

[WEBSITE OF CDW21](#)

Because the next CoDa Association conference, CoDaWork 2021, has been postponed to 2022 due to the Covid pandemics, the Toulouse team has decided to organize for the first time an online only event in 2021. It will consist in five invited talks and will take place on Thursday June 17, 2021, 2:00 pm to 5:00 pm (GMT+2).

One of the goals is to foster the interest in the study of compositional data and their applications, to offer a forum of discussion for people concerned with the statistical treatment and modeling of compositional data or other constrained data sets, and the interpretation of models or applications involving them. Compositional data are typically defined as vectors of positive components representing parts of a whole which carry relative information. Frequently, these compositional vectors have constant sum, usually 100% or 1. These conditions render most classical statistical techniques incoherent on compositions, as they were devised for random variables with the real space as sample space. Aitchison introduced the log-ratio approach to analyze compositional data back in the eighties. His solution was based on transforming the data vector with some log-ratio transformations and applying classical techniques to the scores so obtained. This became the foundation of modern compositional data analysis, nowadays based on an appropriate geometric structure for the simplex and an appropriate representation of the sample space of compositional vectors.

Examples of application abound in different fields of sciences: chemistry, earth and environmental sciences, life sciences and medicine with genome and microbiome research, economy, social statistics and social sciences. The following themes have been selected for the five invited talks of 2021 online CoDa-Day: geosciences, social sciences, microbiome, time-use, and theory.

SCHEDULE

02:00 pm - Introduction

02:05 pm – 02:35 pm - **Alessandra Menafoglio**

"Object Oriented Spatial statistics in Bayes spaces: from distributional data to phase-amplitude variability"

02:35 pm – 03:05 pm - **Berta Ferrer Rosell**

"Compositional analysis of tourism-related content"

03:05 pm - 03:35 pm - **Jennifer McKinley**

"Compositional analysis using balances of geochemical environmental toxins to explore potential associations with chronic kidney disease"

03:35 pm - 04:05 pm - **Luz Calle Rosingana**

"Variable selection in microbiome compositional data analysis"

04:05 pm - 04:35 pm - **Zeljko Pedisic**

"Should compositional data analysis be used in time-use epidemiology"

04:35 pm – 04:40 pm - Conclusion

From 04:40 pm - **Chat rooms**



INVITED SPEAKERS

Invited speaker 1: [Alessandra Menafoglio](#) - [Homepage](#)

Title: “Object Oriented Spatial statistics in Bayes spaces: from distributional data to phase-amplitude variability”

Abstract: In the presence of increasingly massive and heterogeneous spatial data, geostatistical modeling of distributional observations plays a key role. Choosing the ‘right’ embedding space for these data is of paramount importance for their statistical processing, to account for their nature and inherent constraints. The Bayes space theory is a natural embedding space for (spatial) distributional data, and was successfully applied in varied settings. In this presentation, we shall discuss the state-of-the-art methods for the analysis and prediction of georeferenced distributional data, when their spatial dependence cannot be neglected. We shall consider the viewpoint of object-oriented spatial statistics (O₂S₂), a system of ideas for the analysis of complex data with spatial dependence. In the broad context of O₂S₂, a focus will be also made on the analysis of spatial functional data (not necessarily distributional) characterized by both phase and amplitude variability, showing that the Bayes space theory provides a natural embedding for the so-called warping functions, that are used to assess the possible phase variability of the data. All the theoretical developments will be illustrated through their application on real environmental data, highlighting the intrinsic challenges of O₂S₂ and of the Bayes spaces approach.

Invited speaker 2: [Berta Ferrer Rosell](#) - [Homepage](#)

Title: “Compositional analysis of tourism-related content”

Abstract: CoDa has started to be used in several fields of Social Sciences, which often face compositional research questions. For instance, in marketing or communication, typical research questions are related either to the distribution of a whole (market share, product portfolio, consumer spending distribution) or to the relative importance (e.g., advertising content or style, preferred product attributes, dominance of contents posted on different sources). In the particular field of tourism, how tourism destinations allocate resources to their product portfolios or how tourists visiting the destination distribute their tourism budget in activities with different economic impacts at the destination (undertaking different type of cultural activities, consuming at local restaurants, etc.) are questions researchers would be interested in studying. Another example are the contents posted by managers in many platforms, which may affect both consumers’ choices and satisfaction. This talk focusses on the analysis of content posted by destinations, companies, tourists, media, travel guides on several sources such as social media (e.g., Facebook), corporative websites, user opinion platforms (e.g., Trip Advisor), and also considering the type of content posted (e.g., destination image attributes, complaints, emotions), as well as, considering the type of language used. It shows how the CoDa tools (compositional distance, biplot and MANOVA) have been applied to respond to tourism researchers’ interests.

