

ISBIS 2025 CONFERENCE

October 2-3, 2025
University of Amsterdam

ISBIS
INTERNATIONAL SOCIETY FOR BUSINESS AND INDUSTRIAL STATISTICS

UNIVERSITY OF AMSTERDAM

International Statistical Institute

Table of Contents

TABLE OF CONTENTS **2**

ABOUT THE CONFERENCE **3**

SPONSORS **4**

SCIENTIFIC PROGRAM COMMITTEE **5**

LOCAL ORGANIZING COMMITTEE **6**

KEYNOTE SPEAKER: DICK DEN HERTOG **7**

SCHEDULE **8**

- Thursday October 2 8
- Friday October 3 9

ABSTRACTS (THURSDAY OCTOBER 2) **10**

- Keynote Session: Dick de Hertog 10
- Invited Session: International Society for Bayesian Analysis-Industrial Statistics Section 11
- Invited Session: Multivariate Methods in Quality and Process Modelling 12
- Invited Session: Statistical Learning and AI for Societal Challenges 13
- Invited Session: Modern Developments in Time Series I 14
- Invited Session: European Network for Business and Industrial Statistics 15
- Invited Session: Modern Developments in Time Series II 16
- Contributed Session: Statistical & Machine Learning Methodologies 17
- Contributed Session: Applications, Tools & Societal Impact 19

ABSTRACTS (FRIDAY OCTOBER 3) **21**

- Invited Session: *Applied Stochastic Models in Business and Industry* 21
- Invited Session: Advances in Statistical Classification 22
- Invited Session: AI and Machine Learning (ɣ-BIS) 23
- Invited Session: Modeling Complex Data with Business and Industrial Applications 24
- Invited Session: Advanced Statistical Modeling (ɣ-BIS) 25
- Invited Session: Networks and Where to Find Them 26
- Invited Session: Statistical Risk Modelling (ɣ-BIS) 27

MAPS **28**

- Venue Map (Google) 28
- Venue Map (Floorplan) 29
- Banquet Map 30
- Venue to Banquet Directions 31

About the Conference

The International Society for Business and Industrial Statistics (ISBIS) an international society dedicated to the promotion of applications, research, and best practices in business and industrial statistics worldwide; it facilitates technology transfer and fosters communications among members and practitioners around the world. Additionally, ISBIS is part of a family of international associations that operate under the umbrella of the International Statistical Institute (ISI) and this conference is a satellite event of the 65th ISI World Statistics Congress which will be held October 5-9, 2025 in The Hague, The Netherlands.

The 2025 ISBIS Satellite Conference aims to bring together researchers and practitioners from around the world who use statistics in business or industrial contexts. The conference promotes communication among researchers and practitioners to enable and ensure the development and widespread use of novel insights and methodology. The conference focus is on recent advancements in methodology as well as on best practices and innovative applications. As such, the technical program focuses on statistical methodology and creative problem solving to address scientific, industrial, and business challenges, drawing upon advances from the fields of statistics, machine learning, and data science.

The Local Organizing Committee and Scientific Program Committee hope that you enjoy the 2025 ISBIS Satellite Conference and your time in Amsterdam!

Sponsors

The 2025 ISBIS Conference acknowledges the generous financial contributions from the following organizations.







Scientific Program Committee



Nathaniel Stevens
(Chair)
University of Waterloo



Paulo Canas Rodrigues
Federal University of Bahia



Inez Zwetsloot
University of Amsterdam



Daniel Jeske
University of California Riverside



Nalini Ravishanker
University of Connecticut



Luca Frigau
Università degli Studi di Cagliari



Ozan Kocadagli
Mimar Sinan Fine Arts University



Xingdong Feng
Shanghai University of Finance and Economics

Local Organizing Committee



Inez Zwetsloot
(Chair)
University of Amsterdam



Rob Goedhart
University of Amsterdam



Jiaqi Qiu
University of Amsterdam



Yunming Hui
University of Amsterdam

Keynote Speaker: Dick den Hertog



Dr. Dick den Hertog is professor of Operations Research at University of Amsterdam. His research interests cover various fields in prescriptive analytics, in particular linear and nonlinear optimization. In recent years, his main focus has been on robust optimization, and recently he started research on Optimization with Machine Learning. He is also active in applying the theory in real-life applications. In particular, he is interested in applications that contribute to a better society. He received the INFORMS Franz Edelman Award twice: in 2013 for his research on optimal flood protection, and in 2021 for his research on optimizing the food supply chain for the UN World Food Programme. He has been Visiting Professor at MIT for several years now. He is Science-to-Impact Director of the Analytics for a Better World Institute, that he co-founded in 2022.

Schedule

Thursday October 2

	THURSDAY OCTOBER 2	
8:30 - 9:00	**Registration & Coffee** Location: Chirurgisch Theater (A1.01)	
9:00 - 9:15	**Welcome** (Location: Chirurgisch Theater A1.01)	
9:15 - 10:30	**Dick den Hertog Keynote** Location: Chirurgisch Theater (A1.01)	
10:30 - 11:00	**Coffee Break**	
11:00 - 12:15	**ISBA-IS Session** Location: Chirurgisch Theater (A1.01) Organizer: Tahir Ekin | Chair: Tahir Ekin Speakers: Nalini Ravishanker, Silvia Columbu, Thomas A. Mazzuchi	**Multivariate Methods in Quality and Process Modelling** Location: Ruyszaal (D2.08) Organizer: Karel Kupka | Chair: Karel Kupka Speakers: Gejza Dohnal, Huebnerová Zuzana, Karel Kupka
12:15 - 13:15	**Lunch** Location: Chirurgisch Theater (A1.01)	
13:15 - 14:30	**Statistical Learning & AI for Societal Challenges** Location: Chirurgisch Theater (A1.01) Organizer: Luca Frigau | Chair: Matilde Bini Speakers: Giulia Contu, Daniel Jeske, Rahim Mahmoudvand	**Modern Developments in Time Series I** Location: Ruyszaal (D2.08) Organizer: Paulo Canas Rodrigues | Chair: Paulo Canas Rodrigues Speakers: Paulo Canas Rodrigues, Liang-Ching Lin, Ricardo Ehlers
14:30 - 15:00	**Coffee Break**	
15:00 - 16:15	**ENBIS Session** Location: Chirurgisch Theater (A1.01) Organizer: Antonio Lepore | Chair: Antonio Pievatolo Speakers: Inez Zwetsloot, Federico Cortese, Davide Forcina	**Modern Developments in Time Series II** Location: Ruyszaal (D2.08) Organizer: Xingdong Feng | Chair: Yayi Yan Speakers: Yi He, Yingqian Lin, Nathaniel Stevens
16:15 - 16:45	**Break**	
16:45 - 18:00	**Statistical & Machine Learning Methodologies** Location: Chirurgisch Theater (A1.01) Chair: Yunming Hui Speakers: Guilherme Alves Silva, Thiago Gentil Ramires, Marcus A. Agustin, Antonio Pievatolo, Sotiris Bersimis	**Applications, Tools & Societal Impact** Location: Ruyszaal (D2.08) Chair: Jiaqi Qiu Speakers: Ana María Gómez Lamus, Maria Teresa Cappella, Petteri Baer, Cheyenne Amoroso, Eglantina Zyka
19:00 - 21:00	**Banquet @ Indrapura Restaurant** Rembrandtplein 40 1017 CV Amsterdam 020 623 7329	

Plenary	Invited
Contributed	Meals & Social

Friday October 3

	FRIDAY OCTOBER 3	
8:30 - 9:00	**Registration & Coffee** Location: Chirurgisch Theater (A1.01)	
9:00 - 10:30	**ASMBI Session** Location: Chirurgisch Theater (A1.01) Organizer: Nalini Ravishanker | Moderator: Nalini Ravishanker Speaker: Tahir Ekin Discussants: Kathy Ensor, Ozan Kocadagli	
10:30 - 11:00	**Coffee Break**	
11:00 - 12:15	**Advances in Statistical Classification** Location: Chirurgisch Theater (A1.01) Organizer: Daniel Jeske | Chair: Daniel Jeske Speakers: Mia Hubert, Nairanjana Dasgupta, Marco Ortu	**AI and Machine Learning (y-BIS)** Location: Ruyszaal (D2.08) Organizer: Ozan Kocadagli | Chair: Ozan Kocadagli Speakers: Ozan Kocadagli, Atike Ahrabi, Ozge Korukci
12:15 - 13:15	**Lunch** Location: Chirurgisch Theater (A1.01)	
13:15 - 14:30	**Modeling Complex Data with Business and Industrial Applications** Location: Chirurgisch Theater (A1.01) Organizer: Nalini Ravishanker | Chair: Kathy Ensor Speakers: Balaji Raman, Kimberly Sellers, Shibasish Dasgupta	**Advanced Statistical Modeling (y-BIS)** Location: Ruyszaal (D2.08) Organizer: Ozan Kocadagli | Chair: Ozan Kocadagli Speakers: Yunmeng Hui, Burcu Ece Alp, Michael Trequattrini
14:30 - 15:00	**Coffee Break**	
15:00 - 16:15	**Networks and Where to Find Them** Location: Chirurgisch Theater (A1.01) Organizer: Simon Trimborn | Chair: Simon Trimborn Speakers: Simon Trimborn, Wanjie Wang, Lars de Graaf	**Statistical Risk Modeling (y-BIS)** Location: Ruyszaal (D2.08) Organizer: Ozan Kocadagli | Chair: Ozan Kocadagli Speakers: Supurna Biswas, Gulder Kemalbay
16:15 - 16:30	**Closing Remarks** (Location: Chirurgisch Theater A1.01)	

