
CLADAG2017



Book of Short Papers

Editors: Francesca Greselin,
Francesco Mola and Mariangela Zenga

©2017, Universitas Studiorum S.r.l. Casa Editrice
via Sottoriva, 9 - 46100 Mantova (MN)
P. IVA 02346110204
E-book (PDF version) published in September 2017
ISBN 978-88-99459-71-0

This book is the collection of the Abstract / Short Papers submitted by the authors of the International Conference of The CLAssification and Data Analysis Group (CLADAG) of the Italian Statistical Society (SIS), held in Milan (Italy), University of Milano-Bicocca, September 13-15, 2017.

Euro 9,00

Keynotes

Statistical models for complex extremes

Antony Davison,

Institute of Mathematics,

Ecole Polytechnique Federale de Lausanne, Switzerland

Classified Mixed Model Prediction

J. Sunil Rao,

Division of Biostatistics,

Department of Public Health Sciences, University of Miami,
Florida

An URV approach to cluster ordinal data

Roberto Rocci,

Dipartimento di Economia e Finanza,

Università degli studi di Tor Vergata, Rome, Italy

Invited sessions

Clustering and Dimensionality Reduction

Mixture models for simultaneous classification and reduction of three-way data

Roberto Rocci, Maurizio Vichi, *Monia Ranalli*

High-dimensional Clustering via Random Projections

Laura Anderlucci, *Francesca Fortunato*, Angela Montanari

Clustering and Structural Equation Modeling

Mario Fordellone, Maurizio Vichi

Hidden Markov Models for Longitudinal Data

Package LMest for Latent Markov Analysis of longitudinal categorical data

Francesco Bartolucci, Silvia Pandolfi, Fulvia Pennoni

Dynamic sequential analysis of careers

Fulvia Pennoni, Raffaella Piccarreta

Multivariate hidden Markov regression models with random covariates

Antonio Punzo, Salvatore Ingrassia, *Antonello Maruotti*

Analysis of partially ordered data in socio-economics

Comparing three methodological approaches for synthesizing complex phenomena

Katia Iglesias, Christian Suter, Tugce Beycan, B.P. Vani

New posetic tools for the evaluation of financial literacy

Marco Fattore, Mariangela Zenga

Poset theory and policy making: three case studies

Enrico di Bella

Classification models in Economics and Business

Poland on Global Consumer Markets – Multilevel Segmentation of Countries on the basis of Market Potential Index

Adam Sagan, Eugene Kąciak

Hidden Variable Models for Market Basket Data

Harald Hruschka

Sensitivity Analysis in Corporate Bankruptcy Prediction

Barbara Pawełek, Jozef Pocięcha

Advances in Functional data analysis

Growth processes in forensic entomology: a functional data perspective

Davide Pigoli, John Aston, Frédéric Ferraty

Density based classification methods for functional data

Enea Bongiorno, Aldo Goia

Permutation methods for multi-aspect local inference on functional data

Alessia Pini, Lorenzo Spreafico, Simone Vantini, Alessandro Vietti

New results in Robust estimation

A proposal for robust functional clustering based on trimming and constraints

Luis Angel García-Escudero, Diego Rivera-García, Joaquín Ortega and Agustín Mayo-Iscar

Trimming in probabilistic clustering

Gunter Ritter

Covariance matrices of robust estimators in regression

Silvia Salini, Fabrizio Laurini, Marco Riani, Andrea Cerioli

Innovative applications of multidimensional scaling

Preference judgments of curvature and angularity
in architectural façades

Giuseppe Bove, Nicole Ruta, Stefano Mastandrea

Changes in couples' breadwinning patterns and
wife's economic role in Japan

Miki Nakai

Individual differences in brand switching

Akinori Okada, Hiroyuki Tsurumi

Advances in Robust methods

Weighted likelihood estimation of multivariate location and
scatter

Luca Greco, Claudio Agostinelli

Efficient robust methods for multivariate data via monitoring

Antony C. Atkinson

A new robust estimator of multilevel models based on the
forward search approach

Aldo Corbellini, Luigi Grossi, Fabrizio Laurini

Big Data - Big Knowledge

Flexible Inference for FMRI Data

Aldo Solari

Opinion Mining and City Branding

Federico Neri, Roberto Grandi - Integris

From predictive to reactive approach: how not to be biased from the past in volatile contexts

