



17th Conference of the
International Federation
of Classification Societies

Classification and Data Science in the Digital Age

Book of Abstracts

v



IFCS 2022

Book of Abstracts

17th Conference of the International Federation of
Classification Societies

Classification and Data Science
in the Digital Age

PORTO, PORTUGAL

Title: Classification and Data Science in the Digital Age - Book of Abstracts IFCS
2022
Authors: CLAD - Associação Portuguesa de Classificação e Análise de Dados
Cover design: exclamação!
Printed in Portugal by Instituto Nacional de Estatística
ISBN: 978-989-98955-9-1
N. DL: 500335/22
Number of copies: 350

Preface

Welcome to the 17th conference of the IFCS, IFCS 2022, held in Porto, Portugal, from July 19 to July 23, 2022 and the first IFCS conference held in Portugal. It is a joint organisation of the Portuguese Association for Classification and Data Analysis, CLAD, and the Faculty of Economics of the University of Porto, FEP-UP.

IFCS 2022 is preceded by two half-day tutorials, one on *Analysis of Data Streams* by João Gama, and another on *Categorical Data Analysis and Visualization* by Rosaria Lombardo and Eric J Beh, features four keynote speakers, five invited and seventy contributed sessions organised in specific topics. The Benchmarking Challenge, the Awards Session and the Presidential Address are also noteworthy. Overall, the call for papers attracted 280 submissions, representing 42 countries and 578 different authors. The authors come from five continents, being the largest representation from Europe (68%), followed by North America (12%). Additionally to the rich scientific program the LOC has organised a number of social appealing events that will be memorable.

The 17th conference of the IFCS would not have been possible without the support of many individuals and organisations. We owe special thanks to the authors of all the submitted papers, the members of the program committee, and the reviewers for their contributions to the success of the conference. Finally, we acknowledge the institutional and industrial sponsors that contributed to the organisation of the conference. In particular, we thank all those at FEP-UP who enthusiastically supported the conference from the very start, contributing to its success.

This book contains the abstracts corresponding to all the presentations at the conference. It is organised in seven parts, according to the type of session. Within each session the abstracts are ordered according to the programme. The book includes also an author index.

It has been a pleasure and an honor to organise and host IFCS 2022 in Porto. It is our wish that all participants enjoy the scientific program as well as the the social events and the city of Porto and Portugal.

July 2022

The Local Organising Committee

Organisation

Scientific Programme Committee

Ex-Officio

Paula Brito	Co-Chair	Universidade do Porto, Portugal
José G. Dias	Co-Chair	ISCTE- Instituto Universitário de Lisboa
Angela Montanari	IFCS President	Università di Bologna, Italy
Berthold Lausen	IFCS Past President	University of Essex, UK

Representatives of IFCS Member Societies

Theodore Chadjipadelis	GSDA
Brian Franczak	CSNA
Krzysztof Jajuga	SKAD
Hyunjoong Kim	KCS
Simona Korenjak-Černe	SSS
Koji Kurihara	JCS
Francesco Mola	CLADAG
Ahmed Moussa	MCSO
Fionn Murtagh	BCS
Mohamed Nadif	SFC
Mark de Rooij	VOC
Niël le Roux	MDAG
Eva Boj del Val	SEIO
Arthur White	IPRCS
Adalbert Wilhelm	GfKI
Javier Trejos Zelaya	SoCCCAD

Additional Members

Agustín Mayo-Íscar	Geoffrey McLachlan
André Carvalho	Glòria Mateu-Figueras
Anuska Ferligoj	Hans Kestler
Carlos Soares	José Antonio Vilar Fernández
Cinzia Viroli	Karell Bertet
Eyke Hüllermeier	Laura M. Sangalli
Francisco de Carvalho	Laura Palagi

Lazhar Labiod
 Luis Ángel García Escudero
 Panduranga Nagabhushan
 Pedro Campos
 Peter Filzmoser
 Rosanna Verde
 Rosaria Lombardo

Salvatore Ingrassia
 Satish Singh
 Sibel Kazak
 Veronica Piccialli
 Vladimir Batagelj

Reviewers

Adalbert Wilhelm
 Agustín Mayo-Íscar
 Alípio Jorge
 André C. P. L. F. de Carvalho
 Ann Maharaj
 Anuška Ferligoj
 Arthur White
 Berthold Lausen
 Brian Franczak
 Carlos Soares
 Christian Hennig
 Conceição Amado
 Eva Boj del Val
 Francesco Mola
 Francisco de Carvalho
 Geoff McLachlan
 Gilbert Saporta
 Glòria Mateu-Figueras
 Hans Kestler
 Hélder Oliveira
 Hyunjoong Kim
 Jaime Cardoso
 Javier Trejos
 Jean Diatta
 José A. Lozano
 José A. Vilar
 José Matos