Plenary	Invited
Contributed	Meals & Social

Abstracts (Thursday October 2)

Keynote Session: Dick de Hertog

Thursday October 2 (9:15am-10:30am)

Room: Chirurgisch Theater (A1.01)

Analytics for a Better World

Dick den Hertog, University of Amsterdam

This keynote will showcase the transformative potential of analytics in enabling non-governmental organizations (NGOs) to make a greater impact on the UN's Sustainable Development Goals (SDGs). The Analytics for a Better World Institute (https://analyticsbetterworld.org/) harnesses and advances cutting-edge analytics techniques to unlock this potential. Through real-world applications, this keynote will illustrate how data-driven solutions are revolutionizing humanitarian and environmental efforts, including:

1. The optimization of the food supply chain for the World Food Programme (WFP). By leveraging advanced optimization models, WFP is now able to deliver critical food aid to millions more people.
2. An innovative optimization model for The Ocean Cleanup that strategically steers their vessels to accelerate the removal of plastic from the world's oceans, contributing to the fight against marine pollution.
3. Innovative projects to optimize geospatial accessibility to healthcare services, such as optimizing primary healthcare facility locations in Timor-Leste, determining stroke center placements in Vietnam, planning COVID-19 test centers in Nepal, and identifying ideal water well sites in Sudan. These efforts are driven by collaborations with leading organizations such as the World Bank, World Health Organization, Amref, and the American Red Cross.

In addition to presenting these impactful projects, this keynote will touch on the novel research challenges they have sparked, some of which remain open and ripe for exploration. Critical enablers for analytics to achieve meaningful, scalable impact within the NGO sector will be discussed, setting the stage for further innovation and collaboration.

Invited Session: International Society for Bayesian Analysis-Industrial Statistics Section

Thursday October 2 (11:00am-12:15pm)

Room: Chirurgisch Theater (A1.01)

Organizer: Tahir Ekin, Texas State University

Chair: Tahir Ekin, Texas State University

Forecasting Robust Gaussian Process State Space Models for Assessing Intervention Impact in IoT Time Series

Nalini Ravishanker, University of Connecticut

This talk describes a robust Gaussian prior process state space modeling approach to assess the impact of an intervention in an IoT stream of internal temperatures measured by an insurance firm to address the risk of pipe-freeze hazard in a building. The robustness is achieved by assuming a scale mixture of normal distributions instead of the normal distribution in order to accommodate heavy-tailed behavior or anomalous observations in the time series. Gaussian process (GP) prior models provide flexibility by placing a non-parametric prior on the functional form of the model. By treating the pipe-freeze hazard alert as an exogenous intervention, we use the robust Gaussian prior process state space model for Bayesian fitting and forecasting the internal temperatures. By comparing the forecasts with future data, we assess with a high level of confidence whether an alerted customer took preventive action to prevent pipe-freeze loss. This is joint work with Patrick Toman and Nathan Lally (Hartford Steam Boiler) and Sanguthevar Rajasekaran (CSE, University of Connecticut).

On the statistical comparison of populations: A new Bayesian approach

Silvia Columbu, University of Cagliari

We introduce a novele Bayesian approach for comparing two populations in terms of different parameter functions. Such approach considers the application of the Bayesian Discrepancy Measure (BDM), a measure of evidence that allows the assessment of precise hypotheses in parametric models on the basis of the evidence observed in the posterior distribution. The method is highly flexible, accommodating different distributions and parameter transformations. Possible applications explored are the comparison of coefficients of variation, means, variances and skewness in models of diverse nature. The BDM has several advantageous properties: it is invariant to reparametrizations, does not suffer from the Jeffreys-Lindley paradox, and is asymptotically consistent under regularity conditions making it a robust tool for evaluating hypotheses in the Bayesian framework. This methodology can be applied to address problems that are seldom or never discussed in the literature. This is joint work with Francesco Bertolino, Mara Manca, and Monica Musio (University of Cagliari).

Developing Prior and Predictive Distributions for Project Risk

Thomas A. Mazzuchi, The George Washington University

In this talk we discuss the development of prior and predictive distributions for project risk using the well-known risk matrix format to aid project managers in their assessment and tracking of project risks. To develop the prior, bounded distributions are used to express the expert judgement for potential cost and risk and a Generalized Diagonal Band copula is used to express the known dependence between these two quantities. We discuss various forms of feedback used to train experts on the appropriateness of their choices for distribution and copula parameter choices. Once established, the joint prior distribution for cost and probability can be used to generate cost risk distributions given the suspected number of risks to be tracked. Given a specific set of risks, the cost distribution for the specific set can be developed and the effect of mitigation strategies on the set can be determined. The talk will provide an overview of the theory and illustrative examples. This is joint work with Shahram Sarkani (The George Washington University).

Invited Session: Multivariate Methods in Quality and Process Modelling

Thursday October 2 (11:00am-12:15pm)

Room: Ruyszaal (D2.08)

Organizer: Karel Kupka, University of Liberec

Chair: Karel Kupka, University of Liberec

Stability of weight-updated systems for multivariate Statistical Process Control

Gejza Dohnal, Czech Technical University

In many modern manufacturing and industrial processes, multiple quality variables are measured simultaneously. These variables are often correlated, making traditional univariate SPC methods (such as Shewhart charts) insufficient. Traditional multivariate SPC methods like Hotelling's T^2, PCA, or MEWMA assume linearity and normal distribution, which often do not hold in real-world processes. Neural networks offer flexible, nonlinear modeling capabilities that overcome these limitations. This provides powerful alternatives and enhancements to SPC due to their ability to model complex, nonlinear processes and detect subtle patterns in data. In adaptive SPC, weight-updated systems—such as adaptive neural networks or online learning algorithms—play a vital role. These systems make process monitoring flexible, real-time, and responsive to change. They can adapt to process variations, improve prediction accuracy, support real-time monitoring, and enhance robustness and flexibility. This makes them particularly valuable in modern manufacturing environments, where processes are complex, variable, and data rich. A crucial feature of weight-updated systems is their stability. Stability in adaptive systems refers to the ability of a network to produce consistent and bounded outputs in response to small changes in input, parameters, or over time. In this contribution, we will discuss a condition for the stability of In-Parameter-Linear Nonlinear Neural Architectures (IPLNNAs), which are widely used in weight-updated systems.

Multivariate nonparametric diagnostics of turboprop engines condition

Huebnerová Zuzana, Brno University of Technology

Monitoring the condition of aircraft turboprop engines requires the careful observation of numerous physical parameters recorded during each flight. Continuous data collection and analysis enable the early detection of potential issues, such as incipient damage or fouling, often well before they are identified through routine maintenance procedures. Given the complex interdependencies among engine parameters, combined with the effects of varying atmospheric conditions and diverse flight profiles, thorough pre-processing and examination of the data are essential steps in any meaningful statistical assessment. This paper investigates the application of a depth-based nonparametric method for identifying anomalous measurements in this challenging context. Such anomalies may result from erroneous recordings or reflect actual changes in engine condition. The challenges of implementing this approach for engine diagnostics are discussed, and its performance is compared to that of a regression-based technique. This is joint work with Juracka Jaroslav.

Spectral unmixing in Raman spectroscopy to identify and monitor waste polymers

Karel Kupka, University of Liberec

Monitoring industrial waste becomes major part of quality control. The identification of waste polymers, including microplastics and complex mixtures, is crucial for environmental monitoring and recycling efforts. Raman or infrared spectroscopy, combined with spectral unmixing techniques, provides a powerful tool to resolve overlapping spectral components and characterize polymeric materials. This paper explores advanced methodologies for rotated spectral decomposition of chemical signals, including chemometric approaches and machine learning, to enhance fast polymer identification and classification with subsequent stability analysis using multivariate control charts. This is joint work with Jiri Militky and Dana Kremenakova.

Invited Session: Statistical Learning and AI for Societal Challenges

Thursday October 2 (1:15pm-2:30pm)

Room: Chirurgisch Theater (A1.01)

Organizer: Luca Frigau, Università degli Studi di Cagliari

Chair: Matilde Bini, Università Europea di Roma

Exploring the Relationship Between Doctors Availability and Mortality in Italy: A Machine Learning Approach using Multivariate Regression Trees

Giulia Contu, University of Cagliari

We investigate how the percentages of medical specialist impact mortality rates and how they vary across different areas and time periods. The analysis is focused on Italian provinces during two different years, 2011 and 2019, to identify differences and similarities across geographical areas and time periods. We propose a new approach based on tree-based methodology that we will call Machine learning Enhanced Regression of ITAlian mortality rates (MERITA). It is based on a tree-based method called Multivariate Regression Tree (MRT), which is an extension of univariate regression trees. In this study, the MRT is used to identify the relationship between the percentage of active doctors by specialization per 10,000 inhabitants and the Age-adjusted Mortality Rate (AMR) for specific disease and gender. We define nine models. Two focus on the overall AMR by gender. Seven are defined with specific disease-gender-AMR as response variables. All include the availability of active doctors and other socio-economics aspects as covariates. Our findings highlight that the percentages of medical specialists have a significant impact on the AMRs and how it changes across geographic areas and over time, providing insights on which specialists are most in need and where.

