ICORS2026

20-24 JULY 2026

İstanbul

INTERNATIONAL CONFERENCE ON ROBUST STATISTICS 2026

Book of Abstracts



20-24 JULY

İstanbul / Türkiye

icors2026.ankara.edu.tr

ICORS2026
20-24 JULY 2026
Istanbul

Table of Contents

Preface ..... *viii*

Committees ..... *x*

Conference Chairs x

Scientific Program Committee x

Local Organizing Committee xii

Program ..... *xiii*

Monday, 20 July 2026 xiii

Tuesday, 21 July 2026 xvi

Wednesday, 22 July 2026 xx

Thursday, 23 July 2026 xxii

Friday, 24 July 2026 xxv

Keynote Presentations *1*

Independent Component Analysis by Robust Distance Correlation 2

Robust Joint Estimation of Survival Components to Assess Risks and Benefits of Cancer Screening 3

Robustness Features of Invariant Coordinates Selection 4

Robustifying the Bayesian Approach 5

Several Studies in Label Shift 6

Robustness of the Optimal Transport Map 7

Invited Session 1 (IS1) *8*

High-dimensional separability testing in matrix data beyond normality 9

Robust tensor-on-tensor regression 10

A trimming and constraints based robust clustering proposal for matrix-variate data 11

Invited Session 2 (IS2) *12*

Robust estimation for functional joint models with longitudinal and survival outcomes 13

A novel robust approach to spatially structured scalar-on-function quantile regression 14

Multivariate two-sample tests for functional data with outliers 15

Projection depth for functional data **18**

Invited Session 3 (IS3) ***20***

Multivariate central limit theorem in Wasserstein distance **21**

Bayesian Credibility Estimation Using Concomitants of Order Statistics in the Framework of Generalized Linear Mixed Models **22**

Invited Session 4 (IS4) ***24***

Cellwise and Casewise Robust Multivariate **25**

Regression with Inference **25**

Fast and Scalable Cellwise-Robust Ensembles for High-Dimensional Data **26**

Robust inference for nonparametric regression via bootstrap **27**

On Robust Estimation for Heteroscedastic Nonlinear Regression Models **28**

Invited Session 5 (IS5) ***29***

Robust, Sub-Gaussian Mean Estimators in Metric Spaces **30**

Stability and instability of mean-field variational inference **31**

A Frequentist Theory for M-posteriors: Asymptotic and Robustness Properties **32**

Invited Session 6 (IS6) ***33***

What are outliers? **34**

Automatic choice of *k* and *c* in model-based clustering: The R package 'tclust' . **35**

Robust Statistics Through Monitoring: Outliers and Transformations **36**

Robust trimmed clustering for outlier detection in complex settings **37**

Invited Session 7 (IS7) ***39***

A Flexible Heteroscedastic Semiparametric Model for Skewed and Heavy-Tailed Data **41**

Robust Joint Location–Scale Modeling via M-Quantile Estimation **43**

Robust Particle Swarm Optimization for Regression via M-Estimation **44**

Invited Session 8 (IS8) ***45***

The advantage of autoregression rank scores invariance in inference **46**

A signed-rank estimator in nonlinear regression models when covariates and errors are dependent **47**

Adaptive rank scores guided by L-moments for robust inference in regression–autoregression models **48**

M-Quantile Regression with Autoregressive Errors 49

Invited Session 9 (IS9) *50*

New M-estimator of the leading principal component 51

Robust Principal Components Analysis via Subspace Discrepancy Minimization 52

Adaptive lasso-regularized spectral generalized kotz-type robust matrix regression 53

Invited Session 10 (IS10) *54*

Robust Distributional Regression Models for Interval-valued Data 55

Clustering interval-valued data using L1-Wasserstein distance 57

Robust Principal Component Analysis of Symbolic Matrix-valued Data 59

Multivariate Fuzzy Partitioning Methods using Medoids 61

Invited Session 11 (IS11) *63*

Multivariate functional Mahalanobis distance with application to clustering ... 64

Outlier Detection and Robust Estimation of Time Series Model using a Penalised Approach 65

Cellwise outlier detection in clustering: from model-based to fuzzy approaches 66

Invited Session 12 (IS12) *68*

Robust Estimation via Maximum Mean Discrepancy 69

Efficient Differentially Private Regression Inference for Longitudinal Data 70

A Statistical Approach to Image Classification via Deformable Templates 71

Invited Session 13 (IS13) *72*

Cellwise Robust Discriminant Analysis 73

Robustness, clustering and outlier detection under heavy tails 74

Blockwise, cellwise, and casewise robust multiblock PCA 76

Invited Session 14 (IS14) *77*

Provable Robustness to Spurious Correlations via Invariant Data for Robust Finetuning 78

CGL-LIME: A Robustified LIME with Copula-Based Sampling and Gradient-Weighted Kernels 80

From Robust Mixed Models to Temporal Deep Learning: A Five-Paradigm Robustness Benchmark on Contaminated fNIRS Data 81

Robustness of Exponential Family and M-Estimation Based Empirical Likelihood in Linear Regression **84**

Invited Session 15 (IS15) ***85***

Cellwise-robust imputation for mixed continuous and categorical data **86**

Robust Estimation of Circular Dispersion: Theory and an Application to Outlier Detection **88**

A Systematic Comparison of Clustering Methods Under Casewise, Cellwise, and Mixed Data Contamination **89**

Invited Session 16 (IS16) ***91***

When Distribution-Free Falls Short for Prediction Intervals under Right-Censored Covariate **92**

Tweedie compound Poisson mixed models for cross-classified data with left censoring **93**

Quantile Graph Discovery through QuACC: Quantile Association via Conditional Concordance **95**

Invited Session 17 (IS17) ***96***

Robust Inference for Mixtures of Multivariate Wrapped Normal Distributions: The Weighted Likelihood Approach **97**

Optimizing Rényi's alpha-divergences: a variational inference perspective **98**

Bayesianism and Robustness: Predictively Oriented (Pr0) Posteriors **99**

Contributed Session 1 (CS1) ***100***

Comprehensive Evaluation of METHoD-NO for Benchmark Optimization and Neural Network Training **101**

GWO-Optimized Hybrid Deep Learning Approach for Lung Cancer Detection: High-Accuracy Classification on CT Images **103**

Bootstrap-Based Robust Portfolio Optimization under Regime Heterogeneity: A Homogeneity-Driven Approach **105**

On inference using M-estimators under contamination **106**

Hybrid Sliding Window Recursive Least Squares for Battery State of Health Prediction **107**

Contributed Session 2 (CS2) ***108***

Functional canonical correlation for discriminant analysis **109**

An ECM-Metropolis Algorithm for Parameter Estimation in Skew Laplace Linear Mixed Models **110**

Robust Anomaly Detection via a Bayesian Autoencoder with Heavy-Tailed Priors **112**

Uncertainty-Aware Statistical Learning for Robust Clinical Risk Analytics **114**

Contributed Session 3 (CS3) ***115***

The threshold breakdown point **116**

Robust and explainable outlier detection for random surfaces **117**

Breakdown points of simplicial and *k*-hull medians **118**

When is the Tukey median unique? **120**

Contributed Session 4 (CS4) ***121***

Forecasting Multivariate Time Series under Predictive Heterogeneity: AValidation-Driven Clustering Framework **122**

The unit-t_2 distribution: a parametric approach to median regression on the unit interval **123**

Robust XGBoosting for Regression **124**

Contributed Session 5 (CS5) ***125***

Ranked Set Sampling based ANOVA under Non-Normality **126**

Robust Tests for One-Sample and Two-Sample Problems under Ranked Set Sampling: An Application to Real Data **127**

Sequential Conformal Prediction for Electricity Price Forecasting **128**

Contributed Session 6 (CS6) ***130***

Massive parallelization of projection-based depths **131**

Robust smoothing splines with discontinuities **132**

Robustness Problems in Relaxed and Entropy Regularized Portfolio Optimization **133**

Simulation-Based Robust Inference for Response Misclassification Errors with Self-Reported Responses **135**

Contributed Session 7 (CS7) ***136***

Application of Robust SUR Modeling and Variable Selection to Air Pollution Data in Ankara **137**

The Marshall–Olkın Weıbull Process: Methodological Framework and Applıcatıons to Envıronmental Data **138**

Robust Regression for Agricultural Yield Estimation Under Influential Observations: Evidence from Konya Districts **140**

Analysis of Installed Power Levels of Solar Power Plants in Turkiye Using Spatial and Robust Spatial Regression **141**

Comparing Deletion Methods and Imputation Methods in Seemingly Unrelated Regression Models for Air Quality Data **143**

Contributed Session 8 (CS8) ***145***

Robust Reliability Demonstration Test Plans **146**

Projecting Multidimensional Poverty to 2030: A Scenario-Based Microsimulation Framework **149**

Achieving Nearly Deterministic Precision in Dental Implant Identification **152**

Robust Circular–Circular Correlation Coefficient via Huber and Tukey Weighting **155**

Contributed Session 9 (CS9) ***156***

TSRClust: Two-Stage Robust Clustering **157**

SIIE: sliced inverse interval estimation for robust sufficient dimension reduction **158**

Robust Feature Clustering around Latent Components **159**

Robust Sparse Principal Component Regression with Alternative Loss Functions: A Framework for High-Dimensional Settings **161**

Contributed Session 10 (CS10) ***163***

Graph Robust Adaptive LASSO for Network-Aware Variable Selection under Contamination **164**

Poster Session ***166***

A Novel Computational Tool for AMD Diagnosis **167**

Robust Assessment of Benford's Law under Skewed Generalized Error Distributions via Nonparametric Combination **169**

Spherical Projection Depth **170**

Methodology for the Longitudinal Analysis of thrips Counts in Chrysanthemum Cultivation **171**

Amortized Neural Clustering of Time Series based on Statistical Features **173**

Novel alignment-free bioinformatics methods **174**

Author Index ***176***

Preface

Dear Colleagues,

It is my great pleasure, on behalf of the Organizing Committee, to welcome you to the International Conference on Robust Statistics held in İstanbul, Türkiye, from 20 to 24 July 2026.

Since its first meeting in Vorau, Austria, in 2001, ICORS has grown into a distinguished international forum for researchers and practitioners working in robust statistics, data analysis, and related fields. Over the years, the conference has travelled across countries and continents, bringing together established scholars, early career researchers, graduate students, and professionals from a wide range of disciplines. Beyond the formal scientific programme, ICORS has always been valued for its open and collegial atmosphere, in which discussions begun during a session or coffee break often develop into lasting collaborations and friendships.

ICORS 2026 marks the 26th meeting in this remarkable series. It also represents a particularly meaningful return to Türkiye, eighteen years after ICORS 2008 was held in Antalya. We are honored to continue this tradition in İstanbul, a city that has connected peoples, cultures, and ideas for centuries.

Situated on the shores of the Bosphorus and extending across both Europe and Asia, İstanbul offers an inspiring setting for an international scientific gathering. Its character has been shaped by many civilizations and enriched by the Roman, Byzantine, Ottoman, and modern Republican periods. Yet İstanbul is not merely a city of history. It is also a dynamic contemporary metropolis where tradition and innovation, East and West, and the ancient and the modern coexist. This spirit of connection and exchange closely reflects the values at the heart of ICORS.

The importance of robust statistical thinking continues to grow as modern data become increasingly large, complex, heterogeneous, dependent, and vulnerable to contamination or model misspecification. The contributions collected in this Book of Abstracts reflect the vitality and breadth of the field. They range from the theoretical foundations of robust inference to contemporary developments in regression, machine learning, high-dimensional and functional data analysis, outlier detection, statistical learning, visualization, and interdisciplinary applications. We hope these contributions will stimulate constructive discussion, encourage the exchange of new ideas, and open pathways for future research.

I would like to express my sincere gratitude to our keynotes and invited speakers, session organizers, authors, participants, and members of the scientific and organizing committees. I also gratefully acknowledge the valuable support of our partner institutions, sponsors, and everyone whose dedication has made ICORS 2026 possible.

May this conference be remembered not only for the quality of its scientific contributions, but also for the conversations, collaborations, and friendships it inspires.

Welcome to İstanbul, welcome to Türkiye, and welcome to ICORS 2026.

Olcay Arslan

Chair, ICORS 2026

Committees

Conference Chairs		
Olcay Arslan	*Professor*	Department of Statistics, Ankara University
Peter Rousseeuw	*Professor Emeritus*	Statistics and Data Science, KU Leuven
Peter Filzmoser	*Professor*	Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology
Aylin Alın	*Professor*	Department of Statistics, Dokuz Eylul University
F. Sevinc Kurnaz	*Associate Professor*	Department of Statistics, Yıldız Technical University & Case Western Reserve University
Yetkin Tuac	*Assistant Professor*	Department of Statistics, Ankara University
Yesim Guney	*Assistant Professor*	Department of Statistics, Ankara University

Scientific Program Committee	
Abdullah Fırat Özdemir University of Dokuz Eylul	**Luis Angel Garcia Escudero** University of Valladolid
Agustin Mayo-Iscar University of Valladolid	**M. Rosário Oliveira** University of Lisbon
Ana Maria Bianco University of Buenos Aires	**Marc Hallin** Université Libre de Bruxelles
Anand N. Vidyashankar George Mason University	**Marianthi Markatou** University at Buffalo

Andrea Cerioli
University of Parma

Ayanendranath Basu
Indian Statistical Institute

Aylin Alin
Dokuz Eylul University

Baris Surucu
Middle East Technical University

Birdal Senoglu
Ankara University

Christophe Croux
KU Leuven

Claudio Agostinelli
University of Trento

David E. Tyler
Rutgers University

Domenico Perrotta
European Commission (JRC)

Garth Tarr
University of Sydney

Graciela Boente
University of Buenos Aires

Maria-Pia Victoria-Feser
University of Bologna

Mehmet Yılmaz
Ankara University

Meral Cetin
Hacettepe University

Mia Hubert
KU Leuven

Olcay Arslan
Ankara University

Omer Lutfi Gebizlioglu
Kadir Has University

Peter Filzmoser
Vienna University of Technology

Peter Rousseeuw
Professor Emeritus

Ricardo A. Maronna
National University of La Plata

Shoja Chenouri
University of Waterloo

Stefan Van Aelst
KU Leuven

Jana Jureckova Charles University	**Valentin Todorov** UNIDO
Klaus Nordhausen University of Helsinki	**Victor J. Yohai** University of Buenos Aires
Liangyuan Hu Rutgers University	**Xuming He** University of Washington
	Yanyuan Ma Penn State University

Local Organizing Committee		
Fatma Zehra Dogru	Professor	Giresun University
Sukru Acitas	Professor	Eskisehir Technical University
Yakup Murat Bulut	Associate Professor	Eskisehir Osmangazi University
Fulya Gokalp Yavuz	Associate Professor	Purdue University
Ozlem Kaymaz	Associate Professor	Ankara University
Senay Ozdemir	Assistant Professor	Afyon Kocatepe University
Onur Toka	Assistant Professor	Hacettepe University
Abdullah Yalcinkaya	Assistant Professor	Ankara University
Gokhan Uyar	Lecturer (PhD)	Ankara University
Ozge Ustahuseyin	Research Assistant	Ankara University
Serdar Cihat Goren	Head of Quality Group (PhD)	Turkish Statistical Institute
Elifnur Boyraz	Student	Ankara University
M. Enes Tasova	Student	Ankara University
Eylul Zor	Student	Ankara University
Senanur Serce	Student	Ankara University

Program

Monday, 20 July 2026

08:00–09:00	**Registration**

Büyük Salon (Main Hall)	
09:00–10:00	**Opening Ceremony**
10:00–10:20	**Coffee Break**
10:20–11:20	**Keynote 1** **Peter Rousseeuw (KU Leuven, Belgium)** ***Independent Component Analysis by Robust Distance Correlation*** Chair: Elvezio Ronchetti
11:20-11:30	**Conference Group Photo**
11:30–12:45	**IS1 – Robustness for Matrix-valued Data** Chair: Peter Filzmoser (TU Wien, Austria) Marcus Mayrhofer (TU Wien, Austria) *High-dimensional separability testing in matrix data beyond normality* Mia Hubert (KU Leuven, Belgium) *Robust tensor-on-tensor regression* Agustin Mayo-Iscar (University of Valladolid, Spain) *A trimming and constraints based robust clustering proposal for matrix-variate data*
12:45–14:20	**Lunch Break (Butik Otel Personel Yemekhanesi)**
14:20–16:00	**IS6 – Advances in Robustness for Complex Data** Chairs: Luis Angel García Escudero (University of Valladolid, Spain) and Agustin Mayo Iscar (University of Valladolid, Spain) Christian Hennig (University of Bologna, Italy) *What are outliers?* Valentin Todorov (United Nations Industrial Development Organization, Austria) *Automatic choice of k and c in model-based clustering: The R package tclust* Aldo Corbellini (University of Parma, Italy) *Robust statistics through monitoring: outliers and transformations*

	Luis Angel García-Escudero (University of Valladolid, Spain) *Robust trimmed clustering for outlier detection in complex settings*
16:00–16:20	**Coffee Break**
16:20–17:35	**IS5 – Geometric and Bayesian Perspectives in Robustness** Chair: Alberto González-Sanz (Columbia University, United States) Zoraida Rico (Bocconi University, Italy) *Robust, Sub-Gaussian Mean Estimators in Metric Spaces* Shunan Sheng (Columbia University, United States) *Stability and instability of mean-field variational inference* Cynthia Rush (Columbia University, United States) *Frequentist Theory for M-posteriors: Asymptotic and Robustness Properties*

Kırmızı Salon (Red Hall)	
11:30–12:45	**IS12 – Robustness, Privacy and Statistical Learning** Chair: Marco Avella-Medina (Columbia University, United States) Badr-Eddine Chérief-Abdellatif (CNRS, France) *Robust Estimation via Maximum Mean Discrepancy* Getoar Sopa (Columbia University, United States) *Efficient Differentially Private Regression Inference for Longitudinal Data* Sophie Langer (Ruhr University Bochum, Germany) *A Statistical Approach to Image Classification via Deformable Templates*
12:45–14:20	**Lunch Break (Butik Otel Personel Yemekhanesi)**
14:20–16:00	**IS2 – Robust Methods for Functional Data Analysis** Chair: Aylin Alın (Dokuz Eylül University, Turkiye) Ufuk Beyaztas (Marmara University, Turkiye) *Robust estimation for functional joint models with longitudinal and survival outcomes* Müge Mutiş (Yıldız Technical University, Turkiye) *A novel robust approach to spatially structured scalar-on-function quantile regression* Zdeněk Hlávka (Charles University, Czech Republic) *Multivariate two-sample tests for functional data with outliers* Filip Bočinec (Charles University, Czech Republic) *Projection depth for functional data*
16:00–16:20	**Coffee Break**

16:20–17:35	**IS11 – Recent Developments in Robustness** Chairs: Mia Hubert (KU Leuven, Belgium) and Stefan Van Aelst (KU Leuven, Belgium)) Peter Filzmoser (TU Wien, Austria) *Multivariate functional Mahalanobis distance with application to clustering* Yangzhuoran Fin Yang (Maastricht University, Netherlands) *Outlier Detection and Robust Estimation of Time Series Model using a Penalised Approach* Giorgia Zaccaria (University of Milano-Bicocca, Italy) Cellwise outlier detection in clustering: from model-based to fuzzy approaches

Yeşil Salon (Red Hall)	
14:20–16:00	**CS1 – Applications of Robust Methods-1** Chair: Fulya Gokalp Yavuz (Purdue University, United States) Burak Dilber (Dokuz Eylül University, Turkiye) *Comprehensive Evaluation of METHoD-NO for Benchmark Optimization and Neural Network Training* Yasemin Gunter (Halic University, Turkiye; Yildiz Technical University, Turkiye) *GWO-Optimized Hybrid Deep Learning Approach for Lung Cancer Detection: High-Accuracy Classification on CT Images* Gulder Kemalbay (Yildiz Technical University, Turkiye) *Bootstrap-Based Robust Higher-Order Moment Portfolio Optimization under Regime Heterogeneity: A Homogeneity-Driven Estimation Approach* Jens Klooster (University of Groningen, Netherlands) *On inference using M-estimators under contamination* Nuri Berk Ural (Cukurova University, Turkiye) *Hybrid sliding window recursive least squares for battery state of health prediction*

17:35-18:15	**Steering Comittee Meeting**

18:15–20:00	**Welcome Cocktail (Davutpaşa Congress Center)**

Tuesday, 21 July 2026

Büyük Salon (Main Hall)	
09:00–10:00	**Keynote 2** **Karen Kafadar (University of Virginia, United States)** ***Robust Joint Estimation of Survival Components to Assess Risks and Benefits of Cancer Screening*** Chair: Mia Hubert
10:00–10:20	**Coffee Break**
10:20–12:00	**IS10 – Symbolic Data Analysis meets Robust Statistics: Models, Methods and Applications (Hybrid Session)** Chair: Maria do Rosário Oliveira *(CEMAT and Instituto Superior Técnico, Universidade de Lisboa, Portugal)* Eufrasio Lima Neto (Federal University of Paraíba, Brazil) [Virtual] *Robust distributional regression models for interval-valued data* M. Rosário Oliveira (University of Lisbon, Portugal) [Virtual] *Clustering interval-valued data using L1-Wasserstein distance* Pedro Duarte Silva (Catholic University of Portugal, Portugal) *Robust Principal Component Analysis of Symbolic Matrix-valued Data* Renata Maria Cardoso Rodrigues Souza (Federal University of Pernambuco, Brazil) *Multivariate Fuzzy Partitioning Methods using Medoids*
12:00–13:30	**Lunch Break (Butik Otel Personel Yemekhanesi)**
13:30–14:30	**Keynote 3** **David E. Tyler (Rutgers University, United States)** ***Robustness Features of Invariant Coordinates Selection*** Chair: Karen Kafadar
14:35-14:45	**ICORS 2027 Annocument**
14:45–15:05	**Coffee Break**
15:05-16:45	**IS8 – Robust Inference in Combined Regression and Autoregression Models (Hybrid Session)** Chairs: Jana Jureckova (Charles University, Czech Republic), Olcay Arslan (Ankara University, Turkiye), and Yeşim Güney (Ankara University, Turkiye) Jana Jureckova (Charles University, Czech Republic) [Virtual] *The Advantage of Autoregression Rank Scores Invariance in the Inference* Hira L. Koul (Michigan State University, United States) [Virtual]

	A signed-rank estimator in nonlinear regression models when covariates and errors are dependent Jan Picek (Technical University of Liberec, Czech Republic) *Adaptive rank scores guided by l-moments for robust inference in regression–autoregression models* Yetkin Tuaç (Ankara University, Turkiye) *M-Quantile regression with autoregressive errors*
16:45–17:05	**Coffee Break**
17:05-18:20	**IS13 – Analysis of Complex Data** Chair: Domenico Perrotta (University of Milan, Italy) Can Hakan Dağıdır (KU Leuven, Belgium) [Virtual] *Cellwise Robust Discriminant Analysis* Andrea Cerioli (University of Parma, Italy) *Robustness, clustering and outlier detection under heavy tails* Arthur Daisomont (KU Leuven, Belgium & Joint Research Centre,Italy) *Blockwise, cellwise, and casewise robust multiblock PCA*

Kırmızı Salon (Red Hall)

10:20-12:00	**IS7 – Robust Regression for Complex Data** Chairs: Olcay Arslan (Ankara University, Turkiye) and Fatma Zehra Doğru (Giresun University, Turkiye) Olcay Arslan (Ankara University, Turkiye) *Exploring Expectile Regression Through Likelihood and EM Algorithm Perspectives* Fatma Zehra Doğru (Giresun University, Turkiye) *A Flexible Heteroscedastic Semiparametric Model for Skewed and Heavy-Tailed Data* Yeşim Güney (Ankara University, Turkiye) *Robust Joint Location–Scale Modeling via M-Quantile Estimation* Elgiz Askeroğlu (Giresun University, Turkiye) *Robust Particle Swarm Optimization for Regression via M-Estimation*
12:00–13:30	**Lunch Break (Butik Otel Personel Yemekhanesi)**
15:05-16:45	**CS3 – Robust Medians, Breakdown Points or Outlier Detection** Chair: Stanislav Nagy (Charles University, Czech Republic) Marco Avella Medina (Columbia University, United States) *The threshold breakdown point*

	Leopold Micheler (TU Wien, Austria) *Robust and explainable outlier detection for random surfaces* Erik Mendroš (Charles University, Czech Republic) *Breakdown points of simplicial and k-hull medians* Stanislav Nagy (Charles University, Czech Republic) *When is the Tukey median unique?*
16:45-17:05	**Coffee Break**
17:05-18:20	**IS3 – Uncertainty and Learning Dynamics for Causal and Personalized Decisions with Modern AI/ML** Chair: Liangyuan Hu (Rutgers School of Public Health, United States) Xiao Fang (Chinese University of Hong Kong, Hong Kong) *Martingale central limit theorem in Wasserstein distance with applications* Omer Lütfi Gebizlioglu (Kadir Has University, Turkiye) *Bayesian Credibility Estimation Using Concomitants of Order Statistics in the Framework of Generalized Linear Mixed Models*

Yeşil Salon (Green Hall)	
10:20-12:00	**CS2 – Robust Inference for Dependent, Bayesian or Categorical Data** Chair: F. Sevinc Kurnaz (Yıldız Technical University, Turkiye & Case Western Reserve University, United States) Xin Dang (University of Mississippi, United States) *Functional canonical correlation for discriminant analysis* Tuğba Kapucu (Middle East Technical University, Turkiye) *An ECM-Metropolis Algorithm for Parameter Estimation in Skew Laplace Linear Mixed Models* Mutlu Altuntaş (Sinop University, Turkiye) *Robust anomaly detection via a bayesian autoencoder with heavy-tailed priors* Maral Karbaschi (Alzahra University, Iran) *Uncertainty-aware statistical learning for robust clinical risk analytics*
15:05-16:45	**CS4 – Robust Modeling and Monitoring for Dependent Data** Chair: Onur Toka (Hacettepe University, Turkiye) Ziling Ma (King Abdullah University of Science and Technology, Saudi Arabia) *Forecasting Multivariate Time Series under Predictive Heterogeneity: A Validation-Driven Clustering Framework* Talha Arslan (Van Yüzüncü Yıl University, Turkiye) *The unit-t2 distribution: a parametric approach to median regression on the unit interval*

	Iris Aragon Mladosich (KU Leuven, Belgium) *Robust XGBoosting for Regression*
16:45-17:05	**Coffee Break**

Wednesday, 22 July 2026

Büyük Salon (Main Hall)	
09:00–10:00	**Keynote 4** **Elvezio Ronchetti (University of Geneva, Switzerland)** ***Robustifying the Bayesian Approach*** Chair: Peter Rousseeuw
10:05-10:55	**Turkish Statistical Institute (TurkStat) Presentation: Official Statistics in Türkiye**
10:55-11:15	**Coffee Break**

10:55-11:40	**Poster Session** Piotr Waz (Medical University of Gdańsk, Poland) *A Novel Computational Tool for AMD Diagnosis* Gianfranco Piscopo (University of Naples Federico II, Italy) *Robust Assessment of Benford's Law under Skewed Generalized Error Distributions via Nonparametric Combination* Igor Böhm (Charles University, Czech Republic) *Spherical Projection Depth* Sandra Esperanza Melo (The National University of Colombia, Colombia) *Methodology for the Longitudinal Analysis of thrips Counts in Chrysanthemum Cultivation* Angel Lopez Oriona (King Abdullah University of Science and Technology, Saudi Arabia) *Amortized Neural Clustering of Time Series based on Statistical Features* Dorota Bielinska Waz (Medical University of Gdańsk, Poland) *Novel alignment-free bioinformatics methods*

Büyük Salon (Main Hall)	
11:40-12:55	**CS5 – Robust Methods for Sampling and Forecasting** Chair: Yetkin Tuaç (Ankara University, Turkiye) Özge Gürer (Ankara University, Turkiye) (Virtual) *Ranked Set Sampling based ANOVA under Non-Normality* Özge Gürer (Ankara University, Turkiye) (Virtual) *Robust Tests for One-Sample and Two-Sample Problems under Ranked Set*

	Sampling: An Application to Real Data Antonio Panico (University of Parma, Italy) *Sequential Conformal Prediction for Electricity Price Forecasting*

Kırmızı Salon (Red Hall)	
11:40-12:55	**IS15 – Advances in Robust Methodologies** Chair: F. Sevinc Kurnaz (Yıldız Technical University, Turkiye & Case Western Reserve University, United States) Matthias Templ (University of Applied Sciences and Arts Northwestern Switzerland, Switzerland) *Cellwise-robust imputation for mixed continuous and categorical data* Houyem Demni (University of Cassino and Southern Lazio, Italy) *Robust Estimation of Circular Dispersion: Theory and an Application to Outlier Detection* Onur Toka (Hacettepe University, Turkiye) *A Systematic Comparison of Clustering Methods Under Casewise, Cellwise, and Mixed Data Contamination*

13:30–19:00	**City Tour (Lunch boxes will be provided to participants)**

Thursday, 23 July 2026

Büyük Salon (Main Hall)	
09:00–10:00	**Keynote 5** **Yanyuan Ma (Penn State University, United States)** ***Several Studies in Label Shift*** Chair: Peter Filzmoser
10:00–10:20	**Coffee Break**
10:20–12:00	**IS14 – Robust Solutions in Machine Learning Applications (Hybrid Session)** Chair: Fulya Gokalp Yavuz (Purdue University, United States) David Iseri Inouye (Purdue University, United States) [Virtual] *Provable Robustness to Spurious Correlations via Invariant Data for Robust Finetuning* Burcu Koca Guler (Middle East Technical University, Turkiye) *CGL-LIME: A Robustified LIME with Copula-Based Sampling and Gradient-Weighted Kernels* Serenay Çakar (Middle East Technical University, Turkiye) *From robust mixed models to temporal deep learning: a five-paradigm robustness benchmark on contaminated fNIRS data* Şenay Özdemir (Afyon Kocatepe University, Turkiye) *Robustness of exponential family and m-estimation based empirical likelihood in linear regression*
12:00–13:30	**Lunch Break (Butik Otel Personel Yemekhanesi)**
13:30–14:45	**IS17 – Stability in Inference from Different Prospective (Hybrid Session)** Chairs: Claudio Agostinelli (University of Trento, Italy) and Anand Vidyashankar (George Mason University, United States) Giulia Bertagnolli (Free University of Bozen-Bolzano, Italy) [Virtual] *Robust Inference for Mixtures of Multivariate Wrapped Normal Distributions: The Weighted Likelihood Approach* Kamelia Daudel (ESSEC Business School, France) *Optimizing Rényi's alpha-divergences: a variational inference perspective* Badr-Eddine Chérief-Abdellatif (CNRS, France) *Bayesianism and Robustness: Predictively Oriented (Pr0) Posteriors*
15:10-15:30	**Coffee Break**
15:30-16:50	**CS8 – Applications of Robust Methods-3** Chair: Özlem Kaymaz (Ankara University, Turkiye)

	Tobias Leopold (Esslingen University of Applied Sciences, Germany) *Robust Reliability Demonstration Test Plans* Nabamallika Dehingia (United Nations Development Programme, United States) (Virtual) *Projecting Multidimensional Poverty to 2030: A Scenario-Based Microsimulation Framework* Gokalp Tepekoylu (Medipol University, Turkiye) (Virtual) *Achieving Nearly Deterministic Precision in Dental Implant Identification* Onur Camli (Erzurum Technical University, Turkiye) (Virtual) *Robust Circular–Circular Correlation Coefficient via Huber and Tukey Weighting*

Kırmızı Salon (Red Hall)

10:20–12:00	**CS6 – Robust Methods for Non-Standard and Complex Data** Chair: Maria-Pia Victoria-Feser (University of Bologna, Italy) Leonardo Leone (Télécom Paris IPP, France) *Massive parallelization of projection-based depths* Kelly Ramsay (York University, Canada) *Robust smoothing splines with discontinuities* Peter Ruckdeschel (Carl Ossietzky University of Oldenburg, Germany) *Robustness Problems in Relaxed and Entropy Regularized Portfolio Optimization* Maria-Pia Victoria-Feser (University of Bologna, Italy) *Simulation-Based Robust Inference for Response Misclassification Errors with Self-Reported Responses*
12:00–13:30	**Lunch Break (Butik Otel Personel Yemekhanesi)**
13:30–14:45	**IS9 – Robust Principal Components** Chairs: Gabriela Cohen Freue (University of British Columbia, Canada) and David Tyler (Rutgers University, United States) Una Radojicic (Technical University of Vienna, Austria) *New M-estimator of the leading principal component* Shojaeddin Chenouri (University of Waterloo, Canada) *Robust Principal Components Analysis via Subspace Discrepancy Minimization* Yetkin Tuaç (Ankara University, Turkiye) *Adaptive Lasso-regularized Spectral Generalized Kotz-type Robust Matrix Regression*
15:10-15:30	**Coffee Break**

15:30-16:50	**CS9 – Robust Regression, Clustering, and Dimension Reduction** Chair: Şükrü Acıtaş (Eskişehir Technical University, Turkiye) Hasan Bulut (Ondokuz Mayıs University, Turkiye) *TSRClust: Two-Stage Robust Clustering* Han Ming Wu (National Chengchi University, Taiwan) *SIIE: sliced inverse interval estimation for robust sufficient dimension reduction* Ndeye Niang (CNAM Cédric, France) *Robust Feature Clustering around Latent Components* Ismail Yenilmez (Eskişehir Technical University, Turkiye) *Robust Sparse Principal Component Regression with Alternative Loss Functions: A Framework for High-Dimensional Settings*

Yeşil Salon (Green Hall)	
10:20–12:00	**CS7 – Applications of Robust Methods-2** Chair: Meral Çetin (Hacettepe University, Turkiye) Serhan Tunçel (Hacettepe University, Turkiye) *Application of Robust SUR Modeling and Variable Selection to Air Pollution Data in Ankara* Özge Ustahüseyin (Ankara University, Turkiye) *The marshall–olkin weibull process: methodological framework and applications to environmental data* Senol Celik (Bingöl University, Turkiye) *Robust Regression for Agricultural Yield Estimation Under Influential Observations: Evidence from Konya Districts* Gunseli Aytac Cankurtaran (Ankara University, Turkiye) *Analysis of Installed Power Levels of Solar Power Plants in Turkey Using Spatial and Robust Spatial Regression* Onur Toka (Hacettepe University, Turkiye) *Comparing deletion methods and imputation methods in seemingly unrelated regression models for air quality data*

18:30-23:00	**Gala Dinner**

ICORS2026
20-24 JULY 2026
Istanbul

Friday, 24 July 2026

Büyük Salon (Main Hall)	
09:00–10:00	**Keynote 6** **Alberto Gonzalez-Sanz (Columbia University, United States)** ***Robustness of the Optimal Transport Map*** Chair: Agustin Mayo-Iscar
10:00–10:20	**Coffee Break**
10:20–12:00	**IS16 – Robust Estimation, Testing and Prediction** Chair: Yanyuan Ma (Penn State University, United States) Tanya Garcia (University of North Carolina at Chapel Hill, United States) *When Distribution-Free Falls Short for Prediction Intervals under Right-Censored Covariate* Renjun Ma (University of New Brunswick, Canada) *Tweedie compound Poisson mixed models for cross-classified data with left censoring* Ying Wei (Columbia University, United States) *Quantile Graph Discovery through QuACC: Quantile Association via Conditional Concordance*
12:00	**Closing Remarks**

Kırmızı Salon (Red Hall)	
10:20–12:00	**IS4 – Robustness in Complex Models** Chairs: Graciela Boente (University of Buenos Aires, Argentina) and Ana Maria Bianco (University of Buenos Aires, Argentina) Fabio Centofanti (KU Leuven, Belgium) *Cellwise and casewise robust multivariate regression with inference* Gabriela Cohen Freue (University of British Columbia, Canada) *Fast and Scalable Cellwise-Robust Ensembles for High-Dimensional Data* Stefan Van Aelst (KU Leuven, Belgium) *Robust inference for nonparametric regression via bootstrap* Conceição Amado (University of Lisbon, Portugal) *On Robust Estimation for Heteroscedastic Nonlinear Regression Models*

Yeşil Salon (Green Hall)	
10:20–12:00	**CS10 – Robust Penalized Regression** Chair: Abdullah Yalçınkaya (Ankara University, Turkiye Ismail Yenilmez (Eskişehir Technical University, Turkiye) *Graph Robust Adaptive LASSO for Network-Aware Variable Selection under Contamination*

ICORS2026
20-24 JULY 2026
Istanbul

Keynote Presentations

Keynote 1	**Peter J. Rousseeuw (KU Leuven, Belgium)** *Independent Component Analysis by Robust Distance Correlation*
Keynote 2	**Karen Kafadar (University of Virginia, United States)** *Robust Joint Estimation of Survival Components to Assess Risks and Benefits of Cancer Screening*
Keynote 3	**David E. Tyler (Rutgers University, United States)** *Robustness Features of Invariant Coordinates Selection*
Keynote 4	**Elvezio Ronchetti (University of Geneva, Switzerland)** *Robustifying the Bayesian Approach*
Keynote 5	**Yanyuan Ma (Penn State University, United States)** *Several Studies in Label Shift*
Keynote 6	**Alberto Gonzalez-Sanz (Columbia University, United States)** *Robustness of the Optimal Transport Map*

Independent Component Analysis by Robust Distance Correlation

Peter J. Rousseeuw
Section of Statistics and Data Science, KU Leuven, Belgium
Email: peter.rousseeuw@kuleuven.be

Independent component analysis (ICA) is a powerful tool that attempts to decompose a multivariate signal or distribution into fully independent sources, not just uncorrelated ones like PCA does. ICA is harder to do, but it has many important applications. Unfortunately, most approaches to ICA are not robust against outliers. Here we propose a robust ICA method called PICARD, which estimates the components by minimizing a robust measure of dependence between multivariate random variables. The dependence measure used is the distance correlation (dCor). In order to make it more robust we first apply a new transformation called the bowl transform, which is bounded, one-to-one, continuous, and maps far outliers to points close to the origin. This preserves the crucial property that a zero dCor implies independence. PICARD estimates the independent sources sequentially, by looking for the component that has the smallest dCor with the remainder. We prove that PICARD is strongly consistent. Its robustness is investigated by a simulation study, in which it generally outperforms its competitors. The method is illustrated on three applications, including the well-known cocktail party problem.