Koji Kurihara
 Krzysztof Jajuga
 Laura Palagi
 Laura Sangalli
 Lazhar Labiod
 Luis Ángel García-Escudero
 Luis Teixeira
 M. Rosário Oliveira
 Margarida G. M. S. Cardoso
 Mark de Rooij
 Michelangelo Ceci
 Mohamed Nadif
 Niël Le Roux
 Paolo Mignone
 Patrice Bertrand
 Pedro Campos
 Pedro Duarte Silva
 Pedro Ribeiro
 Peter Filzmoser
 Rosanna Verde
 Rosaria Lombardo
 Salvatore Ingrassia
 Satish Singh
 Simona Korenjak-Černe
 Theodore Chadjipadelis
 Veronica Piccialli
 Vladimir Batagelj

Local Organising Committee

Paula Brito (Chair)	Universidade do Porto
Adelaide Figueiredo	Universidade do Porto
Carlos Abreu Ferreira	Instituto Politécnico do Porto
Carlos Marcelo	INE
Conceição Rocha	INESC TEC
Fernanda Figueiredo	Universidade do Porto
Fernanda Sousa	Universidade do Porto
Jorge Pereira	Universidade do Porto
Maria Eduarda Silva	Universidade do Porto
Paulo Teles	Universidade do Porto
Pedro Campos	Universidade do Porto
Pedro Duarte Silva	Universidade Católica - Porto
Sónia Dias	Instituto Politécnico de Viana do Castelo

Off-target Predictions in CRISPR-Cas9 Gene Editing Using Machine Learning	265
Ali Mertcan Kose and Ozan Kocadağlı	
Comparison of k-mer and Alignment-based Pre-processing Approaches for Machine Learning Based Functional Annotation with 16S rRNA Data	266
Rafal Kulakowski, Adi Lausen, Etienne Low-Decarie, and Berthold Lausen	
An Ultrametric Model for Clustering and Dimensionality Reduction	267
Giorgia Zaccaria	
Combining Latent Class Analysis and Multiple Correspondence Analysis	268
Alice Barth	
Simultaneous Factorial Reduction and Clustering for Three-mode Data Sets: a Comparison	269
Prosper Ablordeppey, Adelaide Freitas, Maurizio Vichi, and Giorgia Zaccaria	
Clustering Intensive Longitudinal Data Through Mixture Multilevel Vector-autoregressive Modeling	270
Anja Ernst, Marieke Timmerman, Feng Ji, Bertus Jeronimus, and Casper Albers	
Mispecification Tests for Hidden Markov Models Based on a New Class of Finite Mixture Models	271
Francesco Bartolucci, Silvia Pandolfi, and Fulvia Pennoni	
Natural Cubic Smoothing Splines for Latent Class Identification in Longitudinal Growth Trajectories	272
Katerina M. Marcoulides and Laura Trinchera	
Supervised Classification via Neural Networks for Replicated Point Patterns	273
Kateřina Pawlasová, Iva Karafiátová, and Jiří Dvořák	
Reliability Assessment of Ancient Stone Arch Bridge Applying ANN models, Case Study: Leça Railway Bridge	274
Edward A. Baron, Ana Margarida Bento, José Campos e Matos, Rui Calçada, and Kenneth Gavin	
Application of Artificial Intelligence (AI) in Flood Risk Forecasting	275
Minh Quang Tran, Ana Margarida Bento, Elisabete Teixeira, Hélder Sousa, and José Campos e Matos	
Logistic Regression with Sparse Common and Distinctive Covariates . . .	276
Soogeun Park, Eva Ceulemans, and Katrijn Van Deun	

Mispecification Tests for Hidden Markov Models Based on a New Class of Finite Mixture Models

Francesco Bartolucci, Silvia Pandolfi, and Fulvia Pennoni

In the context of longitudinal data, we show that a general class of hidden Markov (HM, [1]) models may be equivalent to a class of finite mixture (FM, [3]) models based on an augmented set of components and suitable constraints on the conditional response probabilities, given these components. We formulate a misspecification test for the latent structure of an HM model comparing maximum likelihood values of the two models for the same data, and when the number of possible latent state sequences is excessive, we propose a multiple version of this test including the Bonferroni correction. The procedure is simple since it is based on the output of the Expectation-Maximization estimation algorithm [2]. The properties of this testing procedure are evaluated through a simulation study. An empirical application illustrates it through data from the National Longitudinal Survey of Youth, in which we jointly consider wages and years of experience after labour force entry. We show that the proposed testing procedure may also be used as an alternative model selection criterion for the number of latent states of an HM model to those usually employed.

Keywords: expectation-maximization algorithm; likelihood ratio test; model selection; multiple testing

References

1. Bartolucci, F., Farcomeni, A., Pennoni, F.: *Latent Markov Models for Longitudinal Data*. Boca Raton FL: Chapman & Hall/CRC. Taylor & Francis (2013)
2. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Stat. Soc.* **39**, 1–38, (1977)
3. McLachlan, G., Lee, S.X., Rathnayake, S.I.: Finite mixture models. *Annu. Rev. Stat. Appl.* **10**, 355–378 (2019)

Francesco Bartolucci and Silvia Pandolfi

Department of Economics, University of Perugia, Via A. Pascoli, 20, 06123 Perugia, Italy,
e-mail: francesco.bartolucci@unipg.it, e-mail: silvia.pandolfi@unipg.it

Fulvia Pennoni

Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Via degli Arcimboldi, 8, 20126 Milano, Italy, e-mail: fulvia.pennoni@unimib.it