How Artificial Intelligence Can Help Mitigate Bias in Student Evaluations of Teaching

Daniel Jeske, University of California Riverside

This talk addresses bias and classification challenges in the analysis of written comments from student evaluations of teaching (SET). While bias in numerical SET scores is well-documented, less attention has been given to potential biases in open-ended written responses. Written comments from two campuses within the University of California system are analyzed by gender, ethnicity, tenure status, and course type (STEM vs. non-STEM). Findings are that while some combinations of factors favor higher proportions of positive comments, these advantages do not consistently align with gender or race alone. These results suggest that written comments may be less biased and thus more suitable for evaluating instructors compared to numerical scores. To facilitate their use, a machine learning algorithm is developed that predicts whether a written comment is positive, negative, or mixed. Through the incorporation of a neutral zone to address conditional misclassification rates, the algorithm provides a risk-controlled method for efficiently extracting insights from the often-substantial volume of comments. Together, these studies contribute to advancing discussions on equitable and effective teaching evaluation practices in higher education.

Statistical Approaches to Measuring the Relationship between Customer and Employee Satisfaction with Unequal Sample Sizes

Rahim Mahmoudvand, Bu-Ali Sina University and UniCa - Università degli Studidi Cagliari

The relationship between customer satisfaction and employee satisfaction plays a crucial role in organizational success, yet measuring this relationship poses statistical challenges—particularly when the available data consists of two samples of differing sizes. This talk explores the correlation between these two key constructs from a statistical perspective, reviewing existing methods and proposing robust techniques for analyzing satisfaction data under practical constraints. We discuss approaches for dealing with unequal sample sizes, such as aggregation strategies, hierarchical models, and correlation measures suited for paired and unpaired data. Emphasis is placed on the implications of data structure, sampling design, and potential biases that arise in real-world applications. By focusing on the methodological foundations, this talk provides insights for researchers and practitioners interested in quantitatively evaluating the service-profit chain and making informed decisions based on satisfaction metrics.

Invited Session: Modern Developments in Time Series I

Thursday October 2 (1:15pm-2:30pm)

Room: Ruyszaal (D2.08)

Organizer: Paulo Canas Rodrigues, Federal University of Bahia

Chair: Paulo Canas Rodrigues, Federal University of Bahia

Multivariate reconciliation for hierarchical time series

Paulo Canas Rodrigues, Federal University of Bahia

Time series can often be structured hierarchically based on characteristics such as geographic regions. These are known as hierarchical time series. To ensure coherent forecasts, where the forecasts respect the same aggregation structure as the data, predictions are typically generated for all levels of the hierarchy. Several methods have been developed to guarantee this coherence by adjusting a set of initial, or base, forecasts. This adjustment process is known as forecast reconciliation. While most existing approaches focus on univariate time series, multivariate time series can also exhibit hierarchical structures. However, current reconciliation methods are limited to handling one variable at a time. As a result, reconciliation must be applied separately to each variable, ignoring the potential correlations between them. To overcome this limitation, this paper proposes a multivariate reconciliation methodology that ensures coherent forecasts while accounting for inter-variable relationships. The method was evaluated through simulation studies under different correlation scenarios across series and variables and using various base forecasting models. It was also applied to real-world employment data on admissions and dismissals in Brazil. Results show that the proposed multivariate reconciliation approach yields strong performance in both simulated and practical settings. This is joint work with Ana Caroline Pinheiro, Rodrigo de Souza Bulhões, and Rob J. Hyndman.

Comprehensive interval-valued time series model with application to the S&P 500 index and PM2.5 level data analysis

Liang-Ching Lin, National Cheng Kung University

In this study, we develop comprehensive symbolic interval-valued time-series models, including auto-interval-regressive, interval-valued moving average, auto-interval-regressive moving average, and heteroscedastic volatility models. These models can be flexibly combined to adapt more effectively to various situations. To make inferences regarding these models, likelihood functions were derived, and maximum likelihood estimators were obtained. To evaluate the performance of our methods empirically, Monte Carlo simulations and real data analyses were conducted using the S&P 500 index and PM2.5 levels of 15 stations in southern Taiwan. In the former case, it was found that the proposed model in predicting the interval prices outperforms all other existing methods. In the latter case, the residuals deduced from the proposed models provide more intuitively appealing results than conventional vector autoregressive models. Overall, our findings strongly confirm the adequacy of the proposed model. This is joint work with Hsiang-Lin Chien and Hao Sung.

Stochastic volatility in mean using hidden Markov models

*Ricardo Ehlers, University of São Paulo*In this paper, we investigate the Stochastic Volatility in Mean (SVM) model with heavy tails in the family of scale mixture of Normal distributions (SVM-SMN). The likelihood function is approximated by iterated numerical integration and a fast evaluation of the approximate likelihood using recursive techniques routinely applied in Hidden Markov models (HMM). This procedure requires less computational time in simulations and estimates as compared to Markov chain Monte Carlo methods, even when the sample size increases. Simulation studies are conducted to investigate the frequentist properties of Bayesian estimators. Then real data analysis of stock market indexes from US, Europe, China and Latin America was performed. The main results indicate that volatility negatively impacts returns, suggesting that the volatility feedback effect is stronger than the effect related to the expected volatility. This is joint work with Bruno Holtz and Carlos Abanto-Valle.

Invited Session: European Network for Business and Industrial Statistics

Thursday October 2 (3:00pm-4:15pm)

Room: Chirurgisch Theater (A1.01)

Organizer: Antonio Lepore, Università degli Studi di Napoli Federico II

Chair: Antonio Pievatolo, CNR IMATI

Explainable AI signal diagnosis for AI-based statistical process monitoring

Inez Zwetsloot, University of Amsterdam

AI techniques have excelled in many fields, including statistical process monitoring (SPM). Although existing AI-based SPM methods show a promising detection ability in high-dimensional scenarios, post-signal diagnosis for black-box models is often ignored. The lack of explainability makes it difficult to take appropriate actions after a signal. Explainable AI (XAI) techniques, like feature attribution and counterfactual explanations, can help decision-makers investigate these signals. This expository talk provides a perspective on research and use of XAI in statistical process monitoring (SPM). We introduce XAI terminology and methods and focus on how these XAI techniques provide opportunities for enhancing signal diagnosis within monitoring schemes. We review the existing literature and illustrate the use of XAI for signal diagnosis with a case study. We provide our perspective on the challenges and give a number of research ideas. Our primary goal is to introduce the opportunities of XAI for signal diagnosis to the SPM community. This is joint work with Jiaqi Qiu.

Extending Statistical Jump Models: From Mixed-Type Data to Fuzzy Temporal Clustering

Federico P. Cortese, CNR IMATI

Statistical jump models provide a framework for modeling complex dynamics by allowing transitions (“jumps”) between simpler sub-models. These models assume that the evolution of a system is governed by an unobserved discrete latent process that switches between states or regimes. Their key advantages include: (1) interpretability via state-conditional parameter; (2) flexibility in handling high-dimensional datasets; and (3) computational efficiency in estimation. We present recent extensions of this framework to enhance applicability in real-world settings. First, we introduce a statistical jump model capable of handling mixed-type data and missing values. Second, we propose a fuzzy jump model that incorporates uncertainty in state assignments through soft rather than hard clustering, enabling more nuanced representations of transitional dynamics. Together, these models offer a robust and flexible approach to temporal clustering, identifying meaningful regimes in complex systems. Applications span a wide range of fields, including financial time-series analysis, environmental monitoring, and the study of asteroid dynamics. This is joint work with Elisa Maria Alessi and Antonio Pievatolo.

Online Detection of Covariance Changes in Multichannel Profiles

Davide Forcina, University of Naples

This paper introduces a control chart designed to detect covariance shifts in multichannel profiles, a type of multivariate functional data where vector-valued functions (channels) exhibit interdependencies and common structural features. Such profiles often represent the quality characteristics in modern industrial systems. While most existing methods concentrate on detecting changes in the mean profiles, timely identification of shifts in the covariance structure is equally critical for effective statistical process monitoring (SPM). However, this task is complicated by the high dimensionality and the unpredictable nature of potential shift patterns. The proposed SPM scheme employs penalized likelihood ratio tests with varying penalty parameters, allowing it to adapt to a wide range of shift scenarios. The method reframes covariance monitoring to detect changes in the precision matrix and capture conditional dependencies among profiles through functional graphical models. A Monte Carlo simulation and a real-world case study are conducted to evaluate the proposed chart’s performance and practical relevance in comparison to existing techniques. This is joint work with Christian Capezza, Antonio Lepore, and Biagio Palumbo.

Invited Session: Modern Developments in Time Series II

Thursday October 2 (3:00pm-4:15pm)

Room: Ruyszaal (D2.08)

Organizer: Xingdong Feng, Shanghai University of Finance and Economics

Chair: Yayi Yan, Shanghai University of Finance and Economics

Uniform Fixed-k Inference of Explosive Drift

Yi He, University of Amsterdam

We develop a uniform test for detecting short-lived, locally explosive drifts in high-frequency financial price data. In contrast to the existing literature, we consider local estimation windows, possibly overlapping, consisting of a fixed k number of observations. We establish a coupling theory for the supremum t-statistics using extreme value theory. We show that the test statistics across overlapping windows of fixed length are tail independent, despite sharing common observations. As a result, the limiting distribution resembles that of independent windows, allowing critical values to be easily obtained via Monte Carlo methods. Numerical analysis, including simulation studies and applications to real financial data, demonstrates excellent size control of our fixed-k approach across the full range of k, compared with existing methods that apply only to very large k, and reasonably strong power in detecting explosive drifts in financial markets.