Keywords: *Algorithm, Bowl transform, Cocktail party problem, Multivariate data, Outliers*

References

[1] Leyder, S., Raymaekers, J., Rousseeuw, P.J., Van Deuren, T., Verdonck, T. (2025). Independent Component Analysis by Robust Distance Correlation, preprint, arXiv:2505.09425

Robust Joint Estimation of Survival Components to Assess Risks and Benefits of Cancer Screening

Karen Kafadar
Department of Statistics, University of Virginia, Charlottesville, VA, USA [2]Second and Third
Email: kk3ab@virginia.edu

Length biased sampling (LBS) arises when items are sampled in proportion to their values on a random variable of interest. For example, older units may be more likely to be sampled simply because they have been in service for a longer period of time. The effect of this sampling bias on the mean is well known when the length-biased-sampled random variable, say Y, is observable.

A more difficult situation arises when Y is not observed, but the outcome of another random variable, Z, is observed and is correlated with Y. This scenario arises in evaluating screening programs: screening identifies cases during the preclinical phase, the duration of which is unobserved but is correlated with the clinical duration. Length-biased preclinical durations are more likely to be screen-detected than shorter ones and also may have better prognosis, irrespective of screening. Survival is further biased by the effects of lead time and overdiagnosis. We demonstrate the implications of these biases, propose a survival time model that incorporates them, and offer a robust approach to jointly estimating the components of survival, including extended benefit time, in screening programs. The approach is illustrated using data from six actual randomized cancer screening trials.

This work is conducted in collaboration with Dr. Philip C. Prorok, National Cancer Institute.

Robustness Features of Invariant Coordinates Selection

David E. Tyler
Department of Statistics, Rutgers University, New Brunswick, NJ
Email: dtyler@stat.rutgers.edu

When sampling from a multivariate normal distribution, the sample mean vector and sample covariance matrix provide a sufficient summary of the data set. To protect against nonnormality, and in particular against longer tailed distributions and outliers, one can replace the sample mean vector and covariance matrix with robust estimates of multivariate location and scatter. Outliers can often be detected by examining the corresponding robust Mahalanobis distances. Such an approach is appropriate if the bulk of the data arises from a multivariate normal distribution or more generally from an elliptically symmetric distribution. However, if the data arises otherwise, then different location/scatter estimates do no estimate the same population quantities, but rather are reflecting different aspects of the underlying distribution.

Invariant Coordinate Selection [1] is a general multivariate method based on the idea that examining differences between scatter estimates may uncover interesting structures in multivariate data, ones which may not be apparent from a plot of robust Mahalanobis distances. ICS is based on the eigenvalue-eigenvector decomposition of one estimate of scatter relative to another. An important property of this decomposition is that the corresponding eigenvectors generate an affine invariant coordinate system (ICS) for the multivariate data. This leads to new affine equivariant multivariate statistical and graphical methods. By plotting the data with respect to this new invariant coordinate system, various data structures can be revealed. For example, under certain independent component analysis models, which are popular within computer science and engineering disciplines, the invariant coordinates correspond to the independent components. Also, if the data arises from a mixture of elliptical distributions, then a subset of the invariant coordinates correspond to Fishers linear discriminant subspace, even though the class identification of the data points are unknown.

The goal of this talk is to review ICS, to discuss its robustness features, and to explain why the method is able to be resistant to outlier. In particular, it is noted that ICS passes the “wheelbarrow” test for multivariate robustness.

Keywords: *Clustering, Discriminant analysis, ICA, ICS, M-estimation, Unmixing*.

References

[1] Tyler, D.E., Critchley, F., Dümbgen, L. and Oja, H. (2009). Invariant co-ordinate selection (with discussion). *JRSS-B,* 71(3) pp. 549–592.

Robustifying the Bayesian Approach

Elvezio Ronchetti[1]*
[1]Research Center for Statistics and GSEM, University of Geneva, Switzerland
*Corresponding Author Email: Elvezio.Ronchetti@unige.ch

In the past several decades there has been an important development of the theory and applications of robust statistics. This has taken place mainly within the frequentist framework, while fewer results have concerned the Bayesian approach. Since robust statistics deals with deviations from ideal models and develops statistical procedures which are still reliable and reasonably efficient in a neighborhood of the model, the issue of the stability of inference in the presence of small deviations from the assumptions should clearly concern both approaches. This is even more important nowadays, where the analysis and the modelling of complex data are required in many fields, in particular for the development of AI technology. Fortunately, in the past decade with the development of powerful algorithms, the robustness issue has gained importance within the Bayesian framework.
In this talk we discuss some of these recent developments by focusing on two main aspects. First, we outline the transfer of some fundamental ideas and tools from the classical theory of robust statistics (including M-estimation and testing and Huber's minimax theory) to the Bayesian setup. One implication of this transfer is the recommendation to replace exact likelihoods with Huber's least favorable distributions when sampling from posterior distributions. Secondly, we discuss the difficulty of obtaining exact finite sample results in Bayesian robustness, while outlining a proposal, which aims to combine asymptotic guarantees with exact finite sample bounds. Finally, we briefly illustrate how the Bayesian filter can be robustified.

Keywords: *Asymptotic guarantees, robust Bayesian filter, finite sample bounds, least favorable distributions, minimax theory*

ICORS2026
20-24 JULY 2026
Istanbul

Several Studies in Label Shift

Yanyuan Ma
Department of Statistics, Eberly College of Science, The Pennsylvania State University, University Park, PA, USA
Email: yzm63@psu.edu

We provide an introduction to label shift problems. In the context of discrete response, we study the importance weights confidence set problem by a paradigm shift from traditional inversion-based inference to a direct matrix constraint framework. We use this framework to characterize a joint confidence region and extract marginal intervals via linear programming, deriving provably tighter bounds for importance weights while maintaining exact finite-sample validity. In the context of continuous response, we study the estimation and inference of a general target population characteristic by developing doubly and singly robust estimators as well as the efficient estimator. Many ongoing and future developments will be discussed too.

ICORS2026
20-24 JULY 2026
Istanbul

Robustness of the Optimal Transport Map

Alberto González-Sanz
Department of Statistics, Columbia University, New York, USA
Email: ag4855@columbia.edu

Optimal transport (OT) was originally formulated as a resource allocation problem. More recently, it has found many applications in statistics and data science, including statistical testing, multivariate distribution functions, and variational inference. A central object in OT is the transport map T, which sends a reference measure μ to a target measure P, typically representing the data distribution. In this talk, we discuss recent advances on the robustness of the transport map T evaluated at a point u.

First, we study its robustness in terms of the breakdown point, namely, the smallest level of contamination of the target measure P that can force T(u) to take arbitrarily extreme values. This result holds for transport maps induced by strictly convex costs. In particular, it implies that the transport-based median introduced by Hallin et al. (2021) has breakdown point 1/2. More generally, points on the transport-based depth contour of order $\tau \in [0, 1/2]$ have breakdown point τ, showing that multivariate transport depth has the same robustness as its univariate counterpart. The argument relies on a connection between the breakdown point of transport maps and the Tukey depth of points in the reference measure.

In the second part of the talk, we study the influence function of the transport map. We show that the influence function of T(u) is unbounded and has a pole-type singularity when the perturbation is located near T(u). This suggests that the most harmful perturbations of the target measure P are those concentrated near T(u). We will briefly outline the proof, which uses PDE methods, and conclude with a list of open problems.

Invited Session I (IS1)

IS1 – Robustness for Matrix-valued Data
Chair: Peter Filzmoser (TU Wien, Austria)

Marcus Mayrhofer (TU Wien, Austria)
High-dimensional separability testing in matrix data beyond normality

Mia Hubert (KU Leuven, Belgium)
Robust tensor-on-tensor regression

Agustin Mayo-Iscar (University of Valladolid, Spain)
A trimming and constraints based robust clustering proposal for matrix-variate data

High-dimensional separability testing in matrix data beyond normality

Marcus Mayrhofer[1]*, Una Radojičić[1], Tomas Masak[2]
[1]Computational Statistics, TU Wien, Vienna, Austria
[2]Department of Finance Accounting and Statistics, WU Vienna, Vienna, Austria
*Corresponding Author Email: marcus.mayrhofer@tuwien.ac.at

Separability is an important structural assumption often placed on the covariance when working with matrix-variate data, because it greatly simplifies both the interpretation and the computation of subsequent covariance-based statistical tasks. Yet testing the separability assumption is difficult in the high-dimensional regime. We propose testing separability by recasting the problem as a sphericity test after whitening the data using the separable maximum-likelihood estimate of the covariance. The test is calibrated via Monte Carlo simulation, yielding finite-sample-level control. Furthermore, we prove high-dimensional consistency of the test under dense alternatives. To reduce reliance on distributional assumptions, we introduce an angular version of the test based on radial normalization after whitening. We demonstrate the test's practical utility, statistical power, and computational efficiency of the proposed tests in a large simulation study and on real-world data.
Keywords: *Monte Carlo simulations, separable covariance matrix, hypothesis testing, matrix-valued data*

Robust tensor-on-tensor regression

Medhi Hirari[1], Fabio Centofanti[1], Mia Hubert[1]*, Stefan Van Aelst[1]
[1]Section of Statistics and Data Science, KU Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium
*Corresponding Author Email: mia.hubert@kuleuven.be

Tensor-on-tensor (TOT) regression is an important tool for the analysis of tensor data, aiming to predict a set of response tensors from a corresponding set of predictor tensors [1]. However, standard TOT regression is sensitive to outliers, which may be present in both the response and the predictor. It can be affected by casewise outliers, which are observations that deviate from the bulk of the data, as well as by cellwise outliers, which are individual anomalous cells within the tensors. The latter are particularly common due to the typically large number of cells in tensor data.

This paper introduces a novel robust TOT regression method, named ROTOT, that can handle both types of outliers simultaneously and can cope with missing values as well [2]. This method uses a single loss function to reduce the influence of both casewise and cellwise outliers in the response. The outliers in the predictor are handled using a robust Multilinear Principal Component Analysis method. Graphical diagnostic tools are also proposed to identify the different types of outliers detected. The performance of ROTOT is evaluated through extensive simulations and further illustrated using the Labeled Faces in the Wild dataset, where ROTOT is applied to predict facial attributes.

Keywords: *Tensor data; Tensor regression; Robust statistics; Casewise outliers; Cellwise outliers; Anomaly detection*

References

[1] Lock, E. F. (2018). Tensor-on-tensor regression. Journal of Computational and Graphical Statistics, 27 (3), 638–647.
[2] Hirari, M., Centofanti, F., Hubert, M., Van Aelst, S. (2026). Robust tensor-on-tensor regression. arXiv preprint, https://arxiv.org/abs/2603.25911.

A trimming and constraints based robust clustering proposal for matrix-variate data

L.A. García-Escudero[1] and A. Mayo-Iscar[1*] and M. Mayrhofer[2]
[1]Universidad de Valladolid, Valladolid, Spain
[2]Vienna University of Technology, Vienna, Austria
*Corresponding Author Email: agustin.mayo.iscar@uva.es

We propose a robust clustering estimator for data arising from a mixture of matrix normal distributions. On the one hand, we adopt a Kronecker-product covariance structure based on row and column covariances. On the other hand, the robustness of the estimator relies on the joint application of trimming, to avoid the influence of units coming from contaminating sources, and eigenvalue constraints, to ensure a well-posed estimation problem and to avoid spurious solutions.
Our proposal, MTCLUST, generalizes the Matrix Minimum Covariance Determinant [1], a robust estimator of mean and covariance for matrix-valued data, and TCLUST [2], a robust model based clustering proposal for multivariate data. We provide empirical evidence demonstrating the performance of this methodology on both simulated and real datasets.
Keywords: *Matrix-variate data – Model-based clustering – Robustness – tclust*

References

[1] Mayrhofer, M., Radojičić, U., & Filzmoser, P. (2025). Robust covariance estimation and explainable outlier detection for matrix-valued data. Technometrics, 67(3), 516-530.
[2] García-Escudero L.A., Gordaliza A., Matrán C. & Mayo-Iscar A. A general trimming approach to robust cluster analysis. Annals of Statistics, 2008, 36(3), 1324–1345.

Invited Session 2 (IS2)

IS2 – Robust Methods for Functional Data Analysis
Chair: Aylin Alın (Dokuz Eylül University, Turkiye)

Ufuk Beyaztas (Marmara University, Turkiye)
Robust estimation for functional joint models with longitudinal and survival outcomes

Müge Mutiş (Yıldız Technical University, Turkiye)
A novel robust approach to spatially structured scalar-on-function quantile regression

Zdeněk Hlávka (Charles University, Czech Republic)
Multivariate two-sample tests for functional data with outliers

Filip Bočinec (Charles University, Czech Republic)
Projection depth for functional data

ICORS2026
20-24 JULY 2026
Istanbul

Robust estimation for functional joint models with longitudinal and survival outcomes

Ufuk Beyaztas
Department of Statistics, Marmara University, Kadikoy, Istanbul,
Corresponding Author Email: ufuk.beyaztas@marmara.edu.tr

Functional joint models provide a principled framework for analyzing longitudinal trajectories together with time-to-event outcomes, while accounting for their association through shared latent subject-specific effects. However, existing functional joint modeling pipelines typically rely on standard dimension reduction techniques and Gaussian longitudinal errors, making them sensitive to contamination in both the response measurements and the functional covariates, features that are common in biomedical and wearable-sensor studies. We propose a fully robust functional joint modeling approach that addresses these limitations through a two-stage strategy. First, functional predictors entering the longitudinal and survival submodels are mapped to a stable finite-dimensional representation using robust functional principal component analysis, yielding robust scores and basis functions under outlying or noisy functional observations. Second, conditioning on these robust scores, we estimate the joint model via a robust adjusted-profile h-likelihood (APHL) that employs a Student-t likelihood for the longitudinal process and a Weibull proportional hazards specification for the time-to-event process, linked by a shared random intercept and an association parameter. The APHL formulation incorporates a Laplace adjustment to profile out random effects while retaining computational scalability and avoiding the sensitivity of full likelihood maximization under heavy-tailed errors. We develop a numerically stable iterative algorithm combining Student-t EM-type weighting with Newton updates for the association and baseline parameters. Simulation studies based on a functional joint model design demonstrate that the proposed method improves estimation accuracy of functional coefficient functions and scalar parameters under response outliers and functional covariate contamination, while remaining competitive in clean settings. The proposed framework provides a robust, interpretable, and computationally efficient alternative for functional joint modeling in the presence of heterogeneous measurement errors and outliers. We further demonstrate the practical utility of the proposed method by applying it to the Alzheimer's Disease Neuroimaging Initiative dataset.

Keywords: *Functional joint modeling, h-likelihood, outlier and measurement-error robustness, robust functional principal component analysis, student-t longitudinal model, Weibull proportional hazards*.

A novel robust approach to spatially structured scalar-on-function quantile regression

Müge Mutiş[1]*
[1]Yildiz Technical University, Istanbul, Turkiye, 34220
*muge.mutis@yildiz.edu.tr

This study proposes a robust estimation approach for a scalar-on-function quantile regression model incorporating a spatially lagged response term. The approach begins with a robust projection mechanism based on the quantile relationship between the response and the functional covariate. This mechanism generates data-adaptive projection directions while limiting the influence of atypical observations on the finite-dimensional representation through an adaptive weighting structure. The resulting model is then estimated within a spatial quantile regression framework, where the spatial lag term is incorporated into the estimation process through fitted values obtained from the covariates and their spatial derivatives. The parameters are determined by optimizing the quantile loss function while accounting for observation weights, thereby also controlling potential biases that may arise through the spatial network channel. This integrated structure enables more stable and reliable estimation of both the functional effect and the spatial autoregressive parameter. Monte Carlo experiments and the empirical application demonstrate that the proposed approach exhibits stable and competitive performance across different scenarios and provides notable improvements in estimation accuracy and inferential performance, particularly in the presence of contamination.

Keywords: *Functional data, spatial dependence, quantile regression, robust estimation*

Multivariate two-sample tests for functional data with outliers

Zdeněk Hlávka[1]*
[1]Charles University, Faculty of Mathematics and Physics, Sokolovská 49/83, Prague, Czechia
*Corresponding Author Email: hlavka@karlin.mff.cuni.cz

We are interested in the robustness of two-sample tests for functional data. As an illustration, we use the phoneme data set (Figure 1A), i.e., log-periodograms of digitized speech [1] and investigate first 15 or 40 observations for phonemes *aa* and *ao* split into five segments. For the last two segments, two-sample permutation tests comparing mean functions (Fmax test), covariance operators (SQ test) or characteristic functionals [2] do not detect any statistically significant difference. The same two-sample tests detect significant difference of covariance operators in segments 1–2 and mean functions in segments 1–3 (for n=40) or 2–3 (for n=15). Although the so-called *nonparametric combination functions* (e.g. Fisher, Tippet, or Liptak) can be used to address the issue of multiple testing (caused by combining Fmax and SQ tests), the combined permutation test does not provide any clear information concerning the reasons for rejecting the null hypothesis.

(A) (B) (C)

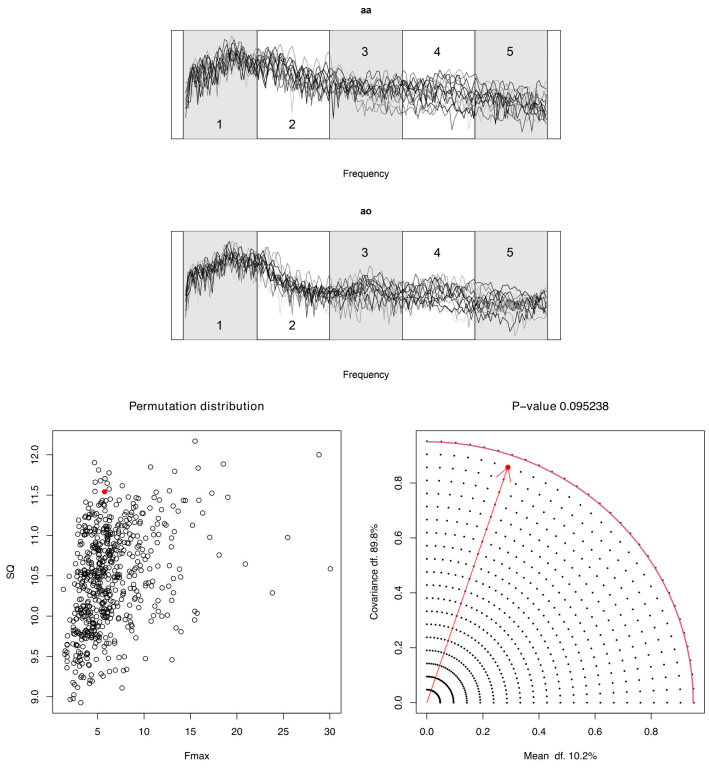


Figure 1: Log periodograms for two phonemes (A), test statistic T= (Fmax, SQ)^T (red dot) with the corresponding bivariate permutation distribution (B), and a

representation of the observed test statistic (red arrow) on a spherical product-type grid (C).

To improve the interpretability of a multivariate permutation test, Hlávka et al [3] suggested to transform the multivariate permutation distribution (plotted in Figure 1B for the bivariate test statistic $T = (Fmax, SQ)^T$ in segment 1 and n = 15) to a suitable spherical grid (Figure 1C) using the *empirical measure transportation* approach [4]. In Figure 1C, the red arrow denotes the value $F_T = (F_1, F_2)^T$ of the so-called empirical center-outward distribution function [4] corresponding to the observed test statistic. The permutation test p-value can then be calculated simply as the probability mass of the permutation distribution that is more extreme than the observed test statistic from the center-outward perspective, i.e., as $1 - \|F_T\|$. Moreover, $F_1^2 + F_2^2 = \|F_T\|^2$ can be interpreted as a *nonconformity score* of the observed data set with respect to the null hypothesis (note that we reject the null hypothesis on significance level α if and only if $F_1^2 + F_2^2 > (1 - \alpha)^2$) and, therefore, the squared coordinates F_1^2 and F_2^2 measure the importance of the two marginal test statistics for the test decision. At the same time, the test result can be displayed graphically. In Figure 1C, the red circle denotes the boundary of the critical region (for F_T) and the red arrow points to F_T. The upward pointing arrow suggests large contribution of the marginal test SQ (equality of covariance operators) but $\|F_T\| < 1 - \alpha = 0.95$ and the bivariate test statistic T is not significant.

In Figure 2A, we plot the results of all 10 bivariate permutation tests obtained for the five segments and sample sizes 15 and 40. For n = 15, we see statistically significant differences in segments 2 and 3, due mostly to differences in mean functions. For n=40, the test results are very similar in segments 2 and 3 but we observe a significant difference also in segment 1, with a somewhat larger contribution of the Fmax text ($F_1^2 / \|F_T\|^2 = 68.5\%$ and $F_2^2 / \|F_T\|^2 = 31.5\%$).

(A) (B)

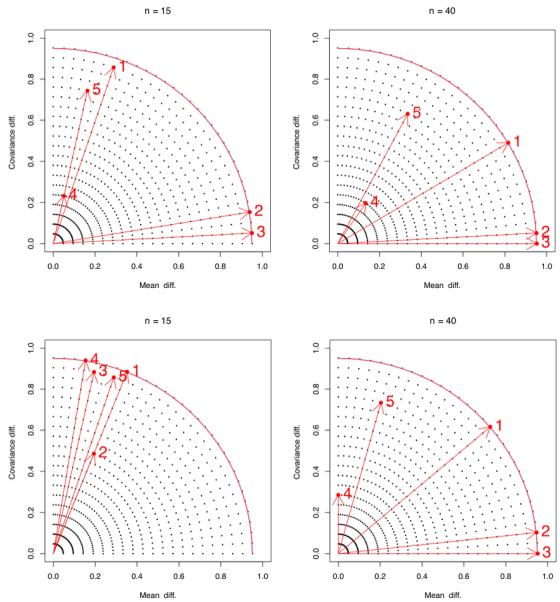


Figure 2: Bivariate permutation test results obtained for five segments and two sample sizes for the phoneme data set (A) and results for the same data set with an artificial additive outlier (B).

In Figure 2B, we use the same plots to visualize the effect of a single large additive outlier. For the smaller sample size, we observe that the differences in segments 2 and 3 become insignificant and the differences in segments 1 and 4 become significant, always with a large contribution of the SQ test. This suggests that the outlier affects the Fmax and SQ tests differently, decreasing the power of the Fmax test (equality of mean functions) and inflating the probability of type I error for the SQ test (equality of covariance operators). Interestingly, this outlier effect almost completely disappears for the larger sample size n=40.

Apart of discussing and investigating the effect of outliers on multivariate permutation two-sample tests for functional data, we will also propose suitable robust modifications and establish small sample properties of the resulting robust two-sample tests in a simulation study.

Keywords: *Empirical measure transport, functional data, multivariate permutation test, nonconformity score, robustness, two-sample test*

References

[1] Ferraty F., Vieu P. (2003). Curve discrimination: a nonparametric functional approach. Comput Stat Data Anal, 44(1-2), 161–173.
[2] Hlávka Z., Hlubinka D., Koňasová K. (2022). Functional ANOVA based on empirical characteristic functionals. J Multivar Anal, 189, 104878.
[3] Hlávka Z., Hlubinka D., Hudecová Š. (2025). Multivariate quantile-based permutation tests with application to functional data. J Comput Graph Stat, 34, 1276–1290.
[4] Hallin M., del Barrio E., Cuesta-Albertos J., Matrán C. (2021). Distribution and quantile functions, ranks and signs in dimension *d*: A measure transportation approach. Ann Stat, 49, 1139–1165.

ICORS2026
20-24 JULY 2026
Istanbul

Projection depth for functional data

Filip Bočinec[1]*, Stanislav Nagy[1], Hyemin Yeon[2]
[1]Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic, 12116
[2]Department of Mathematical Sciences, Kent State University, Kent, OH, USA, 44242
*Corresponding Author Email: bocinec@karlin.mff.cuni.cz

Data depth is a powerful non-parametric concept that provides a center-outward ordering of data points and plays a key role in robust multivariate analysis. Among the many depth notions developed over the years, projection depth stands out as one of the most fundamental due to its strong theoretical properties and intuitive geometric interpretation. For a point $x \in \mathbf{R}^d$ and a unit direction v, the outlyingness with respect to a random vector $X \sim P_X$ is defined as

$$O_v(x; P_X) = \frac{|\langle x, v\rangle - \mathrm{med}[\langle X, v\rangle]|}{\mathrm{MAD}[\langle X, v\rangle]}.$$

The classical projection depth [1] is then obtained by taking the infimum over all possible directions

$$D(x; P_X) = \inf_{v \in S}(1 + O_v(x; P_X))^{-1}).$$

However, the classical projection depth suffers from degeneracy in infinite-dimensional spaces, making it zero almost everywhere.

To overcome this fundamental limitation, we introduce a novel **Regularized Projection Depth (RPD)** for data in a Hilbert space $\mathcal{H}$. We incorporate regularization by appropriately constraining the set of admissible unit projection directions $v \in S = \{y \in \mathcal{H}: \|y\| = 1\}$. For a regularization parameter $\beta > 0$, we define the set of regularized directions as

$$V_\beta = \{v \in S: \mathrm{MAD}[\langle X, v\rangle] \geq \beta\}.$$

The RPD of $x \in \mathcal{H}$ is then defined by evaluating the outlyingness over this restricted set

$$D_\beta(x; P_X) = \inf_{v \in V_\beta}(1 + O_v(x; P_X))^{-1}).$$

By excluding directions along which the projections are too concentrated, we formally guarantee that $D_\beta(x; P_X) > 0$, allowing the depth to remain non-degenerate in functional settings. Crucially, unlike many existing functional depths, RPD is fully non-parametric and requires no moment assumptions on the underlying distribution.

In this talk, we present a comprehensive overview of the RPD, bridging its theoretical foundations with its practical utility. Based on our theoretical work [2], we investigate the properties of the RPD within a general Hilbert space framework. Furthermore, we illustrate its practical applicability in statistical tasks such as functional outlier detection

(effectively identifying shape outliers), functional classification, and two-sample problems [3].

Keywords: *functional data analysis, statistical depth, projection depth, robust statistics*

References

[1] Zuo, Y. (2003). Projection-based depth functions and associated medians. Ann. Statist. 31, 1460-1490.
[2] Bočinec, F., Nagy, S., Yeon, H. (2026+). Projection depth for functional data: Theoretical properties. Manuscript under review.
[3] Bočinec, F., Nagy, S., Yeon, H. (2026+). Projection depth for functional data: Computation and applications. Manuscript under review.

Invited Session 3 (IS3)

IS3 – Uncertainty and Learning Dynamics for Causal and Personalized Decisions with Modern AI/ML
Chair: Liangyuan Hu (Rutgers School of Public Health, United States)

Xiao Fang (Chinese University of Hong Kong, Hong Kong)
Martingale central limit theorem in Wasserstein distance with applications

Omer L. Gebizlioglu (Kadir Has University, Turkiye)
Bayesian Credibility Estimation Using Concomitants of Order Statistics in the Framework of Generalized Linear Mixed Models

Multivariate central limit theorem in Wasserstein distance

Xiao Fang[1]*, Zi-Yao Su[1]
[1]Department of Statistics and Data Science, The Chinese University of Hong Kong, Hong Kong SAR, China
*Corresponding Author Email: xfang@cuhk.edu.hk

We obtain new bounds in Wasserstein distance for multivariate martingale central limit theorems. They are stronger than the Yurinskii-type coupling results in the literature. We give applications to reinforcement learning, the stochastic gradient descent algorithm, and Bayesian inference. The main result is proved using recent advances in Stein's method. This is joint work with Zi-Yao Su.

Keywords: *Multivariate central limit theorem, martingale, Wasserstein distance, reinforcement learning, stochastic gradient descent algorithm, Bayesian inference*

Bayesian Credibility Estimation Using Concomitants of Order Statistics in the Framework of Generalized Linear Mixed Models

Omer L. Gebizlioglu*[1], Serap Yorubulut[2]

[1]Kadir Has University, Faculty of Economics, Administrative and Social Sciences, Department of International Trade and Finance, Istanbul, Turkiye, 34083
[2]Kırıkkale University, Faculty of Science and Letters, Department of Statistics, Kırıkkale, Turkiye, 71451

One of the most important topics in actuarial science is the determination of risk premiums. This is a matter of correct pricing for expected loss amounts that may stem from each and every insurance policy within the insurance portfolios of insurance companies. It is also a crucial corporate management issue for these companies that there should be a general balance between the claimed loss payments and earned premium incomes at both individual policy and collective portfolio levels. In this regard, credibility theory [1] proposes the most widely used of the systems developed to make an allocation process about loss payments and premium incomes more fair, realistic and healthy for insurance companies and their policy holders. Credibility theory treats an individual's true risk parameter as an unknown quantity that varies across a population. The models that credibility theory offers aim to find the optimal balance between a specific risk's own experience, in terms of individual data, and the collective experience, in terms of population data, by setting a credibility premium which is generally expressed as "$P = z$ {individual experience} + (1-z) {collective experience}" where z stands as a credibility factor such that $0 \leq z \leq 1$. More effective extensions of these models have come from the Bayesian credibility theory [2] which leads to the estimation of risk premiums by blending an individual's historical claims data with broader portfolio statistics. It applies Bayes' theorem to update risk beliefs and balancing individual experience against group averages using a credibility weight z. Hence, the final premium calculation of this approach resolves into a weighted average of two key components; a prior (collective) estimate of expected losses of broader group or portfolio and the individual experience that expresses actual observed loss history of a specific policyholder. In this paper we present a Bayesian credibity estimation approach by employing order statistics and their concomitants for credibility estimation within the framework of generalized linear mixed models (GLMM) [3]. We show that order statistics and their concomitants [4] have an important place in the modern actuarial modelling for the robustification of model estimations. So, we use concomitants along with order statistics to provide auxiliary ranking information, variance reduction and improved robustness in the Bayesian GLMM applications for better risk classifications and more reliable credibility premium estimations within the context of their link function specifications, fixed effects and random effects components, covariates and random effect design vectors.

Keywords: *Bayesian credibility, credibility, concomitants, GLMM, risk premium, order statistics*

References

[1] Bühlman, H., Gisler, A. (2005). A Course in Credibility Theory and its Applications. Springer.
[2] D'Hooge, L., Goovaerts, M.J. (2014). Bayesian Inference in Credibility Theory. Cambridge University Press, https://doi.org/10.1017/S05150036100009314.
[3] Antonio, K, Beirlant, J. (2007). Actuarial Statistics with Generalized Linear Mixed Model. Insurance Mathematics and Economics, 40, 57-76, Elsevier.
[4] David, H.A., Nagaraja, H. N. (2003). Order Statistics, 3rd Ed., Wiley

Invited Session 4 (IS4)

IS4 – Robustness in Complex Models
Chairs: Graciela Boente (University of Buenos Aires, Argentina) and Ana Maria Bianco (University of Buenos Aires, Argentina)

Fabio Centofanti (KU Leuven, Belgium)
Cellwise and casewise robust multivariate regression with inference

Gabriela Cohen Freue (University of British Columbia, Canada)
Fast and Scalable Cellwise-Robust Ensembles for High-Dimensional Data

Stefan Van Aelst (KU Leuven, Belgium)
Robust inference for nonparametric regression via bootstrap

Conceição Amado (University of Lisbon, Portugal)
On Robust Estimation for Heteroscedastic Nonlinear Regression Models

Cellwise and Casewise Robust Multivariate Regression with Inference

Fabio Centofanti[1]*, Mia Hubert[1], Peter J. Rousseeuw[1]
[1]Section of Statistics and Data Science, Department of Mathematics, KU Leuven, Belgium
*Corresponding Author Email: fabio.centofanti@kuleuven.be

Multivariate linear regression is a fundamental statistical task, but classical estimators such as ordinary least squares are highly sensitive to outliers. These may occur as casewise outliers that affect entire observations, or as outlying cells, which are individual contaminated entries in the predictor and/or response matrix. Moreover, modern datasets frequently contain missing values and are high-dimensional. To address these challenges, we propose the cellwise multivariate regression (cellMR) estimator, a robust regression method that simultaneously accommodates casewise and cellwise outliers, missing data, and high dimensionality. The approach builds on a cellwise robust covariance estimator and uses ridge regularization for numerical stability. We further introduce cellBoot, a novel bootstrap-based inference procedure tailored to the cellMR framework. Relying on indirect inference, cellBoot provides asymptotically valid confidence intervals that are robust to casewise and cellwise contamination. We derive influence functions of the regression estimator and prove the asymptotic validity of the cellBoot confidence intervals. Simulations and a real genomics application illustrate the strong finite-sample performance of the proposed methods.

Keywords: *Anomaly detection; Casewise outliers; Cellwise outliers; Confidence intervals; Indirect Inference.*

Fast and Scalable Cellwise-Robust Ensembles for High-Dimensional Data

Anthony Christidis[1], Jeyshinee Pyneeandee[2], Gabriela Cohen Freue[2]*
[1]Department of Biomedical Informatics, Harvard Medical School, Boston, USA
[2]Department of Statistics, University of British Columbia, Vancouver, BC, Canada
*Corresponding Author Email: name@institution.edu

High-dimensional data are often affected by cellwise contamination, in which individual entries rather than full observations are corrupted, challenging standard variable selection approaches. At the same time, recent ensemble methods have introduced deterministic frameworks that partition the predictor space to address high collinearity. However, these methods were not designed to accommodate cellwise contamination, leaving a critical methodological gap. To bridge this gap, we propose the Fast and Scalable Cellwise-Robust Ensemble (FSCRE) algorithm, which extends existing robust Least-Angle Regression (LARS) methodology from the casewise to the cellwise contamination setting and integrates it within an ensemble architecture to build multiple predictive models. Through extensive simulations and a bioinformatics application, we demonstrate FSCRE's competitive performance in variable selection precision, recall, and predictive accuracy across various contamination scenarios. This work provides a unified framework connecting cellwise-robust estimation with high-performance ensemble learning, with an implementation available on CRAN.

Keywords: Robust Statistics; Cellwise Contamination; High-Dimensional Data; Variable Selection; Ensemble Learning; Correlation Outliers.

Robust inference for nonparametric regression via bootstrap

Stefan Van Aelst[1]*, Matias Salibian-Barrera[2]
[1] Department of Mathematics, KU Leuven, Celestijnenlaan 200B, Leuven, Belgium, 3001
[2] Department of Statistics, University of British Columbia, 2207 Main Mall, Vancouver BC, Canada, V6T1Z4
*Corresponding Author Email: stefan.vanaelst@kuleuven.be

We consider the non-parametric regression model where the response depends on the predictor through some unknown smooth function f. The errors are assumed to be independent and identically distributed for some symmetric distribution with center zero. In this talk we consider robust penalized spline estimators where the loss function provides robustness against potential outliers and the smoothness is regularized by the penalty term. The estimators can be rewritten in a penalized weighted least squares form which is convenient to develop an iteratively weighted least squares algorithm for their computation. Asymptotic results are not sufficiently developed for such robust spline-based estimators to construct robust inference based on asymptotic theory. Therefore, we explore the use of bootstrap methods to develop robust inference. To obtain desirable results, the bias due to smoothing should be taken into account when bootstrapping nonparametric regression estimators. In this talk we explore bootstrap methods to develop robust inference based on robust spline-based estimators and investigate their performance.

Keywords: *Nonparametric regression, splines, robust inference, bootstrap*

On Robust Estimation for Heteroscedastic Nonlinear Regression Models

Conceição Amado[1]*, Ana M. Bianco[2], Graciela Boente[2], Isabel M. Rodrigues[1]

[1], CEMAT and Departamento de Matemática, Instituto Superior Técnico, Universidade Lisboa, Portugal
[2]Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Ciudad Universitaria, Pabellón 0+∞, 1428, Buenos Aires, Argentina
*Corresponding Author Email: conceicao.amado@tecnico.ulisboa.pt

This work addresses robust estimation in heteroscedastic regression settings, with a particular focus on situations where nonlinearity and heteroscedasticity occur simultaneously. Although under heteroscedasticity, robust estimators computed assuming errors with homogeneous variance remain consistent but inefficient, their interaction with nonlinear model structures create additional challenges, especially in the detection and treatment of outliers.

