluster analysis. *Commun Stat Theory Methods* **3**, 1–27 (1974)
5. Caruso, G., Gattone, S.A., Fortuna, F., Di Battista, T.: Cluster analysis as a decision-making tool: a methodological review. In: Bucciarelli, E., Chen, S.H., Corchado, J. (eds.) *Decision Economics: In the Tradition of Herbert A. Simon's Heritage. Advances in Intelligent Systems and Computing*, Vol. 618, pp. 48–55. Springer International Publishing, Berlin (2017)
6. Chiou, J. M., Li, P. L.: Functional clustering and identifying substructures of longitudinal data. *J R Stat Soc Series B Stat Methodol* **69** Issue 4, 679–699 (2007)
7. Hitchcock, D.B., Greenwood, M.C.: Clustering Functional Data. In: Henning, C., Meila, M., Fionn, M., Rocci, R. (eds) *Handbook of Cluster Analysis*, pp. 265-289. Chapman and Hall/CRC, New York (2015)
8. Gattone, S.A., Rocci, R.: Clustering Curves on a Reduced Subspace. *J Comput Graph Stat* **21**, 361–379 (2012)
9. Jacques, J., Preda, C.: Functional data clustering: a survey. *Adv Data Anal Classif* **8**, 231–255 (2014a)
10. Jacques, J., Preda, C.: Model based clustering for multivariate functional data. *Comput Stat Data Anal* **71**, 92–106 (2014b)
11. Peng J. and Müller H-G. Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. *The Annals of Applied Statistics*, 2(3):1056–1077, 2008

12. Schmutz, A., Jacques, J., Bouveyron, C., Chéze, L., Martin, P.: Clustering multivariate functional data in group-specific functional subspaces. *Comput Stat* **35**, 1101–1131 (2020)
13. Tibshirani, R., Walther, G., Hastie, T.: Estimating the number of clusters in a data set via the gap statistics. *J R Statist Soc B* **63**, Part 2, 411–423 (2001)
14. Vichi, M., Kiers, H.A.L.: Factorial K-means analysis of two-way data. *Computational Statistics & Data Analysis* **37**, 1, 49–64 (2001)
15. Vichi, M., Vicari, D., Kiers, H.A.L.: Clustering and dimension reduction for mixed variables. *Behaviormetrika* **46**, 243–269 (2019)
16. Yamamoto, M.: Clustering of functional data in a low-dimensional subspace. *Adv Data Anal Classif* **6**, 219–247 (2012)
17. Yamamoto, M., Terada, Y.: Functional Factorial k-Means Analysis. *Comput Stat Data Anal* **79**, 133–148 (2014)
18. Wang, J.: consistent selection of the number of clusters via crossvalidation. *Biometrika* **97**, 4, 893–904 (2010)