Spurious Quantile Regressions and Variable Selection in Quantile Cointegrations

Yingqian Lin, Shanghai University of Finance and Economics

Quantile regression is an effective tool in modeling conditional distribution of economic and financial time series. This paper first investigates the spurious quantile regression phenomenon involving processes moderately deviated from a unit root (PMDURs) through numerical experiments. The main findings include large quantile correlation coefficient estimates, divergent significance test statistics, high R2 and highly autocorrelated residuals indicated by very low Durbin-Watson statistics, complementing those discovered in spurious mean regressions (Lin and Tu, 2020). Unlike the recently proposed balanced regression (Ren et al., 2019; Lin and Tu, 2020) for the mean regression, this paper pioneers the use of the quantile partial correlation (Li et al., 2015, QPC) as a simple-to-implement robust inference measure for spurious quantile regressions. This is benefited from our finding that the induced QPC estimator is asymptotically normal and free of nuisance parameters. For the correlated but non-cointegrated quantile regressions and quantile cointegrations, the QPC estimator converges to a nonzero value in probability. Consequently, the proposed inference procedure can serve to perform variable selection in quantile regressions/cointegrations with PMDURs. Finally, the finite sample properties of the robust method are demonstrated through both Monte Carlo and real data examples.

Real-Time Change Detection in Large-Scale Dynamic Networks

Nathaniel Stevens, University of Waterloo

In this talk we present a real-time change detection method for monitoring large, dynamic networks with community structure. We model the propensity for communication within and between communities to incorporate the structure of the underlying network. Our focus on communities makes our method scalable to large-scale networks and we use a window-based approach to accommodate network dynamics and a changing node set over time. We monitor deviations from the underlying model, where unexpectedly large deviations indicate potential changes. We benchmark our method using a simulation study with networks of 10,000 nodes and demonstrate its flexibility and detection accuracy. We apply the proposed method to a Reddit network defined by discussions on the r/WallStreetBets subreddit about the stock GME, which experienced one of the most notorious short squeezes in market history as retail investors took on hedge funds to drive GameStop's stock price sky-high. Our method is able to identify changes in the network well before the short squeeze.

Contributed Session: Statistical & Machine Learning Methodologies

Thursday October 2 (4:45pm-6:00pm)

Room: Chirurgisch Theater (A1.01)

Chair: Yunming Hui, University of Amsterdam

Identifying clusters through the GAMLSS regression modeling framework

Guilherme Alves Silva, Federal University of Technology – Paraná

Generalized Additive Models for Location, Scale, and Shape (GAMLSS) constitute a highly flexible class of regression models capable of accommodating non-normal response distributions. Nevertheless, their performance may be compromised when dealing with multimodal or heavily skewed data, especially in the presence of latent structures. This survey introduces c-GAMLSS, a novel extension of the GAMLSS framework that integrates clustering techniques into the modeling process. By identifying latent subgroups through clustering algorithms (e.g., k-means, fuzzy clustering, or model-based approaches), c-GAMLSS incorporates a latent categorical covariate that enhances the modeling of the distributional parameters of the response variable. A comprehensive set of simulation studies and four real-world applications—spanning actuarial science, environmental data, synthetic datasets, and volcanic eruption records—demonstrate that c-GAMLSS consistently surpasses traditional methods such as Gaussian mixture models, cluster-weighted models, and models designed for complex bimodal distributions, in terms of both goodness-of-fit and predictive performance. Moreover, the framework maintains the flexibility of GAMLSS by accommodating varying covariates, smoothing functions, and a wide range of distribution families. Overall, c-GAMLSS represents a robust and versatile alternative for modeling complex data structures across disciplines such as biostatistics, finance, and environmental science.

Evaluating the Performance of Highly Complex Location Models Versus GAMLSS

Thiago Gentil Ramires, Federal University of Technology - Paraná

This survey presents a comprehensive discussion on regression models, with a particular emphasis on those within the location model class. The primary motivation stems from the observation that simpler distributions, due to their interpretability, can, in certain contexts, yield more effective results than overly complex alternatives. For example, the Reverse Gumbel (RG) distribution, when employed within the Generalized Additive Models for Location, Scale, and Shape (GAMLSS) framework, enables the modeling of multiple distributional parameters such as the mean, mode, variance, and others. This flexibility facilitates a more nuanced understanding of the response variable. To illustrate this point, three real-world datasets are analyzed, comparing several complex location models against the RG distribution implemented under the GAMLSS framework. The results support the argument that a relatively simple distribution, when paired with a sophisticated regression structure, can outperform more intricate location models in both interpretability and predictive performance.

System Reliability of a Modified Littlewood Software Reliability Model with a Counting Process Framework

Marcus A. Agustin, Southern Illinois University Edwardsville

We consider a system of m software where each software failure follows the Littlewood (1980) software reliability model with a modified debugging scheme. Upon the failure of a software, the defect that caused failure is removed with probability p. In addition, there is a probability r of adding one new defect to the software. The system is observed over a finite testing period, and the system reliability is assessed based on the number and type of software failures observed. Estimators of relevant parameters are derived, and the large-sample properties of the system reliability estimator are established.

Reinforcement learning for battery predictive replacement

Antonio Pievatolo, CNR-IMATI (CNR Institute for Applied Mathematics and Information Technologies)

We consider the problem of predictive replacement of lithium-ion batteries, that is the substitution of a battery when its capacity is still sufficient for the intended application, but it is predicted that it will be no longer so before the next maintenance action can be performed. This task can rely on a degradation model for the battery capacity, which at

any given usage time is able to extrapolate when the capacity threshold to declare a failure will be reached. The ability of the model to accurately represent the capacity decay towards the end of battery life is particularly important, because this is when the uncertainty on which action to take becomes greater. We explore the application of reinforcement learning to the selection of an optimal replacement policy, with particular attention to those methods that combine learning the optimal policy both from observed data (real experiences) and from Monte Carlo simulation from the degradation model, also including degradation model updates. The reinforcement learning approach is compared to predictive replacement based only on the degradation model. This work is part of the PRIN 2022 project no. 2022WBN75S - E3DM - Experimental Design and Maintenance, a Decision-Making approach driven by Degradation Models.

A Comparison of Unsupervised Statistical Learning Models for Insurance Fraud Detection

Sotiris Bersimis, University of Piraeus

Insurance fraud has significant economic and ethical implications. For example, in the motor insurance sector, studies suggest that fraudulent claims account for 10% - 20% of the total claims submitted in Central and Eastern Europe. Consequently, insurance fraud detection is a vital, albeit complex task that is made harder by the fact that historical information regarding past claims' fraudulent status, which is used as data labels, are missing, biased, or completely unavailable. This fact presents an opportunity for the utilization of unsupervised machine learning that does not require the existence of labels. In this work the performance of a wide array of unsupervised machine learning models, most of which have not yet been utilized in the insurance sector, are evaluated and compared using appropriate simulation data. Specifically, a variety of simulation algorithms that are found in the literature are used for the generation of different scenarios of outliers which are assumed to represent fraudulent claims.

Contributed Session: Applications, Tools & Societal Impact

Thursday October 2 (4:45pm-6:00pm)

Room: Ruyszaal (D2.08)

Chair: Jiaqi Qiu, University of Amsterdam

Metaverse Approaches for Analytics Tools Development

Ana María Gómez Lamus, University School Colombian of Engineering Julio Garavito

In contexts, such as healthcare, finance, industry, and education, among others, developing solutions for supporting the explainability of results and contributing to knowledge generation is a challenge and face the daunting of transitioning from proof of concepts to generalizable and secure production implementations. We created a testing environment to support the execution of applications and the development of tools for Statistical Analysis and Data Science. Using a framework to give some guidance as to how to think about approaching this problem with Statistical Engineering, we designed a WebXR platform with 3 modules for performing exploratory analysis, dimensionality reduction, and predictive analysis on high-dimensional data, with excellent deep and flexibility. This solution combines VR/AR elements with the real-world problem environment to enable immersive interactions for support the visualization, comprehension, and understanding of a data-driven problem, reinforcing important abstract concepts in statistics, such as the loss of structural relationships or the low of explainability in results. This platform can allow experts, scientists, and academics to further their understanding of problems and decision-making, experimenting with tools and techniques in a controlled, secure, and private environment, toward the construction of scalable solutions with flexibility for developments can be with AI generative, agents and more.

Trust or Deception? An SEM Analysis of Consumer Attitudes on Sustainability Claims in the Food Industry

Maria Teresa Cappella, Università degli Studi della Tuscia

Sustainability has become a central pillar in corporate communication, largely driven by increasing consumer awareness and regulatory pressures (Yadav & Pathak, 2017). However, the proliferation of green marketing has also fostered unfair practices such as greenwashing, where vague or misleading environmental claims are used to gain consumer trust and competitive advantage (Dahl, 2010; Zhang et al., 2018). These practices risk eroding consumer confidence and compromising the credibility of truly sustainable businesses. Our study investigates the attitudinal factors that shape consumers' interpretation of food product labels and their capacity to identify greenwashing, particularly in distinguishing regulatory-compliant sustainability claims from non-compliant ones across products with varying environmental impacts. Drawing on the Elaboration Likelihood Model (Petty & Cacioppo, 1986), we theorize that environmental involvement, skepticism toward green advertising, and environmental knowledge negatively affect green purchase intention, with perceived greenwashing mediating these relationships. Using data from Italian consumers and a Structural Equation Modeling (SEM) approach (Bollen, 1989), this study explores how attitudinal characteristics enhance the detection of greenwashing on food packaging and influence green purchasing intention. Findings will offer practical implications for improving transparency and trust in green marketing strategies.