Building on existing robust approaches for models with heteroscedastic errors, we develop two iterative estimation procedures tailored to nonlinear regression. These procedures integrate weighted MM-type strategies to control the impact of high-leverage points, together with robust methods for modeling the variance structure.

The performance of the proposed methods is evaluated through a simulation study and illustrated using a real data application, highlighting their practical relevance and effectiveness.

Keywords: *Heteroscedastic errors, · MM-estimators, nonlinear regression, robust estimation*

Invited Session 5 (IS5)

IS5 – Geometric and Bayesian Perspectives in Robustness
Chair: Alberto González-Sanz (Columbia University, United States)

Zoraida F. Rico (Bocconi University, Italy)
Robust, Sub-Gaussian Mean Estimators in Metric Spaces

Shunan Sheng (Columbia University, United States)
Stability and instability of mean-field variational inference

Cynthia Rush (Columbia University, United States)
Frequentist Theory for M-posteriors: Asymptotic and Robustness Properties

Robust, Sub-Gaussian Mean Estimators in Metric Spaces

Zoraida F. Rico

Bocconi University

Estimating the mean of i.i.d. random variables under heavy tails and adversarial contamination is a fundamental problem in robust statistics. For samples taking values in a general metric space, the natural extension of the mean is the Fréchet mean.
We introduce a robust estimator inspired by high-dimensional trimmed means and establish non asymptotic deviation bounds under heavy tails and possible contamination. Our guarantees recover the optimal rates known in Euclidean spaces Applications include uniformly convex Banach spaces and metric spaces with curvature bounded from below.
This talk is based on the paper “Robust, Sub-Gaussian Mean Estimators in Metric Spaces”, joint work with Daniel Bartl (NUS), G´abor Lugosi (ICREA/UPF), and Roberto I. Oliveira (IMPA).

Stability and instability of mean-field variational inference

Shunan Sheng[1]*, Bohan Wu[1], Alberto González-Sanz[1], Marcel Nutz[2]

[1]Columbia University, Department of Statistics,
[2]Columbia University, Departments of Statistics and Mathematics,
*Corresponding Author Email: ss6574@columbia.edu

Variational inference (VI) is widely used as a scalable alternative to Markov chain Monte Carlo (MCMC) for approximating complex, high-dimensional posterior distributions. From the perspective of robust statistics, a central question is whether such approximations are stable: how sensitive are variational solutions to small perturbations of the target measure that depends on the underlying statistical model, likelihood, prior?

In this talk, I will discuss the robustness of mean-field variational inference (MFVI) through the lens of optimal transport. The main message is that stability depends sharply on the geometry of the target distribution. When the target is strongly log-concave, MFVI is quantitatively stable under perturbations of the target measure, yielding robustness guarantees for approximate Bayesian inference. In contrast, even simple non-log-concave targets, such as mixtures of two Gaussians, can lead MFVI to exhibit instability known as mode collapse, revealing a fundamental lack of robustness to multimodality.

The talk is based on joint work with Alberto González-Sanz, Marcel Nutz, and Bohan Wu.

Keywords: *Mean-Field; Variational Inference; Optimal Transport*

A Frequentist Theory for M-posteriors: Asymptotic and Robustness Properties

Cynthia Rush[1]*, Juraj Marusic[1], Marco Avella Medina[1]
[1]Columbia University, 1255 Amsterdam Avenue, New York, NY, 10025
*Corresponding Author Email: cynthia.rush@columbia.edu

An M-estimator is a general class of estimators in statistics that are defined as the minimizer of an objective function, typically derived from a loss or score function. In this talk, I will introduce a theoretical framework for a wide class of generalized posteriors that can be viewed as the natural Bayesian posterior counterpart of the class of M-estimators in the frequentist world and we refer to the members of this class as M-posteriors. I will discuss asymptotic normality of the M-posteriors under mild conditions on the M-estimation loss and the prior, showing that M-posteriors contract in probability around a normal distribution centered at the M-estimators, which provides frequentist consistency and suggests some degree of robustness depending on the reference M-estimator. Moreover, I will formalize the robustness properties of the M-posteriors by providing a new characterization of the posterior influence function and a novel definition of breakdown point adapted for posterior distributions.
Keywords: *posterior influence function, breakdown point, M-estimation*

Invited Session 6 (IS6)

IS6 – Advances in Robustness for Complex Data
Chairs: Luis Angel García Escudero (University of Valladolid, Spain) and Agustin Mayo Iscar (University of Valladolid, Spain)

Christian Hennig (University of Bologna, Italy)
What are outliers?

Valentin Todorov (United Nations Industrial Development Organization, Austria)
Automatic choice of k and c in model-based clustering: The R package tclust

Aldo Corbellini (University of Parma, Italy)
Robust statistics through monitoring: outliers and transformations

Luis Angel García-Escudero (University of Valladolid, Spain)
Robust trimmed clustering for outlier detection in complex settings

What are outliers?

Christian Hennig[1]*
[1]Dipartimento di Scienze Statistiche, University of Bologna, Via delle belle arti 41, 40126 Bologna
*Corresponding Author Email: christian.hennig@unibo.it

There is much talk about outliers in data, and the identification of outliers is often important in practice. I will talk about the problem of defining the outlier identification problem. It is possible to define outliers data analytically (as implied for example by the definition of boxplots), but in order to assess formally whether and to what extent any method does a good job identifying outliers, a model-based definition is required. Outliers can be defined with respect to a model for non-outliers. A widespread approach is to investigate contamination mixtures where the majority of observations is assumed to be generated by the non-outlier model (often but not necessarily the Gaussian distribution), and outliers are generated by a "contaminating" minority component. Davies and Gather [1] argued that outliers should be defined as unlikely observations based on the reference distribution for non-outliers alone; not allowing any "contamination component" to generate "outliers" that are not actually "outlying" w.r.t. the reference distribution. Besides discussing these concepts, their implications, and the choices the researcher needs to make in order to apply them, I will discuss how outliers can be defined w.r.t. a reference distribution, following Davies and Gather, in such a way that "true outliers" are defined for general distributions not necessarily involving the reference distribution as a mixture component, which is arguably too restrictive for real data generating processes.

If time allows I will comment on specific problems such as distributional skewness, outliers in cluster analysis, nonparametric outlier definition based on low density regions.

Keywords: *Outlier region, contamination mixture, outlier identification, boxplot, reference distribution, skewness, cluster analysis, density*

References

[1] Davies, L., & Gather, U. (1993) The Identification of Multiple Outliers. Journal of the American Statistical Association, 88(423), 782–792.

Automatic choice of *k* and *c* in model-based clustering: The R package 'tclust'

Valentin Todorov[1*] and Agustín Mayo-Iscar [2] and Luis Angel García-Escudero [2] and Marco Riani [3]
[1]United Nations Industrial Development Organization (retired), Vienna, Austria
[2]Universidad de Valladolid, Valladolid, Spain
[3]University of Parma, Parma, Italy
*Corresponding Author Email: valentin@todorov.at

The **tclust** package for the **R** environment for statistical computing implements different robust non-hierarchical clustering algorithms where trimming plays a key role [2; 5]. The current 2.0 version of the package implements a number of new features, like improved initialization routines, parallel computing, maximization of constrained mixture loglikelihood, robust linear grouping and others, but the most exciting new feature is the automatic selection of the number of groups *k*, the constraint constant *c* and possibly the trimming factor *alpha*. The implementation is based on the new approaches for selecting *k*, *c* and *alpha* proposed by [1] and [4]. A new methodology for constrained parsimonious model-based clustering which includes fourteen parsimonious models that are often applied in model-based clustering when assuming normal components as limit cases is extended to cope with the presence of atypical observations [3]. Useful graphical tools for summarizing the clustering output and assisting with the choice of the number of groups and the other parameters are provided. A series of examples based on real and simulated data with different degrees of overlap will be given to demonstrate the performance of the new implementation of the different proposals.

Keywords: *Model-based clustering – Robustness – R – parsimonious – tclust*

References

[1] Andrea Cerioli, Luis Angel Garc´ıa-Escudero, Agustın Mayo-Iscar, and Marco Riani. Finding the number of normal groups in model-based clustering via constrained likelihoods. *Journal of Computational and Graphical Statistics*, **27**(2):404–416, 2018. doi: 10.1080/10618600.2017.1390469.
[2] Luis A. Garcıa-Escudero and Agustın Mayo-Iscar. Robust clustering based on trimming. WIREs Computational Statistics, 16(4):e1658, 2024. doi: 10.1002/wics.1658.
[3] Luis A. Garc´ıa-Escudero, Agust´ın Mayo-Iscar, and Marco Riani. Constrained parsimonious model-based clustering. Statistical Computing, 32(2), 2022. doi: 10.1007/s11222-021-10061-3.
[4] Luis Angel Garcıa-Escudero, Christian Hennig, Agustın Mayo-Iscar, Gianluca Morelli, and Marco Riani. Choice of trimming proportion and number of clusters in robust clustering based on trimming. Journal of Computational and Graphical Statistics, 0(0):1–13, 2025. doi:10.1080/10618600.2025.2554675.
[5] V. Todorov. The R package ecosystem for robust statistics. Wiley Interdisciplinary Reviews: Computational Statistics, 16(6):e70007, 2024. doi: 10.1002/wics.70007.

Robust Statistics Through Monitoring: Outliers and Transformations

Anthony C. Atkinson[1], Marco Riani[2], Aldo Corbellini[2]*, Domenico Perrotta[3], Valentin Todorov[4]

[1]LSE, London School of Economics and Political Science, London, UK
[2]Department of Economics and Management, University of Parma, Parma, Italy
[3]Joint Research Centre, European Commission, Ispra, Italy
[4]Statistics Division, United Nations Industrial Development Organization, Vienna, Austria
*Corresponding Author Email: aldo.corbellini@unipr.it

Robust methods protect inference against contamination and can indicate subsets of the data with distinct structures. But applied conclusions depend on the robustness–efficiency tuning. The monitoring approach replaces a single robust fit with a sequence of fits indexed by subset size or breakdown point [1] and tracks residuals, coefficients and test statistics along the sequence. Following the order of the verbal presentation, we describe reproducible software, the monitoring with robust regression in a large banking study, finite-sample size and power of outlier tests, and monitoring-based choice of Box–Cox transformations [3] via fan plots.

The message we want to convey is that monitoring gives applied robust analysis a reproducible reporting language. Instead of quoting only a coefficient table with t-statistics, an outlier list or a selected transformation, the analyst can report the interval of breakdown points or subset sizes over which the conclusion remains stable [2]. This is particularly useful when the data contain more than one population, because a change in the trajectories can reveal where a robust fit changes from protecting the main group to absorbing a secondary structure. In this sense, monitoring is both a robustness device and an exploratory diagnostic for model adequacy.

Overall, monitoring turns robust statistics into a diagnostic workflow. It links outlier detection, tuning of robustness and transformation choice within a single graphical and computational framework. The approach is implemented in reproducible software through the FSDA toolbox for MATLAB, and the examples considered here emphasize that robustness should be reported dynamically: conclusions are most credible when they produce stable results that persist across a meaningful interval of subset sizes or breakdown-point values.

Keywords: *Robust statistics; monitoring; Forward Search; breakdown point; outlier detection; transformations*

References

[1] Atkinson, A.C., Riani, M., Corbellini, A., Perrotta, D., Todorov, V. (2025). Robust Statistics through the Monitoring Approach: Applications in Regression. Springer, Heidelberg.
[2] Riani, M., Cerioli, A., Atkinson, A.C., Perrotta, D. (2014). Monitoring robust regression. Electronic Journal of Statistics, 8, 642–673.
[3] Riani, M., Atkinson, A.C., Corbellini, A. (2023), Automatic robust Box–Cox and extended Yeo–Johnson transformations in regression. Statistical Methods & Applications, 32, 75–102.

Robust trimmed clustering for outlier detection in complex settings

Luis Angel García-Escudero[1]*, Agustín Mayo-Iscar[1], Lucia Trapote[1]
[1]Universidad de Valladolid, Facultad de Ciencias, Paseo Belén 2, Valladolid 47011, Spain
*Corresponding Author Email: lagarcia@uva.es

Outliers, or atypical observations, pose a significant challenge in statistical analysis. Even a few unusual data points can distort estimates and lead to misleading statistical conclusions. Robust methods have been developed to reduce the influence of such observations, ensuring reliable inference while maintaining efficiency when the data largely conform to assumed models.

While often seen as problematic, outliers can also carry valuable information and be of intrinsic interest. In many applied settings, detecting atypical observations is not merely a preparatory step but a central analytical goal, as they may signal important events, novel findings, or remarkable changes. Developing effective methods to identify and display these anomalies is a major focus of current statistical research.

Real-world datasets do not always originate from a single homogeneous population; instead, multiple overlapping groups often underlie the data. In such contexts, clustering provides a natural framework for uncovering latent structure. However, these methods are typically sensitive to outliers, which can severely distort cluster identification. This motivates the use of robust clustering approaches that can simultaneously accommodate heterogeneity and resist the influence of atypical observations.

Robust clustering methods, particularly those based on trimming [1], offer a practical solution for simultaneously dealing with heterogeneity in the data and the possible presence of anomalies. Among these methods, TCLUST [2] stands out as an interesting approach with well-established theoretical and practical properties. However, TCLUST encounters difficulties when handling problems of increasing dimensionality, a situation that is becoming increasingly common in modern statistics. A possible solution is to consider methods that simultaneously address dimensionality reduction and clustering. In this context, and still within the trimming framework, RLG [3] is a useful alternative for higher-dimensional settings. Nevertheless, RLG may be somewhat simplistic and presents certain limitations that can be mitigated by combining it with TCLUST. In this talk, we present some numerical and graphical tools based on this combination of TCLUST and RLG, aimed at exploring atypical observations in settings with moderate to high dimensionality, beyond those where TCLUST and RLG are ideally suited. High-dimensional data introduce additional challenges. As dimensionality increases, there is also an increasing need for approaches capable of trimming not only unusual observations (row-wise trimming) but also atypical individual entries (cell-wise trimming). Methods have been introduced to integrate clustering with strategies to handle cellwise contamination (see [4] and [5]).

Finally, for outlier detection under heterogeneity, it is sensible to initially discard a conservative fraction of the data, possibly larger than necessary. This high conservative initial trimming has the undesired effect of shrinking the dispersion estimators of the different components fitted by the clustering procedure. In fact, while estimating cluster centers and shapes is relatively straightforward, determining the clusters' scales (which can be interpreted as "size" parameters of the components' scatter matrices) is well known to present greater challenges [6]. However, accurate estimation of cluster scale parameters is essential for setting thresholds and measuring atypicality. It can be easily seen that scale estimation is closely interrelated with assessing the (unknown) level of contamination and the total weights of the cluster components. These quantities can be jointly estimated by following sequential iterative procedures [7], using a controlled sequence of trimming levels combined with robustified Mahalanobis distances. The radius process [8] provides a theoretical framework for these procedures and has been extended to accommodate distributions with heavier tails [9].

Keywords: *Clustering, robustness, trimming, diagnostic tools, high-dimension, cell-wise trimming*

References

[1] García-Escudero, L.A., Mayo-Iscar, A. (2024). Robust clustering based on trimming. WIREs Computational Statistics 16(4): e1658.
[2] García-Escudero, L.A., Gordaliza, A., Matrán, C., Mayo-Iscar, A. (2008). A general trimming approach to robust cluster analysis. The Annals of Statistics 36(3),1324-1345.
[3] García-Escudero, L.A., Gordaliza, A., San Martin, R., Van Aelst, S., Zamar, R. (2009). Robust linear clustering. Journal of the Royal Statistical Society Ser. B, 71(1), 301-318.
[4] García-Escudero, L. A., Rivera-García, D., Mayo-Íscar, A., Ortega, J. (2021). Cluster analysis with cellwise trimming and applications for the robust clustering of curves. Information Sciences 573, 100-124.
[5] Zaccaria, G., García-Escudero, L. A., Greselin, F., Mayo-Iscar, A. (2025). Cellwise outlier detection in heterogeneous populations. Technometrics, 67(4), 643-654
[6] García-Escudero, L.A., Gordaliza, A. (2007). The importance of the scales in heterogeneous robust clustering. Computational Statistics & Data Analysis 51(9), 4403-4412.
[7] Dotto, F., Farcomeni, A., García-Escudero, L. A., Mayo-Iscar, A. (2018). A reweighting approach to robust clustering. Statistics and Computing, 28(2), 477-493.
[8] García-Escudero, L.A., Gordaliza, A. (2005). Generalized radius processes for elliptically contoured distributions. Journal of the American Statistical Association 100(471), 1036–1045.
[9] Barabesi, L., Cerioli, A., García-Escudero, L.A., Mayo-Iscar, A., Perrotta, D., Torti, F. (2026). Robust distances and multivariate outlier detection under heavy tails. Statistics and Computing 36:89.

Invited Session 7 (IS7)

IS7 – Robust Regression for Complex Data
Chairs: Olcay Arslan (Ankara University, Turkiye) and Fatma Zehra Doğru (Giresun University, Turkiye)

Olcay Arslan (Ankara University, Turkiye)
Exploring Expectile Regression Through Likelihood and EM Algorithm Perspectives

Fatma Zehra Doğru (Giresun University, Turkiye)
A Flexible Heteroscedastic Semiparametric Model for Skewed and Heavy-Tailed Data

Yeşim Güney (Ankara University, Turkiye)
Robust Joint Location–Scale Modeling via M-Quantile Estimation

Elgiz Askeroğlu (Giresun University, Turkiye)
Robust Particle Swarm Optimization for Regression via M-Estimation

Exploring Expectile Regression Through Likelihood and EM Algorithm Perspectives

Olcay Arslan
oarslan@ankara.edu.tr
Ankara University, Faculty of Science, Department of Statistics, Ankara, Türkiye

In this paper, we study parameter estimation in expectile regression via a likelihood-based approach and the Expectation-Maximization (EM) algorithm. By connecting the Asymmetric Least Squares (ALS) loss function to a likelihood framework through the two-piece skew normal distribution, we develop a maximum likelihood approach for parameter estimation in expectile regression. Using the scale mixture representation of the two-piece skew normal distribution, we further construct an EM algorithm for computing the parameter estimates. Since, except for the symmetric case corresponding to least squares (LS) regression, closed-form estimators for expectile regression are generally unavailable, numerical optimization procedures are required. The integration of the ALS loss function with the EM algorithm through the two-piece skew normal distribution provides a novel and practical framework that advances expectile regression from both theoretical and computational perspectives. To evaluate the proposed methodology, we present several numerical studies, including simulation experiments and real data applications, illustrating the capability of expectile regression to effectively model data exhibiting skewness, heavy-tailedness, and heteroskedasticity. The results from both the simulation studies and the real data examples demonstrate that the proposed method and algorithm perform effectively and reliably.

Keywords: *ALS objective function; expectile regression; asymmetric normal distribution; heteroscedasticity*

A Flexible Heteroscedastic Semiparametric Model for Skewed and Heavy-Tailed Data

Fatma Zehra Doğru[1]*, Olcay Arslan[2]
[1] Department of Data Science and Analytics, Giresun University, Giresun, Türkiye, 28200
[2] Department of Statistics, Ankara University, Ankara, Türkiye, 06100
*Corresponding Author Email: fatma.dogru@giresun.edu.tr

Partially linear models (PLMs) provide a flexible semiparametric framework by combining parametric and nonparametric components within a unified regression structure. Nevertheless, classical PLMs typically rely on normality and homoscedasticity assumptions, which are frequently violated in practice. Many real-world data sets exhibit skewness, heavy tails, and non-constant variance structures, leading to unreliable inference and reduced estimation efficiency under conventional modelling approaches.

To address these challenges, this study proposes a flexible heteroscedastic partially linear model based on the skew Laplace normal (SLN) distribution (HPLM-SLN). The proposed framework simultaneously accommodates skewness, heavy-tailed behavior, and heteroscedasticity through an explicit scale modelling structure. The proposed methodology extends recent robust PLM formulations in [1] by explicitly incorporating heteroscedasticity and likelihood-based inferential procedures. Parameter estimation is carried out via maximum likelihood estimation using an expectation conditional maximization (ECM) algorithm [2].

The inferential and robustness properties of the proposed HPLM-SLN model are investigated through extensive Monte Carlo simulation studies under multiple data-generating mechanisms, including nonlinear structures, varying scale formulations, multiple covariates, and model misspecification settings. Comparative analyses are performed against alternative PLM and HPLM formulations based on normal and skew-normal error distributions, as well as a Box-Cox transformed PLM-SLN approach. In addition, a likelihood ratio (LR) test is developed to assess the homogeneity of the scale parameter, and the empirical distribution of the LR statistic is examined under different sample sizes.

Simulation results indicate that the proposed methodology provides reliable estimation, stable inferential performance, and improved robustness in the presence of skewness, heavy tails, heteroscedasticity, and extreme observations. The practical utility of the proposed framework is further illustrated through an application to ragweed pollen concentration data, where the model demonstrates enhanced flexibility and improved goodness-of-fit compared with competing approaches.

Keywords: *Heteroscedasticity, likelihood ratio test, model diagnostics, partially linear model, skew Laplace normal distribution*

References

[1] Doğru, F.Z., Arslan, O. (2026). Robust parameter estimation in partially linear models with skewed and heavy-tailed error distributions. Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics, in press.
[2] Meng, X.L., Rubin, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. Biometrika, 80(2), 267-278.

Robust Joint Location–Scale Modeling via M-Quantile Estimation

Yeşim Güney[1]*, Yetkin Tuaç[1], Olcay Arslan[1]
[1]Ankara University, Faculty of Science, Department of Statistics, Ankara, Türkiye, 06100
*Corresponding Author Email: ydone@ankara.edu.tr

Scale heterogeneity and the presence of outliers often limit the reliability of classical regression models, leading to biased inference and poor predictive performance. In many applied settings, the response variable is influenced by explanatory variables not only through its central tendency but also through its variability, motivating the use of joint location–scale models. Such models offer a more comprehensive representation of complex data structures.

In this study, we consider a robust joint location–scale framework based on the Generalized Asymmetric Least Informative Distribution (GALID), which relaxes the normality assumption and accommodates asymmetric error distributions. Parameter estimation within this framework is naturally connected to M-quantile regression, providing robustness against outliers while maintaining computationally efficiency.

We propose an M-quantile–based joint location–scale model and introduce a penalized M-quantile approach for variable selection in both the location and scale components. The effectiveness of the proposed method is demonstrated through simulation studies and a real data application.

Keywords: *Generalized asymmetric least informative distribution, heteroscedasticity, joint location and scale models, robust regression, M-quantile regression.*

Robust Particle Swarm Optimization for Regression via M-Estimation

Elgiz Askeroğlu[1]*, Fatma Zehra Doğru[2]
[1]Giresun University, Institute of Science, Giresun, Türkiye, 28200
[2]Giresun University, Faculty of Arts and Sciences, Department of Data Science and Analytics, Giresun, Türkiye, 28200
*Corresponding Author Email: elgiz_askeroglu@hotmail.com

Regression analysis of real-world data is frequently challenged by outliers, heteroscedasticity, and structural irregularities that can bias parameter estimates and weaken predictive performance. While classical estimation methods such as Least Squares (LS) are optimal under standard assumptions, they are highly sensitive to anomalous observations. Although artificial intelligence–based optimization algorithms, particularly Particle Swarm Optimization (PSO) originally introduced by [1], are widely used for nonlinear regression problems, standard PSO implementations typically minimize mean squared error (MSE) and therefore remain vulnerable to outliers. To overcome this limitation, this study proposes a robust artificial intelligence–based regression framework that integrates M-estimation principles [2] into a modified PSO algorithm. Specifically, Huber and Tukey biweight loss functions are incorporated into the PSO fitness evaluation stage, replacing the conventional MSE objective function. This modification reduces the influence of extreme residuals and enhances the stability and robustness of parameter estimation. Unlike traditional robust regression methods based on iterative reweighting schemes, the proposed framework directly estimates model parameters through a unified global optimization approach.

The performance of the proposed robust PSO variants (PSO-Huber and PSO-Tukey) is evaluated through both simulation studies and a real data application. Their estimation accuracy and robustness are assessed using multiple performance criteria and systematically compared with their classical counterparts. The results demonstrate that the robust PSO variants provide improved reliability and resistance to outliers while maintaining competitive predictive accuracy. Overall, the study highlights the methodological and practical advantages of combining M-estimation theory with metaheuristic optimization techniques for robust regression modelling.

Keywords: *Artificial Intelligence, Least Squares, M-Estimator, Particle Swarm Optimization, Robust regression.*

References

[1] Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. Proceedings of the IEEE International Conference on Neural Networks, 4(1), 1942-1948.

[2] Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. Annals of Statistics, 1(5), 799-821.

Invited Session 8 (IS8)

IS8 – Robust Inference in Combined Regression and Autoregression Models (Hybrid Session)
Chairs: Jana Jureckova (Charles University, Czech Republic), Olcay Arslan (Ankara University, Turkiye), and Yeşim Güney (Ankara University, Turkiye)

Jana Jureckova (Charles University, Czech Republic) [Virtual]
The Advantage of Autoregression Rank Scores Invariance in the Inference

Hira L. Koul (Michigan State University, United States) [Virtual]
A signed-rank estimator in nonlinear regression models when covariates and errors are dependent

Jan Picek (Technical University of Liberec, Czech Republic)
Adaptive rank scores guided by l-moments for robust inference in regression–autoregression models

Yetkin Tuaç (Ankara University, Turkiye)
M-Quantile regression with autoregressive errors

ICORS2026
20-24 JULY 2026
Istanbul

The advantage of autoregression rank scores invariance in inference

Jana Jureckova

Institute of Information Theory and Automation of CAS, Prague, Czech Republic.
Author Email: jureckova@utia.cas.cz

In the practical problems, we often measure a physical or economic entity Z, whose values can be influenced by its past values. Such situation is described with the autoregression model, with generally unknown autoregression parameters: $Y_t = \theta_1 Y_{t-1} + \cdots + \theta_p Y_{t-p} + Z_t$, $t = 1, \ldots, n$, where $Y_{t-1}, \ldots, Y_{t-p}$ are the past measurements, while Z_t is the value of the interest. We repeat the experiment n-times with observations $Y_{t1}, \ldots, Y_{tn}$ and wish to make an inference on $Z_{t1}, \ldots, Z_{tn}$. For instance, we need to estimate the quantile function of Z. In another situation, Z_t can follow a linear regression model $Z_t = \beta_0 + \boldsymbol{X\beta} + u_t$, $t = 1, \ldots, n$, while the autoregression is a nuisance. Then the problem is to estimate the parameter $\boldsymbol{\beta}$. Another time we compare two experiments, for instance the measurements of the amounts of CO_2, emitted from two diferent sources in the region. The measurements form two autoregressive sequences of orders p, q and leading to Z_t, Z_t^*. Then we want to verify the independence of two series $\{Z_1, \ldots, Z_n\}$ and $\{Z_1^*., \ldots, Z_n^*.\}$ n g by a rank test of independence, e.g. by the Spearman rank correlation coefficient. Our tools for such inference are the autoregression quantiles and the autoregression rank scores for the model. Because the AR rank scores are invariant with respect to changes of the autoregression parameters, the AR rank scores for Y_t coincide with those for Z_t. We can replace the unknown ranks of Z_t with the regression rank scores, what considerably helps to solve the situation even with the unobservable Z_t.

A signed-rank estimator in nonlinear regression models when covariates and errors are dependent

Hira L. Koul[1]*, Palaniappan Vellaisamy[2]
[1] Michigan State University Statistics and Probability A435 Wells Hall East Lansing, Michigan USA
[2] Department of Mathematics Indian Institute of Technology Bombay Powai Mumbai-400076, India.
*Corresponding Author Email: KOUL@msu.edu

This talk will first discuss asymptotic relative efficiency (ARE) of a signed rank estimator in an errors in variables linear regression model with known Gaussian distributions of the measurement error, the predicting covariate and its surrogate. The ARE of this estimator relative to the bias corrected least squares estimator at a Gaussian regression error distribution is shown to increase to infinity as the measurement error variance increases to infinity. This is a kind of robustness property of this estimator against large measurement error variance. Given this motivation, we then derive asymptotic normality of an analog of this signed rank estimator in a class of nonlinear semi- parametric regression models where the predicting random covariate vector is possibly dependent on the regression error and where the regression error distribution need not be known.

Keywords: *Asymptotic normality, measurement error, rank-based estimation.*

Adaptive rank scores guided by L-moments for robust inference in regression–autoregression models

Jan Picek [1]*
[1]Department of Mathematics, Technical University of Liberec, Studentská 1402/2 Liberec 461 17 Czech Republic
*Corresponding Author Email: jan.picek@tul.cz

We consider combined regression–autoregression models, motivated by time series with heavy tails, outliers and leverage effects. These models combine the influence of external covariates with autoregressive dependence on past observations, and therefore require inference methods that remain reliable under nonstandard innovation distributions.

We propose a robust rank-based inference approach in which rank scores are calibrated using l-moment characteristics of the residuals. For a candidate parameter vector, residuals are computed and standardized by an l-scale estimate, providing stability under heavy-tailed innovations. L-moment ratios, namely l-skewness and l-kurtosis [1], computed from the standardized residuals are then used to adapt the score function, yielding a data-driven weighting scheme in Jaeckel's dispersion criterion [2].

The resulting R–L estimator is obtained by minimizing this dispersion with adaptive scores, implemented through a simple reweighting iteration initialized by a Wilcoxon-score R-fit. Consistency and √n-asymptotic normality are discussed under standard regularity conditions, with the adaptive step treated as a plug-in procedure. Simulation results indicate that the proposed approach improves the mean squared error of regression coefficient estimates compared with ordinary least squares and classical fixed-score R-estimators, especially under contamination and heavy-tailed innovations.

Keywords: *Robust rank-based inference, L-moments, regression–autoregression models, Jaeckel dispersion, adaptive scores*

References

[1] Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society: Series B, 52(1), 105–124.
[2] Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. The Annals of Mathematical Statistics, 43(5), 1449–1458.

M-Quantile Regression with Autoregressive Errors

Yetkin Tuaç[1]*, Yeşim Güney[1], Olcay Arslan[1]
[1]Ankara University, Faculty of Science, Department of Statistics, Ankara, Türkiye, 06100
*Corresponding Author Email: ytuac@ankara.edu.tr

M-quantile regression provides a flexible and robust alternative to classical regression by allowing the analysis of different parts of the conditional distribution of a response variable. Although this methodology has been widely studied under the independence assumption, serial dependence, such as autoregressive error terms, has not gained much attention in the literature. In many real applications, however, regression errors are correlated over time, and ignoring this structure may reduce efficiency and distort inference.

In this study, we consider M-quantile estimation [1] in linear regression models with autoregressive errors. More specifically, the error term is assumed to follow an AR(p) process, enabling the model to capture serial correlation while preserving the robustness properties of M-quantile methods. The proposed framework extends standard M-quantile regression to situations where observations are ordered in time or where residual dependence cannot be neglected.

An iterative estimation procedure is employed for the joint estimation of regression coefficients and autoregressive parameters. The method is expected to provide a robust modeling strategy in the presence of both outliers and autocorrelation. Its finite-sample performance can be examined through simulation studies under different sample sizes, dependence levels, and innovation distributions. Comparisons with conventional procedures may illustrate the practical advantages of incorporating autoregressive dependence into the M-quantile framework. Overall, the proposed approach offers a useful extension of robust regression methodology for dependent data settings.

Keywords: *M-quantile regression, autoregressive errors, robust estimation, dependent data.*

References

[1] Breckling, J., & Chambers, R. (1988). M-Quantiles. Biometrika, 75(4), 761–771.

Invited Session 9 (IS9)

IS9 – Robust Principal Components
Chairs: Gabriela Cohen Freue (University of British Columbia, Canada) and David Tyler (Rutgers University, United States)

Una Radojicic (Technical University of Vienna, Austria)
New M-estimator of the leading principal component

Shojaeddin Chenouri (University of Waterloo, Canada)
Robust Principal Components Analysis via Subspace Discrepancy Minimization

Yetkin Tuaç (Ankara University, Turkiye)
Adaptive Lasso-regularized Spectral Generalized Kotz-type Robust Matrix Regression

New M-estimator of the leading principal component

Una Radojicic[1]*, Joni Virta[2], Marko Voutilainen[2]
[1] Vienna University of Technology, Wiedner Hauptstrasse 7-9, 1040, Vienna, Austria
[2] University of Turku, FI-20014 Turun yliopisto, Finland
*Corresponding Author Email: una.radojicic@tuwien.ac.at

We study the minimization of the non-convex and non-differentiable objective function $v \to E(\| X - v \| \| X + v \| - \| X \|^2)$ in $\mathbf{R}^p$. In particular, we show that its minimizers recover the first principal component direction of elliptically symmetric X under specific conditions. The stringency of these conditions is studied in various scenarios, including a diverging number of variables p. We establish the consistency and asymptotic normality of the sample minimizer. We propose a Weiszfeld-type algorithm for optimizing the objective and show that it is guaranteed to converge in a finite number of steps. The results are illustrated with two simulations. The presented work is based on [1].

Keywords:, *Elliptical distribution, M-Estimation, Principal Component Analysis, Robustness, Spatial Median*

References

[1] Virta, Joni, Una Radojicic, and Marko Voutilainen. "New M-estimator of the leading principal component." arXiv preprint arXiv:2510.02799 (2025).

Robust Principal Components Analysis via Subspace Discrepancy Minimization

Shojaeddin Chenouri*[1], Varnan Sarangian[1], Christopher G. Small[1]
[1]Department of Statistics and Actuarial Science, University of Waterloo, Canada
*Corresponding Author Email: schenouri@uwaterloo.ca

We study a family of robust two-stage estimators for principal component analysis and subspace recovery. The procedure can be viewed as a generalization of geometric medians: the sample is partitioned into disjoint blocks, a classical PCA computation is performed within each block, and the resulting subspace estimates are aggregated by minimizing an empirical risk defined by a specified discrepancy function. For a large collection of subspace discrepancies, we derive non-asymptotic excess-risk bounds and parametric consistency rates for these estimators that are independent of the ambient data dimension. We complement this analysis with a detailed robustness theory that quantifies the rates at which point-level contamination propagates through the block partitioning to displace the subsample estimator. We further identify both benefits and limitations of the method through an information-theoretic lens. On the one hand, compressing each block's data into a subspace confers a form of *intrinsic robustness* by attenuating gross-error contamination in some settings. On the other hand, this compression of inter-block magnitude information induces a non-identifiability obstruction once the block-level contamination radius exceeds an explicit threshold. We show that in this regime, magnitude-retaining estimators may still recover the population eigenspace. We conclude with simulation studies and applications to image processing to illustrate the practical performance of the estimator.
Keywords: *Robust PCA, Geometric median, Median-of-means, Manifold optimization, High-dimensional statistics*

Adaptive lasso-regularized spectral generalized kotz-type robust matrix regression

Yetkin Tuaç[1]*, Olcay Arslan[1]
[1]Ankara University, Faculty of Science, Department of Statistics, Ankara, Türkiye, 06100
*Corresponding Author Email: ytuac@ankara.edu.tr

This study develops an adaptive lasso-regularized spectral generalized Kotz-type low-rank matrix regression framework for robust face reconstruction and recognition under severe illumination variation and mixed corruption. Building on matrix-decomposition formulations used in mixed-noise face analysis, the main contribution of the present work is the introduction of a new spectral generalized Kotz-type model for the structured low-rank component and a generalized double exponential (GDE) model for the sparse contamination component, together with optimization algorithms for their practical estimation. The proposed formulation combines classwise matrix representation with adaptive lasso regularization to obtain a parsimonious and discriminative reconstruction while accommodating both structured and localized image degradation. A penalized optimization scheme is developed to estimate the regression coefficients and latent error components in a unified framework. The empirical evaluation is conducted on the Extended Yale B face database under subset-based training and testing settings motivated by the mixed-noise recognition literature, with experiments designed to assess performance under strong illumination changes and additional sparse corruption. Overall, the proposed approach aims to provide a more flexible probabilistic and computational framework for robust face reconstruction and recognition in settings where classical linear representation methods are challenged by simultaneously structured and sparse noise.