Federico Stefanato, Marco Cagna - Waterdata

Big data and Design of experiments

Passive and active observation: experimental design issues in big data

Henry Wynn

Optimal design of experiments in the presence of covariate information

Peter Goos

Is it possible a design of experiment with puzzling dynamic data?

Giacomo Aletti

Robust Clustering

Robust clustering tools based on optimal transportation

Eustasio del Barrio

Advances in robust clustering for regression structures

Domenico Perrotta, Francesca Torti, Andrea Cerioli, Marco Riani

Robustness aspects of DD-classifiers for directional data

Giuseppe Pandolfo

Classification and Visualization

Explorative visualization techniques for imbalanced classification tasks

Adalbert Wilhelm

Calibrated cluster validity for comparing the quality of clusterings

Christian Hennig

Visual Tools for Interactive Clustering of UE State Members via Metabolic Patterns

Massimo Aria, Carmela Iorio, Roberta Siciliano, Michele Staiano

Designing clinical trials

The rise of early phase clinical trials

Nancy Flourney

Adaptive dose-finding designs to identify multiple doses that achieve multiple response targets

Adrian Mander

A new design strategy for hypothesis testing under response adaptive randomization

Maroussa Zagoraiou, Alessandro Baldi Antognini, Alessandro Vagheggin

Advances in Credit Risk modelling

Incorporating heterogeneity and macroeconomic variables into multi-state delinquency models for credit cards

Viani Djeundje, Jonathan Crook

Scoring models for P2P lending platforms: an evaluation of predictive performance

Paolo Giudici, Branka Hadij-Misheva

Advances in risk measurement in a distressed banks scenario

Mauro Bernardi, Roy Cerqueti, Arsen Palestini

Heterogeneity and new statistical models

Fitting Cluster-Weighted Models in R

Angelo Mazza, *Antonio Punzo*, Salvatore Ingrassia

Outcome evaluation in healthcare: The Multilevel Logistic Cluster Weighted Model

Paolo Berta, Fulvia Pennoni, Veronica Vinciotti

Mixture model under overlapping clusters: an application to network data

Saverio Ranciati, *Veronica Vinciotti*, Ernst Wit

A world of data

Active and passive measurement: a paradigm change

Giorgio Licastro - GFK Eurisko

Data Science approach and challenges in private sectors

Rocco Michele Lancellotti - Data Reply

Analytics Data LAB: The power of Big Data Investigation and Advanced Analytics to maximize the Data Capital

Roberto Falcinelli - Oracle

Advances in Biostatistics

Regression models for the restricted residual mean time for right-censored and left-truncated data

Giuliana Cortese, Thomas Scheike

Estimating mediation effects in epigenomic studies

Vera Djordjilovic

Statistical challenges in single-cell RNA sequencing

Davide Risso, Fanny Perraudau, Svetlana Gribkova, Sandrine Dudoit, Jean-Philippe Vert

Advances in Ordinal and Preference data

Zero inflated CUB models for the evaluation of leisure time activities

Maria Iannario, Rosaria Simone

Constrained consensus bucket order

Antonio D'Ambrosio, Carmela Iorio, Roberta Siciliano

Ensemble methods for Ranking data

Antonella Plaia, *Mariangela Sciandra*, Roberta Murò

Inequality indices and their decomposition

Contributions from macro-regions and from income components to the Zenga Index $I(Y)$: an application to data from Poland and Italy