Developing Ambassadors of Statistics

Petteri Baer, Statistics Finland, United Nations Economic Commission for Europe (UNECE) Statistical Division

Organizing training sessions and explanatory materials on the web on statistical items provides organizations making use of statistical information and individuals participating in these training courses with a deeper and broader understanding of statistics. Providing training services brings along numerous benefits evidence-based decision-making structures in numerous fields of all societies and for the providers of Official Statistics themselves, usually the National Statistical Institute. Especially important these activities are for Business, Economic and Financial Statistics, as building good relations to a wide set of corporate, companies and experts in the fields of economy, as practically all important decisions today are very strongly linked to economic issues, funding and

resources – be it about investments, allocation of economic resources, the efficiency of data collections or even most simple decision making. This presentation counts up and highlights the benefits of organizing proactive communication activities and short training courses on statistical items based on the over 40 years' long experience of Statistics Finland and to some extent also of the UNECE's work in the field of statistics.

A Statistical Dynamic Decomposition Model for the Environmental Kuznets Curve

Cheyenne Amoroso, Universidade da Coruña

The environmental Kuznets curve (EKC) is of key importance to economic-environmental policy making. It links economic development, often measured by per capita income, to environmental quality, often measured by per capita carbon dioxide emissions. When the EKC is depicted for a large group of countries world-wide, it shows that the two variables grow together until a certain inflection point, after which the growth of pollutants levels off and then overturns. As motivated by its paradigm of "grow first and clean later", we propose a dynamic statistical longitudinal model that decomposes the EKC into an economic-growth effect and a green-economy effect, while both effects are allowed to gradually change over time. The new dynamic evolution of the two effects can account, for example, for the increases in efficiency of production and in supply of green technology that also becomes cheaper. We present the statistical dynamic EKC decomposition model and discuss its properties, the estimation of the nonlinear EKC components and the empirical findings for a dataset of more than 200 countries and of over more than 30 years of observations in time. We reveal that the EKC evolves over time significantly. This is joint work with Carolina García-Martos and Siem Jan Koopman.

Factors Influencing Employment Opportunities: Exploring the Impact of Education and Socio-Demographic Characteristics (In Albania)

Eglantina Zyka, University of Tirana

This study investigates the interplay between education and socio-demographic factors in shaping individuals' chances of employment in Albania. Utilizing a mixed-methods approach—combining quantitative analysis of microdata from the Albania Demographic and Health Survey (ADHS) 2008–2009 with qualitative interviews—the research aims to uncover how variables such as educational attainment, age, gender, and socio-economic background influence employability. The findings indicate that higher education generally increases the likelihood of employment, but this relationship is significantly moderated by other socio-demographic factors. The study highlights persistent structural inequalities, showing that women and individuals from disadvantaged socio-economic backgrounds face disproportionate barriers to employment, even when educational levels are comparable. Data analysis was conducted using SPSS, employing statistical techniques such as regression analysis and chi-square tests to assess both individual and combined effects of the influencing variables. The study provides valuable insights for policymakers and stakeholders, emphasizing the need for inclusive policies that promote equal access to quality education and address structural barriers in the labor market. The findings contribute to evidence-based strategies for enhancing employment outcomes and fostering a more equitable and sustainable labor market in Albania.

Abstracts (Friday October 3)

Invited Session: *Applied Stochastic Models in Business and Industry*

Friday October 3 (9:00am-10:30am)

Room: Chirurgisch Theater (A1.01)

Organizer: Nalini Ravishanker University of Connecticut

Moderator: Nalini Ravishanker University of Connecticut

Fraud Analytics and Adversarial Statistical Learning: Models, Challenges, and Practical Advances

Tahir Ekin, Texas State University

Fraud detection systems face increasing challenges from sophisticated adversarial strategies that exploit vulnerabilities especially in financial, healthcare, and cybersecurity domains. Traditional fraud analytics methods often struggle to detect adversaries who strategically manipulate data and algorithms to evade detection. This talk explores the intersection of fraud analytics and adversarial statistical learning, focusing on the integration of adversarial risk analysis within a Bayesian decision-theoretic framework to enhance fraud detection in adversarial environments. We explore computationally efficient approaches to adversarial statistical decision-making, with an emphasis on impact of data manipulation on statistical and machine learning models. Additionally, we discuss the role of agentic frameworks in modeling strategic interactions between fraudsters and detection systems and the use of synthetic data for stress-testing fraud detection systems. Overall, such frameworks could enable adaptive and semi-autonomous decision-making in high-stakes adversarial settings in practice. We illustrate these concepts with examples of robust anomaly detection methods applied to healthcare fraud detection in adversarial environments. This paper aims to bridge theoretical advancements in adversarial risk analysis with practical applications in business and industry, fostering the development of resilient fraud detection strategies through innovative stochastic modeling and data-driven adversarial learning techniques

Discussion 1

Kathy Ensor, Rice University

Discussion 2

Ozan Kocadagli, Mimar Sinan Fine Arts University

Invited Session: Advances in Statistical Classification

Friday October 3 (11:00am-12:15pm)

Room: Chirurgisch Theater (A1.01)

Organizer: Daniel Jeske, University of California Riverside

Chair: Daniel Jeske, University of California Riverside

Class Maps for Visualizing Classification Results

Mia Hulbert, KU Leuven

Classification is a major tool of statistics and machine learning. A classification method first processes a training set of objects with given classes (labels), with the goal of afterward assigning new objects to one of these classes. When running the resulting prediction method on the training data or on test data, it can happen that an object is predicted to lie in a class that differs from its given label. This is sometimes called label bias and raises the question whether the object was mislabeled. The proposed class map reflects the probability that an object belongs to an alternative class, how far it is from the other objects in its given class, and whether some objects lie far from all classes. The goal is to visualize aspects of the classification results to obtain insight in the data. The display is constructed for discriminant analysis, the k-nearest neighbor classifier, support vector machines, logistic regression, and coupling pairwise classifications. It is illustrated on several benchmark datasets, including some about images and texts.

Neutral Zones: Scrutinizing Machine Learning Algorithms

Nairanjana Dasgupta, Washington State University

As computing takes center stage, Machine Learning (ML) and/or Artificial Intelligence (AI) methods are being widely used in traditional classification problems. However, no matter how sophisticated the algorithm is, classification faces the challenges of forcing objects/items to a class even when two objects have highly similar features. Forcing a definitive classification decision can lead to suboptimal decisions. This is exacerbated when we use black-box methods often associated with ML and AI. Neutral zone classifiers have become an alternative option (Jeske et al, 2018). In this talk we talk about how to apply neutral zones to outcomes from ML and AI algorithms. Our idea uses concepts from inference like r-power (Dasgupta et al, 2016) and ranking to come up with a neutral zone in situations where decisions are made based on ML methods. We will develop the methods and have an algorithm for constructing neutral zones for Artificial Neural Network (ANN), Multiple Layer Perceptron (MLP) and Convoluted Neural Network (CNN) but can be extended to most ML classification methods. We will use an example of drug delivery using CNN to illustrate our results.

Topic-Augmented Generation: Topic-Driven Chunk Classification and Re-Ranking for Enhanced Retrieval-Augmented Generation

Marco Ortu, University of Cagliari

Retrieval-Augmented Generation (RAG) systems rely critically on the relevance of retrieved context to produce accurate and coherent outputs. We propose Topic-Augmented Generation (TAG), a novel framework that incorporates topic modeling into the retrieval pipeline to improve the semantic alignment of retrieved chunks prior to generation. We exploit topic modeling as a semantic classification technique for chunk of text and as a re-ranking method during the retrieval phase of RAG. First, each document chunk is assigned a topic distribution via Topic Modeling technique, such as Latent Dirichlet Allocation. Next, at query time, we compute the similarity between the query's inferred topic distribution and that of each candidate chunk. We then re-rank the top-K chunks retrieved by conventional embedding-based methods according to topic proximity, promoting chunks more thematically aligned with the query. Finally, the generation component of the RAG model consumes this reordered context. TAG preserves the flexibility of existing RAG architectures while enhancing retrieval precision through an orthogonal, topic-based signal. Empirical evaluation on public corpora such as question-answering and summarization benchmarks, demonstrates that TAG yields a substantial increase in retrieval precision and improves downstream generation quality, as measured by ROUGE and human relevance assessments. These results underscore the effectiveness of topic-driven chunks' classification and re-ranking as a complementary mechanism for advancing RAG performance.

Invited Session: AI and Machine Learning (y-BIS)

Friday October 3 (11:00am-12:15pm)

Room: Ruyszaal (D2.08)

Organizer: Ozan Kocadagli, Mimar Sinan Fine Arts University

Chair: Ozan Kocadagli, Mimar Sinan Fine Arts University

Revolutionizing Healthcare: The Role of Artificial Intelligence-Based Early Diagnosis

Ozan Kocadagli, Mimar Sinan Fine Arts University

Artificial Intelligence (AI) is redefining early diagnostic practices in healthcare by enabling data-driven precision, speed, and scalability. This study investigates the scientific evolution and clinical integration of AI-based diagnostic systems, focusing on their capacity to detect subclinical patterns across diverse data modalities such as medical imaging, electronic health records, and genomic profiles. Through advanced machine learning techniques—including deep learning and natural language processing—AI enhances diagnostic accuracy and supports timely clinical decision-making. The paper presents real-world case studies demonstrating AI's performance and utility in hospital settings, while also addressing challenges related to model interpretability, ethical considerations, and regulatory compliance. A key contribution lies in the exploration of explainable AI frameworks and future directions such as federated learning and adaptive diagnostic models. These innovations mark a significant shift toward intelligent, patient-centered care.