Keywords: *Adaptive lasso, spectral generalized Kotz-type model, low-rank matrix regression, robust face reconstruction, mixed-noise face recognition.*

Invited Session 10 (IS10)

IS10 – Symbolic Data Analysis meets Robust Statistics: Models, Methods and Applications (Hybrid Session)
Chair: Maria do Rosário Oliveira *(CEMAT and Instituto Superior Técnico, Universidade de Lisboa, Portugal)*

Eufrasio Lima Neto (Federal University of Paraíba, Brazil) [Virtual]
Robust distributional regression models for interval-valued data

M. Rosário Oliveira (University of Lisbon, Portugal) [Virtual]
Clustering interval-valued data using L1-Wasserstein distance

Pedro Duarte Silva (Catholic University of Portugal, Portugal)
Robust Principal Component Analysis of Symbolic Matrix-valued Data

Renata Maria Cardoso Rodrigues Souza (Federal University of Pernambuco, Brazil)
Multivariate Fuzzy Partitioning Methods using Medoids

Robust Distributional Regression Models for Interval-valued Data

Eufrasio Lima Neto [1]*, Sónia Dias[2], Paula Brito[3]
[1]Federal University of Paraíba, João Pessoa, Brazil
[2]Polytechnic Institute of Viana do Castelo, Viana do Castelo, Portugal
[2]University of Porto, Porto, Portugal
*Corresponding Author Email: eufrasio@de.ufpb.br

The Robust Interval Distributional Regression (RiDR) model represents a step forward in Symbolic Data Analysis (SDA) [1], evolving from the structural foundations of the ID model [2]. While traditional regression models often decompose intervals into separate linear regressions for bounds or midpoints and ranges, the ID framework treats intervals as holistic entities by representing them through quantile functions. By assuming a specific distribution within each interval - typically a Uniform distribution - the model accounts for internal variability that is lost in classical point-based summaries.

The primary innovation of this framework lies in its mathematical constraints, which ensure that predicted outputs are always mathematically valid intervals. By incorporating both the quantile function of the independent variables and their symmetric counterparts, the model allows for both direct and inverse linear relationships through parameters a_j and b_j, $j = \{1, \ldots, p\}$. Crucially, the model imposes non-negativity constraints (a_j, $b_j > 0$), which guarantees that the predicted range can never be negative, thus ensuring that the lower bound never exceeds the upper bound.

Secondly, to address the sensitivity of standard least-squares methods to atypical data (outliers), the RiDR model integrates robust principles from the Exponential-type kernel robust regression (iETKRR) model [3]. It employs an objective function which uses Gaussian kernels [4, 5] to measure the similarity between observed and predicted values. Through an iterative re-weighting process, the model identifies and penalizes outliers by assigning them near-zero weights in an iterative way, effectively reducing their influence on parameter estimation in the midpoint and/or the range spaces. At each iteration, given these fixed weights, the model's parameters are updated by solving a constrained quadratic programming problem that minimizes the objective function.

This combination of distributional theory and robust kernel weighting allows the RiDR model to maintain high precision even in the presence of leverage points, X-space and Y-space outliers. Furthermore, the model remains highly flexible; it reduces to classical linear regression when applied to degenerate intervals (real numbers) and the weights settled as one, and can be adapted to alternative distributions, such as the Symmetric Triangular distribution, making it a versatile tool for complex aggregated datasets.

Finally, the RiDR model and the non-robust counterpart were applied to real interval-valued data set [2] and the performances compared. The models were applied to the original data and the contaminated data set, where one observation from the original

data set was replaced by an outlier. The results demonstrated that the RiDR model outperform the non-robust counterpart based on different metrics.

Keywords: *Interval-valued Variables, Mallow's Distance, Outlier, Kernel, Robust Regression*

References

[1] Billard, L., Diday, E.: Symbolic Data Analysis: Conceptual Statistics and Data Mining. John Wiley & Sons, Ltd, New Jersey (2006).

[2] Dias, S., Brito, P.: Off the beaten track: A new linear model for interval data. European Journal of Operational Research (3), 1118–1130 (2017).

[3] Lima Neto, E.A., De Carvalho, F.A.T.: An exponential-type kernel robust regression model for interval-valued variables. Information Sciences (454), 419–442 (2018).

[4] Mercer, J.: Functions of positive and negative type and their connection with the theory of integral equations. Philosophical Transactions of the Royal Society A (209), 415–446 (1909).

[5] Müller, K.R., Mika, S., Rätsh, G.R., Tsuda, K., Schölkopf, B.: An introduction to kernel-based learning algorithms. IEEE Trans. Neural Networks (12), 181–202 (2001).

Clustering interval-valued data using L1-Wasserstein distance

M. Rosário Oliveira[1]*, Sara Suarez-Fernandez[2], Diogo Pinheiro[1]
[1] Department of Mathematics, Instituto Superior Técnico, Lisbon, Portugal
[2] Department of Statistics and O.R. and M.T., University of Oviedo, Oviedo, Spain
*Corresponding Author Email: rosario.oliveira@tecnico.ulisboa.pt

Interval-valued data frequently arises in many real-world contexts where microdata are aggregated into macrodata, or where uncertainty and imprecision are inherent, leading to observations represented by intervals rather than single numerical values. Such data structure may exist by their own right or from aggregating original data due to security or convenience reasons.

In this work, we consider the modeling framework introduced in Oliveira et al. [1], that links the macrodata with the underlying microdata distributions associated with each interval. Within this setting, we study the L1-Wasserstein distance and investigate its role as a comparison measure for interval-valued variables. This distance is defined originally in terms of the quantile functions of the microdata distributions associated to each interval. We present an explicit form of this distance, facilitating their use in data analysis.

The performance of the L1-Wasserstein distance is evaluated through partitional clustering algorithms, including centroid-based and medoid-based approaches. The analysis also includes a comparison with Mallow's distance (L2-Wasserstein), which has been previously studied as a basis for defining location and association measures [2], and as the underlying metric in Fisher discriminant analysis for interval-valued data [3]. The evaluation is based on simulation studies and further supported by real-data experiments.

The empirical behavior and robustness of the clustering methods using this metric under contamination is examined through a structured simulation study in which three different scenarios are considered: outliers affecting the interval centers, the interval ranges, and a mixed configuration combining both components. The experimental design varies key structural factors including the number of clusters, the number of variables, the distributions, and the number of observations, among others. This analysis allows us to identify settings in which the L1 formulation exhibit improved performance, as well as scenarios where both approaches behave similarly.

Overall, the study contributes to a clearer understanding of the behavior of the L1-Wasserstein metric when working with interval-valued variables. In particular, it characterizes its performance and robustness in clustering settings under contamination.

Keywords: *SDA, interval-valued data, L1 Wasserstein distance, Robustness, Clustering*
Acknowledgments
The authors express their gratitude for the financial support. This work was supported by FCT — Fundação para a Ciência e a Tecnologia through the project UID/04621/2025 (https://doi.org/10.54499/UID/PRR/04621/2025).

Suarez-Fernandez is supported by the Spanish Ministry of Science and Innovation under project PID2022-139886NB-I00 and the Severo Ochoa Programme of the Principality of Asturias (Grant No. BP24-214).

References

[1] Oliveira, M. R., Azeitona, M., Pacheco, A., & Valadas, R. (2022). Association measures for interval variables. *Advances in Data Analysis and Classification*, *16*(3), 491-520.

[2] Oliveira, M. R., Pinheiro, D., & Oliveira, L. (2024). Location and association measures for interval-valued data based on Mallows' distance. *arXiv preprint arXiv:2407.05105*.

[3] Pinheiro, D., Oliveira, M. R., Kravchenko, I., & Oliveira, L. (2025). Interval Fisher's Discriminant Analysis and Visualisation. *arXiv preprint arXiv:2512.11945*.

Robust Principal Component Analysis of Symbolic Matrix-valued Data

Marcus Mayrhofer[1], A. Pedro Duarte Silva[2*], Paula Brito[3], Peter Filzmoser[1]

[1]Institute of Statistics and Mathematical Methods in Economics, TU Wien, Austria
[2]Universidade Católica Portuguesa, Católica Porto Business School & CEGE, Portugal
[3]Fac. Economia, Universidade do Porto & LIAAD-INESC TEC, Portugal
*Corresponding Author Email: psilva@ucp.pt

We consider numerical distributional data, where for each unit S_i, i=1,…,n and each variable Y_j , j=1,…,p, a distribution is recorded, representing the corresponding observed variability. In our model, each distribution is represented by a set of indicators: a central statistic C, and the logarithm transformation of inter-quantile ranges, for a chosen set of quantiles $\phi_1,\ldots,\phi_q$, denoted by R^*_h, h=1,…,m, where m=q−1 is the number of considered intervals. Typical cases consist in using the median, or else the midpoint, as central statistics, and quartiles, or other equally-spaced quantiles [1]; interval-valued data are represented by midpoints and log-ranges (m=1) [2]. Furthermore, we consider alternative structures of the variance-covariance matrix. In the most general formulation we allow for non-zero correlations among all central statistics and log-ranges; for distributional variables there are however other cases of interest, whether the variables, the central statistics, and the different log-ranges, are or are not correlated between or among themselves, leading to five different configurations in the distributional data case, and four configurations for interval-valued data.
In this work we consider these data as matrix-valued data, represented as a tensor of dimension n×p×q, following [3], $X \sim ME(M,\Sigma_{var},\Sigma_{ind},g)$, where

- Σ_{var} is p×p and gathers variances and covariances between the variables Y_j
- Σ_{ind} is q×q and gathers variances and covariances between the considered q=m+1 indicators, C, $R^*_1,\ldots, R^*_m$
- $g(z) = \exp(-z/2) / (2\pi)^{pq/2}$

In this model, the global covariance matrix Σ is written as $\Sigma = \Sigma_{ind} \otimes \Sigma_{var}$. This implies that we assume that covariances between the different indicators are constant across variables, thereby obtaining a more parsimonious model and reducing the number of parameters to be estimated. The different covariance configurations correspond to setting Σ_{ind} and/or Σ_{var} as block-diagonal matrices. The Matrix Minimum Covariance Determinant (MMCD) method [3] accounts for the matrix-variate data structure and robustly estimates the mean matrix, as well as the row-wise Σ_{var} and column-wise Σ_{ind} covariance matrices.

Given n units $S_{1,\ldots,}S_n \in R^{p\times q}$, let $\hat{\Sigma}_{var} \in R^{p\times p}$ and $\hat{\Sigma}_{ind} \in R^{q\times q}$ denote mode-wise estimators of the variables' and indicators' covariance matrices, respectively, obtained via either the matrix maximum likelihood estimator [4] or the matrix minimum

covariance determinant method [3]. PCA is then performed independently on each estimator via eigen decomposition,

$\hat{\Sigma}_{var} U = U \Lambda_{var}$ $\hat{\Sigma}_{ind} V = V \Lambda_{ind}$ (1)

yielding variable-space loadings $U \in R^{p \times k}$ and indicator-space loadings $V \in R^{q \times \ell}$.

The matrix PCA scores are then

$Z_i = U^T (S_i - \hat{M}) V \in \mathbb{R}^{k \times l}$ $i = 1,...,n$ (2)

where $\hat{M} \in R^{p \times q}$ is an estimate of the mean matrix, providing a bilinear low-rank representation that captures variance along both matrix dimensions.

This framework was applied to a sample of 1544 distributional data units of Spotify track data, collected from the Spotify's public API, and aggregated by artist. The original microdata consisted of 170 663 track observations on 19 music attributes. For the purpose of this analysis we have considered nine numerical attributes (valence, acousticness, danceability, duration, energy, liveness, loudness, speechiness, and tempo) and nine distributional indicators (the median, and the log-ranges of the intervals defined by the min, max and the $q_{0.125}$, $q_{0.25}$, $q_{0.375}$, $q_{0.5}$, $q_{0.75}$, and $q_{0.875}$ quantiles) for each distributional data unit.

The results lead to low-rank representations with interpretable and interesting components for both matrix dimensions. Furthermore, the comparison between the PCA results based on maximum likelihood and MMCD covariance estimates highlighted the impact of unit outliers, and the importance of robustly estimating covariance matrices in symbolic matrix-valued data.

Keywords: *Histogram data, Matrix-valued data, Robust covariance estimation*

References

[1] Brito, P., Duarte Silva, A. P. (2025). Parametric models for distributional data. Advances in Data Analysis and Classification, 19, 1119–1146.

[2] Brito, P., Duarte Silva, A. P. (2012). Modelling interval data with Normal and Skew-normal distributions. Journal of Applied Statistics, 39(1), 3-20.

[3] Mayrhofer, M., Radojičić, U., Filzmoser, P. (2025). Robust covariance estimation and explainable outlier detection for matrix-valued data. Technometrics, 67(3), 516-530.

{4] Dutilleul, P. (1999). The MLE algorithm for the matrix normal distribution. Journal of Statistical Computation and Simulation, 64(2):105–123.

Multivariate Fuzzy Partitioning Methods using Medoids

Thomaz C.C. Araújo[1], Renata M.C.R. Souza[1]*, Bruno A. Pimentel[2]
[1]Centro de Informática, Av. Jornalista Aníbal Fernandes, s/n – Cidade Universitária., Recife - PE- Brazil.
[2]Instituto de Computação, Universidade Federal de Alagoas Av. Lourival Melo Mota, S/N - Tabuleiro do Martins, Maceió - AL, CEP 57072-90, AL, Brazil.
*Corresponding Author Email: rmcrs@cin.ufpe.br

Clustering Analysis is a fundamental tool in machine learning used to uncover underlying structures in unlabeled datasets by grouping similar instances together. While partitional methods like Fuzzy C-Means (FCM) are widely adopted for their ability to model overlapping classes through membership degrees, they face significant limitations in real-world scenarios [1]. Specifically, FCM is highly sensitive to noise and outliers because it relies on mean-based prototypes (centroids). Furthermore, standard fuzzy algorithms typically employ global distance metrics that treat all variables as equally important, which degrades performance in high-dimensional datasets containing irrelevant or correlated features. To address these challenges simultaneously, this work proposes the Multivariate Fuzzy C-Medoids (MFCMd) algorithm.

The MFCMd method integrates the robustness of medoids [2] - actual observations from the dataset - with a multivariate membership mechanism [3] that performs local soft feature selection. Unlike centroids, a medoid is restricted to be an instance with the minimal aggregate dissimilarity to other members of the cluster, significantly improving interpretability and robustness against extreme values. The algorithm minimizes an objective function that accounts for distances between data points and cluster medoids across individual variables, enabling the estimation of feature relevance during the clustering process. Two distance-based variants are explored: MFCMd-E, utilizing the Euclidean distance, and MFCMd-C, utilizing the Cityblock distance.

A significant contribution of this work is the extension of classical sum-of-squares-based interpretation indices [4] to the multivariate fuzzy context. These indices decompose total global dispersion into intra-group and inter-group components at the cluster and variable levels. This decomposition allows for a granular understanding of which variables contribute most to the distinction of specific groups.

The computational complexity of MFCMd is $O(c^2p^2n + cpn^2)$ per iteration, reflecting the trade-off required to achieve simultaneous robustness to outliers and feature weighting. Experimental evaluations were conducted on seven synthetic configurations and seven real-world datasets, including Iris, Wine, and Cervical Cancer. Results demonstrated that MFCMd consistently outperforms traditional methods such as FCM, Fuzzy C-Medoids (FCMd), and the Gustafson-Kessel (GK) algorithm [5] in complex, non-spherical scenarios. Specifically, in high-dimensional or

heterogeneous spaces, the ability to assign variable-specific weights allows the model to suppress noisy attributes and focus on features that define the cluster structures.

In the Wine dataset analysis, the MFCMd-C variant successfully isolated prototypes with sharp chemical signatures that Euclidean-based methods failed to capture. Furthermore, using the proposed interpretation indices for variable selection led to significant performance improvements, with increases of up to 38.5% in the Adjusted Rand Index (ARI) [6] compared to using the full feature set. These findings confirm that the combination of medoid-based prototypes and multivariate learning provides a superior framework for uncovering complex, anisotropic patterns in modern data analysis.

Keywords: *Clustering analysis, Fuzzy C-Medoids, Multivariate Fuzzy Clustering, Robust Clustering, Feature Weighting, Interpretation Indices.*

References

[1] Bezdek, J. C. (1987). Pattern recognition with fuzzy objective function algorithms. Springer.
[2] Pinheiro, D. N. (2020). Convex fuzzy k-medoids clustering. Fuzzy Sets and Systems, 389, 66-92.
[3] Pimentel, B. A., & Souza, R. M. C. R. (2013). A multivariate fuzzy c-means method. Applied Soft Computing, 13(4), 1592-1607.
[4] Celeux, G., et al. (1989). Classification Automatique des Données. Vol 1.
[5] Gustafson, D. E., & Kessel, W. C. (1978). Fuzzy clustering with a fuzzy covariance matrix. IEEE Conference on Decision and Control, 761-766.
[6] Hubert, L., & Arabie, P. (1985). Comparing Partitions. Journal of Classification, 2(1), 193-218.

Invited Session 11 (IS11)

IS11 – Recent Developments in Robustness
Chairs: Mia Hubert (KU Leuven, Belgium) and Stefan Van Aelst (KU Leuven, Belgium))

Peter Filzmoser (TU Wien, Austria)
Multivariate functional Mahalanobis distance with application to clustering

Yangzhuoran Fin Yang (Maastricht University, Netherlands)
Outlier Detection and Robust Estimation of Time Series Model using a Penalised Approach

Giorgia Zaccaria (University of Milano-Bicocca, Italy)
Cellwise outlier detection in clustering: from model-based to fuzzy approaches

Multivariate functional Mahalanobis distance with application to clustering

Jeremy Oguamalam[1], Peter Filzmoser[1]*, Tomas Masak[2], Una Radojičić[1]
[1]Institute of Statistics and Mathematical Methods in Economics, TU Wien, Vienna, Austria
[2]Institute for Statistics and Mathematics, WU Wien, Vienna, Austria
*Corresponding Author Email: Peter.Filzmoser@tuwien.ac.at

The increasing availability of multivariate functional data across diverse scientific domains highlights the importance of precise analyses of such data structures. Unlike univariate functional data, multivariate functional observations not only exhibit temporal dependence but also between-component correlation. A common simplifying assumption, used to make the estimation of the covariance more tractable, is the separability of the covariance operator, which assumes that the correlation across time and between the components can be uncoupled.

Recently, the α-Mahalanobis distance [1,2] was introduced to the univariate functional setting. Building on this idea, this work introduces the regularized multivariate Mahalanobis distance (RMMD) as an extension of this metric to the multivariate functional case under a separable covariance structure. The incorporation of an appropriately chosen regularization operator assures that the RMMD of a multivariate stochastic process can be calculated as the sum of univariate α-Mahalanobis distances of its scaled and de-correlated components.

We demonstrate the utility of the RMMD in the context of distance-based clustering of multivariate functional data. We show strong clustering performances in various simulation settings, and real-data examples.

Keywords: *Functional data analysis, Mahalanobis distance, cluster analysis*

References

[1] Berrendero, J. R., Bueno-Larraz, B., and Cuevas A. (2020). On Mahalanobis distance in functional settings. Journal of Machine Learning Research, 21 (9), 1–33.
[2] Oguamalam, J., Radojičić, U., and Filzmoser, P. (2024). Minimum regularized covariance trace estimator and outlier detection for functional data. Technometrics, 66 (4), 588–599.

Outlier Detection and Robust Estimation of Time Series Model using a Penalised Approach

Yangzhuoran Fin Yang[1]*, Ines Wilms[2], Jakob Raymaekers[3]
[1] Department of Data Analytics and Digitalisation, Maastricht University, Maastricht, the Netherlands
[2] Department of Quantitative Economics, Maastricht University, Maastricht, the Netherlands
[3] Department of Mathematics, University of Antwerp, Antwerp, Belgium
*Corresponding Author Email: yangzhuoran.yang@maastrichtuniversity.nl

Outliers in time series data pose challenges for both accurate outlier detection and reliable parameter estimation, which can in turn affect forecasting performance. We adopt a penalised regression approach to reliably estimate ARMA models in the presence of outliers, which are identified through mean shift parameters assigned to each time period. Regularisation of the mean shift parameters ensures that nonzero values are retained only at the locations of outliers. We develop different model specifications for innovative outliers, where the effect propagates into future observations, and additive outliers, where the effect is confined to the current period. Estimation is carried out using proximal gradient descent to achieve computational efficiency, and the effect of outlier identification is examined using both soft and hard thresholding. Simulations show that our method outperforms existing robust estimation approaches in terms of outlier identification and model estimation.
Keywords: *Outlier detection, time series model, penalised regression*

Cellwise outlier detection in clustering: from model-based to fuzzy approaches

Giorgia Zaccaria[1]*, Lorenzo Benzakour[2], Luis A. García-Escudero[3], Francesca Greselin[2], Agustín Mayo-Íscar[3]

[1]Department of Economics, Management and Statistics, University of Milano-Bicocca, Via Bicocca degli Arcimboldi 8, Milan, Italy, 20100

[2] Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Via Bicocca degli Arcimboldi 8, Milan, Italy, 20100

[2] Department of Statistics and Operational Research, University of Valladolid, Paseo de Belén 7, Valladolid, Spain, 47011

*Corresponding Author Email: giorgia.zaccaria@unimib.it

Real data often contain outliers, which typically refer to entire cases or rows of a data matrix and are therefore called casewise or rowwise outliers. In recent years, a novel paradigm has been introduced to account for contamination in individual cells of a data matrix, referred to as cellwise outliers [1]. Within the cellwise paradigm, potentially contaminated cells can either be discarded from parameter estimation, as is commonly done in trimming approaches for handling casewise outliers, or "corrected" using reliable information available for each unit in the data matrix [2, 3 for an overview]. In this talk, we present robust clustering methodologies for handling cellwise contamination within a maximum likelihood framework. We illustrate the potential of cellGMM, a cellwise Gaussian Mixture Model [4], for model-based clustering, where cells flagged as contaminated are imputed before parameter estimation via an Expectation-Maximization (EM) algorithm [5]. Thus, potentially outlying cells are treated as missing data, which can be naturally handled within cellGMM, and parameter estimation is performed without discarding units, instead relying on a *completed* data matrix. Additionally, the rationale behind cellGMM has been extended to a fuzzy clustering approach, called cellFCLUST [6], with the advantage of allowing explicit control over the degree of fuzziness in cluster assignments. Although cellGMM, as a Gaussian mixture model, provides "soft" clustering through posterior probabilities and thus captures membership uncertainty, it does not offer the same flexibility as fuzzy clustering in tuning fuzziness. cellFCLUST is also estimated within a maximum likelihood framework through an EM-inspired algorithm with an additional C-step for cellwise outlier identification. The performance of the proposed methodologies is illustrated through both simulated and real data sets.

Keywords: *Cellwise contamination, robust clustering, constrained estimation, missing data, EM algorithm*

References

[1] Alqallaf, F., Van Aelst, S., Yohai, V. J., and Zamar, R. H. (2009). Propagation of Outliers in Multivariate Data. The Annals of Statistics, 37, 311-331.

[2] Raymaekers, J., and Rousseeuw, P. J. (2024). The Cellwise Minimum Covariance Determinant Estimator. Journal of the American Statistical Association, 119, 2610-2621.
[3] Raymaekers, J., and Rousseeuw, P. J. (2024). Challenges of Cellwise Outliers, Econometrics and Statistics, 38, 6-25.
[4] Zaccaria, G., García-Escudero, L. A., Greselin, F., Mayo-Íscar, A. (2025). Cellwise outlier detection in heterogeneous populations. Technometrics, 67(4), 643-654.
[5] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B (Statistical Methodology), 39, 1-38.
[6] Zaccaria, G., Benzakour, L., García-Escudero, L. A., Greselin, F., Mayo-Íscar, A. (2026). Robust fuzzy clustering with cellwise outliers. International Journal of Approximate Reasoning, 195, 109698, doi: 10.1016/j.ijar.2026.109698.

Invited Session 12 (IS12)

IS12 – Robustness, Privacy and Statistical Learning
Chair: Marco Avella-Medina (Columbia University, United States)

Badr-Eddine Chérief-Abdellatif (CNRS, France)
Robust Estimation via Maximum Mean Discrepancy

Getoar Sopa (Columbia University, United States)
Efficient Differentially Private Regression Inference for Longitudinal Data

Sophie Langer (Ruhr University Bochum, Germany)
A Statistical Approach to Image Classification via Deformable Templates

Robust Estimation via Maximum Mean Discrepancy

Badr-Eddine Chérief-Abdellatif

CNRS, Paris, France
*Corresponding Author Email: badr-eddine.cherief-abdellatif@cnrs.fr

In this talk, we investigate the statistical properties of parametric estimators based on the Maximum Mean Discrepancy (MMD). We show that these estimators exhibit strong robustness properties, in the sense that they converge to the best approximation of the data distribution within the chosen model, under minimal assumptions on the data-generating process. We discuss applications including adversarial contamination, temporal dependence, distribution shift, and missing data problems.

Keywords: *Robustness, Maximum mean discrepancy, Kernels*

Efficient Differentially Private Regression Inference for Longitudinal Data

Getoar Sopa[1]*, Marco Avella Medina[1], Cynthia Rush[1]
[1]Columbia University, New York, New York, USA
*Corresponding Author Email: gs3222 @columbia.edu
Keywords: *Linear Regression, Differential Privacy*
Abstract: Differential privacy provides a rigorous framework for releasing statistical analyses while protecting individual-level information. In longitudinal regression, this protection must apply to an individual's entire trajectory, making user-level privacy the natural requirement. We study efficient estimation and inference for fixed-effects longitudinal linear regression under user-level Gaussian differential privacy. We propose private ordinary least squares as well as feasible generalized least squares procedures based on both distributed local regression estimates and pooled estimating equations. We establish finite-sample utility bounds and asymptotic normality under short-range temporal dependence, identifying when privacy noise is asymptotically negligible. Our results characterize regimes where distributed or pooled private regression is preferable, depending on covariate heterogeneity, panel length, sample size, and coefficient magnitude. We also develop private heteroskedasticity and autocorrelation consistent covariance estimators with finite sample privacy noise corrections for valid confidence intervals. Simulations and a longitudinal data example demonstrate the predicted efficiency gains and near-nominal coverage.

References

[1] Sopa, G., Avella Medina, M., Rush, C. (2026). Differentially private inference for longitudinal linear regression. *arXiv preprint arXiv:2601.10626*, 2026.
[2] Sopa, G., Avella Medina, M., Rush, C. (2026). A differentially private and efficient regression estimator for longitudinal data. *Working paper*

A Statistical Approach to Image Classification via Deformable Templates

Sophie Langer[1]*, Juntong Chen[2], Johannes Schmidt-Hieber[3]
[1]Ruhr Universität Bochum, Germany
[2]Xiamen University, China
[3]University of Twente, Netherlands
*Corresponding Author Email: s.langer@rub.de

Modern statistical theory for deep learning often models randomness through additive noise. In image classification problems, however, variability typically arises from geometric deformations such as changes in shape, position, orientation, and other object-specific characteristics.

In this talk, I present the deformation-based statistical framework for image classification. The model captures image variability through stochastic deformations of underlying templates and allows a theoretical analysis of both alignment-based classifiers and convolutional neural networks. A key feature of the framework is that increasing image resolution improves the classification problem, contrasting with the curse-of-dimensionality phenomena commonly encountered in high-dimensional statistics.

We will also briefly discuss how this framework relates naturally to questions in robust statistics, including robustness with respect to geometric perturbations, model misspecification, and distributional shifts.

Invited Session 13 (IS13)

IS13 – Analysis of Complex Data (Hybrid Session)
Chair: Domenico Perrotta (University of Milan, Italy)

Can Hakan Dağıdır (KU Leuven, Belgium) [Virtual]
Cellwise Robust Discriminant Analysis

Andrea Cerioli (University of Parma, Italy)
Robustness, clustering and outlier detection under heavy tails

Arthur Daisomont (KU Leuven, Belgium & Joint Research Centre,Italy)
Blockwise, cellwise, and casewise robust multiblock PCA

Cellwise Robust Discriminant Analysis

Fabio Centofanti[1], Can Hakan Dağıdır[1]*, Mia Hubert[1] Peter J. Rousseeuw[1]
[1]KU Leuven, Celestijnenlaan 200B, Leuven (Heverlee), Belgium, 3001
*Corresponding Author Email: canhakan.dagidir@kuleuven.be

Classical discriminant analysis (DA) is based on the arithmetic mean and empirical covariance matrix of each class, both of which are sensitive to outliers in the data. In the past the focus was on casewise outliers, that is, datapoints that lie far away. But nowadays there is increasing interest in cellwise outliers, that are outlying entries in the data matrix. Removing an entire case because of one or a few outlying cells would lose much information. Cellwise robust methods aim to detect the outlying cells and to preserve the information in the other cells. We propose a DA method based on cellwise robust estimators of location and covariance of the training classes that can also handle NA's. Its main novelty is in the prediction on test data, that may contain outlying cells and NA's as well. We focus on quadratic DA, but also cover the setting of linear DA. The new cellLDA and cellQDA methods perform well in simulation. The approach is illustrated on some real data sets.
Keywords: *cellwise outliers, discriminant analysis, classification*

Robustness, clustering and outlier detection under heavy tails

Andrea Cerioli[1]*, Luis Angel Garcia-Escudero[2], Agustin Mayo-Iscar[2], Domenico Perrotta[3], Francesca Torti[3]
[1]Department of Economics and Management, Via Kennedy 6, 43125 Parma, Italy
[2]Department of Statistics and Operational Research, Paseo de Belén 7, 47011 Valladolid, Spain
[3]Joint Research Centre, Via Fermi 2479, 21027 Ispra (VA), Italy
*Corresponding Author Email: andrea.cerioli@unipr.it

Elliptical heavy-tailed distributions, such as the Student-*t* distribution, have long been advocated as "robust" models for multivariate data in many fields. The underlying rationale of this research stream is that robustness should be achieved by letting the classical maximum-likelihood estimators accommodate extreme observations naturally arising from the process under investigation [1]. However, there is growing recognition that contamination might also occur under heavy tails and other non-Gaussian scenarios, for example in the presence of clusters, heterogeneity or changes of regime that affect the non-Gaussian model assumed to be true for the majority of the data [2,3,4,5,6,7]. In our framework the data generating process for the observable *p*-variate random vector X, with distribution function denoted as F_X, is thus defined through the contamination model

$$F_X = (1 - \varepsilon)F_Y + \varepsilon F_Z,$$

where F_Y is the distribution function of random vector Y representing the postulated null model for the data, while F_Z is the distribution function of a contaminant-model component and $\varepsilon \in [0, 0.5)$ is the contamination rate. The form of F_Z is usually left unspecified, except for some (more or less) informal assumption of separation from F_Y.

Recent developments suitable for an elliptical heavy-tail scenario assume that Y is distributed according to the *p*-variate Student-*t* law. This assumption yields a simple and computationally tractable version of the consistency factor required for high-breakdown estimation of the scatter matrix of Y when a trimming approach is adopted [8], e.g. through the Minimum Covariance Determinant estimator. The same assumption then leads to principled algorithms both for estimating the degrees of freedom of Y and for detecting the contaminant observations from Z [9], based on the theory of the generalized radius process [10]. The main goal of the present work is to highlight some relevant challenges of high-breakdown estimation under heavy tails which need to be addressed for practical application of the methodology. These challenges include both the case of heterogeneous populations, leading to a robust clustering algorithm based on trimming and constraints along the lines of TCLUST [11], and evaluation of the small sample bias of the estimator of the scatter matrix of Y in the spirit of [12].

Keywords: *Contamination, Multivariate Student-t, Minimum Covariance Determinant, Outlier Detection, Robust Clustering*

References

[1] Peel, D., McLachlan, G.J. (2020). Robust mixture modelling using the *t* distribution. Stat. Comput., 10, 339-348.
[2] Croux, C., Haesbroeck, G. (1999). Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. J. Multivar. Anal., 71, 161–190.
[3] Cuesta-Albertos, J.A., Matràn, C., Mayo-Iscar, A. (2008). Trimming and likelihood: Robust location and dispersion estimation in the elliptical model. Ann. Stat., 36, 2284–2318.
[4] Schreurs, J., Vranckx, I., Hubert, M., Suykens, J., Rousseeuw, P. (2001). Outlier detection in non-elliptical data by kernel MRCD. Stat. Comput., 31, 66.
[5] Berenguer-Rico, V., Johansen, S., Nielsen, B. (2023). A model where the least trimmed squares estimator is maximum likelihood. J. R. Stat. Soc. Ser. B, 85, 886–912.
[6] Mayrhofer, M., Radojičić, U., Filzmoser, P. (2025). Robust Covariance Estimation and Explainable Outlier Detection for Matrix-Valued Data. Technometrics, 67, 516–530.
[7] Guney, Y., Arslan, O. (2025). Robust parameter estimation and variable selection in regression models for asymmetric heteroscedastic data. J Appl Stat., 52, 2559–2596.
[8] Barabesi, L., Cerioli, A., Garcìa-Escudero, L.A., Mayo-Iscar, A. (2023). Consistency factor for the MCD estimator at the Student-*t* distribution. Stat. Comput., 33,132.
[9]. Barabesi, L., Cerioli, A., Garcìa-Escudero, L.A., Mayo-Iscar, A., Perrotta, D., Torti, F. (2026). Robust distances and multivariate outlier detection under heavy tails. Stat. Comput., 36, 89.
[10] Garcìa-Escudero, L.A., Gordaliza, A. (2005). Generalized radius processes for elliptically contoured distributions. J. Am. Stat. Assoc. 100, 1036–1045.
[11] Garcìa-Escudero, L.A., Mayo-Iscar, A. (2024). Robust clustering based on trimming. WIREs Comput. Stat., 16, e1658.
[12] Pison, G., Van Aelst, S., Willems, G. (2002). Small sample corrections for LTS and MCD. Metrika, 55, 111–123.

Blockwise, cellwise, and casewise robust multiblock PCA

Arthur Daisomont[12]*, Fabio Centofanti[2], Mia Hubert[2]
[1]Joint Research Center, European Commission, Ispra, Italy
[2] Department of Mathematics, KU Leuven, Leuven, Belgium
*Corresponding Author Email: arthur.daisomont@kuleuven.be

Gross Domestic Product (GDP) and its growth have long been misused as indicators of wellbeing [1]. Emerging alternatives use a large number of variables to assess other areas of wellbeing such as sustainability and inclusiveness. However, it can be unrealistic to condense this complex data into a single indicator. Principal Component Analysis offers a way of extracting insights from such high-dimensional data.

Often, as in the case of wellbeing, the variables can be partitioned into blocks that correspond to different areas or domains. One must then consider the family of multiblock PCA methods, which simultaneously compute a PCA decomposition at the global and block levels while satisfying some constraints between the two. However, existing methods are not robust against cellwise and casewise outliers.

In this work, we first introduce a new type of outlier, the blockwise outlier, which originates from the block structure of the variables. We then propose bloccPCA (BLock-Oriented Cellwise and Casewise PCA), the first multiblock PCA method that is robust against blockwise, cellwise, and casewise outliers. The method minimizes two objective functions to compute loadings at the global and block levels, respectively. In simulations, bloccPCA performs similarly to its non-robust counterpart on clean data, and to its cellwise and casewise robust equivalent [2] on contaminated data, with an improved blockwise robustness overall.

We apply bloccPCA to the data from the SIWB initiative [3] and provide insights into sustainable and inclusive wellbeing within the EU, supported by graphical tools to visualize outliers.

Keywords: *Principal subspace, outlier detection, wellbeing, cellwise outliers, casewise outliers*

References

[1] Costanza, R., Hart, M., Kubiszewski, I., Talberth, J. (2014). A Short History of GDP: Moving Towards Better Measures of Human Well-being. The Solutions Journal, *5*(1), 91-97.
[2] Centofanti, F., Hubert, M., Rousseeuw, P.J. (2026). Robust principal components by casewise and cellwise weighting. Technometrics, appeared online. https://doi.org/10.1080/00401706.2026.2643216
[3] European Commission: Joint Research Centre, Benczur, P., Boskovic, A., Giovannini, E., Pagano, A., Sandor, A.-M. (2025). Measuring sustainable and inclusive wellbeing: a multidimensional dashboard approach. Publications Office of the European Union, Luxembourg. https://data.europa.eu/doi/10.2760/4186342, JRC140456.

Invited Session 14 (IS14)

IS14 – Robust Solutions in Machine Learning Applications (Hybrid Session)
Chair: Fulya Gokalp Yavuz (Purdue University, United States)

David Iseri Inouye (Purdue University, United States) [Virtual]
Provable Robustness to Spurious Correlations via Invariant Data for Robust Finetuning

Burcu Koca Guler (Middle East Technical University, Turkiye)
CGL-LIME: A Robustified LIME with Copula-Based Sampling and Gradient-Weighted Kernels

Serenay Çakar (Middle East Technical University, Turkiye)
From robust mixed models to temporal deep learning: a five-paradigm robustness benchmark on contaminated fNIRS data

Şenay Özdemir (Afyon Kocatepe University, Turkiye)
Robustness of exponential family and m-estimation based empirical likelihood in linear regression

Provable Robustness to Spurious Correlations via Invariant Data for Robust Finetuning

Ruqi Bai[1], Yao Ji[2], Mingyu Kim[1], Easton Currie[1], Zeyu Zhou[1], David Iseri Inouye [1]*
[1]Purdue University, West Lafayette, IN, USA.
[2]Georgia Institute of Technology, Atlanta, GA, USA.
*Corresponding Author Email: dinouye@purdue.edu

While adapting pre-trained foundation models via linear probing is a ubiquitous paradigm, the resulting models remain highly vulnerable to out-of-distribution (OOD) shifts [1] driven by spurious correlations [2, 3]. Traditional Domain Generalization (DG) methods attempt to solve this by learning invariant representations [4], but they frequently suffer from optimization instabilities, lack rigorous finite-sample guarantees, or underperform simple Empirical Risk Minimization (ERM) [5].