Invited speaker 3: [Jennifer McKinley - Homepage](#)

Title: “Compositional analysis using balances of geochemical environmental toxins to explore potential associations with chronic kidney disease”

Abstract: Digital spatial data can be used to explore potential relationships between naturally occurring geogenic elements in soil and water, potentially toxic elements (PTEs) and long term cumulative exposure which may be linked to chronic disease. However, spatial data sets such as geochemical survey data pose many challenges for exploratory data analysis. Geochemical data are compositional in nature in that they convey relative information. As a result, compositional data analysis (CoDA) methods are frequently used to extract information from geochemical data by treating log ratio or equivalently transformed data instead of analysing the raw constant sum values. Using an urban soil geochemistry database of total element concentrations of potentially toxic elements (PTEs), and the UK Renal Registry which provided Standardised Incidence Rates (SIRs) of Chronic Kidney Disease (CKD), a statistical relationship was found between CKD of uncertain aetiology (CKDu) and environmental toxins. This study investigates the use of a compositional balance approach to determine the relative influence of geogenic and anthropogenic sources on the urban geochemistry signature. Joint with Ute Mueller(Edith Cowan University), Raimon Tolosana Delgado(TU Bergakademie Freiberg), Peter M. Atkinson(Queen’s University Belfast - TU Bergakademie Freiberg), Ulrich Offerdinger(Queen’s University Belfast), Siobhan F. Cox(Queen’s University Belfast), Rory Doherty(Queen’s University Belfast), Damian Fogarty(Lancaster University), J.J. Egozcue(Belfast Health Trust) and V. Pawlowsky-Glahn(U. Polit cnica de Catalu a (UPC))

Invited speaker 4: [Luz Calle Rosingana - Homepage](#)

Title: “Variable selection in microbiome compositional data analysis”

Abstract: Though variable selection is one of the most relevant tasks in microbiome analysis, e.g. for the identification of microbial signatures, many studies still rely on methods that ignore the compositional nature of microbiome data. The applicability of compositional data analysis methods has been hampered by the availability of software and the difficulty in interpreting their results. In this talk I will focus on three methods for variable selection that acknowledge the compositional structure of microbiome data: selbal, a forward selection approach for the identification of compositional balances, and clr-lasso and codalasso, two penalized regression models for compositional data analysis. I will discuss the link between these methods and their advantages and limitations for variable selection in the context of microbiome studies.

Invited speaker 5: [Zeljko Pedisic - Homepage](#)

Title: “Should compositional data analysis be used in time-use epidemiology?”

Abstract: Time use is a health-related factor that every single person is inevitably exposed to every day, 24 hours a day. From a public health perspective, it is therefore important to investigate determinants, incidence, distribution, and effects of health-related time-use patterns in the population, and methods for preventing unhealthy time use and achieving the optimal distribution of time for population health. This presentation will include: [i] a summary of the framework for Viable Integrative Research in Time-Use Epidemiology (VIRTUE); [ii] examples of the use of compositional data analysis (CoDA) in time-use research; and [iii] highlights from a recent debate between a strong proponent and a vigorous opponent of the use of CoDA. The presentation will conclude with recommendations on what to do to facilitate (or discourage) the use of CoDA in time-use epidemiology.



VITA



Alessandra Menafoglio is a Senior Assistant Professor in Statistics at the Department of Mathematics of the Politecnico di Milano, within the laboratory for modeling and scientific computing (MOX). She obtained her PhD in Mathematical Models and Methods in Engineering in 2015 at Politecnico di Milano. Her research interests focus on the study of innovative statistical models and methods for the analysis of complex and large data (such as curves and images), with particular emphasis on spatially distributed complex data and functional compositional data. In the last years, she focused on the problems of modeling, prediction (kriging) and stochastic simulation for very general types of data, in the context of Object Oriented Spatial Statistics, with application to Earth and Environmental Sciences. Further scientific interests regards the field of statistical process control in modern industrial production (e.g., in additive manufacturing), when the response to be monitored is a complex shape, or a high-dimensional object. Her doctoral thesis was awarded in 2016 with the "Eni Award, Debut in Research Prize". Recently, she was awarded the 2019 Andrei Borisovich Vistelius Research Award by the IAMG.