Michele Zenga, *Alina Jedrzejczak*, Igor Valli

Joint decomposition by subpopulations and sources of the point and synthetic Bonferroni inequality measures

Michele Zenga, *Igor Valli*

Transfers between sources and units in Zenga's inequality index decomposition

Alberto Arcagni

Advances in Classification and Clustering of complex Data

Combined methods in multi-label classification algorithms

Luca Frigau, Claudio Conversano, Francesco Mola

Validation of Experiments Involving Image Segmentation of Botanic Seeds

Jaromir Antoch, Claudio Conversano, Luca Frigau, Francesco Mola

Time Series Clustering for Portfolio Selection

Carmela Iorio, Antonio D'Ambrosio

Preferences in freshly graduates recruiting

Academic disciplines as perceived by entrepreneurs

Luigi Fabbris, *Manuela Scioni*

University and tourism. Graduates' profiles for the tourism sector

Antonio Giusti, Laura Grassini, Manuela Scioni

The effect of the firm size in the selection of recruitment for new graduate

Franca Crippa, Paolo Mariani, Andrea Marletta, *Mariangela Zenga*

Network Analysis with applications on biological, financial and social networks

Co-authorship Network in Statistics: methodological issues and empirical results

Susanna Zaccarin, Maria Prosperina Vitale, Domenico De Stefano

Network inference in genomics

Ernst Wit

Spatial modeling of brain connectivity data

Daniele Durante, Emanuele Aliverti

Contributed sessions

Classification of Multiway and Functional Data

A generalized Mahalanobis distance for the classification of functional data

Andrea Ghiglietti, Francesca Ieva, Anna Maria Paganoni

Classification methods for multivariate functional data with applications to biomedical signals

Andrea Martino, Andrea Ghiglietti, Anna M. Paganoni

A new Biclustering method for functional data: theory and applications

Jacopo Di Iorio, Simone Vantini

A leap into functional Hilbert spaces with Harold Hotelling

Alessia Pini, Aymeric Stamm, *Simone Vantini*

Sampling Designs and Stochastic models

Statistical matching under informative probability sampling

Daniela Marella, Danny Pfeffermann

Goodness-of-fit test for discrete distributions under complex sampling design

Pier Luigi Conti

Structural learning for complex survey data

Daniela Marella, *Paola Vicard*

The size distribution of Italian firms: an empirical analysis

Anna Maria Fiori, Anna Motta

Robust statistical methods

New proposal for clustering based on trimming and restrictions

Luis Angel García Escudero, Francesca Greselin, *Agustin Mayo Iscar*

Wine authenticity assessed via trimming

Andrea Cappelletto, Francesca Greselin

Robust and sparse clustering for high-dimensional data

Sarka Brodinova, Peter Filzmoser, Thomas Ortner, Maia Zaharieva, Christian Breiteneder

M-quantile regression for multivariate longitudinal data

Marco Alfo', *Maria Francesca Marino*, Maria Giovanna Ranalli, Nicola Salvati, Nikos Tzavidis

New proposals in Clustering methods

Reduced K-means Principal Component Multinomial Regression for studying the relationships between spectrometry and soil texture

Pietro Amenta, *Antonio Lucadamo*, Antonio Pasquale Leone

Comparing clusterings by copula information based distance

Marta Nai Ruscone

Fuzzy methods for the analysis of psychometric data

Isabella Morlini

Inverse clustering: the paradigm, its meaning, and illustrative examples

Jan W. Owsinski, Jaroslaw Stanczak, Karol Opara, Slawomir Zadrozny

Big data mining and classification

The importance of the minorities' viewpoints: Rare Event Sampling Technique on Sentiment analysis supervised algorithm

Marika Arena, *Anna Calissano*, Simone Vantini

A generalized K-means algorithm for multivariate big data with correlated components

Giacomo Aletti, *Alessandra Micheletti*

Big data process analysis: from data mining to process mining

Massimiliano Giacalone, Carlo Cusatelli, Roberto Casadei, Angelo Romano, Vito Santarcangelo