To address this fundamental challenge, we introduce Geometric Robustness Invariant Training (GRIT). By leveraging a small dataset of noisy invariant pairs—samples that share underlying semantics but differ in spurious attributes—GRIT estimates the spurious linear subspace of the feature embedding and forces the linear classifier to be orthogonal to this spurious subspace. Notably, this framework generalizes robustness in causal modeling: latent causal models with linear observations satisfy our spurious shift assumption—which generalizes classical covariate shift—while the concept of invariant pairs maps to spurious counterfactuals.

Crucially, we provide a rigorous finite-sample statistical analysis of this framework. We prove that the out-of-distribution test risk explicitly decomposes into in-domain risk, a covariate shift term, and a spurious subspace misalignment term. Our theoretical guarantees demonstrate that the spurious term shrinks at a rate of $O(1/\sqrt{k})$ with k noisy invariant pairs, ensuring provable generalization even when the data pairs contain sub-Gaussian measurement noise. Empirical evaluations on CLIP embeddings [6] confirm that GRIT significantly improves worst-group accuracy over ERM and strong DG baselines on challenging benchmarks like Waterbirds [3] and ColoredMNIST [4].

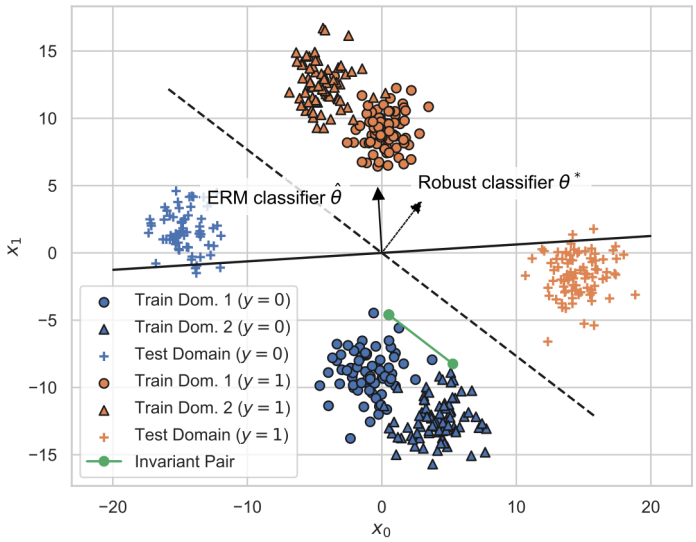


Figure 1: GRIT Illustration - While ERM $\hat{\theta}$ (with solid line decision boundary) fitted on the training domains (circles and triangles) is not robust to shifts in the spurious feature in the unseen test domain (pluses), the robust linear classifier $\hat{\theta}^*$ (with dashed line decision boundary) can be obtained via our GRIT method by enforcing orthogonality to the difference vector of a *single* clean invariant pair (green line). Color indicates the label and shape indicates domain. Noisy pairs would require more than one pair but the error shrinks as $O(1/\sqrt{k})$ where k is the number of noisy invariant pairs.

Keywords: *Spurious Correlations, Out-of-Distribution Robustness, Causal Modeling, Domain Generalization, Invariant Data*

References

[1] Koh, P. W., Sagawa, S., Marklund, H., Xie, S. M., Zhang, M., Balsubramani, A., ... & Liang, P. (2021, July). Wilds: A benchmark of in-the-wild distribution shifts. In International Conference on Machine Learning (ICML) (pp. 5637-5664).

[2] Kumar, A., Raghunathan, A., Jones, R. M., Ma, T., & Liang, P. (2022). Fine-Tuning can Distort Pretrained Features and Underperform Out-of-Distribution. In International Conference on Learning Representations (ICLR).

[3] Kirichenko, P., Izmailov, P., & Wilson, A. G. (2023). Last layer re-training is sufficient for robustness to spurious correlations. In The Eleventh International Conference on Learning Representations (ICLR).

[4] Arjovsky, M., Bottou, L., Gulrajani, I., & Lopez-Paz, D. (2019). Invariant risk minimization. arXiv preprint arXiv:1907.02893.

[5] Gulrajani, I., & Lopez-Paz, D. (2021). In search of lost domain generalization. In International Conference on Learning Representations.

[6] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In International Conference on Machine Learning (ICML) (pp. 8748-8763).

CGL-LIME: A Robustified LIME with Copula-Based Sampling and Gradient-Weighted Kernels

Burcu Koca Guler[1]*, Ceylan Yozgatlıgil[1], Fulya Gokalp Yavuz[1,2]
[1] Department of Statistics, Middle East Technical University, Ankara, Turkiye, 06800
[2] The Data Mine, Purdue University, IN, USA
*Corresponding Author Email: bkoca@metu.edu.tr

As increasingly complex and powerful AI models have emerged, the need to understand and explain their behavior has grown, giving rise to the field of eXplainable AI (XAI). In the existing AI literature, Local Interpretable Model-agnostic Explanations (LIME), one of the most widely used post-hoc explanation methods alongside SHAP, provides local interpretations by fitting a surrogate model to randomly perturbed samples. Despite its pioneering role in the development of local explanation techniques, LIME suffers from limited reproducibility. This limitation arises from the way neighborhood samples are generated, which can produce unrealistic observations. Consequently, the correlation between the variables may not be adequately preserved. This can reduce the fidelity of the surrogate model in the neighborhood of the observation being explained. In this study, we propose a robust extension of LIME that incorporates copula-based sampling and gradient-informed weighting. These modifications aim to generate more realistic samples and yield more robust local explanations. According to the well-established benchmark metrics, simulation studies demonstrate that the proposed method remains highly consistent with the standard LIME framework while providing improved performance.

Keywords: *Machine Learning, Explainable AI, LIME, Black-box Model, Model-Agnostic Interpretation*

From Robust Mixed Models to Temporal Deep Learning: A Five-Paradigm Robustness Benchmark on Contaminated fNIRS Data

Serenay Çakar[1]*, Ceylan Yozgatlıgil[1], Fulya Gokalp Yavuz[1,2]

[1]Department of Statistics, Middle East Technical University, Ankara, Turkiye
[2]The Data Mine, Purdue University, West Lafayette, IN, USA
*Corresponding Author Email: serenay@metu.edu.tr

Functional Near-Infrared Spectroscopy (fNIRS) provides a non-invasive window into brain activity but is highly vulnerable to motion artifacts, sensor displacement, and physiological noise. These sources of contamination pose significant challenges for predictive modeling, particularly in real-world settings where signal quality is often compromised. Moreover, fNIRS data exhibit a complex hierarchical structure, with measurements collected across multiple channels nested within subjects and repeated over time. Subject-specific modeling and robust settings are therefore not merely a statistical refinement but a practical necessity for reliable inference in this domain.

This study presents a comprehensive cross-paradigm robustness benchmark evaluating several models across five methodological families: pure statistical models for repeated measures (including Linear Mixed Model (LMM) and its robust counterpart), hybrid statistical-machine learning (ML) models, pure ML algorithms, deep learning architectures, and a temporal deep learning model. The benchmark targets prediction of standardized oxygenated hemoglobin change (ΔHbO) using an fNIRS dataset comprising 30 subjects and 20 channels across three experimental conditions [1]. LMM analysis established a significant hierarchical structure with nested random intercepts for Channel within Subject: the overall nested random-effects structure was highly significant. The nested channel-within-subject random intercept specifically contributed an additional significant improvement, while fixed effects explained negligible variance. This result motivates models capable of capturing individual-level temporal dynamics. Robustness was assessed via a Contaminate-Then-Split (CTS) protocol in which synthetic outliers of two magnitudes (±10 SD, ±15 SD) were injected at three contamination rates prior to temporal train-test splitting.

Results reveal three principal findings. First, contamination magnitude governs predictive degradation more strongly than rate. Second, the Robust LMM demonstrates the most stable performance across all scenarios, validating M-estimation theory in practice [2]. Among pure ML models, Support Vector Regression proves most robust owing to the implicit outlier tolerance of its $\varepsilon-$insensitive loss function, whereas Extreme Gradient Boosting exhibits systematic overfitting in every scenario, achieving the lowest training error yet among the highest test errors across all conditions. Third, the proposed Dilated Temporal Network with Learned Random Effects (DilTNet-RE; see Figure 1), which combines dilated 1-D convolutions with learned Subject, Channel, Subject and Channel interaction embeddings, achieves the lowest test MAE on clean data, outperforming competing models by explicitly capturing slow hemodynamic autocorrelation across sequential trials that static models discard by treating observations independently [3]. DilTNet-RE maintains this ranking across all contaminated scenarios with a training-test error pattern consistent with good generalisation, though its margin over competing models remains modest.

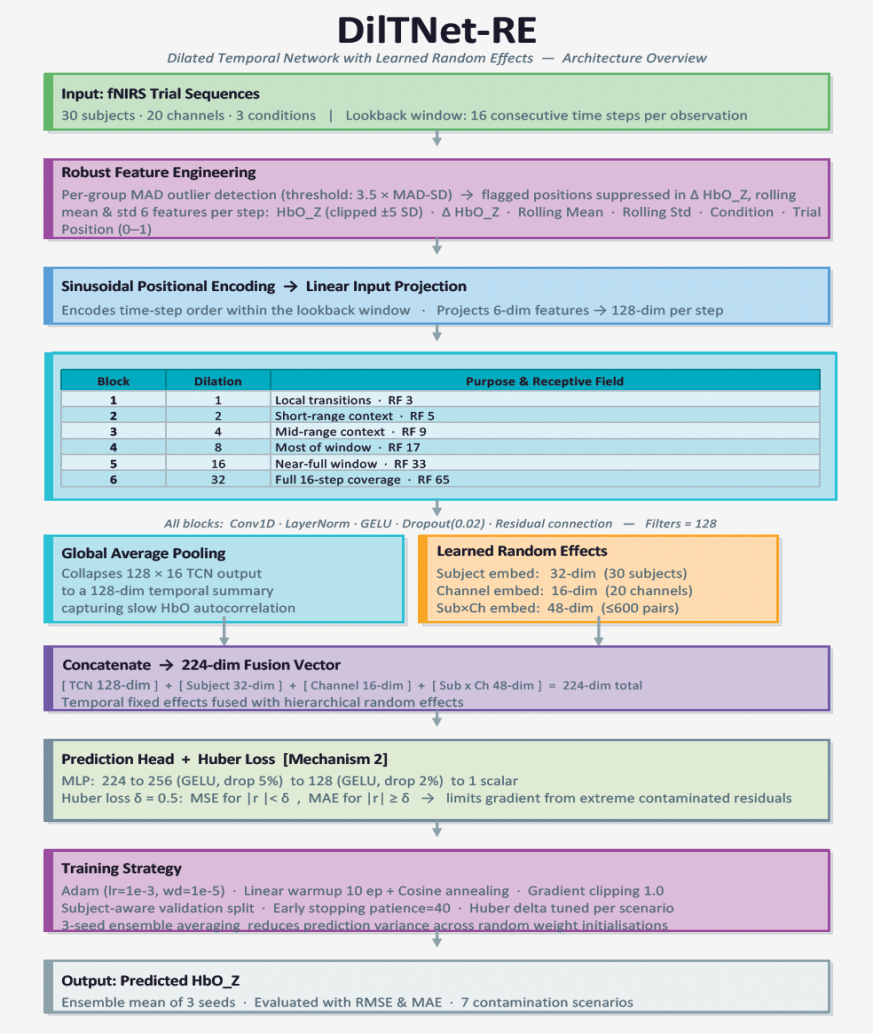


Figure 1: Architecture of the Dilated Temporal Network with Learned Random Effects (DilTNet-RE). Input sequences of 16 time steps with 6 features per step pass through positional encoding, a 6-block dilated convolutional stack (dilations 1–32), and global average pooling. The temporal summary is concatenated with learned Subject, Channel, and Subject×Channel embeddings before a two-layer prediction head produces the standardized HbO estimate.

Keywords: *functional near-infrared spectroscopy, random effects, robust mixed models, M-estimation, contamination benchmarking, temporal deep learning*

References

[1] Bak, S., Park, J., Shin, J., & Jeong, J. (2019). Open-access fNIRS dataset for classification of unilateral finger-and foot-tapping. *Electronics*, *8*(12), 1486.
[2] Koller, M. (2016). robustlmm: an R package for robust estimation of linear mixed-effects models. *Journal of statistical software*, *75*, 1-24.

[3] Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.

Robustness of Exponential Family and M-Estimation Based Empirical Likelihood in Linear Regression

Şenay ÖZDEMİR[1]*, Olcay ARSLAN[2]
[1] Afyon Kocatepe University, Department of Data Science and Analytics, Afyonkarahisar, Turkiye, 03200
[2] Ankara University, Department of Statistics, Degol Street, Besevler, Ankara, Turkiye, 06100
*Corresponding Author Email: senayozdemir@aku.edu.tr

The empirical likelihood method is a parameter estimation method that can be used with small sample sizes, possesses the flexibility of non-parametric methods, and exhibits the advantages of parametric likelihood methods [1]. Although this method inherently assigns a probability value to each observation, the constraints on the relevant parameter are based on classical methods—OLS for regression model parameters—therefore estimates are affected by data contamination and/or the presence of outliers. To cope with such situations, it is preferable to write the problem in the form of an exponential family or to use M-estimation-based constraints. While the exponential family formulation enhances the flexibility and numerical stability of the dual optimization process across various data structures, M-estimation-based constraints provide a direct defense mechanism by bounding the influence of severe outliers [2, 3]. In this study, it is aimed to compare different scenarios in terms of breakdown points using MSEs calculated to determine which method is more robust in estimating linear regression parameters under various sample sizes and contamination rates.

Keywords: *Empirical likelihood, Exponential family, M-estimation, Breakdown point*

References

[1] Owen, A. (1991). Empirical likelihood for linear models. *The Annals of Statistics*, 1725-1747.S
[2] Corcoran, S. A. (2000). Empirical exponential family likelihood using several moment conditions. Statistica Sinica, 545-557.
[3] Özdemir, Ş., and Arslan, O. (2020). Combining empirical likelihood and robust estimation methods for linear regression models. Communications in Statistics-Simulation and Computation, 51(3), 941-954.

Invited Session 15 (IS15)

IS15 – Advances in Robust Methodologies
Chair: F. Sevinc Kurnaz (Yıldız Technical University, Turkiye & Case Western Reserve University, United States)

Matthias Templ (University of Applied Sciences and Arts Northwestern Switzerland, Switzerland)
Cellwise-robust imputation for mixed continuous and categorical data

Houyem Demni (University of Cassino and Southern Lazio, Italy)
Robust Estimation of Circular Dispersion: Theory and an Application to Outlier Detection

Onur Toka (Hacettepe University, Turkiye)
A Systematic Comparison of Clustering Methods Under Casewise, Cellwise, and Mixed Data Contamination

ICORS2026
20-24 JULY 2026
Istanbul

Cellwise-robust imputation for mixed continuous and categorical data

Matthias Templ
Institute of Data Science, School of Engineering, FHNW University of Applied Sciences and Arts Northwestern Switzerland
Corresponding Author Email: matthias.templ@fhnw.ch

Cellwise contamination — in which individual cells of a data matrix deviate from the assumed model independently — is increasingly recognised as a realistic error model for modern datasets. While recent methods for cellwise outlier detection and robust estimation have advanced substantially (DDC [2], cellMCD [3]), all existing cellwise approaches are restricted to fully continuous, complete data. No method currently handles cellwise contamination and missing values jointly for mixed continuous and categorical data. In practice, datasets in survey research, medical records, and administrative databases routinely contain a mix of continuous and categorical variables.

We propose three integrated cellwise-robust imputation methods for mixed data. The flagship method, cellIRMI, extends the iterative regression imputation framework [1] by replacing row-level robustness weights with per-cell weights in the design matrix. Each continuous predictor cell receives a weight reflecting its estimated probability of being clean; categorical predictors receive unit weights. The cell-weighted IRWLS engine alternates between psi-weight updates from regression residuals and cell-weight updates from robust standardisation, enabling simultaneous detection and imputation across variable types. Two alternative methods are also proposed: cellM, a single-regression variant with a formal consistency theorem under the independent cellwise contamination model, and cellEM, an EM algorithm with latent contamination indicators that inherits standard EM convergence guarantees.

A simulation study at dimensions p = 12 (6 continuous + 6 categorical) with block-correlated contamination (shift = 3 SD), MAR missingness, and strong categorical group effects (2–4 SD) demonstrates that the proposed methods provide competitive inference quality. Using RMSE of regression coefficients as the primary metric, cellM matches the two-step DDC + imputeRobust baseline [2] across all contamination rates ($\varepsilon = 0$ to 0.20), while cellEM shows advantages at low contamination where subtle outliers evade hard-threshold detection. At $\varepsilon = 0.20$ (91% of rows affected), all three proposed methods are competitive with DDC + imputeRobust and substantially outperform standard mice. The key advantage of the cellwise methods emerges when categorical variables carry strong predictive information for the continuous variables — a scenario where DDC [2], which ignores categoricals, loses detection power. We discuss the relationship between cellwise breakdown points [4, 5] and the empirical robustness observed in our simulation.

The methods are implemented as imputeCellIRMI(), imputeCellM(), and imputeCellEM() in the R package VIM, available on CRAN (soon).

Keywords: *cellwise contamination, robust imputation, mixed data, missing values, iterative regression, cell weights*

References

[1] Templ, M., Kowarik, A., Filzmoser, P. (2011). Iterative stepwise regression imputation using standard and robust methods. Computational Statistics & Data Analysis, 55(10), 2793–2806.
[2] Rousseeuw, P.J., Van den Bossche, W. (2018). Detecting deviating data cells. Technometrics, 60(2), 135–145.
[3] Raymaekers, J., Rousseeuw, P.J. (2024). The cellwise minimum covariance determinant estimator. Journal of the American Statistical Association, 119(545), 576–588.
[4] Alqallaf, F., Van Aelst, S., Yohai, V.J., Zamar, R.H. (2009). Propagation of outliers in multivariate data. The Annals of Statistics, 37(1), 311–331.
[5] Agostinelli, C., Leung, A., Yohai, V.J., Zamar, R.H. (2015). Robust estimation of multivariate location and scatter in the presence of cellwise and casewise contamination. TEST, 24(3), 441–461.

Robust Estimation of Circular Dispersion: Theory and an Application to Outlier Detection

Houyem Demni[1]*, Mia Hubert[2], Giovanni C. Porzio[1], Peter J. Rousseeuw [2]
[1]University of Cassino and Southern Lazio, Cassino 03043, Italy
EUt+ Data Science Research Group, European University of Technology, EU
[2]Section of Statistics and Data Science, KU Leuven, Belgium
*Corresponding Author Email: Houyem.demni@unicas.it

Circular variables arise in the analysis of directional and periodic phenomena and they are encountered across diverse scientific domains, including biology, environmental science, meteorology, and geosciences. Typical examples include wind directions, animal movement orientations, and time-of-day measurements. A fundamental challenge in circular data analysis is the reliable estimation of dispersion in the presence of outliers or contaminated observations, which can substantially compromise classical methods. This motivates the development of robust estimation techniques specifically designed for circular settings.

In this work, we propose a general framework for the robust estimation of circular dispersion. We extend three established robust dispersion measures from the linear domain to the circular context, ensuring resistance to anomalous observations while maintaining efficiency under uncontaminated conditions. The robustness properties of the proposed estimators are rigorously assessed through influence functions and relative bias curves, providing formal guarantees of stability under contamination.

Building on these dispersion measures, we further derive robust estimators for key parameters of commonly used circular distributions, including the concentration parameter of the von Mises distribution and the dispersion parameter of the wrapped normal distribution. Their theoretical properties are thoroughly investigated, offering insight into the trade-off between efficiency and robustness.

Finally, leveraging the best-performing estimator, we introduce a robust anomaly detection rule for circular data. This approach is complemented by a novel visualization tool, the circular violin plot, which facilitates the identification of outlying observations. The practical utility of the proposed framework is demonstrated through the analysis of real datasets.

Keywords: *Breakdown value, Circular Median Absolute Deviation, Circular Least Median Spread, Circular Least Trimmed Standard Deviation, Directional data, Outlier detection.*

[1] Demni, H., Porzio, G. C., Hubert, M., & Rousseeuw, P. (2026). Robust measures of dispersion for circular data with an anomaly detection rule. *arXiv preprint*, 2603.01237.

A Systematic Comparison of Clustering Methods Under Casewise, Cellwise, and Mixed Data Contamination

Onur Toka[1]*
[1]Hacettepe University, Beytepe Campus, Department of Statistics, Ankara, Turkiye, 06800
*Corresponding Author Email: onur.toka@hacettepe.edu.tr

The robust statistics literature has sought to mitigate the effects of outliers through a variety of methodological developments, with foundational contributions covering robust estimation of location and scatter, linear regression, principal component analysis, clustering, and classification. In general, robust methods have demonstrated considerable success in handling casewise outliers, as evidenced by trimming-based and reweighting approaches reviewed comprehensively in the literature. However, cellwise outliers, which affect only isolated cells within an observation in a sparse and unpredictable manner, have been shown to propagate through multivariate procedures in ways that classical robust methods fail to accommodate [1]. Addressing this challenge, researchers have proposed methods and algorithms to combat both cellwise and casewise outliers simultaneously, including approaches for detecting deviating data cells [2], robust principal component analysis under mixed contamination [3], robust covariance estimation [4,5], robust variable selection [6,7], and robust hierarchical clustering.

Clustering algorithms serve as a fundamental tool for uncovering latent structure in data; however, it has been established that they are sensitive to both types of outliers, with adverse consequences for the number of clusters selection, cluster assignment, and the reliability of cluster validity indices. Nevertheless, a systematic and comprehensive simulation framework that jointly examines clustering performance in the presence of casewise, cellwise, and mixed outlier structures remains largely absent from the literature. In this study, we conduct an extensive simulation study evaluating the robustness of clustering algorithms under different contamination mechanisms. The experiments distinguish between casewise, cellwise, and mixed contamination, while varying contamination proportion, contamination magnitude, dimensionality, number of clusters, and cluster separation. Clustering performance is evaluated using the Adjusted Rand Index, internal clustering indices, stability measures, and robustness degradation profiles.

Keywords: *Cellwise outliers, Casewise outliers, Robust clustering, Robustness degradation curves, Simulation study*

References

[1] Alqallaf, F., Van Aelst, S., Yohai, V. J., & Zamar, R. H. (2009). Propagation of outliers in multivariate data. *The Annals of Statistics*, *37*(1), 311–331.
[2] Rousseeuw, P. J., & Van den Bossche, W. (2018). Detecting deviating data cells. *Technometrics*, *60*(2), 135–145.

[3] Hubert, M., Rousseeuw, P. J., & Van den Bossche, W. (2019). MacroPCA: An all-in-one PCA method allowing for missing values as well as cellwise and rowwise outliers. *Technometrics*, *61*(4), 459–473.
[4] Raymaekers, J., & Rousseeuw, P. J. (2021). Handling cellwise outliers by sparse regression and robust covariance. *Journal of Data Science, Statistics, and Visualisation*, *1*(3).
[5] Agostinelli, C., Leung, A., Yohai, V. J., & Zamar, R. H. (2015). Robust estimation of multivariate location and scatter in the presence of cellwise and casewise contamination. *TEST*, *24*(3), 441–461.
[6] Toka, O., Çetin, M., & Arslan, O. (2021). Robust regression estimation and variable selection when cellwise and casewise outliers are present. *Hacettepe Journal of Mathematics and Statistics*, *50*(1), 289–303.
[7] Cappozzo, A., Greselin, F., & Murphy, T. B. (2021). Robust variable selection for model-based learning in presence of adulteration. *Computational Statistics & Data Analysis*.

Invited Session 16 (IS16)

IS16 – Robust Estimation, Testing and Prediction
Chair: Yanyuan Ma (Penn State University, United States)

Tanya Garcia (University of North Carolina at Chapel Hill, United States)
When Distribution-Free Falls Short for Prediction Intervals under Right-Censored Covariate

Renjun Ma (University of New Brunswick, Canada)
Tweedie compound Poisson mixed models for cross-classified data with left censoring

Ying Wei (Columbia University, United States)
Quantile Graph Discovery through QuACC: Quantile Association via Conditional Concordance

When Distribution-Free Falls Short for Prediction Intervals under Right-Censored Covariate

Kihyun Han[1], Yanyuan Ma[1], Tanya P. Garcia[2*]
[1]Penn State University, Department of Statistics, Pennsylvania State University, State College, PA 16802
[2]Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27517
*Corresponding Author Email: tpgarcia@email.unc.edu

Prediction intervals provide a data-driven range for an unmeasured outcome at a prespecified probability level, and are essential for prognosis and patient monitoring in clinical studies. Unstable prediction intervals---ones that vary substantially in length or coverage rate from study to study---cannot support reliable clinical decisions about patient screening, resource allocation, or intervention timing. A common source of this instability is a right-censored time-to-event covariate: the event has not yet occurred by the end of the study, so the covariate value needed for prediction is unknown. No existing method constructs prediction intervals directly from a right-censored time-to-event covariate, and adapting distribution-free methods to this setting produces exactly this instability. We develop a semiparametric prediction method that incorporates an outcome model, a model for the time-to-event covariate, and a model for the censoring time into the estimation of the prediction interval length. The method achieves the smallest possible variance in interval length estimation---a formal improvement over distribution-free methods---and remains consistent even when the model for the time-to-event covariate or the model for the censoring time is misspecified. Simulation studies confirm substantially more stable interval lengths and coverage rates than distribution-free methods across censoring rates. In a Huntington disease study with 77% censoring, our method achieves reliable coverage with stable interval lengths, while distribution-free methods produce either persistent undercoverage or intervals too wide to be informative.
Keywords: *Semiparametric, Doubly Robust, Efficient*

Tweedie compound Poisson mixed models for cross-classified data with left censoring

Jingyu Cui[1], Renjun Ma[2]*, M. Tariqul Hasan[2]

[1]Department of Biostatistics, Yale University School of Public Health, Connecticut, USA
[2]Department of Mathematics and Statistics, University of New Brunswick, New Brunswick, Canada
*Corresponding Author Email: renjun@unb.ca

Environmental and health data are often subject to left censoring at detection limit. Examples of such data are water pollutants in fish, tumor size due to radioactivity treatments, lesion depth related to ultrasound and precipitation. These censored data are usually right-skewed continuous, but with a point mass at the detection limit. Our current study was motivated by an environmental monitoring study in Ontario where mercury concentration was measured with a detection limit on 2753 fish from 91 lakes with maximum of 7 species surveyed between 2011-2016 [1]. In this study, we introduce Tweedie compound Poisson models with partially crossed lake and species distribution-free random effects to analyze mercury concentration data. Our model estimation has been done through an orthodox best linear unbiased predictors (BLUP) approach [2]. Our analysis results are robust against distributional assumptions about random effects.

Tweedie compound Poisson models with partially crossed random effects

Let Y_k represent the response for the kth ($k \in 1, 2, \ldots, N$) observation, so the response vector can be expressed as $\boldsymbol{Y} = (Y_1, \ldots, Y_k, \ldots, Y_N)^T$. In addition, $\boldsymbol{U} = (U_1, \ldots, U_i, \ldots, U_I)^T$ and $\boldsymbol{V} = (V_1, \ldots, V_j, \ldots, V_J)^T$ represent two independent random effects vectors. The model is specified as follows:

Assumption 1: Lake-specific random effect (U_i):

$$U_1, U_2, \ldots, U_i, \ldots, U_{91} \text{ are iid with } E(U_i) = 1 \text{ and } Var(U_i) = \sigma^2.$$

Assumption 2: Species-specific random effect (V_j) :

$$V_1, V_2, \ldots, V_j, \ldots, V_7 \text{ are iid } E(V_j) = 1 \text{ and } Var(V_j) = \tau^2.$$

Assumption 3. Given random effects U and V, the components of Y are conditionally independent and follow the Tweedie's compound Poisson distribution, that is

$$Y_k \mid \boldsymbol{U}, \boldsymbol{V} \sim \mathrm{Tw}_p\{\mu_k U_i V_j, \rho^2 (U_i V_j)^{1-p}\},$$

where Y_k is the difference between kth observation of mercury concentration and detection limit.

Orthodox best linear unbiased predictor

Let $\mathbf{W} = (U_1V_1, \cdots, U_1V_j, \cdots, U_1V_J, \cdots, U_iV_j, \cdots, U_IV_1, \cdots, U_IV_j, \cdots, U_IV_J)^T$. The orthodox best linear unbiased predictors of random effects **U**, **V** and **W** are given by

$$\widehat{\mathbf{U}} = E[\mathbf{U}] + Cov(\mathbf{U}, \mathbf{Y})Cov^{-1}(\mathbf{Y})\{\mathbf{Y} - E[\mathbf{Y}]\}$$

$$\widehat{\mathbf{V}} = E[\mathbf{V}] + Cov(\mathbf{V}, \mathbf{Y})Cov^{-1}(\mathbf{Y})\{\mathbf{Y} - E[\mathbf{Y}]\}$$

$$\widehat{\mathbf{W}} = E[\mathbf{W}] + Cov(\mathbf{W}, \mathbf{Y})Cov^{-1}(\mathbf{Y})\{\mathbf{Y} - E[\mathbf{Y}]\}$$

Optimal estimating function for regression parameters

$$\psi(\beta) = \sum_{k=1}^{N} X_k' \frac{\mu_k(\beta)^{1-p}}{\rho^2}\left(Y_k - \mu_k(\beta) \cdot \widehat{U_i V_j}(\beta)\right) = 0$$

Moment estimation of dispersion parameters

$$\hat{\sigma}^2 = \frac{1}{I}\sum_{i=1}^{I}(\widehat{U}_i - 1)^2 + bias\ correction$$

$$\hat{\tau}^2 = \frac{1}{J}\sum_{j=1}^{J}(\widehat{V}_j - 1)^2 + bias\ correction$$

$$\hat{\rho}^2 = \frac{1}{N}\sum_{k=1}^{N}\frac{(Y_k - \mu_k \widehat{U_i V_j})^2}{\mu_k^p} + bias\ correction$$

Our orthodox BLUP algorithm iterates between updating the regression parameter estimates, updating random-effect predictors and updating dispersion parameters. This orthodox BLUP approach not only enjoys the desirable theoretical properties such as robustness, consistent and optimal estimators, but also is computationally more efficient than its competitors [3]. (Acknowledgement: The authors thank Dr. Thomas Johnston from Ontario Ministry of Natural Resources for providing us the data.)

Keywords: *Best linear unbiased predictor, crossed random effects, detection limit, semicontinuous data,* zero-inflation

References

[1] Tang, R. W., Johnston, T. A., Gunn, J. M., & Bhavsar, S. P. (2013). Temporal changes in mercury concentrations of large-bodied fishes in the boreal shield ecoregion of northern Ontario, Canada. Science of the total environment, 444, 409-416.

[2] Ma, R., & Jørgensen, B. (2007). Nested generalized linear mixed models: an orthodox best linear unbiased predictor approach. Journal of the Royal Statistical Society Series B: Statistical Methodology, 69(4), 625-641.

[3] Ma, R., Krewski, D., & Burnett, R. T. (2003). Random effects Cox models: a Poisson modelling approach. Biometrika, 90(1), 157-169.

Quantile Graph Discovery through QuACC: Quantile Association via Conditional Concordance

Zain Khan[1], Daniel Malinsky[1], Martin Picard[1], Alan A. Cohen[1], Columbia SOH Group[1], Ying Wei[1]*
[1]Columbia University, New York, NY 10032
*Corresponding Author Email: yw2148@cumc.columbia.edu

Graphical structure learning is an effective way to assess and visualize cross-biomarker dependencies in biomedical settings. Standard approaches to estimating graphs rely on conditional independence tests that may not be sensitive to associations that manifest at the tails of joint distributions, i.e., they may miss connections among variables that exhibit associations mainly at lower or upper quantiles. In this work, we propose a novel measure of quantile-specific conditional association called QuACC: Quantile Association via Conditional Concordance. For a pair of variables and a conditioning set, QuACC quantifies agreement between the residuals from two quantile regression models, which may be linear or more complex, e.g., quantile forests. Using this measure as the basis for a test of null (quantile) association, we introduce a new class of quantile-specific graphical models. Through simulation we show our method is powerful for detecting dependencies that manifest at the tails of distributions. We apply our method to biobank data from All of Us and identify quantile-specific patterns of conditional association in a multivariate setting.

Keywords: *graphical models, conditional independence, tail dependence, quantile regression*

Invited Session 17 (IS17)

IS17 – Stability in Inference from Different Prospective (Hybrid Session)
Chairs: Claudio Agostinelli (University of Trento, Italy) and Anand Vidyashankar (George Mason University, United States)

Giulia Bertagnolli (Free University of Bozen-Bolzano, Italy) [Virtual]
Robust Inference for Mixtures of Multivariate Wrapped Normal Distributions: The Weighted Likelihood Approach

Kamelia Daudel (ESSEC Business School, France)
Optimizing Rényi's alpha-divergences: a variational inference perspective

Badr-Eddine Chérief-Abdellatif (CNRS, France)
Bayesianism and Robustness: Predictively Oriented (Pr0) Posteriors

Robust Inference for Mixtures of Multivariate Wrapped Normal Distributions: The Weighted Likelihood Approach

Giulia Bertagnolli[1]*, Luca Greco[2], Claudio Agostinelli[3]
[1]Free University of Bozen-Bolzano, Piazza Università, 1, 39100, Bozen-Bolzano, Italy
[2]University G. Fortunato, Viale Raffaele Delcogliano, 12, 82100, Benevento, Italy
[3]University of Trento, Via Sommarive, 14, 38123 Trento, Italy
*Corresponding Author Email: giulia.bertagnolli@unibz.it

We propose a robust extension of finite mixture models of multivariate Wrapped Normal distributions[1] for the analysis of heterogeneous circular data on a p-dimensional torus. While finite mixtures of Wrapped Normal distributions provide a flexible framework for modelling multivariate circular or directional observations, standard maximum likelihood estimation may be severely affected by atypical observations, contamination, or model misspecification. To address this issue, we develop a weighted likelihood estimation procedure in which the contribution of each observation to the likelihood is adaptively weighted according to its agreement with the assumed model and with the mixture component membership. As for the non-robust case, parameter estimation is carried out through a nested weighted EM-type algorithm that exploits the unwrapping representation of circular data. As in the classical approach, the outer E-step accounts for the unobserved component memberships, whereas the inner E-step handles the latent vectors of wrapping coefficients and includes the additional weight for robustness. The proposed robust procedure preserves the flexibility of the original model while improving stability and resistance to outlying observations. The finite-sample performance of the method is investigated through Monte Carlo simulation studies under both clean and contaminated scenarios, and its practical usefulness is illustrated on real multivariate circular datasets.

Keywords: *robust statistics, circular data, weighted likelihood, mixture models, Wrapped Normal distributions*

References

[1] Greco, L., Novi Inverardi, P. L., & Agostinelli, C. (2023). Finite mixtures of multivariate wrapped normal distributions for model based clustering of p-torus data. Journal of Computational and Graphical Statistics, 32(3), 1215-1228.

Optimizing Rényi's alpha-divergences: a variational inference perspective

Kamélia Daudel[1]*
[1]First Author's Affiliation, ESSEC Business School
*Corresponding Author Email: daudel@essec.edu

This talk focuses on the methodological and computational challenges of optimizing Rényi's α-divergences within the framework of Variational Inference (VI). While standard VI relies almost exclusively on the Kullback-Leibler divergence, moving to the broader family of Rényi's α-divergences introduces optimization difficulties that require dedicated theoretical analyses as well as variance-reduction strategies [1,2].

Keywords: *Rényi's alpha-divergences, importance sampling, Monte Carlo methods, variational inference*

References

[1] K. Daudel and F. Roueff (2026). Learning with Importance Weighted Variational Inference: Asymptotics for Gradient Estimators of the VR-IWAE Bound. arXiv preprint arXiv:2410.12035
[2] K. Daudel, MN. Tran, C. Zhang (2026). Importance Weighted Variational Inference without the Reparameterization Trick. arXiv preprint arXiv:2602.01412

Bayesianism and Robustness: Predictively Oriented (Pr0) Posteriors

Badr-Eddine Chérief-Abdellatif

CNRS, Paris, France
*Corresponding Author Email: badr-eddine.cherief-abdellatif@cnrs.fr

In this talk, we will present a new statistical principle that combines the most desirable aspects of both parameter inference and density estimation. This leads us to the PrO posterior, which expresses uncertainty as a consequence of predictive ability. Doing so leads to inferences which predictively dominate both classical and generalised Bayes posterior predictive distributions. Our PrO posteriors adapt to the level of model misspecification: they concentrate around the true model in the same way as Bayes and Gibbs posteriors if the model can recover the data-generating distribution, but do not concentrate toward a single parameter in the presence of non-trivial forms of model misspecification. Instead, they stabilise towards a predictively optimal posterior whose degree of irreducible uncertainty admits an interpretation as the degree of model misspecification — a sharp contrast to how Bayesian uncertainty and its existing extensions behave.