A Comprehensive Review of Chaos Control and Modeling Using Intelligent and Hybrid Methods: Trends, Challenges, and Future Directions

Atike Reza Ahrabi, Islamic Azad University

The study of chaos control and modeling has become increasingly crucial across diverse fields, including engineering, neuroscience, cryptography, and complex dynamic systems, due to the intrinsic sensitivity and nonlinearity of chaotic behaviors. This paper presents a comprehensive review of 70 peer-reviewed studies focusing on intelligent and hybrid methods for chaos control and modeling. The review systematically categorizes key approaches—fuzzy logic (Type-1, Type-2, hybrid), neural networks, evolutionary algorithms, advanced adaptive and robust control, and deep learning—and analyzes their strengths, limitations, and cross-domain applications, such as biomedical signal processing, memristor-based oscillators, secure communication, and FPGA-based implementations. Comparative analysis highlights the dominance of fuzzy systems in uncertainty handling, neural networks in nonlinear prediction, and evolutionary algorithms in optimization, with an emerging shift toward hybrid integrations (e.g., fuzzy-neural-PSO, fuzzy-GA) that mitigate individual drawbacks. Furthermore, we identify research gaps, including the limited fusion of deep reinforcement learning with chaos stabilization, insufficient real-world deployment in medical and transportation systems, and the need for scalable event-triggered mechanisms. Based on these findings, we propose a forward-looking RL–Fuzzy–Event-Triggered framework as a potential pathway for robust, adaptive chaos regulation in complex environments. This review not only consolidates the current state-of-the-art but also provides clear insights into future research trends and actionable opportunities for advancing intelligent chaos control.

Advancing Facial Emotion Recognition through Deep Learning: A CNN-Based Approach

Özge Körükci, University of Bremen

In the era of ubiquitous visual data and intelligent systems, facial emotion recognition (FER) has emerged as a pivotal task at the intersection of computer vision, affective computing, and human-computer interaction. With the proliferation of high-resolution imaging devices and scalable data storage, vast repositories of facial imagery are now available for computational analysis. This study explores the application of deep learning—particularly convolutional neural networks (CNNs) and their advanced variants—for automated emotion classification from facial expressions. Leveraging the FER-2013 benchmark dataset, which encompasses six fundamental emotional states, we implement a robust preprocessing pipeline including grayscale normalization, noise reduction, and data augmentation to enhance model generalization. The CNN architecture is optimized through systematic hyperparameter tuning, encompassing batch size, learning rate, and epoch scheduling. We also discuss the integration of recent innovations such as residual connections, attention mechanisms, and transfer learning to improve recognition accuracy and model interpretability. Our findings underscore the efficacy and limitations of CNN-based FER systems in real-world scenarios, highlighting their potential for deployment in emotion-aware applications such as mental health monitoring, adaptive user interfaces, and social robotics. The study concludes with a discussion on future directions, including multimodal emotion recognition, ethical considerations, and the role of explainable AI in affective computing. This is joint work with Ozan Kocadagli.

Invited Session: Modeling Complex Data with Business and Industrial Applications

Friday October 3 (1:15pm-2:30pm)

Room: Chirurgisch Theater (A1.01)

Organizer: Nalini Ravishanker, University of Connecticut

Chair: Kathy Ensor, Rice University

Latent Gaussian Models for Market Shares

Balaji Raman, Cogitaas AVA

A key performance indicator of brands in FMCG (Fast Moving Consumer Goods) sector is market share. Changes in market shares reflects health of a brand, competitive pressures and finally the revenue. Hence, it is important to model the movement of market shares over time and identify the drivers, especially price and cross-price elasticities. Market shares of different brands within a category are an example of Compositional Data (CoDa). In this talk, we apply some recent ideas in the use of Latent Gaussian Models (LGM) for CoDa through a fast and approximate Bayesian inference for modelling market shares. This is based on a joint work with Joaquin Martinez Minaya (Valencia Polytechnic University).

Computationally Efficient Models for Count Data with Varying Levels of Dispersion

Kimberly Sellers, North Carolina State University

The Poisson distribution is ubiquitous for modeling count data but is limited in application by its assumption of equi-dispersion. Alternative models exist that do not suffer from this challenge but most are (1) limited to either under- or over-dispersed data, (2) only provide estimates of moments, or (3) are computationally prohibitive for large data sets. We review existing methods for count data with varying levels of dispersion, including some new results for generalized Poisson regression. We introduce a latent-variable model based on a discretized log-normal distribution, develop a scalable EM algorithm to estimate it, and provide straightforward likelihood-based theory for performing statistical inference. Finally, we illustrate these methods in simulation and a case study forecasting grocery sales; we find that our latent-variable model performs nearly as well as the leading method (COM-Poisson regression) but at a fraction of the computational cost. This presentation is based on a coauthored work with Easton Huch and Mason Ferlic (both from the University of Michigan), and Candace Berrett (Brigham Young University).

A Statistical Learning Approach to Feature Selection in High Dimensional Low Sample Size Data

Shibasish Dasgupta, TCG CREST

The evolution of data collection and processing techniques has facilitated the generation of extensive high-dimensional datasets. In specialized fields such as biomedical data science, managing datasets characterized by numerous features and limited sample sizes, known as High Dimensional Low Sample Size (HDLSS) data, presents a notable obstacle for statistical methods that hinge on likelihood estimation. Feature selection (FS) assumes a crucial role in mitigating challenges associated with dimensionality, overfitting, and improving model interpretability and accuracy. By streamlining input variables during model creation, FS assists in constructing a more precise predictive model. Our study delved into diverse filter and wrapper techniques, culminating in a personalized feature selection framework tailored for robust variable selection, particularly suited for HDLSS data, incorporating the principle of stability. Our exploration focused on evaluating the efficacy of Minimum-Redundancy-Maximum-Relevance (mRMR) and Support Vector Machine Recursive Feature Elimination (SVM-RFE) methodologies. Through a comparative assessment of these FS strategies, we analyzed their predictive efficacy and stability using Clinical Trial data. This in-depth scrutiny highlights the effectiveness and significance of these FS algorithms in the domain of high-dimensional, low sample size data settings.

Invited Session: Advanced Statistical Modeling (y-BIS)

Friday October 3 (1:15pm-2:30pm)

Room: Ruyszaal (D2.08)

Organizer: Ozan Kocadagli, Mimar Sinan Fine Arts University

Chair: Ozan Kocadagli, Mimar Sinan Fine Arts University

Domain-Informed Negative Sampling Strategies for Dynamic Graph Embedding in Meme Stock-Related Social Networks

Yunming Hui, University of Amsterdam

Social network platforms like Reddit are increasingly impacting real-world economics. Meme stocks are a recent phenomenon where price movements are driven by retail investors organizing themselves via social networks. To study the impact of social networks on meme stocks, the first step is to analyze these networks. Going forward, predicting meme stocks' returns would require predicting dynamic interactions first. This is different from conventional link prediction, frequently applied in e.g. recommendation systems. For this task, it is essential to predict more complex interaction dynamics, such as the exact timing. These are crucial for linking the network to meme stock price movements. Dynamic graph embedding (DGE) has recently emerged as a promising approach for modeling dynamic graph-structured data. However, current negative sampling strategies, an important component of DGE, are designed for conventional dynamic link prediction and do not capture the specific patterns present in meme stock-related social networks. This limits the training and evaluation of DGE models in such social networks. To overcome this drawback, we propose novel negative sampling strategies based on the analysis of real meme stock-related social networks and financial knowledge. Our experiments show that the proposed negative sampling strategies can better evaluate and train DGE models targeted at meme stock-related social networks compared to existing baselines.

Physics Informed Neural Network Method for the Numerical Solution of the Fractional Partial Equations

Burcu Ece Alp, Kultur University

In this study, we propose a physics-informed neural network (PINN) framework to solve fractional partial differential equations (PDEs) with variable coefficients over a finite domain. These types of equations are commonly encountered in modeling diffusion and wave propagation phenomena in science and engineering. The PINN model learns an approximate continuous solution by minimizing a physics-based loss function that incorporates the equation residual, boundary, and initial conditions. The fractional derivative is discretized using the Grünwald-Letnikov scheme, which is used to formulate the residual part of the loss. Numerical experiments demonstrate that the proposed PINN method provides accurate and smooth solutions over the computational space-time domain. This approach offers a promising alternative to traditional numerical methods for solving complex fractional PDEs.