Semiparametric estimation of large conditional variance-covariance and correlation matrices with an application to financial data

Claudio Morana

Advances in model-based clustering

Probabilistic Distance Algorithm generalization to Student's t mixtures

Christopher Rainey, Cristina Tortora, *Francesco Palumbo*

Model-based Clustering of Data with Measurement Errors

Michael Fop, Thomas Brendan Murphy, Lorraine Hanlon

Gaussian Mixture Modeling Under Measurement Uncertainty

Volodymyr Melnykov, Shuchismita Sarkar, Rong Zhengi

A dynamic model-based approach to detect the trend of Statistics from 1970 to 2015

Laura Anderlucci, Angela Montanari, Cinzia Viroli

Bayesian methods and networks

Non parametric Bayesian Networks for measurement error detection

Daniela Marella, Paola Vicard, *Vincenzina Vitale*

Sparse Naïve Bayes Classification

Rafael Blanquero, Emilio Carrizosa, Pepa Ramírez-Cobo, *M. Remedios Sillero-Denamiel*

A Constraint-based Algorithm for Nonparanormal Data

Flaminia Musella, *Paola Vicard*, Vincenzina Vitale

Interventional data and Markov equivalence classes of DAGs

Federico Castelletti, Guido Consonni

Categorical data analysis

Study of context-specific independencies through Chain Stratified Graph Models for categorical variables

Federica Nicolussi, Manuela Cazzaro

Redundancy Analysis Models with Categorical Endogenous Variables: A New Estimation Technique

Gianmarco Vaccà

Mixture of copulae based approach for defining the subjects distance in cluster analysis

Andrea Bonanomi, Marta Nai Ruscone, Silvia Angela Osmetti

Dissimilarity profile analysis for assessing the quality of imputation in cardiovascular risk studies

Nadia Solaro

Data Analysis

Measuring vulnerability: a Structural Equation Modelling approach

Ambra Altimari, *Simona Balzano*, Gennaro Zezza

On the turning point detection in financial time series

Riccardo Bramante, *Silvia Facchinetti*

Optimization of the Listwise Deletion Method

Graziano Vernizzi, Miki Nakai

Discretization of measures: an IRT approach

Silvia Golia

Mixture and Latent Class Models for Clustering

Analysis of university teaching quality merging student ratings with professor characteristics and opinions

Francesca Bassi, Leonardo Grilli, Omar Paccagnella, *Carla Rampichini*, Roberta Varriale

Clustering technique for grouped survival data with a nonparametric frailty term

Francesca Gasperoni, Francesca Ieva, Anna Maria Paganoni, Chris Jackson, Linda Sharples

A latent trajectory model for migrants' remittances: an application to the German Socio-Economic Panel data

Silvia Bacci, Francesco Bartolucci, Giulia Bettin, *Claudia Pigini*

Stepwise latent Markov modelling with covariates in presence of direct effects

Roberto Di Mari, Zsuzsa Bakk

Network analysis

[Non-parametric inference for network-valued data](#)

Ilenia Lovato, Alessia Pini, Aymeric Stamm, Simone Vantini

[Applying network analysis to online news big data](#)

Giovanni Giuffrida, Simona Gozzo and Francesco Mazzeo Rinaldi, *Venera Tomaselli*

[Interval Regression Analysis for the Representation of the Core-Periphery Structure on Large Networks](#)

Carlo Drago

[A Latent Space Model for Multidimensional Networks](#)

Silvia D'Angelo, Thomas Brendan Murphy and Marco Alfò

Advances in LMs, GLMs and PCA

[Bootstrap prediction intervals in linear models](#)

Davide Passaretti, Domenico Vistocco

[Bayesian Variable Selection in Linear Regression Models with non-normal Errors](#)