Keywords: *Robustness, Uncertainty quantification, Bayes, Misspecification, Prediction*

Contributed Session 1 (CS1)

CS1 – Applications of Robust Methods-1
Chair: Fulya Gokalp Yavuz (Purdue University, United States)

Burak Dilber (Dokuz Eylül University, Turkiye)
Comprehensive Evaluation of METHoD-NO for Benchmark Optimization and Neural Network Training

Yasemin Günter (Halic University, Turkiye; Yildiz Technical University, Turkiye)
GWO-Optimized Hybrid Deep Learning Approach for Lung Cancer Detection: High-Accuracy Classification on CT Images

Gulder Kemalbay (Yildiz Technical University, Turkiye)
Bootstrap-Based Robust Higher-Order Moment Portfolio Optimization under Regime Heterogeneity: A Homogeneity-Driven Estimation Approach

Jens Klooster (University of Groningen, Netherlands)
On inference using M-estimators under contamination

Nuri Berk Ural (Cukurova University, Turkiye)
Hybrid sliding window recursive least squares for battery state of health prediction

Comprehensive Evaluation of METHoD-NO for Benchmark Optimization and Neural Network Training

Burak Dilber[1]*, A. Fırat Özdemir[1]
[1]Dokuz Eylul University, Department of Statistics, İzmir, Türkiye

*Corresponding Author Email: dilber.burak@ogr.deu.edu.tr

Metaheuristic optimization algorithms are widely used for solving complex global optimization problems due to their flexibility, derivative-free search mechanisms, and effectiveness in nonlinear search spaces. In this study, the performance of the recently proposed METHoD-NO optimizer is comprehensively investigated on both numerical benchmark problems and artificial neural network training tasks. METHoD-NO is a robust statistics-inspired metaheuristic algorithm whose search mechanism is organized into three complementary phases to achieve a balanced exploration–exploitation process. In the **central tendency phase**, the search is guided by robust location measures so that the population can be directed toward promising regions while reducing sensitivity to extreme or misleading candidate solutions. In the **Harrell–Davis phase**, the algorithm benefits from the Harrell–Davis quantile estimator, which estimates quantiles as weighted combinations of order statistics and thereby provides a smoother and more informative characterization of the population distribution [1]. In the **Navruz–Özdemir quantile estimator phase**, an alternative quantile-based learning mechanism is employed to refine the search dynamics and strengthen the algorithm's ability to adapt to the underlying distributional structure of candidate solutions [2]. By combining these three phases, METHoD-NO aims to preserve population diversity, enhance stability, and maintain effective search performance over complicated optimization landscapes.

In the first part of the study, the global optimization capability of METHoD-NO is evaluated using 23 classical benchmark functions together with the CEC 2017 and CEC 2022 test suites. These benchmark sets include unimodal, multimodal, hybrid, and composition functions with varying difficulty levels and therefore provide a comprehensive basis for performance assessment. The results obtained by METHoD-NO are compared with those of well-known metaheuristic algorithms in terms of solution quality, convergence behavior, and robustness. In addition, the statistical significance of the observed differences is examined using the Wilcoxon rank-sum test, providing a more reliable basis for pairwise algorithm comparisons.

In the second part of the study, the applicability of METHoD-NO to data-driven learning problems is examined through artificial neural network training. In this framework, METHoD-NO is employed to optimize the weights and biases of feed-forward neural networks, and its performance is compared with competing metaheuristic methods on both regression and classification datasets. Since neural network training often involves highly nonlinear and multimodal error surfaces, it is an appropriate test bed for evaluating the search capability of metaheuristic optimizers. The findings are expected to show that METHoD-NO provides competitive and stable

performance not only for numerical global optimization problems but also for artificial neural network training tasks.

Keywords: *METHoD-NO, metaheuristic optimization, global optimization, benchmark functions, artificial neural networks, Wilcoxon test*

References

[1] Harrell, F. E., & Davis, C. E. (1982). A new distribution-free quantile estimator. Biometrika, 69(3), 635–640.

[2] Navruz, G., & Özdemir, A. F. (2020). A new quantile estimator with weights based on a subsampling approach. British Journal of Mathematical and Statistical Psychology, 73(3), 506–521.

GWO-Optimized Hybrid Deep Learning Approach for Lung Cancer Detection: High-Accuracy Classification on CT Images

Yasemin Günter[1,2]*, Fatma Noyan Tekeli[2]
[1]Department of Big Data Analytics, Haliç University, İstanbul, Turkiye, 34060
[2]Department of Statistics, Yıldız Technical University, İstanbul, Turkiye, 34220
*Corresponding Author Email: yasemingunter@halic.edu.tr

The development of methods for the early and accurate diagnosis of lung cancer, which has a high mortality rate worldwide, is of critical importance [1]. As the manual evaluation of Computed Tomography (CT) images is time-consuming and prone to error for radiologists, this has increased the need for high-accuracy automated diagnostic systems [2]. However, limitations in distinguishing complex lesion structures and data imbalance issues necessitate the development of more robust hybrid methods. In this study, a hybrid deep learning model and four different deep learning architectures have been developed for the early detection of lung cancer. Within the hybrid model, the ability of the VGG16 architecture to capture local tissue features and the multi-scale feature extraction capacity of the InceptionV3 architecture were integrated using the feature concatenation method, resulting in a rich 2560-dimensional feature representation [3]. To enhance model performance, the Grey Wolf Optimizer (GWO) algorithm was employed to optimise the hyperparameters of learning rate, batch size, momentum, the number of dense units, and the number of epochs [4, 5]. The GWO-VGG16, GWO-DenseNet, GWO-DenseNet+LSTM, GWO-InceptionV3 and hybrid GWO-VGG16+InceptionV3 models developed within this scope were evaluated comparatively on the IQ-OTH/NCCD dataset, which is widely used in the literature. The dataset consists of a total of 1,097 CT images, comprising 120 benign, 561 malignant and 416 normal images [5]. To address the class imbalance in the dataset and improve generalisation performance on minority classes, data balancing was performed during the training phase using the Synthetic Minority Over-sampling Technique (SMOTE) [6]. According to the analysis results, the GWO-VGG16+InceptionV3 hybrid model demonstrated the highest performance with an accuracy of 94.58%, a specificity of 97.40%, and an AUC value of 0.9929. When examining the individual models; it was found that the GWO-VGG16 and GWO-InceptionV3 models demonstrated similar performance with 93.98% accuracy, whereas the GWO-DenseNet and GWO-DenseNet+LSTM models remained at a 92.77% accuracy level. In particular, the fact that the GWO-InceptionV3 model, despite having the lowest training loss (0.0048), lagged behind the hybrid model in terms of overall performance demonstrates that a low loss value alone does not guarantee high generalisation performance. In conclusion, the integration of hybrid deep learning architectures with GWO-based optimisation offers the potential for high accuracy, balanced classification and a reliable clinical decision support system in the early diagnosis of lung cancer.

Keywords: *CT , Lung Cancer, Deep Learning, VGG16, InceptionV3, Grey Wolf Optimizer.*

References

[1] World Health Organization. (2023). Lung cancer. https://www.who.int/news-room/fact sheets/detail/lung-cancer. Accessed: 29-Apr-2026.
[2] Shatnawi, M. Q., Abuein, Q., & Al-Quraan, R. (2025). Deep learning-based approach to diagnose lung cancer using CT-scan images. Intelligence-Based Medicine, 11, 100188.
[3]Kumaran, Y. S., Jeya, J. J., Mahesh, T. R., Khan, S. B., Alzahrani, S., & Alojail, M. (2024). Explainable lung cancer classification with ensemble transfer learning of VGG16, Resnet50 and InceptionV3 using grad-cam BMC Medical Imaging, 24, 176.
[4] Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey Wolf Optimizer. Advances in Engineering Software, 69, 46-61.
[5] Hossain, M. I., Ahmad, S., Hasan, S., Rahim, M. A., & Shin, J. (2024). Lung cancer detection and classification using LeNet-LSTM model on computed tomography images. Multidisciplinary Science Journal, 7(1), 2025188.
[6] Mohamed Musthafa, M., Manimozhi, I., Mahesh, T. R., & Guluwadi, S. (2024). Optimizing double-layered convolutional neural networks for efficient lung cancer classification through hyperparameter optimization and advanced image pre-processing techniques. BMC Medical Informatics and Decision Making, 24(1), 142.

Bootstrap-Based Robust Portfolio Optimization under Regime Heterogeneity: A Homogeneity-Driven Approach

Gulder Kemalbay[1]*
[1]Department of Statistics, Yildiz Technical University, Istanbul, Turkiye, 34220
*Corresponding Author Email: kemalbay@yildiz.edu.tr

Traditional portfolio optimization models typically rely on point estimates and often assume that information can be pooled across market environments. However, sampling uncertainty and regime-dependent heterogeneity may substantially affect portfolio decisions, particularly when higher-order moments are incorporated into the optimization process [1]. This study develops a bootstrap-based robust portfolio optimization framework that jointly accounts for uncertainty in mean, variance, skewness, kurtosis, and portfolio beta within a regime-dependent setting.

The proposed methodology employs stationary block bootstrap resampling to generate empirical distributions of portfolio moments and systematic risk measures [2]. Robustification is achieved through empirical tail quantiles for mean, variance, skewness, and kurtosis together with bootstrap-calibrated beta constraints. To determine the appropriate estimation strategy, covariance homogeneity is assessed using distribution-free permutation tests, while higher-order moment structures are examined through robust multivariate diagnostics based on MCD, MVE, and M-estimation principles [3].

The methodology is evaluated using stocks from the NASDAQ Clean Edge Green Energy Index (CELS) across multiple bull and bear market sub-periods. Empirical evidence indicates substantial heterogeneity in both covariance and higher-order moment structures across market scenarios. Consequently, moment estimation is performed locally within each scenario rather than through pooled estimation across regimes.

The findings demonstrate that robust portfolio construction should account not only for sampling uncertainty but also for structural heterogeneity when selecting an estimation strategy. The study highlights the value of combining bootstrap-based optimization with robust statistical diagnostics in higher-order moment portfolio analysis.

Keywords: *Robust portfolio optimization, higher-order moments, stationary block bootstrap, homogeneity diagnostics, regime heterogeneity, renewable energy stocks*

References

[1] Jondeau, E., & Rockinger, M. (2006). Optimal portfolio allocation under higher moments. European Financial Management, 12(1), 29-55.
[2] Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. Journal of the American Statistical Association, 89(428), 1303-1313.
[3] Hubert, M., Rousseeuw, P. J., & Van Aelst, S. (2008). High-breakdown robust multivariate methods. Statistical Science, 23(1), 92–119.

On inference using M-estimators under contamination

Jens Klooster[1]*, Bent Nielsen[2]
[1]University of Groningen, Nettelbosje 2, 9747 AE, Groningen, The Netherlands
[2]University of Oxford and Nuffield College, Nuffield College, Oxford OX1 1NF, UK
*Corresponding Author Email: j.klooster@rug.nl

We analyze inferential properties of robust location-scale M-estimators under contamination. If we have a clean, good sample from a normal location-scale model, then the sample average is consistent and efficient and delivers inference with good local power properties. But, if large outliers are added to the sample, then the overall average can be far from the location of the good observations. We consider several popular robust M-estimators for location and investigate to what extent they have properties similar to the now infeasible average of the good observations. We show that the Tukey biweight estimator [1] asymptotically can be nearly as good as the infeasible average in terms of consistency, efficiency and local power. In contrast, the median and the Huber estimator [2] will, in general, be bounded, but inconsistent for the location of the good observations.

Keywords: *Robust procedures, mathematical statistics, inference, contamination, asymptotics.*

References

[1] Beaton, A. E. and Tukey, J.W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. Technometrics, 16:147-185.
[2] Huber, P.J. (1964). Robust estimation of a location parameter. Ann. Math. Statist., 35:73-101.

Hybrid Sliding Window Recursive Least Squares for Battery State of Health Prediction

Nuri Berk Ural[1]*, Mustafa Gök[2]
[1]Department of Statistics, Çukurova University, Adana, Türkiye, 01250
[2]Department of Electrical and Electronics Engineering, Çukurova University, Adana, Türkiye, 01250
*Corresponding Author Email: nberkural@gmail.com

Accurate prediction of battery State of Health (SOH) is essential for the safe and efficient operation of lithium-ion battery systems. However, battery degradation is a dynamic process influenced by operating conditions, aging mechanisms, and environmental factors, making reliable SOH estimation challenging. Traditional Recursive Least Squares (RLS) models provide efficient online learning capabilities but may struggle to adapt to local changes in degradation behavior. Sliding window approaches focus on recent observations but do not fully exploit recursive parameter updating.

This study proposes a Hybrid Sliding Window Recursive Least Squares (SW-RLS) framework for battery SOH prediction. The proposed method combines a forgetting factor based RLS estimator with the local learning capability of sliding windows, allowing the model to emphasize recent observations while continuously adapting model parameters to evolving degradation patterns. Experiments were conducted using the NASA lithium-ion battery aging dataset. After data cleaning and quality control procedures, 30 battery cells were included in the analysis. SOH values were calculated using normalized capacity measurements, while previous cycle SOH, temperature, internal resistance, and charge transfer resistance were used as predictor variables.

The proposed approach was compared with conventional RLS, Sliding Window Regression, Random Forest, and Gradient Boosting models under a rolling prediction framework. Performance was evaluated using RMSE, MAE, MAPE, and R^2 metrics. Friedman rank tests and Wilcoxon signed-rank post-hoc analyses with Bonferroni correction were also used. Results indicate that the Hybrid SW-RLS model improves prediction accuracy compared with conventional RLS and achieves competitive performance relative to machine learning approaches while maintaining low computational complexity. These findings demonstrate the effectiveness of combining recursive estimation with localized learning for battery health monitoring applications.

Keywords: *Batery State of Health, Forgetting Factor, Online Learning, Sliding Window*

References

[1] Plett, G. L. (2011). Recursive approximate weighted total least squares estimation of battery cell total capacity. *Journal of Power Sources*, 196(4), 2319–2331.
[2] Saha, B., & Goebel, K. (2009). Prognostics methods for battery health monitoring using a Bayesian framework. *IEEE Transactions on Instrumentation and Measurement*, 58(2), 291–296.
[3] Moens, V., & Zénon, A. (2019). Learning and forgetting using reinforced Bayesian change detection. *Neural Computation*, 31(6), 1225–1260.

Contributed Session 2 (CS2)

CS2 – Robust Inference for Dependent, Bayesian or Categorical Data
Chair: F. Sevinc Kurnaz (Yıldız Technical University, Turkiye & Case Western Reserve University, United States)

Xin Dang (University of Mississippi, United States)
Functional canonical correlation for discriminant analysis

Tuğba Kapucu (Middle East Technical University, Turkiye)
An ECM-Metropolis Algorithm for Parameter Estimation in Skew Laplace Linear Mixed Models

Mutlu Altuntaş (Sinop University, Turkiye)
Robust anomaly detection via a bayesian autoencoder with heavy-tailed priors

Maral Karbaschi (Alzahra University, Iran)
Uncertainty-aware statistical learning for robust clinical risk analytics

Functional canonical correlation for discriminant analysis

Xin Dang[1]*, Kankanamge Madushanka[2], Yongli Sang[2]
[1] Department of Mathematics, University of Mississippi, University, MS 38677, USA
[2] Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70504, USA *Corresponding Author Email: xdang@olemiss.edu

Canonical correlation analysis (CCA) is a classical tool for quantifying the dependence between two random vectors, $X \in R^p$ and $Y \in R^q$. When X and Y are functional observations, functional canonical correlation analysis (FCCA) extends this framework to characterize the linear association between random functions. In this paper, we develop a novel FCCA framework tailored for discriminant analysis with a categorical response variable Y. The proposed dependence measure is constructed by extending the Gini distance correlation [1] to the functional setting, enabling a rigorous quantification of the association between a functional predictor and a categorical outcome. We establish key theoretical properties of the proposed estimator, including its consistency. Extensive simulation studies demonstrate that the proposed method provides accurate and stable estimation of the underlying dependence structure and exhibits strong performance in discriminant analysis tasks. The practical utility of the approach is further illustrated through a real data application, where a k-sample testing problem is considered.

Keywords: *Canonical correlation analysis; categorical Gini correlation; functional data analysis; multivariate functional data*

References

[1] Dang, X., Nguyen, D., Chen, X. and Zhang, J. (2021). A new Gini correlation between quantitative and qualitative variables. *Scandinavian Journal of Statistics*, 48 (4), 1314-1343.

An ECM-Metropolis Algorithm for Parameter Estimation in Skew Laplace Linear Mixed Models

Tuğba Kapucu[1]*, Fulya Gokalp Yavuz[1,2]
[1]Department of Statistics, Middle East Technical University, Ankara, Turkiye, 06800
[2]The Data Mine, Purdue University, West Lafayette, Indiana, USA, 47906
*Corresponding Author Email: tkapucu@purdue.edu

Modeling both within and between-subject variability in clustered and repeated-measures data allows more accurate parameter estimation by reducing error. Linear mixed models (LMMs) [1] are widely used for this purpose, but their normality assumptions are violated when the data include outliers, skewness, or heavy-tailed structures, leading to less reliable estimates. To better accommodate such features, we introduce the skew Laplace linear mixed model (SL-LMM).

In the classical formulation of LMM, within-subject errors and random effects are typically assumed to follow a normal distribution. In our extension, within-subject errors follow a Laplace distribution, providing robustness against heavy tails, while the random effects follow Arslan's [2] multivariate skew Laplace (MSL) distribution, simultaneously addressing both heavy tails and skewness and offering a more tractable density function than other multivariate skew distributions.

Expressing the skew Laplace distribution as a normal mean–variance mixture introduces a latent scaling variable and allows us to reformulate the proposed model in a hierarchical form, which facilitates the E-step of the ECM algorithm. However, in this setting the conditional expectations of the latent scaling variable required in the E-step are analytically intractable. To overcome this limitation, we combine the ECM algorithm [3] with Metropolis sampling [4] in the E-step. We use Metropolis updates for the latent scaling variable to generate Markov chain Monte Carlo (MCMC) samples that approximate these intractable conditional expectations. Convergence diagnostics confirm adequate mixing of the MCMC chains. Given these approximated expectations, the CM-steps then update the model parameters in closed form, yielding a stable and tractable ECM–Metropolis procedure for the proposed SL-LMM.

To assess the performance of the proposed SL-LMM, 500 Monte Carlo data sets are generated from a specific instance of LMM with a random intercept and a random slope, following the simulation design of Ho & Lin [5]. Results are compared against the classical LMM and the Laplace linear mixed model (L-LMM) [6]. The simulation results show that the proposed SL-LMM yields smaller mean absolute bias and root mean squared error than the competing methods, demonstrating its superiority over the classical LMM and the L-LMM. Together with the multivariate skew Laplace specification and the ECM-Metropolis algorithm provides a flexible and robust framework for parameter estimation in LMMs for skewed and heavy-tailed data sets.

Keywords: *Linear mixed models, Skew Laplace distribution, ECM* algorithm*, Markov chain Monte Carlo, Metropolis algorithm, Robust statistics*

References

[1] Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. Biometrics, 38(4), 963–974.
[2] Arslan, O. (2010). An alternative multivariate skew Laplace distribution: Properties and estimation. Statistical Papers, 51(4), 865–887.
[3] Meng, X.-L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. Biometrika, 80(2), 267–278.
[4] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. The Journal of Chemical Physics, 21(6), 1087–1092.
[5] Ho, H. J. and Lin, T. I. (2010). Robust linear mixed models using the skew t distribution with application to schizophrenia data. Biometrical Journal, 52(4), 449–469.
[6] Yavuz, F. G. and Arslan, O. (2018). Linear mixed model with Laplace distribution (LLMM). Statistical Papers, 59(1), 271–289.

Robust Anomaly Detection via a Bayesian Autoencoder with Heavy-Tailed Priors

Mutlu Altuntaş[1]*
[1]Department of Data Science and Analytics, Faculty of Arts and Sciences, Sinop University, Sinop, Turkiye, 57000
*Corresponding Author Email: maltuntas@sinop.edu.tr

Anomaly detection remains a central problem in statistics and machine learning, as anomalous observations may substantially affect model estimation, prediction accuracy, and decision-making. In recent years, autoencoders (AEs) and variational autoencoders (VAEs) have become popular tools for unsupervised anomaly detection because they learn low-dimensional representations of complex data and identify unusual observations through reconstruction errors. However, these methods are typically based on Gaussian assumptions and may be sensitive to contaminated observations and heavy-tailed data. Although Bayesian autoencoders provide a natural framework for incorporating uncertainty into deep learning models, most existing approaches still rely on Gaussian priors and likelihood functions. In this study, a Robust Bayesian Autoencoder (RBAE) is proposed for anomaly detection under contaminated data conditions. The proposed framework assigns Student-t priors to the encoder and decoder weights and employs a Student-t likelihood for the reconstruction model. This heavy-tailed specification aims to reduce the influence of extreme observations and improve robustness during learning. Unlike conventional autoencoder-based approaches that mainly depend on reconstruction errors, the proposed method performs anomaly detection through posterior predictive distributions. Observations with low posterior predictive probabilities are flagged as anomalies, directly incorporating uncertainty into the detection process. The proposed framework combines robust Bayesian modelling and deep representation learning within a unified anomaly detection framework. Most existing Bayesian autoencoder approaches rely on Gaussian priors and likelihood functions, which may limit robustness in the presence of contaminated or heavy-tailed data. The proposed approach may therefore provide a useful alternative for anomaly detection in contaminated or heavy-tailed settings.

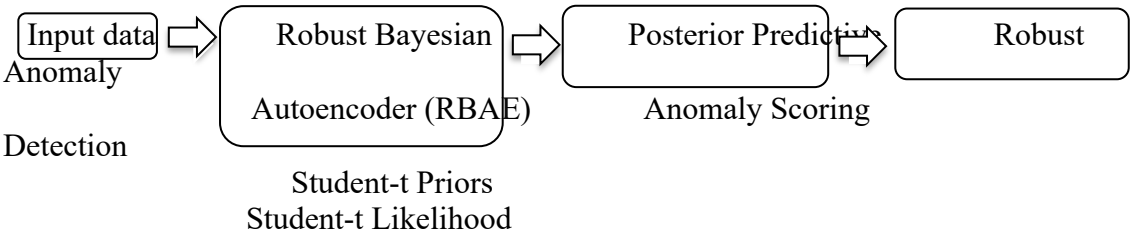


Figure: Conceptual framework of the proposed Robust Bayesian Autoencoder (RBAE)

Keywords: *Robust statistics, bayesian deep learning, anomaly detection, heavy-tailed priors*

References

[1] Yong, B.X. and Brintrup, A. (2022). Bayesian autoencoders with uncertainty quantification: Towards trustworthy anomaly detection. Expert Systems with Applications, 209, 118196.

[2] Egels, P. and Castillo, I. (2025). Posterior and variational inference for deep neural networks with heavy-tailed weights. Journal of Machine Learning Research, 26, 1-58.

Uncertainty-Aware Statistical Learning for Robust Clinical Risk Analytics

Maral Karbaschi[1]*
[1]Department of Mathematical Statistics, Alzahra University, Tehran, Iran
*Corresponding Author Email: maral.karbaschi@gmail.com

Healthcare analytics increasingly relies on machine learning models to support risk assessment, early detection, and data-driven decision-making. However, real-world clinical data are often affected by class imbalance, heterogeneous patient profiles, missing or noisy observations, and changes in data distribution over time. These challenges can reduce model reliability and limit the practical usefulness of predictive systems in healthcare environments. Learning under imbalanced data conditions has been recognized as a major challenge in predictive modeling and statistical learning research [1].

This work investigates an uncertainty-aware statistical learning framework for robust clinical risk analytics. The main objective is to improve the stability and interpretability of machine learning models when applied to imbalanced and evolving healthcare datasets. The framework combines statistical learning principles with adaptive model evaluation strategies to analyze how predictive performance changes under rare-event conditions and distributional shifts.

Rather than focusing only on predictive accuracy, the study emphasizes deployment-oriented evaluation criteria, including sensitivity to minority classes, robustness under data variation, and consistency of model behavior across different patient subgroups. This perspective is particularly important in healthcare applications, where unreliable predictions for rare but clinically important outcomes may have significant consequences. Recent studies have also highlighted the importance of uncertainty quantification in healthcare-oriented machine learning systems [2].

The proposed direction highlights the need for closer integration between statistical reasoning, machine learning, and practical healthcare analytics. By focusing on uncertainty, imbalance, and robustness, this work contributes to the development of more reliable analytical frameworks for clinical decision-support systems.

Keywords: *statistical learning, clinical risk analytics, imbalanced data, uncertainty-aware modeling, healthcare AI, robust prediction*

References

[1] He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263–1284.
[2] Begoli, E., Bhattacharya, T., & Kusnezov, D. (2019). The need for uncertainty quantification in machine-assisted medical decision making. Nature Machine Intelligence, 1(1), 20–23.

Contributed Session 3 (CS3)

CS3 – Robust Medians, Breakdown Points or Outlier Detection
Chair: Stanislav Nagy (Charles University, Czech Republic)

Marco Avella Medina (Columbia University, United States)
The threshold breakdown point

Leopold Micheler (TU Wien, Austria)
Robust and explainable outlier detection for random surfaces

Erik Mendroš (Charles University, Czech Republic)
Breakdown points of simplicial and k-hull medians

Stanislav Nagy (Charles University, Czech Republic)
When is the Tukey median unique?

The threshold breakdown point

Tianjun Ke and Marco Avella Medina*
Department of Statistics, Columbia University, New York, USA
*Corresponding Author Email: marco.avella@columbia.edu

We introduce a novel approach to finite sample robustness that avoids the pessimism of traditional breakdown analyses. We define the *threshold breakdown point*, the smallest contamination fraction needed to induce a prescribed deviation, and the finite sample k-sensitivity, the worst-case deviation that an estimator can incur after k observations are contaminated. We derive these measures for commonly used M-estimators, their standard errors and related test statistics. This allows us to extend the decision breakdown point of Zhang (1996) to obtain general breakdown characterizations for hypothesis testing, and show how these notions correspond to finite sample counterparts of the power and level breakdown functions of He et al (1990). We complement our work with an inferential framework for the threshold breakdown and k-sensitivity that yields consistency and asymptotic normality results, as well as a valid multiplier bootstrap for uncertainty quantification. We illustrate the practical utility of our methods in various numerical examples and an application to a two-sample testing problem for a blood pressure dataset.

Keywords: *breakdown point, M-estimators, hypothesis testing, sensitivity curve, bootstrap inference*

References

[1] Zhang, J. (1996). The sample breakdown points of tests. *Journal of Statistical Planning and Inference*, *52*(2), 161-181

[2] He, X., Simpson, D. G., & Portnoy, S. L. (1990). Breakdown robustness of tests. *Journal of the American Statistical Association*, *85*(410), 446-452.

[3] Donoho, D. L., & Huber, P. J. (1983). The notion of breakdown point. *A Festschrift for Erich L. Lehmann*, *157184*, 1004.

[4] Ke, T., & Avella Medina, M. (2026). The threshold breakdown point. *arXiv preprint arXiv:2605.04317*.

Robust and explainable outlier detection for random surfaces

Leopold Micheler[1]*, Marcus Mayrhofer[1], Una Radojicic[1], Peter Filzmoser[1]
[1]Institute of Statistics and Mathematical Methods in Economics, TU Wien, Wiedner Hauptstraße 8-10, 1040 Vienna, Austria
*Corresponding Author Email: leopold.micheler@tuwien.ac.at

This contribution introduces a new distance measure and a robust method for covariance estimation of random surfaces with a separable covariance structure. The approach links separable random surfaces to the matrix-variate distribution of their basis function representations, which provides a convenient and principled framework for estimation.

Building on this connection, we develop a robust procedure based on the Matrix Minimum Covariance Determinant (MMCD) [1] estimator, combined with a truncated functional Mahalanobis semi-distance to estimate mean and covariance functions. This formulation is designed to remain stable under contamination and to provide reliable distance-based inference for functional data.

To improve interpretability, we extend the Shapley value [2] methodology to the functional setting. This allows us to decompose the proposed distance measure and thus functional outlyingness into contributions from specific spatial and temporal regions. We further introduce a novel formulation for computing these functional Shapley values while preserving their fundamental axiomatic properties.

The resulting framework integrates matrix-variate modeling, robust distance measures, and interpretable decomposition techniques into a unified approach for analyzing random surfaces. We provide theoretical justification alongside empirical results on real-world datasets, demonstrating strong performance in classification and robust outlier detection tasks.

Keywords: *Functional data analysis, Robustness, Classification, Outlier detection*

References

[1] M. Mayrhofer, U. Radojicic and P. Filzmoser (2025). Robust covariance estimation and explainable outlier detection for matrix-valued data. Technometrics, 67(3), 516–530.
[2] L. S. Shapley (1953). A Value for n-Person Games. Contributions to the Theory of Games, Volume II, 307–318.

ICORS2026
20-24 JULY 2026
Istanbul

Breakdown points of simplicial and *k*-hull medians

Erik Mendroš[1]*, Stanislav Nagy[1],
[1]Department of Probability and Mathematical Statistics, Charles University, Prague, Czech Republic
*erikmendros@gmail.com

We analyze the robustness of the bivariate simplicial median by determining its exact breakdown point. First, we refine existing inequalities [1] for the breakdown point with respect to a fixed data set. By combining these tighter bounds with the so called first selection lemma from geometry, we raise the known lower bound [2] for the worst-case asymptotic breakdown point from 0.064 to 0.080. Furthermore, we construct a specific point configuration that can be broken precisely at this contamination level, proving that 0.080 is the exact worst-case breakdown point.

We extend this analysis to the more general *k*-hull depth [3] and its associated medians; for $k = 3$, the bivariate *k*-hull depth reduces to the bivariate simplicial depth. The obtained breakdown points for the first few values of *k* are evaluated in Table 1. It is observed that these values increase monotonically with *k* and converge to 1/3 from below. Consequently, larger values of *k* appear to induce more robust medians.

This limiting value coincides with the breakdown point of the bivariate halfspace median. We provide an explanation for this phenomenon: suitably scaled *k*-hull depth can be represented as an L^{k-1} norm of differences of halfplane probabilities from one-half, whereas halfspace depth is a function of the supremum norm of these same differences. Thus, as *k* goes to infinity, the rescaled *k*-hull depth of a point converges to its halfspace depth. Based on these representations, we additionally derive sharp bounds relating *k*-hull depth to halfspace depth at any depth level.

Finally, because the worst-case breakdown is driven by highly degenerate contaminations, we examine more regular settings, including contaminations in general position and samples drawn from angularly symmetric distributions.

Table 1: Worst-case limiting breakdown points of *k*-hull medians for different values of *k*.

$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$
0.080	0.136	0.174	0.201	0.220	0.235	0.246

Keywords: *Robust estimation, breakdown point, data depth, simplicial median, halfspace depth*

References

[1] Chen, Z. Q. (1995). Bounds for the breakdown point of the simplicial median. Journal of Multivariate Analysis, 55(1), 1-13.

[2] Nagy, S. (2023). Simplicial depth and its median: Selected properties and limitations. Statistical Analysis and Data Mining: The ASA Data Science Journal, 16(4), 374-390.
[3] Mendroš, E., & Nagy, S. (2025). *k*-hull depth: in between simplicial and halfspace depth. Statistical Papers, 66(4), 88.

When is the Tukey median unique?

Stanislav Nagy[1]*, Dušan Pokorný[1]
[1]Charles University, Faculty of Mathematics and Physics, Sokolovská 83, 186 75 Praha, Czechia0
*Corresponding Author Email: nagy@karlin.mff.cuni.cz

The Tukey median (also called the halfspace median, the Simpson point, or the Hotelling point) is an important robust nonparametric location estimator for multivariate data [1]. In the plane, it is well known [2] that the Tukey median is unique if the underlying measure has a Lebesgue density and connected support. Surprisingly, in dimension $d>2$, we show that the same claim is no longer true and formulate a set of conditions that guarantee the uniqueness of the Tukey median.

Keywords: *halfspace depth, halfspace median, Tukey median, Dupin theorem, uniqueness*

References

[1] Donoho, D. L., and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness, Ann. Statist., 20, 1803–1827.
[2] Massé, J. C., and Theodorescu, R. (1994). Halfplane trimming for bivariate distributions, J. Multivariate Anal., 41(2), 81 (2002), no. 2, 188–202.

Contributed Session 4 (CS4)

CS4 – Robust Modeling and Monitoring for Dependent Data
Chair: Onur Toka (Hacettepe University, Turkiye)

Ziling Ma (King Abdullah University of Science and Technology, Saudi Arabia)
Forecasting Multivariate Time Series under Predictive Heterogeneity: A Validation-Driven Clustering Framework

Talha Arslan (Van Yüzüncü Yıl University, Turkiye)
The unit-t2 distribution: a parametric approach to median regression on the unit interval

Iris Aragon Mladosich (KU Leuven, Belgium)
Robust XGBoosting for Regression

Forecasting Multivariate Time Series under Predictive Heterogeneity: AValidation-Driven Clustering Framework

Ziling Ma[1]*, Ángel López-Oriona[1], Hernando Ombao[1], Ying Sun[1]
[1]Statistics Program, CEMSE, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
*Corresponding Author Email: ziling.ma@kaust.edu.sa

We study adaptive pooling under predictive heterogeneity in high-dimensional multivariate time series forecasting, where global models improve statistical efficiency but may fail to capture heterogeneous predictive structure, while naive specialization can induce negative transfer. We formulate adaptive pooling as a statistical decision problem and propose a validation-driven framework that determines when and how specialization should be applied. Rather than grouping series based on representation similarity, we define partitions through out-of-sample predictive performance, thereby aligning data organization with predictive risk, defined as expected out-of-sample loss and approximated via validation error. Cluster assignments are iteratively updated using validation losses for both point (Huber) and probabilistic (pinball) forecasting, improving robustness to heavy-tailed errors and local anomalies. To ensure reliability, we introduce a leakage-free fallback mechanism that reverts to a global model whenever specialization fails to improve validation performance, yielding a no-regret strategy under a strict training–validation–test protocol. Experiments on large-scale traffic datasets demonstrate consistent improvements over strong baselines while avoiding degradation when heterogeneity is weak. Overall, the proposed framework provides a principled and practically reliable approach to adaptive pooling in high-dimensional forecasting problems.

Keywords: *Adaptive pooling, Negative transfer, Validation-driven learning, Robust forecasting, Model selection*

References

[1] López-Oriona, Á., Montero-Manso, P., & Vilar, J. A. (2025). Time series clustering based on prediction accuracy of global forecasting models. Knowledge-Based Systems, 113649.
[2] Montero-Manso, Pablo, and Rob J. Hyndman. "Principles and algorithms for forecasting groups of time series: Locality and globality." International Journal of Forecasting 37, no. 4 (2021): 1632-1653.
[3] Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. International journal of forecasting, 36(3), 1181-1191.

The unit-t_2 distribution: a parametric approach to median regression on the unit interval

Talha Arslan[1]*
[1]Department of Econometrics, Van Yüzüncü Yıl University, Van, Türkiye, 65080.
*Corresponding Author Email: mstalhaarslan@yyu.edu.tr

Distributions having support on the unit interval (0,1) play an essential role in modeling proportional data. Although the beta and Kumaraswamy distributions are well-known, researchers have proposed new unit distributions to fill gaps in the related literature; see Afuecheta et al. [1]. Recently, Arslan [2] introduced a versatile family of unit distributions based on the transformation

$$Y = [1 + \exp(-X)]^{-\frac{1}{\alpha}}$$

where X is a random variable defined on $\Re$ and α is a shape parameter. This transformation allows heavy-tailed distributions defined on $\Re$ to be moved to the interval (0,1). Consequently, it yields unit distributions that serve as heavy-tailed alternatives to the beta and Kumaraswamy models, offering robustness against data points near the boundaries. In this context, Arslan and Yu [3] developed a unit-Cauchy quantile regression model as a robust alternative to classical unit models.

In this study, the unit-t2 (UT2) distribution – derived from the Student's *t*-distribution with 2 degrees of freedom – is introduced by following the methodology in Arslan [2], along with its median-based parametrization UT2(τ, σ). Furthermore, a median-conditional regression model is developed, in which covariates are linked to the conditional median via a strictly monotonic, twice-differentiable link function. The maximum likelihood estimators of the UT2 model parameters have bounded influence functions, and hence the resulting estimators are robust against outliers.

In the application section, the results show that the UT2 model has a higher log-likelihood (34.74) value than the beta (33.80), Kumaraswamy (31.98), and unit-Cauchy (31.17) models. It should also be noted that although the Kumaraswamy model failed to yield statistical significance for one covariate, the UT2 model yielded consistent and significant estimates. Therefore, it is concluded that the UT2 model effectively optimizes the robustness-efficiency trade-off; i.e., it offers a strategic balance between the extreme robustness of unit-Cauchy and the sensitivity of classical unit models, providing superior modeling performance.