Weighted Estimation of Hidden Markov Stochastic Volatility Models

Michael Trequattrini, University of Perugia

This work addresses the short-term forecasting of realized variance, a key task in financial econometrics. Building on existing stochastic volatility models where log-volatility follows an autoregressive latent process, we adopt a hidden Markov model approximation to enable efficient likelihood-based estimation. As a novel contribution, we propose a weighted variant of the estimation approach of the standard model that incorporates observation-specific weights into the likelihood function. These exponentially decaying weights emphasize more recent data, accounting for the time-varying relevance of past observations, an important consideration given the regime-switching nature of financial markets. The decay rate is selected through cross-validation, and inference is carried out via Newton-Raphson optimization, exploiting Gaussian quadrature for the latent state approximation. We assess the predictive performance of our approach through both simulation studies and empirical tests. Synthetic datasets allow us to isolate the impact of temporal weighting, while real financial data, including weekly realized variance for the S&P500 and other indices, provide practical validation. Using a rolling-window procedure to perform forecasting, we find that the weighted estimation approach achieves a consistent improvement of roughly 20% in median absolute error over the unweighted benchmark estimation, demonstrating the value of adaptive weighting in volatility forecasting on real financial datasets.

Invited Session: Networks and Where to Find Them

Friday October 3 (3:00pm-4:15pm)

Room: Chirurgisch Theater (A1.01)

Organizer: Simon Trimborn, University of Amsterdam

Chair: Simon Trimborn, University of Amsterdam

Forecasting Extreme Values in Financial Time Series via Scores

Simon Trimborn, University of Amsterdam

Extreme value prediction in time series analysis remains a significant challenge for forecasting models, particularly in financial markets where tail events can have substantial impact. This paper introduces a novel time-dependent methodology specifically designed for forecasting extreme values in financial time series. Our approach addresses the limitations of conventional models by incorporating temporal dynamics into extreme value prediction. By combining a time dependent Hüsler Reiss model with market informed parameter constraints, we keep the necessary parameters to estimate reasonable to handle the few observations in the tails of financial time series. We demonstrate the efficacy of our approach using an extensive dataset of individual stock components from the S&P 500 index. Our empirical analysis shows improvements in extreme value forecasting across diverse market sectors and varying conditions.

Federated Community Detection in Bipartite Networks from Various Platforms

Wanjie Wang, National University of Singapore

In modern applications such as recommendation systems and e-commerce platforms, user-item interactions are naturally modelled by bipartite networks. When data are collected across multiple servers, privacy concerns and communication constraints render centralized community detection impractical. It motivated the development of federated learning. In this talk, I will present a federated framework for community detection in bipartite networks, where user data reside on disjoint servers, but all servers share a common item set. We introduce a gradient-based power iteration method that approximates leading singular vectors using only local computations and limited inter-server communication. Building on this, we propose the BiFLICKER algorithm, which combines spectral methods with federated learning to recover global community structures while preserving data locality. Theoretical guarantees are established for the accuracy of spectral estimation and community recovery. We validate our method through simulations and real-world analysis of federated movie rating data from Douban, revealing meaningful and interpretable user and movie communities.

Does News Sentiment Walk the Talk? A Network Analysis of Volatility Spillovers around News Events

Lars de Graaff, University of Amsterdam, Erasmus University Rotterdam, Vrije Universiteit Amsterdam

We study whether unscheduled, firm-specific news events reconfigure the volatility-risk network among large U.S. technology stocks in a tight intraday event window based on 5-minute and 1-minute volatility proxies and whether modeling the anticipated reconfiguration improves forecasts. To do so, we examine the three-hour window before and after each timestamped event using a penalized Vector AutoRegressions (VAR) using both element-wise and influencer-group-wise penalisation in the VAR coefficient matrices.

Invited Session: Statistical Risk Modelling (y-BIS)

Friday October 3 (3:00pm-4:15pm)

Room: Ruyszaal (D2.08)

Organizer: Ozan Kocadagli, Mimar Sinan Fine Arts University

Chair: Ozan Kocadagli, Mimar Sinan Fine Arts University

Nonparametric Regression Based Estimation of Conditional Spectral Risk Measures

Supurna Biswas, Indian Statistical Institute

A novel nonparametric method for estimating conditional spectral risk measures (SRMs) is discussed in this work. The nonparametric estimator of conditional SRM is constructed by a plugging-in method where we use the weighted double kernel local linear estimate and the classical Nadaraya-Watson type double kernel estimate of the conditional distribution function. We propose our new nonparametric estimators of conditional SRM by employing these two nonparametric kernel regression estimators. The asymptotic normality and consistency of the proposed estimators are derived under-mixing assumptions. Monte Carlo studies are conducted to assess the performance of the proposed estimators through the calculation of their mean squared error (MSE), which serves as a global discrepancy measure. Our observations reveal that the conditional SRM estimator obtained using the classical Nadaraya-Watson type double kernel estimate outperforms under all the scenarios considered in our simulation study.

Beta-Constrained Markowitz Portfolio Optimization across Different Market Phases

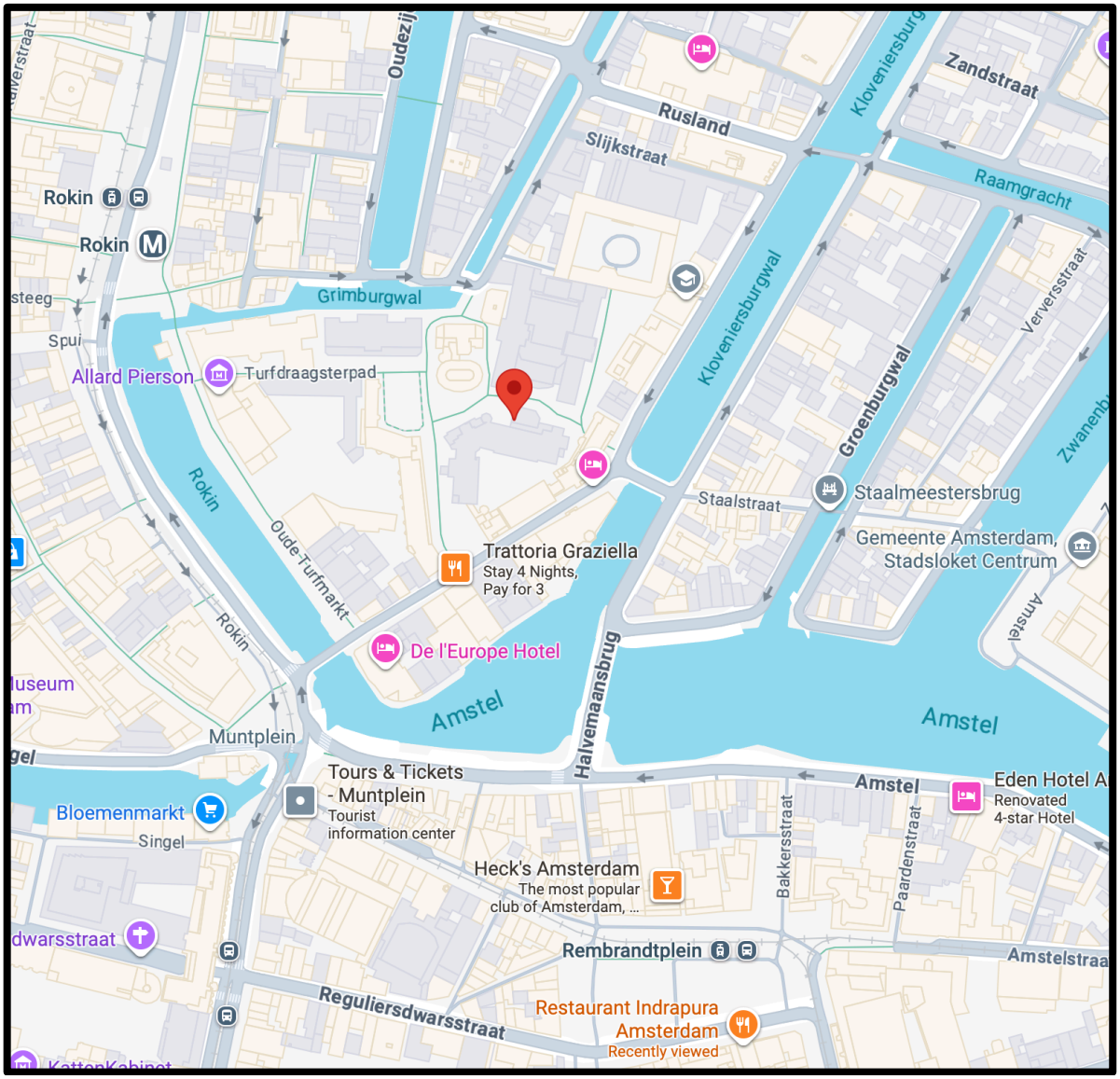
Gulder Kemalbay, Yildiz Technical University

This study proposes a novel extension of the classical Markowitz portfolio optimization model by integrating a Beta-weighting mechanism that adapts to different market phases. Financial assets are categorized based on their Beta coefficients, allowing for differentiated weighting strategies in rising and falling markets. A bi-objective optimization framework is formulated to simultaneously minimize portfolio risk and maximize expected return under these Beta-based constraints. The model is solved using a heuristic algorithm. Empirical findings reveal that Beta-constrained portfolios outperform the traditional Markowitz model across both market phases, with notable improvements in downside risk control during bear periods. The results offer a practical and interpretable investment strategy that balances return and risk based on static market characteristics. This is joint work with Yaren Korkmaz.