She is Associate Editor of the journals Stochastic Environmental Research and Risk Assessment and Applied Computing and Geosciences, and she is part of the Editorial Board of the journal Mathematical Geosciences.



Berta Ferrer Rosell. She holds an international PhD in Tourism (Excellent Cum Laude and Extraordinary Award), a MSc in Tourism Management and Planning and a BA in Tourism. She is currently senior lecturer (Serra Húnter Fellow) at the University of Lleida (Catalonia – Spain) and her current research interests cover the analysis of compositional data on electronic tourism and electronic marketing. She defines herself as CoDa user, feeling greatly comfortable applying the method and thus approaching it to applied tourism (and other social sciences) researches. She was the first worldwide researcher in using CoDa in the field of tourism, to analyse tourist's trip budget composition. Afterwards she moved into the marketing area and e-tourism. In these areas, she has been invited to talk about CoDa in Tourism in several conferences (e.g., COMPSTAT 2018), as well as teaching the session "Introduction to Compositional Data Analysis & Applications to e-tourism" at the European IFITT Masterclass on e-tourism, and co-authored chapters in books related to e-tourism methods and research methods for marketing management with contributions using CoDa. Coordinator of the Joint Degree in Tourism and

Business Management at the University of Lleida, Associate Editor of the European Journal of Tourism Research, board member of the Spanish Association of Scientific Experts in Tourism (AECIT) and chair of the PhDWorkshop at conference ENTER21 (IFITT). Her CoDa research has been published in Q1JCR journals such as Tourism Management, Journal of Destination Marketing & Management, International Journal of Hospitality Management and Annals of Tourism Research.



Jennifer McKinley - School of Natural and Built Environment, Geography, Queen's University Belfast. Jennifer is Director of the Centre for GIS and Geomatics and Director of the Environmental Change and Resilience Research Cluster, QUB. Her research focuses on the application of spatial analysis techniques, including geostatistics, compositional data analysis and Geographical Information Science (GIS), to soil geochemistry, environmental and criminal forensics, human health, geohazards and weathering studies.

Jennifer holds a number of international roles including Councillor of the International Union of Geological Sciences (IUGS) , Past President of the International Association of Mathematical Geoscientists (IAMG), member of the Royal Irish Academy (RIA) Geosciences and Geographical Sciences committee and Communications Officer for the Initiative on Forensic Geology (IFG). Throughout her career Jennifer has taken an active role in promoting equality, diversity and inclusion in academia and has contributed to this globally. Interdisciplinary collaboration and strong partnership working underpins her commitment to high impact global research.



Luz Calle Rosingana. She is Full Professor of Biostatistics and Bioinformatics and Head of the Biosciences Department, University of Vic – Central University of Catalonia. With a background in Mathematics (BSc Mathematics, Universitat de Barcelona, 1986 and PhD in Mathematics, Universitat Politècnica de Catalunya, 1997), she teaches biostatistics, bioinformatics and genetic epidemiology in the Biotechnology degree and is the chair of the Master of Sciences in Omics Data Analysis, where she teaches the course “Statistical and data-mining methods for omics data analysis”. She is the group leader of the Bioinformatics and Medical Statistics Group of the University of Vic (consolidated group 2014SGR-596). Her main research areas are statistical genetics, omics data analysis, microbiome data analysis and survival analysis. She works on the development of new methods for biomarker discovery, identification of genetic risk profiles and construction of dynamic prediction and prognostic models of disease evolution. She is also interested in statistical methods for integration of multi-omics data and compositional data approaches in metagenomics. She is member of several scientific societies: BiostatNet-Spanish National Network in Biostatistics, Catalan Statistical Society, Spanish Society of Statistics and Operational Research, International Biometric Society, International Genetic Epidemiology Society.



Zeljko Pedisic. He leads the Active Living & Public Health research group at the Institute for Health and Sport, Victoria University, Melbourne, Australia. His research is centred around chronic disease prevention and promotion of well-being through healthy use of time. He is one of the authors of the framework for Viable Integrative Research in Time-Use Epidemiology (VIRTUE). He is Secretary of the International Network of Time-Use Epidemiologists (INTUE) and honorary Docent of Quantitative Methods at the University of Zagreb, Croatia. His publication record includes >80 journal articles. As a methodological consultant or supervisor he contributed to four bachelor, 25 master's and 17 PhD research theses. In his free time, he enjoys spending time with family and friends, playing piano, composing music, playing sports, cycling, skiing, and eating pizza