Keywords: *Beta regression, Kumaraswamy regression, Maximum likelihood, Student's t-distribution, Unit distributions.*

References

[1] Afuecheta, E, Okorie, I.E., Jallow, H., and Nadarajah, S. (2025). A review of unit continuous probability distributions. AIMS Mathematics, 10(11), 25939–26057.
[2] Arslan, T. (2023). A new family of unit-distributions: definition, properties, and applications. TWMS J. App. and Eng. Math, 13(2), 782–791.
[3] Arslan, T. and Yu, K. (2025). The unit-Cauchy quantile regression model with variates observed on (0,1): percentages, proportions, and fractions. Hacet. J. Math. Stat., 54(2), 633–655.

Robust XGBoosting for Regression

Iris Aragón Mladosich[1]*, Christophe Croux[1]
[1]KU Leuven, Naamsestraat 69, 3300 Leuven, Belgium
*Email: iris.aragonmladosich@kuleuven.be

XGBoost [1] is a very popular and powerful method for prediction. It iteratively fits simple decision trees to the residuals of the previous step. An efficient and scalable implementation is available. The standard loss function for XGBoost is the quadratic function, but a Huber loss can also be used.

In this paper, we study the robustness of XGBoost and show that its performance can be affected by vertical outliers and leverage points. To address this, we explore alternative loss functions, based on M-, S-, and tau-estimators from robust regression. Our approach builds on ideas from "Robust Boosting for Regression Problems" [2], which introduces robust loss functions in gradient boosting. Our results show that a two-step procedure, referred to as MM-XGBoost, provides the best trade-off between robustness and prediction accuracy.

Keywords: *Robustness, XGBoost, Regression*

References

[1] Chen, T. and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.
[2] Ju, X. and Salibi´an-Barrera, M. (2021). Robust boosting for regression problems. Computational Statistics & Data Analysis, 153, Article 107065.

Contributed Session 5 (CS5)

CS5 – Robust Methods for Sampling and Forecasting
Chair: Yetkin Tuaç (Ankara University, Turkiye)

Özge Gürer (Ankara University, Turkiye) (Virtual)
Ranked Set Sampling based ANOVA under Non-Normality

Özge Gürer (Ankara University, Turkiye) (Virtual)
Robust Tests for One-Sample and Two-Sample Problems under Ranked Set Sampling: An Application to Real Data

Antonio Panico (University of Parma, Italy)
Sequential Conformal Prediction for Electricity Price Forecasting

Ranked Set Sampling based ANOVA under Non-Normality

Özge Gürer[1]*, Birdal Şenoğlu[1]
[1]Ankara University, Faculty of Science, Department of Statistics, Ankara, Turkiye, 06100
*Corresponding Author Email: otanju@ankara.edu.tr

In the classical one-way and two-way analysis of variance (ANOVA) procedures, distribution of the error terms is assumed to be normal and the observations are obtained via simple random sampling (SRS). However, in this study, logistic error distribution is considered, since its symmetric form and heavier tails make it a widely used and plausible alternative to the normal distribution. Also, ranked set sampling (RSS), originally proposed by McIntyre [1], is used as the sampling scheme due to its potential to yield more efficient estimators than SRS without increasing the sample size. Robust and efficient estimators of the unknown model parameters are obtained through modified maximum likelihood (MML) methodology, see Tiku [2], and then test statistics based on them are proposed. The performances of the proposed tests, shortly referred to as RSS based tests, are compared with those of the classical SRS based tests in terms of their sizes and powers through a comprehensive Monte Carlo simulation study. Particular emphasis is also placed on evaluating the robustness properties of RSS based tests in the presence of data anomalies and model misspecifications. In addition, the impact of imperfect ranking, which is frequently encountered in practical RSS applications, is investigated. Simulation results indicate that RSS based tests are more powerful than their SRS based counterparts besides having smaller type I error rates in most of the scenarios. Moreover, the proposed tests possess strong robustness properties to the deviations from the assumed model, and their performance advantage persists even under imperfect ranking conditions. A real data set is also analyzed to demonstrate the implementation of the proposed methodology.

Keywords: *ANOVA, ranked set sampling, logistic distribution, modified likelihood, Monte Carlo simulation, robustness*

References

[1] McIntyre, G.A. (1952). A method for unbiased selective sampling, using ranked sets. Australian journal of agricultural research, 3(4), 385–390.
[2] Tiku, M.L. (1967). Estimating the mean and standard deviation from a censored normal sample. Biometrika, 54(1-2), 155–165.

Robust Tests for One-Sample and Two-Sample Problems under Ranked Set Sampling: An Application to Real Data

Özge Gürer[1]*, Birdal Şenoğlu[1]
[1]Ankara University, Faculty of Science, Department of Statistics, Ankara, Turkiye, 06100
*Corresponding Author Email: otanju@ankara.edu.tr

In this study, ranked set sampling (RSS) based one-sample and two-sample tests are proposed as more efficient and robust alternatives to the corresponding tests existing in literature, see McIntyre [1] in the context of RSS. The error terms are assumed to follow logistic distribution, which is a reasonable alternative to the well-known normal distribution due to its symmetric shape. Logistic distribution has heavier tails than normal distribution and therefore is widely used for modelling data sets having outliers. The unknown model parameters are estimated by using the modified maximum likelihood (MML) methodology proposed by Tiku [2], which is based on the idea of linearization of the nonlinear terms in the likelihood equations and therefore yields closed-forms solutions. MML estimators are asymptotically equivalent to the maximum likelihood (ML) estimators besides being robust. Then, RSS based robust one-sample and two-sample tests based on these estimators are defined. Their performances are evaluated in terms of type I error and power via an extensive Monte Carlo simulation study. Simulation results demonstrate that proposed tests do not only have significantly higher power than their competitors in general but also maintain lower type I error rates. Robustness properties of the RSS based robust one-sample and two-sample tests are also studied to understand how they perform when there exist plausible departures from the assumed model. Based on numerical outcomes, the proposed tests are shown to exhibit greater robustness than their competitors against various violations of the model assumptions, including the presence of outliers, mixture and contamination models, as well as misspecification of the error distribution. Furthermore, the effect of imperfect ranking is investigated, and the findings reveal that the powers of the proposed tests are still remarkably high even in cases of severe errors in ranking when the set size is relatively small. Finally, a real data set is analyzed by using the proposed methodology for illustrative purposes.

Keywords: *t-test, ranked set sampling, logistic distribution, modified likelihood, Monte Carlo simulation, robustness*

References

[1] McIntyre, G.A. (1952). A method for unbiased selective sampling, using ranked sets. Australian journal of agricultural research, 3(4), 385–390.
[2] Tiku, M.L. (1967). Estimating the mean and standard deviation from a censored normal sample. Biometrika, 54(1-2), 155–165.

Sequential Conformal Prediction for Electricity Price Forecasting

Antonio Panico[1]*, Luigi Grossi[1],
[1] Department of Engineering for Industrial Systems and Technologies, University of Parma
*Corresponding Author Email: antonio.panico@unipr.it

The integration of Renewable Energy Sources makes electricity prices highly volatile, rendering deterministic point forecasts inadequate [1], [2]. In this context, uncertainty quantification of predictive models is crucial for decision-making problems [3], [4]. Conformal prediction (CP) offers a theoretically grounded framework for generating valid prediction intervals [5]. However, Inductive Conformal Prediction (ICP) fails in time-series because it relies on the assumption of exchangeability [6]. While several methods have emerged to adapt CP to non-exchangeable sequences, [7], [8], [9], Sequential Predictive Conformal Inference (SPCI) [10] has surfaced as a leading edge-case solution. Despite its strengths, the choice of the underlying "driver" model within the SPCI framework remains under-explored. To bridge this gap, we propose integrating a rolling Dynamic Multiple Quantile (DMQ) model [11]. This model not only calibrates residuals to react immediately to price shocks but also inherently prevents quantile crossing to ensure logical consistency. Furthermore, it maintains high performance even in the presence of market outliers, providing a robust foundation for uncertainty quantification. Evaluated on the German EPEX market, our proposed framework consistently outperforms standard benchmarks. Notably, it is the only model to uniquely pass the Kupiec test across all confidence levels, demonstrating superior calibration and reliability.

Keywords: *Electricity, Price Forecasting, Conformal Prediction, Robust Forecasting, Probabilistic Forecasting*

References

[1] J. Nowotarski and R. Weron, “Recent advances in electricity price forecasting: A review of probabilistic forecasting,” Jan. 01, 2018, *Elsevier Ltd*. doi: 10.1016/j.rser.2017.05.234.

[2] J. Lago, G. Marcjasz, B. De Schutter, and R. Weron, “Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark,” Jul. 01, 2021, *Elsevier Ltd*. doi: 10.1016/j.apenergy.2021.116983.

[3] A. Brusaferri, A. Ballarino, L. Grossi, and F. Laurini, “On-line conformalized neural networks ensembles for probabilistic forecasting of day-ahead electricity prices,” *Appl. Energy*, vol. 398, Nov. 2025, doi: 10.1016/j.apenergy.2025.126412.

[4] C. O’Connor, M. Bahloul, R. Rossi, S. Prestwich, and A. Visentin, “Conformal Prediction for electricity price forecasting in the day-ahead and real-time balancing market,” *Energy and AI*, vol. 21, Sep. 2025, doi: 10.1016/j.egyai.2025.100571.

[5] G. ; Zeni, M. ; Fontana, F. Vantini, G. Zeni, M. Fontana, and S. Vantini, “Conformal Prediction: a Unified Review of Theory and New Challenges.” [Online]. Available: http://mox.polimi.it

[6] R. F. Barber, E. J. Candes, A. Ramdas, and R. J. Tibshirani, “Conformal prediction beyond exchangeability,” Mar. 2023, [Online]. Available: http://arxiv.org/abs/2202.13415

[7] C. Kath and F. Ziel, “Conformal prediction interval estimation and applications to day-ahead and intraday power markets,” *Int. J. Forecast.*, vol. 37, no. 2, pp. 777–799, Apr. 2021, doi: 10.1016/j.ijforecast.2020.09.006.

[8] M. Zaffran, A. Dieuleveut, O. Féron, Y. Goude, and J. Josse, “Adaptive Conformal Predictions for Time Series,” Feb. 2022, [Online]. Available: http://arxiv.org/abs/2202.07282

[9] C. Xu and Y. Xie, “Conformal Prediction for Time Series,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 10, pp. 11575–11587, Oct. 2023, doi: 10.1109/TPAMI.2023.3272339.

[10] C. Xu and Y. Xie, “Sequential Predictive Conformal Inference for Time Series,” in *Proceedings of Machine Learning Research*, 2023, pp. 38707 – 38727. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85174420062&partnerID=40&md5=f9e603c71ef4ef191053ea31d7b529b1

[11] L. Catania and A. Luati, “Semiparametric modeling of multiple quantiles,” *J. Econom.*, vol. 237, no. 2, Dec. 2023, doi: 10.1016/j.jeconom.2022.11.002.

Contributed Session 6 (CS6)

CS6 – Robust Methods for Non-Standard and Complex Data
Chair: Maria-Pia Victoria-Feser (University of Bologna, Italy)

Leonardo Leone (Télécom Paris IPP, France)
Massive parallelization of projection-based depths

Kelly Ramsay (York University, Canada)
Robust smoothing splines with discontinuities

Peter Ruckdeschel (Carl Ossietzky University of Oldenburg, Germany)
Robustness Problems in Relaxed and Entropy Regularized Portfolio Optimization

Maria-Pia Victoria-Feser (University of Bologna, Italy)
Simulation-Based Robust Inference for Response Misclassification Errors with Self-Reported Responses

Massive parallelization of projection-based depths

Leonardo Leone[1,2]*, Pavlo Mozharovskyi[1], David Bounie[2]
[1]LTCI, Télécom Paris, Institut Polytechnique de Paris, Paris, France
[2]CREST, CNRS, Télécom Paris, Institut Polytechnique de Paris, Paris, France
*Corresponding Author Email: leonardo.leone@telecom-paris.fr

Providing a statistically-meaningful center-outward ordering for a data set constitutes a powerful tool of analysis and inference over a range of fields, such as fraud detection being one of them. In the multivariate setting, due to complex character of contemporary data sets, assumptions on the data-generating process should be avoided as much as possible, while maintaining high level of data description is mandatory. While absence of assumptions limits statistical modelling methodology, absence of natural ordering in the Euclidean space impedes univariate tools. Being a non-parametric and robust technique that, in an agnostic way, generalizes distribution function and quantiles to higher dimensions, statistical data depth function comes as a remedy for multivariate anomaly detection.

Despite numerous recent advances in the field of data depth, the computational complexity and time consumption are often returned as critics. In this article, projection depth notions are studied as a fast and efficient anomaly detection method in the multivariate context. Due to the nature of the projection depth algorithm, the growth in big data technologies and hardware (such as multiple-kernel processors and graphics processing units) availability, computation can be optimized by parallelizing its non-sequential constituent without precision loss.

The article introduces a novel methodology for the massive parallelization of projection-based depths, addressing the computational challenges of data depth in high-dimensional spaces. We propose an algorithmic framework based on Refined Random Search (RRS) and demonstrate significant speedup (up to 7,000 times faster) on GPUs. Empirical results on synthetic data show improved precision and reduced runtime, making the method suitable for large-scale applications. The RRS algorithm (and other depth functions) are available in the Python-library data-depth with ready-to-use tools to implement and to build upon this work.

ICORS2026
20-24 JULY 2026
Istanbul

Robust smoothing splines with discontinuities

Kelly Ramsay[1]*, Jairo Diaz-Rodriguez[1], Hanna Jankowski[1]
[1]York University, 5700 Keele Street, Toronto, Ontaro, Canada, M3J 3K1
*Corresponding Author Email: kramsay2@yorku.ca

Classical smoothing splines enforce global smoothness of the fitted function. In many applications, however, the underlying signal exhibits discontinuities, for which standard smoothing splines perform poorly. In addition, real data often contains outliers or corrupted points, which can cause the estimator to perform poorly. We present a robust smoothing spline method that relaxes the global smoothness assumption, allowing for the signal to exhibit discontinuities. We retain the finite-dimensional structure and, to some extent, the computational tractability of classical smoothing splines. The approach combines classical smoothing splines with change-point methods and modern interpretations of Huber loss.
Keywords: *splines, Huber loss, discontinuity, nonparametric, functional data*

Robustness Problems in Relaxed and Entropy Regularized Portfolio Optimization

Peter Ruckdeschel[1]*, Sascha Desmettre[2]
[1]Institute for Mathematics, Carl von Ossietzky University Oldenburg, Ammerlaender Heerstrasse 114 – 118, 26111 Oldenburg, Germany,
[2]Institute of Financial Mathematics and Applied Number Theory, Johannes Kepler University Linz, Science Park 2, 4040 Linz, Austria
*Corresponding Author Email: peter.ruckdeschel@uol.de

Recently, reinforcement learning (RL) approaches have gained attraction in continuous time portfolio optimization, see, e.g. [1].

To do so, the original Merton problem of Stochastic Control [2,3] is relaxed by enlarging the set of possible strategies beyond those adapted to the filtration generated by the observed stock prices, allowing for "randomized" strategies where, instead of finding an optimal portfolio process adapted to the filtration generated by the stock prices, one instead searches for an optimal probability density according to which the portfolios are generated.

This relaxation opens the door for new, continuous time reinforcement learning strategies called *q learning*, see [4], which simultaneously learn the value function and and the (infinitesimal, gene-ralized) Hamiltonian function, called *q function*.

We add a Statistics, and more specifically a Robust Statistics, perspective to these approaches.

Using general Delta method approaches we show that for unknown expected log returns (for the drift) and their volatilities (for the diffusion), the classical Merton approach is unsatisfactory in the sense that learning these unknowns from data has too slow learning rates to be attractive, hence calling for regularization.

Such regularizations have successfully been proposed in, e.g. [5,6], using suitable shrinkage strategies to regularize covariances and expectations of the log returns, which of course could be done robustly as in [6].

In the cited RL approaches, one instead regularizes the problem by imposing an entropy condition to the proposal densities in the relaxed setting. q learning tackles the optimization problem by suitably parametrizing the value function and the Hamiltonian giving some low-dimensional parameter θ and learns θ by some stochastic gradient approach.

Compared to other RL approaches, the novelty in q learning lies in the fact that instead of (unre-strictedly) minimizing the (squared) temporal difference (TD) of the reward function (as in Q learning), one imposes a martingale orthogonality side condition.

Formulated in the language of Robust statistics, q learning produces M regression equations for estimating θ in a nonlinear least squares fit, where one imposes the side condition that the resulting influence functions be conditionally centered.

Cast in this setting, it becomes obvious that using the parametrized value and Hamiltonian functions unchanged in q-learning, one may run into unbounded influence functions and hence into vul-nerabilities by (outlying) log returns.

We propose instead to use off-the-shelf weighting strategies from robust nonlinear regression, see [8,9], and thereby also to enhance diagnostics for the existing q learning approaches by correspond-ing robust regression diagnostics [10,11].

We present first findings as to how effective these robustness enhancements work on original DAX 40 data and in the presence of artificial outliers, giving a complementary approach to [12].

Keywords: *Reinforcement Learning, q Learning, Merton Problem, Portfolio Optimization, Nonlinear Regression, M equations*

References

[1] Wang, H. and Zhou, X.Y. (2020). Continuous-Time Mean-Variance Portfolio Selection: A Reinforcement Learning Framework. *Mathematical Finance*, **30**, 1273-1308. DOI: **10.1111/mafi.12281**

[2] Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, **3**(4), 373–413. DOI: **10.1016/0022-0531(71)2990038-X**

[3] Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, **51**(3), 247–257. DOI: **10.2307/1926560**

[4] Jia, Y., and Zhou, X. Y. (2023). q-Learning in Continuous Time. *Journal of Machine Learning Research*, **24**(161), 1-61. DOI: **10.5555/3648699.3648860**

[5] Ledoit O. and Wolf M. (2003). Improved Estimation of the Covariance Matrix of Stock Returns With an Application to Portfolio Selection. *Journal of Empirical Finance*, **10**(5), 603-621. DOI: **10.1016/S0927-5398(03)00007-0**

[6] Bodnar, T., Okhrin, Y., and Parolya, N. (2022). Optimal shrinkage-based portfolio selection in high dimensions. *Journal of Business & Economic Statistics*, **41**(1), 140-156. DOI: **10.1080/07350015.2021.2004897**

[7] Chen, Y., Wiesel, A. and Hero, A.O. (2011). Robust Shrinkage Estimation of High-Dimensional Covariance Matrices, *IEEE Transactions on Signal Processing*, **59**(9), 4097-4107, DOI: **10.1109/TSP.2011.2138698**

[8] Verboon, P. (1993), Robust nonlinear regression analysis. *British Journal of Mathematical and Statistical Psychology*, **46**: 77-94. DOI: **10.1111/j.2044-8317.1993.tb01003.x**

[9] Riazoshams, H. Midi, H., and Ghilagaber, G. (2018). *Robust Nonlinear Regression: with Applications using R.* Wiley. DOI:**10.1002/9781119010463**

[10] Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust regression and outlier detection*. Wiley. DOI: **10.1002/0471725382**

[11] Rousseeuw, P. J., and Van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association,* **85**(411), 633-639. DOI **10.1080/01621459.1990.10474920**

[12] Desmettre, S., Korn, R., Ruckdeschel, P., Seifried, F. (2015): Robust Worst-Case Optimal Investment. *OR Spectrum*, **37**(3), 677-701. DOI: **10.1007/s00291-014-0370-y**

Simulation-Based Robust Inference for Response Misclassification Errors with Self-Reported Responses

Maria-Pia Victoria Feser[1]*, Yuming Zhang[2], Yanyuan Ma[3], Samuel Orso[4], Mucyo Karemera[4], Stephane Guerrier[4].
[1]Department of Statistical Sciences, University of Bologna, Italy.
[2]School of Public Health, Harvard University, USA.
[3]Department of Statistics, Pennsylvania State University, USA.
[4]Faclty of Sciences, University of Geneva, Switzerland.
*Corresponding Author Email: maria.victoriafeser@unibo.it

Surveys' questionnaires remain indispensable tools in capturing human thoughts and behaviours across multiple disciplines. Such measurement instruments are inherently prone to self-reporting errors, for example in the form of social desirability bias, which can lead to substantial estimation biases if unaddressed. Moreover, in practice, data collected via questionnaires are typically numerous in the number of questions p and not that numerous in the number of participants n. For such data settings, even without self-reporting errors, traditional estimators and associated inferential procedures (like confidence intervals) can suffer from important finite sample biases and bad coverages. This is even more the case, when consistent estimators are developed to include data features such as self-reporting errors. In this paper, we propose a reliable estimation and inference approach for parametric models based on the Just Identified iNdirect Inference estimator (JINI). The key advantage of our approach is that it allows to construct a consistent estimator in a simple manner, while providing strong bias correction guarantees that lead to accurate inference in finite samples. Our approach allows to bypass the analytical difficulties in deriving estimating equations that account for specific data features, which often result in non closed form expressions which are numerically very difficult to solve and the resulting estimators have severe finite sample biases. The properties of the JINI (including consistency, asymptotic normality, and its bias correction property) are studied when the parameter dimension p is allowed to diverge (with $p<n$), which provide the theoretical foundation to explain the advantageous performance of the JINI in data settings with large p relative to n. As a leading example, we consider data on alcohol consumption that suffers from self-reporting errors and possibly also random response misclassification that can be considered as outliers. The data and associated simulation studied are performed using a logistic regression model. The resulting analyses highlight the practical usefulness and excellent performance of the JINI when data contain response misclassification due both to social desirability and random choices.

Keywords: *Logistic Regression, Bounded Score Function, High Dimensions*

Contributed Session 7 (CS7)

CS7 – Applications of Robust Methods-2
Chair: Meral Çetin (Hacettepe University, Turkiye)

Serhan Tunçel (Hacettepe University, Turkiye)
Application of Robust SUR Modeling and Variable Selection to Air Pollution Data in Ankara

Özge Ustahüseyin (Ankara University, Turkiye)
The marshall–olkin weibull process: methodological framework and applications to environmental data

Şenol Çelik (Bingöl University, Turkiye)
Robust Regression for Agricultural Yield Estimation Under Influential Observations: Evidence from Konya Districts

Gunseli Aytac Cankurtaran (Ankara University, Turkiye)
Analysis of Installed Power Levels of Solar Power Plants in Turkey Using Spatial and Robust Spatial Regression

Onur Toka (Hacettepe University, Turkiye)
Comparing deletion methods and imputation methods in seemingly unrelated regression models for air quality data

Application of Robust SUR Modeling and Variable Selection to Air Pollution Data in Ankara

Serhan Tunçel[1]*, Meral Çetin[2], Olcay Arslan[3]
[1,2]Hacettepe University, Faculty of Sciences, Department of Statistics, Ankara, Turkiye
[3]Ankara University, Faculty of Sciences, Department of Statistics, Ankara, Turkiye
*Corresponding Author Email: serhan.tuncel@hacettepe.edu.tr

In developing countries, healthy living conditions increasingly depend on environmental factors. Population growth, along with industrial and agricultural activities, intensifies the presence of pollutants. This study proposes a model to examine the variables affecting air pollution in Ankara province. Air pollution data obtained from meteorological sources were analyzed using Seemingly Unrelated Regression (SUR) models to account for correlated response variables. To address the influence of outliers, Robust SUR estimation was also employed. Variable selection methods were incorporated into the modeling process, considering challenges such as measurement accessibility, cost, and multicollinearity among predictors. Comparing classical and robust approaches revealed the superiority of robust estimation and variable selection methods in environmental data analysis.

Keywords: *SUR, Robust SUR, Air Pollution*

References

[1] Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association, 57(298), 348–368.
[2] Bilodeau, M., & Duchesne, P. (2000). Robust estimation of the SUR model. Canadian Journal of Statistics, 28(2), 277–288.

The Marshall–Olkın Weıbull Process: Methodological Framework and Applıcatıons to Envıronmental Data

Özge Ustahüseyin[1,2], Mehmet Yılmaz[1]

[1] Department of Statistics, Ankara University Ankara, Türkiye, 06800
[2] Department of Statistics, Ankara University Ankara, Türkiye, 06800
*Corresponding Author Email: oustahuseyin@ankara.edu.tr

This study revisits the Marshall–Olkin bivariate Weibull (MOBW) distribution in order to address the modeling gap associated with weak dependence structures, which are often overlooked in the literature. The main contribution of the study is the derivation of refined analytical expressions for the survival copula, the singular component of the joint density, and dependence measures such as Spearman's rho and Kendall's tau. The resulting theoretical framework provides analytical consistency and numerical stability, particularly in low-dependence regimes where conventional estimation procedures frequently exhibit numerical fragility.

Methodologically, the framework is constructed under a first-order stationary Markov chain structure, and parameter estimation is performed using the Maximum Likelihood Estimation (MLE) approach. For model validation, a computationally efficient algorithmic scheme is developed by combining autocorrelation analysis with a bivariate chi-square goodness-of-fit test based on asymptotic theory, thereby avoiding the need for computationally intensive resampling procedures.

In addition, the MOBW structure is evaluated through comparative simulations against elliptical copula models such as the Gaussian and t-copula. The results indicate that the MOBW model can naturally capture dependence patterns driven by a common shock mechanism, which are often difficult to represent using conventional continuous models. In particular, the model is able to accommodate identical consecutive observations (ties) through its singular diagonal structure, preventing the misspecification that may arise when discrete-like clustering appears in continuous-valued series.

Applications to wind speed and temperature datasets demonstrate that, although linear models perform adequately under strong dependence scenarios, the MOBW framework provides a more physically coherent representation of the complex and weak dependence patterns observed during periods of atmospheric stagnation. Beyond meteorological applications, the proposed framework may also be relevant for other weakly dependent systems driven by common-shock mechanisms, including financial contagion and reliability studies. By incorporating the singular mass on the diagonal as a physical characteristic rather than a theoretical artifact, the proposed framework offers a flexible and robust alternative for the analysis of weakly dependent stochastic processes.

Keywords: *Bivariate Marshall-Olkin Distribution, Common Shock Mechanism, Weak Dependence, Survival Copula, Atmospheric Stagnation*

References

[1] Anonymous. 2019. Web site: https://www.forumatmosfer.com/konu-son-40-yilin-yaz-aylari-istatistiksel- degerlendirmesi.html Date of Access:06.04.2024
[2] Anonymous. 2023. Web site:https://www.wunderground.com/history/monthly/tr/çubuk/LTAC/date/2023-4 Date of Access:06.04.2024
[3] Anonymous. 2024. Web site: https://www.wunderground.com/history/monthly/tr/istanbul/LTBA/date/2019-8 Date of Access:26.11.2024
[4] Bayramoglu, I., and Ozkut, M. 2022. Recent Developments About Marshall–Olkin Bivariate Distribution. Journal of Statistical Theory and Practice, 16(4), 58.
[5] Bemis, B. M., Bain, L. J., and Higgins, J. J. 1972. Estimation and hypothesis testing for the parameters of a bivariate exponential distribution. Journal of the American Statistical Association, 927-929.
[6] Csörgő, S., and Welsh, A. H. 1989. Testing for exponential and Marshall–Olkin distributions. Journal of Statistical Planning and Inference, 23(3), 287-300.
[7] Feizjavadian, S. H., and Hashemi, R. 2015. Analysis of dependent competing risks in the presence of progressive hybrid censoring using Marshall–Olkin bivariate Weibull distribution. Computational Statistics & Data Analysis, 82, 19-34.
[8] Gupta, R. C., Kirmani, S. N. U. A., and Balakrishnan, N. 2013. On a class of generalized Marshall–Olkin bivariate distributions and some reliability characteristics. Probability in the Engineering and Informational Sciences, 27(2), 261-275.
[9] Jamalizadeh, A., and Kundu, D. 2013. Weighted Marshall–Olkin bivariate exponential distribution. Statistics, 47(5), 917-928.
[10] Koç, E., ve Şenel, M. C. 2016. Yatay Eksenli Rüzgâr Türbinlerinde Optimum Türbin Parametrelerinin Belirlenmesi-Teorik Yaklaşım. Mühendis ve makina, 57(676), 32-40.
[11] Kundu, D., and Dey, A. K. 2009. Estimating the parameters of the Marshall–Olkin bivariate Weibull distribution by EM algorithm. Computational Statistics & Data Analysis, 53(4), 956-965.
[12] Kundu, D., and Gupta, A. K. 2013. Bayes estimation for the Marshall–Olkin bivariate Weibull distribution. Computational Statistics & Data Analysis, 57(1), 271- 281.
[13] Kundu, D. 2023. A stationary Weibull process and its applications. Journal of Applied Statistics, 50(13), 2681-2700.
[14] Marshall, A. W., and Olkin, I. 1967. A multivariate exponential distribution. Journal of the American Statistical Association, 62(317), 30-44.
[15] Mino, T., Yoshikawa, N., and Suzuki, K. 2003. The mean of lifespan under dependent competing risks with application to mice data 順天堂大学スポーツ健康科学研究=Journal of health and sports science, Juntendo University, (7), 68-74.
[16] Spearman, C. 1904. The proof and measurement of association between two things. The American Journal of Psychology, 15(1), 72–101.
[17] Ustahüseyin, Ö. 2024. Marshall-Olkin İki Değişkenli Weibull Dağılımının İstatistiksel Karakterizasyonu. Yüksek Lisans Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, 73, Ankara.
[18] Weibull, W. 1939. A statistical study of the strength of material. Ing. Vetenskaps Akad. Handl.(Stockholm) Vol, 151, 15.
[19] Wu, K., Wang, L., Yan, L., and Lio, Y. 2021. Statistical inference of left truncated and right censored data from Marshall–Olkin bivariate Rayleigh distribution. Mathematics, 9(21), 2703.

Robust Regression for Agricultural Yield Estimation Under Influential Observations: Evidence from Konya Districts

Şenol Çelik[1]*, Mutlu Altuntaş[2]
[1]Prof. Dr., Bingöl University, Faculty of Agriculture, Department of Animal Science, Biometry and Genetic Unit, Bingöl, Türkiye, 12000
[2]Asst. Prof., Sinop University, Faculty of Arts and Science, Department of Data Science and Analytics, Sinop, Türkiye, 57000
*Corresponding Author Email: senolcelik@bingol.edu.tr

Accurate estimation of agricultural yield is essential for production planning, efficient resource allocation, and sustainable agricultural management. Classical linear regression models are widely used in agricultural data analysis; however, parameter estimates may be substantially affected by influential observations and leverage points, particularly in small-sample datasets. Robust regression methods offer an alternative modeling framework that reduces the impact of such observations and provides more stable and reliable estimates.

This study investigates the performance of classical and robust regression approaches for agricultural yield estimation using district-level carrot production data from Konya Province, Türkiye. Yield (ton per decare) was considered as the response variable, while cultivated area (decare) was used as the explanatory variable. Although exploratory analysis did not reveal any extreme outliers, diagnostic measures, including leverage values and Cook's distance, indicated the presence of observations with relatively high influence on model estimates.

To address this issue, ordinary least squares (OLS) regression and robust regression models based on M- and MM-estimation were fitted and compared in terms of model fit, residual behavior, and parameter stability. The findings demonstrated that both approaches successfully captured the relationship between cultivated area and yield; however, the robust regression model produced more stable parameter estimates and exhibited greater resistance to influential observations. Furthermore, the robust approach provided more reliable predictions in the presence of heterogeneous observations.

The results highlight that influential observations may affect model performance even when extreme outliers are not evident. Therefore, robust regression can serve as a valuable alternative to classical regression methods for agricultural yield estimation and decision-making processes, particularly in datasets characterized by limited sample sizes and leverage effects.

Keywords: Robust regression, MM-estimation, agricultural yield, influential observations, Cook's distance

References

Rousseeuw, P. J., & Leroy, A. M. (1987). Robust regression and outlier detection. John Wiley & Sons.
Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. The Annals of Statistics, 15(2), 642–656. https://doi.org/10.1214/aos/1176350366

Analysis of Installed Power Levels of Solar Power Plants in Turkiye Using Spatial and Robust Spatial Regression

Günseli Aytaç Cankurtaran[1]*, Özlem Türkşen[2]
[1] Ankara University, Institute of Science, Department of Statistics, Ankara, Türkiye, 06100
[2] Ankara University, Faculty of Science, Department of Statistics, Ankara, Türkiye 06100
*Corresponding Author Email: gunselicankurtaran@ankara.edu.tr

Today, demand for renewable energy sources is rapidly increasing in line with energy security concerns and climate change mitigation policies. Solar energy investments in Turkiye have increased significantly in recent years; the number of licensed solar power plants (SPPs) and total installed capacity have expanded considerably. In this study, a point dataset comprising 79 facility-level observations of solar power plants licensed by the Energy Market Regulatory Authority (EMRA) in Turkiye was used. Spatial point data is a data type containing the coordinates (x, y) of events or objects observed at specific locations within the study area. The installed capacity level of the facilities was taken as the dependent variable (Y), while solar potential (X1), elevation (X2), slope (X3), and distance to transformer/grid center (X4) were considered as explanatory variables. The functional relationship between these independent variables and the dependent variable was established. Since the spatial nature of the data may give rise to a dependency structure among observations stemming from geographic proximity, spatial regression models were employed. Spatial regression analysis is a statistical modeling approach developed due to the frequent violation of the independence assumption among observations in spatial data; it aims to estimate the relationship between the dependent variable and explanatory variables more accurately by incorporating neighborhood relationships between spatial units, spatial autocorrelation, and spatial heterogeneity into the model [1,2].

The study necessitated the application of spatial regression models alongside spatially robust regression models based on the GM (Generalized Method of Moments) estimator — which does not require the normality assumption — due to the presence of a prominent outlier observation in the SPP dataset. The Spatial Autoregressive Model (SAR), the Spatial Error Model (SEM), and their robust alternatives based on the Generalized Method of Moments (GM) estimator — Robust SAR and Robust SEM — were estimated; the models were compared within a framework of various performance metrics (AIC, LM Tests, RMSE, etc.) [3,4,5,6,7,8]. The analyses were carried out using the RStudio software [9].

According to the results obtained, it was concluded that the key factors determining installed capacity cannot be explained solely by technical-geographical conditions; socioeconomic and institutional variables that could not be included in the model may play a decisive role. The study empirically demonstrates how MLE and GM estimators differ in small-sample spatial datasets, thereby providing a methodological comparison framework for applied spatial regression analysis.

Keywords: *Spatial Regression, Robust Estimation, Spatial Autocorrelation, Solar Power Plants, Renewable Energy*

References

[1] Anselin, L. (1988). Spatial econometrics: Methods and models. Kluwer Academic Publishers.
[2] Cliff, A. D., and Ord, J. K. (1981). Spatial processes: Models and applications. Pion.
[3] Anselin, L., Bera, A. K., Florax, R., and Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. Regional Science and Urban Economics, 26(1), 77–104.
[4] Kelejian, H. H., and Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. Journal of Real Estate Finance and Economics, 17(1), 99–121.
[5] Kelejian, H. H., and Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review, 40(2), 509–533.
[6] Yildirim, V., and Kantar, Y. M. (2020). Robust estimation approach for spatial error model. Journal of Statistical Computation and Simulation. https://doi.org/10.1080/00949655.2020.1740223
[7] Mastuti, W. C., Djuraidah, A., and Erfiani. (2020). Robust spatial regression model on original local government revenue in Java 2017. Indonesian Journal of Statistics and Its Applications, 4(1), 68–79.
[8] Atikah, N., Afifah, D. L., and Kholifia, N. (2021). Robust spatial regression model in city minimum wages (CMW) in East Java 2018. Advances in Social Science, Education and Humanities Research, 528, 315–322.
[9] Bivand, R. S., Pebesma, E., & Gomez-Rubio, V. (2013). Applied spatial data analysis with R. Springer.

Comparing Deletion Methods and Imputation Methods in Seemingly Unrelated Regression Models for Air Quality Data

Serhan Tunçel[1]*,Onur Toka[1], Meral Çetin[1]
[1]Hacettepe University, Beytepe Campus, Department of Statistics, Ankara, Turkiye, 06800
*Corresponding Author Email: stuncel@hacettepe.edu.tr

Missing data are common in real-world datasets. In environmental monitoring, for instance, sensor malfunctions, maintenance periods, and gaps in spatiotemporal coverage frequently result in incomplete air quality records. The most common response to this problem, deletion methods, discards observations containing at least one missing value, shrinking the effective sample, potentially inducing selection bias, and inflating standard errors as the proportion of missing values grows [1]. Although multiple imputation (MI) has been shown to outperform listwise deletion across a broad range of regression settings [2,3], the literature on imputation strategies tailored specifically to Seemingly Unrelated Regression (SUR) models remains thin. The only simulation study that directly evaluates MI combined with SUR relies on conventional, non-robust imputation routines applied to economic evaluation data [4]. To the best of our knowledge, no published work examines the joint problem of high missingness and cellwise contamination within a SUR framework applied to air quality monitoring [5,6]. This study applies a SUR framework to model two air quality outcomes—PM10 and NO_2—as jointly dependent variables, with eight meteorological and pollution-related indicators (GBB, GMinS, GONN, GTY, GMRH, GSS, GOAB, GOB) serving as independent variables. Two approaches to the substantial missing data in this dataset are compared: deletion methods and imputation methods including robust algorithms. For deletion, both listwise and pairwise deletion are considered. For imputation, the most known imputation methods are considered including robust procedures based on cellwise outlier detection are employed [7,8].