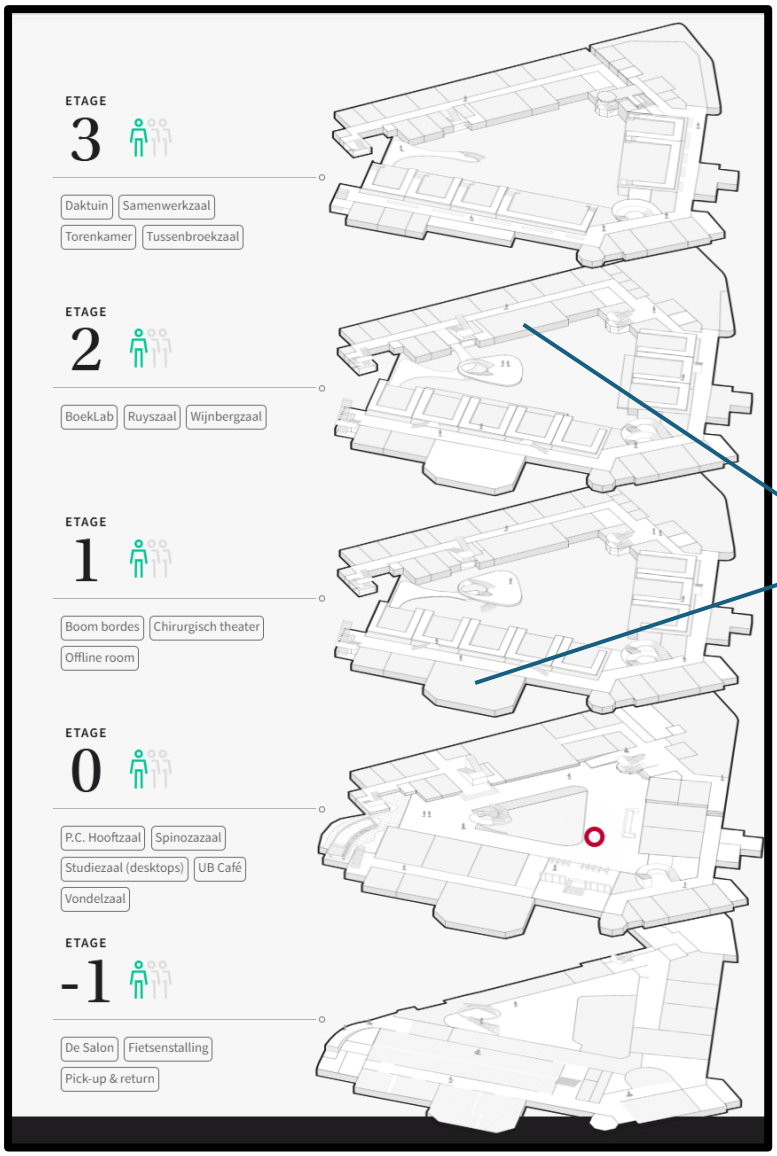
Maps

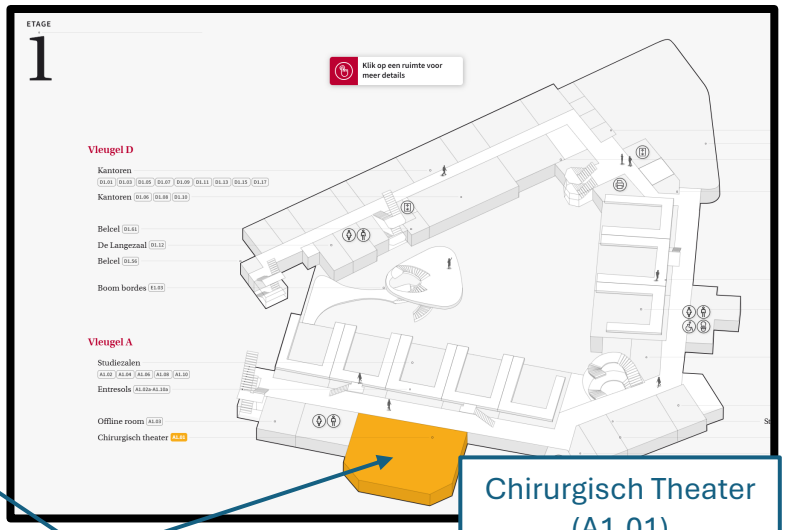
Venue Map (Google)

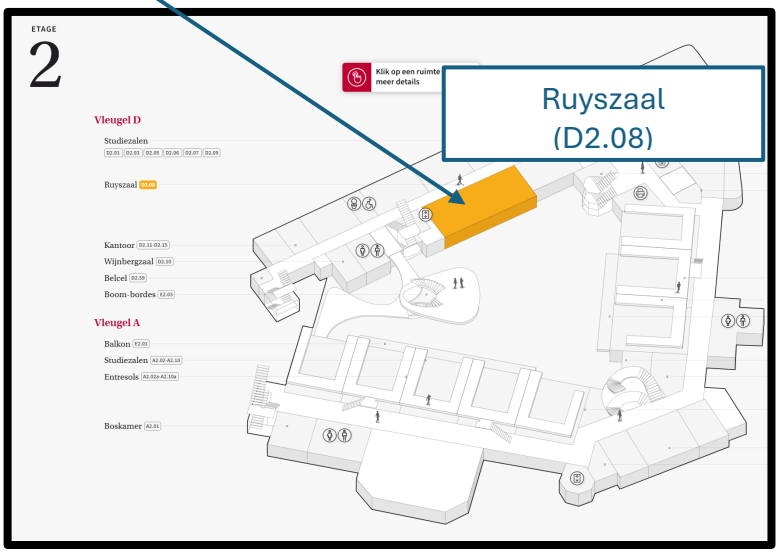
University of Amsterdam Library
Vendelstraat 2, 1012 XX Amsterdam, Netherlands



Venue Map (Floorplan)



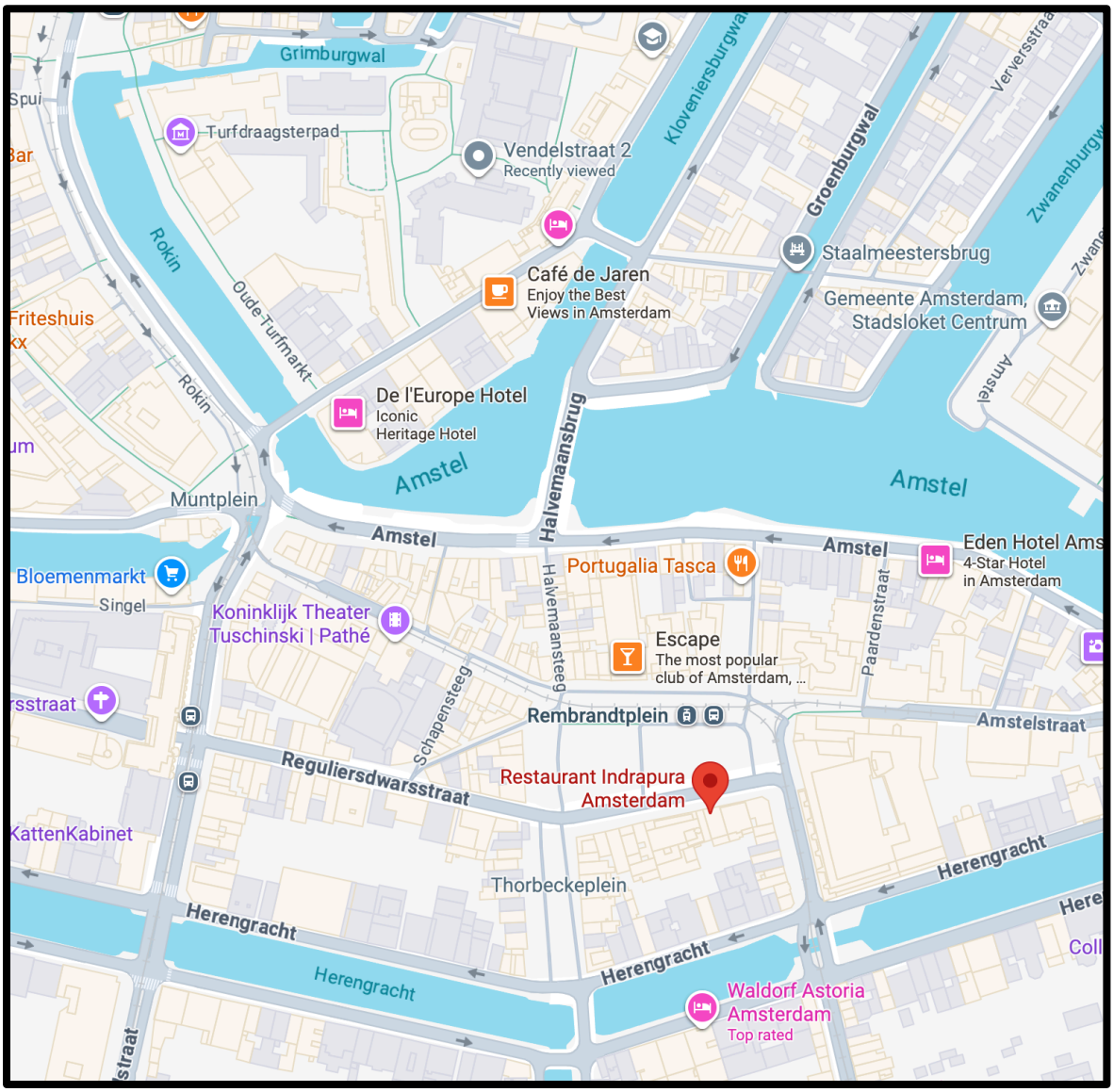




Banquet Map

INDRAPURA
INDONESIAN RESTAURANT

Rembrandtplein 40, 1017 CV Amsterdam, Netherlands



Venue to Banquet Directions