Saverio Ranciati, Giuliano Galimberti and Gabriele Soffritti

[Principal Component Analysis: the Gini Approach](#)

Stéphane Mussard, *Téa Laurent Jérôme Akeywidi Ouraga*

[On Proportional Odds Modelling and Marginal Effects of Sardinian Hotels data](#)

Giulia Contu, Claudio Conversano, Thomas W. Yee

Advances in Classification

Using PAM and DTW for time series classification

Ilaria Lucrezia Amerise

On Support Vector Machines under multiple-cost scenario

Sandra Benítez-Peña, Rafael Blanquero, Emilio Carrizosa,
Pepa Ramírez-Cobo

Macroeconomic forecasting: a non-standard optimisation
approach to the calibration of dynamic factor models

Fabio Della Marra

Classification of Textual Data

Measuring popularity from Twitter

Farideh Tavazoee, Claudio Conversano, Francesco Mola

A Gamification Approach to Text Classification in R

Giorgio Maria Di Nunzio

From unstructured data and word vectorization to meaning:
Text mining in Insurance

Mattia Borrelli, Diego Zappa

Gamlss for Big Data: ROC curve prediction using Twitter data

Paolo Mariani, *Andrea Marletta*, Mariangela Sciandra

Evaluation in Education

Nonparametric mixed-effects model for unsupervised classification in the Italian education system

Chiara Masci, Francesca Ieva, Anna Maria Paganoni, Tommaso Agasisti

Multivariate mixed models for assessing equity and efficacy in education. An analysis over time using EU15 PISA data

Isabella Sulis, Francesca Giambona, Mariano Porcu

A zero-inflated beta regression model for predicting first-year performance in university career

Matilde Bini, Lucio Masserini

Students' satisfaction in higher education: how to identify courses with low-quality teaching

Marco Guerra, *Francesca Bassi*, José G. Dias

Statistical models for complex data

Spatial Survival Models for Analysis of Exocytotic Events on Human beta-cells Recorded by TIRF Imaging

Thi Huong Phan, Giuliana Cortese

Testing different structures of spatial dynamic panel data models

Francesco Giordano, Massimo Pacella, *Maria Lucia Parrella*

Identification of earthquake clusters through a new space-time-magnitude metric

Renata Rotondi, Antonella Peresan, Stefania Gentili, Elisa Varin

A circular density strip plot

Davide Buttarazzi, Giovanni Camillo Porzio

Mixture Models

A special Dirichlet mixture model for multivariate bounded responses

Agnese Maria Di Brisco, Sonia Migliorati

Cluster-Weighted Beta Regression

Marco Alfò, Luciano Nieddu, Cecilia Vitiello

A Special Dirichlet Mixture Model in a Bayesian Perspective

Roberto Ascari, Sonia Migliorati, Andrea Ongaro

Advances in data Analysis

Assessing Heterogeneity in a Matching Estimation of Endogenous Treatment Effect

Maria Gabriella Campolo, Antonino Di Pino, Edoardo Otranto

Template matching for hospital comparison: an application to birth event data in Italy

Massimo Cannas, Paolo Berta, Francesco Mola

On variability analysis of evolutionary algorithm-based estimation

Manuel Rizzo

Poster session

Accounting for Model Uncertainty in Individualized Designs for Discrete Choice Experiments

Eleonora Saggini, Laura Deldossi, Guido Consonni

Financial-literacy: Socio-demographic variables versus environment

Doriana Cuccinelli, Paolo Trivellato, Mariangela Zenga

Joint models for survival and bivariate longitudinal data: a likelihood formulation

Marcella Mazzoleni, Mariangela Zenga

A Spatial and model-based approach to identify the effect of cultural capital on high school dropout. The Italian case.

Stefano Barberis, Enrico Ripamonti

M-quantile Regression in small area estimation: estimation and testing

Annamaria Bianchi, Enrico Fabrizi, Nicola Salvati, Nikos Tzavisis