The two approaches will be evaluated in terms of effective sample size and the proportion of information, equation-specific coefficient estimates and their standard errors, and model fit diagnostics. It is expected that robust imputation will recover a substantially larger share of the available data, yield more reliable standard errors, and produce a richer characterization of inter-pollutant dependence through the SUR covariance matrix [5,9]. This study aims to give an answer conventional MI-SUR has been shown to outperform complete-case SUR in economic evaluation contexts, the present work will demonstrate whether robust imputation is the more appropriate pre-processing choice when data are simultaneously incomplete and cellwise contaminated—conditions that are routine in air quality monitoring.

Keywords: *Seemingly Unrelated Regression, Robust Imputation, Missing Data, Air Quality, Deletion Methods, Cellwise Contamination*

References

[1] Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data*. John Wiley & Sons.
[2] Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of statistical software*, *45*, 1-67.

[3] Alwateer, M., Atlam, E. S., Abd El-Raouf, M. M., Ghoneim, O. A., & Gad, I. (2024). Missing Data Imputation: A Comprehensive. *Journal of Computer and Communications*, *12*, 53-75.
[4] Ben, Â. J., van Dongen, J. M., Alili, M. E., Heymans, M. W., Twisk, J. W., MacNeil-Vroomen, J. L., ... & Bosmans, J. E. (2023). The handling of missing data in trial-based economic evaluations: should data be multiply imputed prior to longitudinal linear mixed-model analyses?. *The European Journal of Health Economics*, *24*(6), 951-965.
[5] Saraceno, G., Alqallaf, F., & Agostinelli, C. (2021). A robust seemingly unrelated regressions for row-wise and cell-wise contamination. *arXiv preprint arXiv:2107.00975*.
[6] Peremans, K., & Van Aelst, S. (2018). Robust inference for seemingly unrelated regression models. *Journal of Multivariate Analysis*, *167*, 212-224.
[7] Raymaekers, J., & Rousseeuw, P. J. (2019). Handling cellwise outliers by sparse regression and robust covariance. *arXiv preprint arXiv:1912.12446*.
[8] Raymaekers, J., & Rousseeuw, P. J. (2024). The cellwise minimum covariance determinant estimator. *Journal of the American Statistical Association*, *119*(548), 2610-2621.
[9] Agostinelli, C., Leung, A., Yohai, V. J., & Zamar, R. H. (2015). Robust estimation of multivariate location and scatter in the presence of cellwise and casewise contamination. *Test*, *24*(3), 441-461.

Contributed Session 8 (CS8)

CS8 – Applications of Robust Methods-3
Chair: Özlem Kaymaz (Ankara University, Turkiye)

Tobias Leopold (Esslingen University of Applied Sciences, Germany)
Robust Reliability Demonstration Test Plans

Nabamallika Dehingia (United Nations Development Programme, United States) (Virtual)
Projecting Multidimensional Poverty to 2030: A Scenario-Based Microsimulation Framework

Gokalp Tepekoylu (Medipol University, Turkiye) (Virtual)
Achieving Nearly Deterministic Precision in Dental Implant Identification

Onur Camli (Erzurum Technical University, Turkiye) (Virtual)
Robust Circular–Circular Correlation Coefficient via Huber and Tukey Weighting

ICORS2026
20-24 JULY 2026
Istanbul

Robust Reliability Demonstration Test Plans

Tobias Leopold
Esslingen University of Applied Sciences, Kanalstr. 33, 73728 Esslingen, Germany
tobias.leopold@hs-esslingen.de

In practice, planning a test to proof the product reliability while the development of a new product is a central task of reliability engineering. Based on a suitable quantitative reliability target, a common approach is to apply a Reliability Demonstration Test (RDT). The basic principle of RDT is based on successful testing when all test items survive the test duration without any failure [1]. Important test parameters can be determined based on the Binomial distribution for the so called success run [2]:

$$1 - CL = R_{min}(t_{target})^n \tag{1}$$

In practice, the test duration often does not correspond to the required product service life. In such cases, a service life ratio LR can be used, which describes the ratio between test duration t_{test} and service life requirement t_{target} [4]:

$$LR = \frac{t_{test}}{t_{target}} \tag{2}$$

Taking (2) into account in (1), the verifiable minimum reliability R_{min} under the assumption of a Weibull-distributed failure characteristic with a shape parameter β is thus as follows [1]:

$$R_{min}(t_{target}) = (1 - CL)^{\frac{1}{LR^{\beta}n}} \tag{3}$$

The shape parameter β is often estimated from failure data of similar products in operating conditions that are as comparable as possible. Another option within a product development project is the possibility to derive the shape parameter β from previous test phases of the product development project. The possible difference of the test specimens from different sample phases from design changes or manufacturing influences must be evaluated with regard to a changed failure characteristic [5].
These uncertainties in the estimation of the shape parameter β consequently lead to an erroneous determination of the necessary sample size or confidence level of an RDT [5], see Fig. 1.

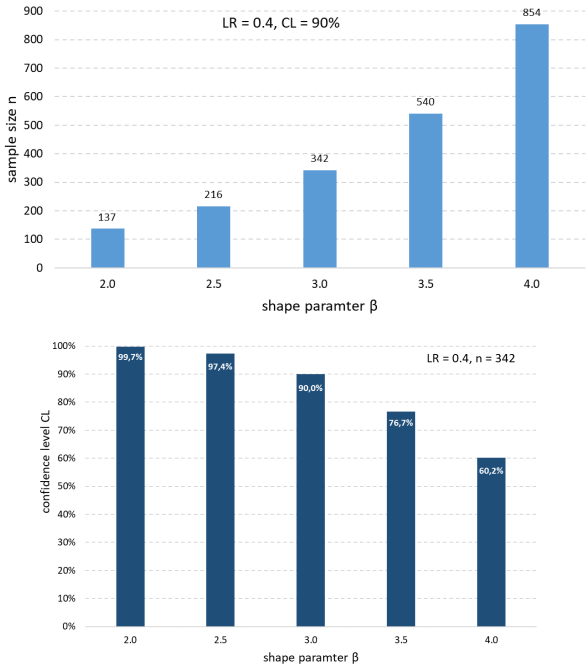


Figure 1: Effects of uncertain input data on sample size and confidence level of a RDT

To improve the robustness of RDT, an enhanced approach of test planning by using Monte Carlo simulations is applied. Based on these simulations, both bias errors and uncertainties in the determination of the shape parameter can be quantified and taken into account on the basis of sample results [5].

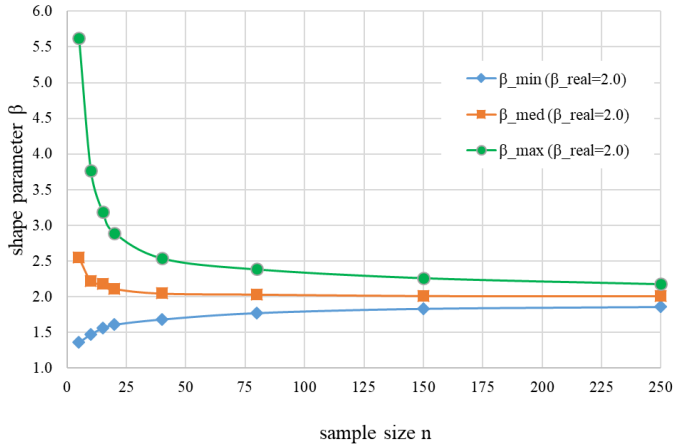


Figure 2: Quantification of uncertainties of shape parameters based on Monte Carlo simulations

Keywords: *Reliability Demonstration Test, Monte Carlo, Weibull, Binomial*

References

[1] H. Guo, E. Pohl, A. Gerokostopoulos, "Determining the Right Sample Size for Your Test: Theory and Application",
Proceedings - Annual Reliability and Maintainability Symposium, 2013.
[2] D. B. Kececioglu, "Reliability & Life Testing Handbook", 2nd ed, PTR Prentice Hall, New Jersey, 1994.

[3] T. Leopold, “Planning and evaluation of reliability demonstration testing with uncertainties”, Proceedings - Probabilistic Safety Assessment & Management conference, 2022.
[4] T. Leopold, “Determination of sample size and test duration of Reliability Demonstration Tests – Challenges and Opportunities”, Proceedings EuroBrake, 2023.
[5] T. Leopold, “Estimation of uncertainties in sample results and their impact on Reliability Demonstration Tests”, Proceedings of the 28th ISSAT International Conference on Reliability and Quality in Design 2023.

Projecting Multidimensional Poverty to 2030: A Scenario-Based Microsimulation Framework

Nabamallika Dehingia

Human Development Report Office, United Nations Development Programme

nabamallika.dehingia@undp.org

The global Multidimensional Poverty Index (MPI), jointly developed by UNDP and the Oxford Poverty and Human Development Initiative (OPHI), operationalizes poverty as a weighted sum of ten binary deprivation indicators across three domains: health, education, and living standards. Estimated from nationally representative household surveys across more than 100 countries, the global MPI is an important empirical tool for monitoring Sustainable Development Goals (SDG) 1.2, which commits to reducing poverty "*in all its dimensions*". However, despite the MPI's role in global poverty monitoring, forward-looking analysis remains underdeveloped; existing approaches rely on aggregate trend extrapolation[1] that is sensitive to sparse data and structural uncertainties, which are conditions endemic to low-income country contexts. This paper introduces a robust, scenario-based microsimulation framework that characterizes plausible future trajectories of multidimensional poverty to 2030, explicitly accounting for data limitations and structural uncertainty in underlying demographic processes. The analysis covered 71 countries that had relevant household survey data available for more than two time periods.

The framework is built around three interlocking statistical components. First, baseline household survey microdata are reweighted to align with projected 2030 population structures, preserving joint household characteristics under demographic uncertainty. Second, Latin Hypercube Sampling over logistic-bounded deprivation trajectories generates 1,200 scenario combinations per country, spanning plausible ranges of demographic and indicator-specific change. Third, intercept-shifted logistic regressions with Monte Carlo simulation translate these macro-level targets into household-level deprivation probabilities, yielding a full distribution of poverty outcomes. Representative scenarios are then extracted via machine learning classification models and diagnosed through ANOVA-based variance decomposition.

In the first phase of the analysis, baseline household survey microdata is reweighted to reflect projected 2030 demographic structures, incorporating age composition, sex distribution, and rural-urban dynamics.[2] This demographic re-weighting preserves the joint distribution of household characteristics while accommodating uncertainty in population trajectories, addressing a significant gap in prior aggregate-level projection methods.

In the second phase, we construct plausible indicator-specific trajectories for each of the ten MPI deprivation indicators at the country level. Annual rates of change are estimated via logistic transformation, which is appropriate for bounded proportions, and projected forward from the most recent observation to 2030. Recognizing that a single growth-rate estimate is unreliable given sparse time series and potential structural breaks,[3] we introduce symmetric ±25% scaling of each estimated growth rate to define upper and lower bounds. A key robustness concern addressed here is the treatment of indicators with historically increasing deprivation rates: we evaluate three methodological options for handling these cases, including floor constraints, zero-growth caps, and a "pandemic tax" adjustment that widens uncertainty intervals for countries whose most recent data predate 2020. Sensitivity analyses confirm that cross-country variation in projected poverty is not driven by the length of inter-survey intervals, though outlier growth rates in 16 countries are identified and treated with additional country-specific rules.

The third phase implements a household-level microsimulation linking macro-level indicator targets to individual household characteristics. For each indicator, weighted logistic regressions model baseline deprivation as a function of demographic covariates: household size, child composition, and rural location, and intercepts are adjusted so that predicted national prevalence matches each scenario's indicator-specific target. Monte Carlo simulation is then applied, drawing 100 stochastic Bernoulli outcomes per household-indicator pair within each scenario, with results averaged to produce stable estimates. Crucially, the ten MPI indicators are treated as independent during scenario generation, avoiding the imposition of macroeconomic coherence assumptions that would be difficult to justify given the complexity and unpredictability of structural change. The 400 indicator scenarios, generated via Latin Hypercube Sampling to ensure stratified coverage of the joint input space, are combined with the three demographic variants to yield 1200 distinct, internally consistent projections per country.

In the fourth phase, representative "*prosperity*" and "*deprivation*" scenarios are identified from the 1200 simulations using random forest classifiers for each country. Scenarios are classified relative to each country's own distribution of simulated outcomes, ensuring that the representative trajectories reflect plausible upper and lower bounds for that country's specific historical context rather than an absolute external standard.

Applied to 71 countries with at least two survey rounds between 2003 and 2023, the framework produces several substantive findings relevant to the SDG period. Global multidimensional poverty remains heavily concentrated in Sub-Saharan Africa under both scenarios. The gap between prosperity and deprivation scenarios, ranging from one to eight percentage points across countries, provides a direct measure of each country's sensitivity to alternative development pathways.

This analysis makes key contributions to the robust statistics literature as applied to development measurement. First, it demonstrates how stratified sampling techniques (Latin Hypercube Sampling) and ensemble machine learning methods (random forests) can be combined with classical microsimulation[4] to navigate high-dimensional uncertainty spaces efficiently and interpretably. Second, it formalizes a set of methodological decision rules for handling problematic time-series features: monotonically increasing trends, outlier growth rates, and pandemic-era discontinuities, within a transparent and replicable robustness framework. Together, these elements constitute a transferable methodological toolkit for robust scenario analysis in settings where data are scarce, structural breaks are plausible, and single-point predictions are epistemically unjustifiable. The framework is designed to be extensible: future iterations could incorporate macroeconomic coherence constraints, Bayesian updating of growth-rate priors as new survey data become available, or explicit conflict-adjustment modules for currently excluded fragile states.

As the 2030 SDG deadline approaches and the limits of deterministic forecasting become increasingly apparent, methods that characterize the full distribution of plausible futures, rather than collapsing uncertainty into a single trajectory, offer the most defensible and policy-relevant basis for global poverty monitoring. Beyond the specific application to multidimensional poverty, the combination of Latin Hypercube Sampling, logistic-bounded uncertainty intervals, and random forest classification represents a broadly applicable template for scenario-based inference under data scarcity.

Keywords: *microsimulation, scenario-based projections, machine learning*

References

1. Alkire, S., Nogales, R., Quinn, N. N., & Suppa, N. (2023). On track or not? Projecting the global Multidimensional Poverty Index. *Journal of Development Economics*, *165*, 103150.
2. UN DESA. (2024). *World Population Prospects 2024: Summary of Results*. Stylus Publishing, LLC.
3. Kalra, N., Hallegatte, S., Lempert, R., Brown, C., Fozzard, A., Gill, S., & Shah, A. (2014). Agreeing on robust decisions: New processes for decision making under deep uncertainty. *World Bank Policy Research Working Paper*, *6906*.
4. Marois, G., Sabourin, P., & Bélanger, A. (2017). *A microsimulation approach for modelling the future human capital of EU28 member countries*

Achieving Nearly Deterministic Precision in Dental Implant Identification

Hüseyin Ünözkan[1]*, Gökalp TEPEKÖYLÜ[2], Gökhan EROKTAY[3]
[1]Department of Industrial Engineering, Mudanya University, Bursa, Turkiye, 16000
[2]School of Dentistry, Medipol University, Istanbul, Turkiye, 34000
[3]Eroktay Dental Clinic, Istanbul, Turkiye, 34000
*Corresponding Author Email: huseyin.unozkan@mudanya.edu.tr

The precise identification of dental implants from radiographic imaging is a cornerstone of clinical assessment, restorative planning, and retrospective forensic records. While deep learning-based automated systems have proliferated, their clinical adoption is frequently hindered by the demand for massive, annotated datasets and their inherent "black-box" nature, which complicates clinical accountability. This study presents a comprehensive comparative analysis of three fundamental feature extraction techniques-Scale-Invariant Feature Transform (SIFT), Oriented FAST and Rotated BRIEF (ORB), and Accelerated-KAZE (AKAZE)-to establish a high-precision, interpretable framework for implant recognition in periapical radiographs. Utilizing a specialized dataset of 600 high-resolution periapical images, the study subjects these algorithms to a rigorous "stress test" framework, simulating clinical artifacts such as additive Gaussian noise, perspective rotations ($\pm 15^0$), and motion blur. Experimental results demonstrate that the SIFT-based approach achieves a deterministic success rate of 100%, outperforming AKAZE (96%) and ORB (91%). Statistical validation using McNemar's test ($p<0.001$) confirms that SIFT's superiority is mathematically significant and robust against acquisition variances. Unlike opaque neural networks, the proposed feature-based method offers "explainability maps," allowing clinicians to visually verify correspondences in thread pitch and apex geometry. These findings suggest that SIFT provides the necessary robustness for a reliable clinical decision support system, offering a transparent and computationally efficient alternative for modern digital dentistry where data scarcity remains a significant challenge.

The Scale-Invariant Feature Transform (SIFT), introduced by Lowe (2004), revolutionized this field by extracting keypoints that are highly resistant to changes in scale, rotation, and affine transformation. On the other hand, the ORB algorithm (Rublee et al., 2011) emerged as a computationally efficient, open-source alternative. By utilizing FAST keypoints and binary BRIEF descriptors, ORB offers significant advantages in terms of processing speed and memory usage, though it sometimes sacrifices the extreme robustness found in SIFT. In dental imaging, multi-scale matching is especially relevant. Implant features such as thread pitch, collar shape, and apex design exhibit consistent morphological signatures across scales. Studies in orthopedics and dental biomechanics (e.g., Wang et al., 2016) affirm the discriminative power of localized pattern comparison when supported by adequate normalization.

To simulate real-world clinical conditions, images were subjected to a "Stress Test" comprising additive Gaussian noise, perspective rotations, and simulated motion blur.

The performance was quantified using the Top-1 Identification Accuracy, defined as the exact matching of a sample to its corresponding database entry. Furthermore, the Distance Ratio (Lowe's Ratio) was calculated to measure the confidence margin between the first and second best matches, providing a quantitative metric for "matching certainty". The proposed SIFT-based identification system was executed on a standard workstation (Intel i5-12400, 16GB RAM). The average feature extraction and matching time per image was recorded at 142.5 milliseconds, making it suitable for real-time clinical integration without specialized hardware.

The proposed methodology is designed to operate as a Clinical Decision Support System (CDSS) for dentists who encounter "orphan" implants with missing patient records. The system follows a four-stage execution pipeline:

1. Image Input: Direct upload of a periapical radiograph.

2. Feature Mapping: Real-time extraction of SIFT keypoints from the implant body.

3. Geometric Verification: Comparison against the local database of annotated implant models using a FLANN-based (Fast Library for Approximate Nearest Neighbors) matching engine.

4. Ranked Reporting: Output of the most likely implant models with a confidence score and visual "matching-lines" for clinician verification.

The comparative results across all 600 images are summarized in Table 1. Our findings indicate a clear hierarchy in descriptor reliability for dental implant recognition.

Table 1: Multi-Metric Comparative Performance Analysis

Metric	SIFT	ORB	AKAZE
Top-1 Identification Accuracy	100%	91%	96%
Avg. Keypoints per Image	842 ± 124	3000 (Fixed)	618 ± 95
Avg. Matching Time (ms)	142.5	18.2	64.7
Lowe's Ratio Margin (μ)	0.42	0.78	0.58
Precision	1.00	0.89	0.94
Recall	1.00	0.91	0.96
F1-Score	1.00	0.90	0.95

To evaluate the statistical significance of the performance disparity between SIFT and ORB, a McNemar's test was performed. Given the 100% success rate of SIFT versus the 91% of ORB, the calculated p-value was found to be 0,001, indicating that SIFT's superiority in periapical implant identification is statistically significant and not due to random variation in image quality or acquisition angle. This study has provided a rigorous evaluation of the clinical resilience and diagnostic precision of SIFT, AKAZE, and ORB descriptors for the automated identification of dental implants in periapical radiographs. Despite the exceptional results, it is acknowledged that the precision of the system is inherently linked to the quality of the initial periapical acquisition. While our stress tests confirm high robustness, extreme underexposure or anatomical overlap may still necessitate secondary imaging.

Keywords: *Periapical radiography, dental implants, feature extraction, implant identification*

References

[1] Lowe, D. G. (2004) "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vis., vol. 60, no. 2, pp. 91-110.
[2] Rublee, E., Rabaud V., Konolige K., and Bradski G. (2011) "ORB: An efficient alternative to SIFT or SURF," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), pp. 2564-2571.
[3] Wang J. et al., (2016) "Hybrid texture–shape feature extraction for dental radiograph analysis," Med. Phys., vol. 43, no. 3, pp. 1138-1148.

Robust Circular–Circular Correlation Coefficient via Huber and Tukey Weighting

Onur Camli[1]*
[1]Department of Mathematics, Faculty of Science, Erzurum Technical University, Erzurum 25100, Turkiye
*Corresponding Author Email: onur.camli@erzurum.edu.tr

Circular (angular) data appear in many applications where observations are directional or periodic, such as wind directions, animal movement paths, and phase measurements. Since these data are defined on a circle, their analysis requires statistical methods that respect their periodic structure rather than standard linear approaches [1, 2]. A key problem in this setting is measuring the relationship between two angular variables. Several circular-circular correlation coefficients have been developed for this purpose [3, 4]. Although these methods are widely used, they are often sensitive to unusual observations. Even a small number of atypical data points can affect the results and reduce the reliability of the estimates. In this study, we focus on this issue and propose robust alternatives for circular-circular correlation. The proposed methods are based on Huber and Tukey weighting schemes, which limit the influence of extreme observations while keeping the main pattern in the data. The performance of the proposed methods is examined through an extensive simulation study under different sample sizes, contamination levels, and contamination structures. Within this framework, the new estimators are also compared with existing methods to better understand their behavior in both clean and contaminated data settings. Overall, this study aims to provide simple and reliable alternatives for correlation analysis in angular data, especially in situations where unusual observations are present.

Keywords: *Circular statistics, Circular-circular correlation, Robust methods, Huber estimator, Tukey bisquare, Contamination*

References

[1] Fisher, N. I. (1993). Statistical analysis of circular data. Cambridge University Press.
[2] Jammalamadaka, S. R., and Sengupta, A. (2001). Topics in circular statistics (Vol. 5). World Scientific.
[3] Jupp, P. E, and K. V. Mardia. (1980). A general correlation coefficient for directional data and related regression problems. Biometrika, 67 (1), 163–73.
[4] Jammalamadaka, S. R., and Y. R. Sarma. (1988). A correlation coefficient for angular variables. In Statistical Theory and Data Analysis II, ed. K. Matusita, 349–64. Amsterdam (North-Holland), Elsevier Science Publishers B. V.

Contributed Session 9 (CS9)

CS9 – Robust Regression, Clustering, and Dimension Reduction
Chair: Şükrü Acıtaş (Eskişehir Technical University, Turkiye)

Hasan Bulut (Ondokuz Mayıs University, Turkiye)
TSRClust: Two-Stage Robust Clustering

Han Ming Wu (National Chengchi University, Taiwan)
SIIE: sliced inverse interval estimation for robust sufficient dimension reduction

Ndeye Niang (CNAM Cédric, France)
Robust Feature Clustering around Latent Components

Ismail Yenilmez (Eskişehir Technical University, Turkiye)
Robust Sparse Principal Component Regression with Alternative Loss Functions: A Framework for High-Dimensional Settings

TSRClust: Two-Stage Robust Clustering

Hasan BULUT[1]*
[1]Department of Statistics, Faculty of Science, Ondokuz Mayıs University, Samsun, Turkiye, 55139
*Corresponding Author Email: hasan.bulut@omu.edu.tr

Robust clustering methods, particularly those based on trimming like the TCLUST [1] or trimming k-means [2] algorithms, are designed to protect the estimation of main cluster parameters by discarding a fraction α of the observations as outliers. While effective, these methods often operate under the implicit assumption that the trimmed set is composed of non-informative stochastic noise. However, in many real-world applications, such as anomaly detection or rare event discovery, the "outlier" set may contain small, high-density structural minorities that are statistically distinct from both the main clusters and pure random noise.

In this study, we propose the Two-Stage Robust Clustering (TSRClust) algorithm to address this "outlier heterogeneity" problem. In the first stage (Macro-Stage), the algorithm employs a robust trimming approach to isolate dominant structures and identify a residual set. In the second stage (Micro-Stage), a density-based scan is performed specifically on the residuals to distinguish between structured minority clusters and genuine noise.

Simulation studies demonstrate that while traditional methods fail to recognize low-frequency clusters (leading to a high information loss), TSRClust successfully recovers these hidden structures, significantly improving the Adjusted Rand Index (ARI) and providing a high Minority Recovery Rate (MRR). The proposed method bridges the gap between robust partitioning and structural discovery, offering a more nuanced treatment of the contamination layer in complex datasets.

Keywords: *Robust clustering, two-stage clustering, density-based scan, structural discovery.*

References

[1] Fritz, H., García-Escudero, L. A., & Mayo-Iscar, A. (2012). tclust: An R package for a trimming approach to cluster analysis. Journal of Statistical Software, 47, 1-26.
[2] Cuesta-Albertos, J. A., Gordaliza, A., & Matrán, C. (1997). Trimmed $ k $-means: an attempt to robustify quantizers. The Annals of Statistics, 25(2), 553-576.

SIIE: sliced inverse interval estimation for robust sufficient dimension reduction

Han-Ming Wu[1]*
[1] Department of Statistics, National Chengchi University, Taipei City, Taiwan, 11605
*Corresponding Author Email: wuhm@g.nccu.edu.tw

Standard slice-based Sufficient Dimension Reduction (SDR) [1] methods typically rely on reducing data slices to point-valued summary statistics, such as conditional means [2], variances, or quantiles, thereby discarding structural information regarding the shape and support of the conditional distribution. To address this limitation, we propose Sliced Inverse Interval Estimation (SIIE), a framework that reimagines the slicing procedure through the lens of Symbolic Data Analysis (SDA) [3]. Instead of compressing slices into points, SIIE aggregates them into interval-valued hypercubes constructed along slice principal directions, thereby preserving topological support information for the high-dimensional covariates while retaining both slice location and spread. We construct a symbolic kernel matrix from these interval summaries and establish Fisher consistency under the stated spread-support conditions, affine invariance, and root-n consistency. The revised simulations and real data analyses locate SIIE's niche rather than claim uniform superiority: SIIE is competitive with existing inverse moment-based methods on mean-dominated problems, while first-moment and local-smoothing methods can be preferable there. Its main advantage appears in detecting symmetric dependencies and heteroscedasticity under heavy-tailed errors, where the conditional mean carries little signal and trimmed symbolic intervals remain informative even when conventional second-moment estimates are unstable. Furthermore, the trimmed-quantile version exhibits enhanced robustness against high-leverage outliers and data contamination relative to the non-robust Min-Max construction and unstable moment-based competitors in these target settings. The method is therefore best viewed as a complementary symbolic SDR tool for spread-driven structure, with limitations in purely mean-driven or very high-dimensional low-sample regimes.

Keywords: *Central sunspace, interval-valued data, inverse regression, robust statistics, symbolic data analysis*

References

[1] Li, K. C. (1991). Sliced inverse regression for dimension reduction. Journal of the American Statistical Association, 86(414), 316-327.
[2] Cook, R. D., and Li, B. (2002). Dimension reduction for conditional mean in regression. The Annals
of Statistics, 30(2):455-474.
[3] Billard, L., and Diday, E. (2003). From the statistics of data to the statistics of knowledge: symbolic data analysis. Journal of the American Statistical Association, 98(462), 470-487.

Robust Feature Clustering around Latent Components

Mouhamadou Lamine Ndao[1], Ndèye Niang[1]*, Gilbert Saporta[1], Tony Sarre[2],
[1]CNAM Cédric, 2 rue Conté 75003 Paris, France
[2]UFR SET, Université de Thiès, Thiès, Sénégal
*Corresponding Author Email: ndeye.niang_keita@cnam.fr

We consider unsupervised analysis of large data sets in which features are assumed to be organized into homogeneous, but unknown, blocks. Feature clustering which identifies blocks and summarizes them using latent components is closely related to reducing the dimensionality of the feature space, thereby facilitating interpretation. Compared to clustering of observations, feature clustering has received far less attention in the literature [1,2]. Existing approaches often mimic standard observation-clustering techniques such as Agglomerative Hierarchical Clustering (AHC), which are conceptually unsatisfactory.

Among the methods specifically designed for feature clustering, based on PCA (Principal Component Analysis), one can find methods from the two classical families of clustering methods: (i) direct partitioning like K-means and the (ii) Hierarchical descendant or divisive clustering. One may mention the CLV (Clustering of Variables around Latent Components) method [2], a K-means-like approach and VARCLUS procedure of SAS software, which has never been the subject of scientific articles or its interpretable extension IDFC [4]. These methods allow organizing features into homogeneous clusters; each associated with the first principal component as centroid. Their associated criteria, based on covariances or correlations between features and the latent components, are naturally sensitive to outliers that can distort the estimated latent components and corrupt the resulting feature partition.

We propose robust extensions of feature clustering methods, based on robust estimations of the covariances and robust PCA [5,6]. The proposed methods are evaluated on simulated data under various contamination schemes and illustrated on real data, demonstrating improved stability of the recovered partition and enhanced interpretability of the latent dimensions compared to the standard approaches.

Keywords: *dimension reduction*, *feature clustering, latent component, robust estimation, robust PCA*

References

[1] Niang, N., Ouattara, M., Saporta, G. (2021). Sparse Divisive Feature Clustering. XXVIII Meeting of the Portuguese Association for Classification and Data Analysis (JOCLAD 2021), pp. 75–76.
[2] Niang, N., Ouattara, M., Saporta, G. (2024). Clustering variables: a survey and some new developments. Sensometrics 2024.

[3] Vigneau, E., Qannari, E.M. (2003). Clustering of variables around latent components. Communications in Statistics – Simulation and Computation, 32(4), 1131–1150.
[4] Ndao, M. L., Youness G., Niang, N., Saporta, G. (2026). Explainable Remaining Useful Life Prediction Using Interpretable Divisive Feature Clustering. Applied Stochastic Models in Business and Industry, 42(2), e70084.
[5] Hubert, M., Rousseeuw, P. J., Vanden Branden, K. (2005). ROBPCA: A new approach to robust principal component analysis. Technometrics, 47(1), 64–79.
[6] Rousseeuw, P.J., Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. Technometrics, 41(3), 212–223.

Robust Sparse Principal Component Regression with Alternative Loss Functions: A Framework for High-Dimensional Settings

Ersin YILMAZ[1], İsmail YENİLMEZ[2]*
[1] Muğla Sıtkı Koçman University, Department of Statistics, Muğla, Türkiye
[2]Eskişehir Technical University, Department of Statistics, Eskişehir, Türkiye
*Corresponding Author Email: ismailyenilmez@eskisehir.edu.tr

Sparse Principal Component Regression via Singular Value Decomposition (SPCRsvd) is an effective approach for simultaneous dimension reduction, variable selection, and prediction [1]. This framework is closely related to sparse PCA and penalized low-rank approximation methods that have been widely studied in high-dimensional statistics [3]. However, the classical SPCRsvd formulation relies on squared-error loss and may suffer from substantial performance degradation in the presence of outliers, leverage observations, or departures from Gaussian assumptions. Such limitations are well documented in high-dimensional regression settings, where robustness becomes a critical requirement [2,6].

To address these issues, this study proposes a robust extension of the SPCRsvd framework by incorporating alternative robust regression loss functions into the supervised sparse dimension-reduction objective. In particular, Huber, Tukey biweight, least absolute deviation (LAD), quantile, Hampel, and correntropy-based loss functions are embedded within a unified optimization framework. These loss functions are widely recognized in robust statistics and regularized estimation literature for their ability to mitigate the influence of heavy-tailed noise and outliers [6]. The resulting formulations preserve the sparse low-rank representation structure of SPCRsvd while improving robustness against different types of contamination.

Estimation is performed using an Alternating Direction Method of Multipliers (ADMM)-based algorithm [4], combined with a multi-start strategy to alleviate initialization sensitivity arising from the nonconvex optimization landscape. The convergence properties of ADMM in nonconvex settings support the validity of the proposed optimization scheme [5].

A comprehensive simulation study is conducted under both low-dimensional and high-dimensional settings with varying contamination rates, leverage structures, sparsity levels, and signal-to-noise ratios. Competing robust formulations are evaluated in terms of prediction accuracy, estimation stability, sparsity recovery, and subspace identification. Results indicate that robust loss functions significantly improve predictive performance and estimation reliability compared to the classical squared-error formulation, particularly under moderate and severe contamination scenarios.

Overall, the proposed framework extends existing sparse PCA regression methodologies [1,3] by integrating robust statistical learning principles [2,6] into a unified SPCRsvd formulation suitable for contaminated high-dimensional data environments..

Keywords: *SPCRsvd, Loss Function, Quantile Regression, Correntropy, ADMM, High-Dimensional Data*

References

[1] Kawano, S. (2021). Sparse principal component regression via singular value decomposition approach. Advances in Data Analysis and Classification, 15, 795–823.
[2] Cai, T., Ma, Z., Wu, Y. (2013). Sparse PCA: optimal rates and adaptive estimation. The Annals of Statistics, 41, 3074–3110.
[3] Zou, H., Hastie, T., Tibshirani, R. (2006). Sparse principal component analysis. Journal of Computational and Graphical Statistics, 15, 265–286.
[4] Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning, 3, 1–122.
[5] Hong, M., Luo, Z. Q., Razaviyayn, M. (2016). Convergence analysis of ADMM for a family of nonconvex problems. SIAM Journal on Optimization, 26, 337–364.
[6] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58, 267–288.

Contributed Session 10 (CS10)

CS10 – Robust Penalized Regression
Chair: Abdullah Yalçınkaya (Ankara University, Turkiye)

Ismail Yenilmez (Eskişehir Technical University, Turkiye)
Graph Robust Adaptive LASSO for Network-Aware Variable Selection under Contamination

Graph Robust Adaptive LASSO for Network-Aware Variable Selection under Contamination

Ersin YILMAZ[1], İsmail YENİLMEZ[2]*
[1]Muğla Sıtkı Koçman University, Department of Statistics, Muğla, Türkiye
[2]Eskişehir Technical University, Department of Statistics, Eskişehir, Türkiye
*Corresponding Author Email: ismailyenilmez@eskisehir.edu.tr

Variable selection in high-dimensional regression becomes substantially more challenging when explanatory variables exhibit complex dependence structures and data are contaminated by outliers. Traditional penalized regression methods such as LASSO perform simultaneous estimation and variable selection but often ignore relationships among predictors and are highly sensitive to atypical observations [1,2]. Although Adaptive LASSO achieves oracle properties under suitable regularity conditions [3], its performance may deteriorate in the presence of leverage points, response outliers, and heavy-tailed distributions [4].

Robust penalized regression methods based on M-estimation and redescending ψ-functions have been proposed to reduce the influence of contamination [5,6]. However, these methods generally treat predictors as independent variables and do not exploit the underlying network structure frequently observed in genomic studies, financial systems, environmental monitoring, and other high-dimensional applications [7]. Ignoring such dependence structures may result in unstable variable selection and reduced predictive performance.

To address these limitations, this study proposes a novel Graph Robust Adaptive LASSO (GR-ALASSO) estimator that integrates robust estimation and network-constrained variable selection within a unified framework. The proposed approach combines a redescending robust loss function with Adaptive LASSO penalization and an additional graph-based regularization term derived from the graph Laplacian matrix. The graph structure is estimated from predictor relationships using robust correlation measures, allowing the model to account for both contamination and predictor connectivity simultaneously.

The proposed objective function encourages sparsity while preserving important structural relationships among predictors. Consequently, correlated variables connected within the network tend to be selected jointly, improving interpretability and selection stability. Robust estimation is performed using redescending ψ-functions, including Tukey Biweight and Welsch functions, combined with an Iteratively Reweighted Least Squares (IRLS) optimization procedure [5,6].

A comprehensive Monte Carlo simulation study will be conducted under various contamination scenarios involving response outliers, leverage points, mixed contamination, and heavy-tailed distributions. Multiple network topologies, including block networks, hub networks, and scale-free structures, will be considered. The proposed method will be compared with LASSO, Adaptive LASSO, Elastic Net, Graph

LASSO, and existing robust penalized regression approaches. Performance evaluation will be based on mean squared error, variable selection accuracy, false discovery rate, true positive rate, and network recovery measures.

The proposed GR-ALASSO framework provides a robust and network-aware alternative for high-dimensional regression analysis. By simultaneously incorporating robustness against contamination and structural information among predictors, the method aims to improve estimation accuracy, variable selection consistency, and model interpretability. The proposed framework is expected to be particularly useful for modern data applications characterized by complex dependence structures and the presence of outliers.

Keywords: *Robust Adaptive LASSO, Graph Regularization, Network-Aware Variable Selection, Robust Regression, High-Dimensional Data, Outlier Detection*

References

[1] Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society: Series B, 58(1), 267–288.
[2] Fan, J., & Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. Journal of the American Statistical Association, 96(456), 1348–1360.
[3] Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. Journal of the American Statistical Association, 101(476), 1418–1429.
[4] Lambert-Lacroix, S., & Zwald, L. (2011). Robust Regression through the Huber's Criterion and Adaptive Lasso Penalty. Electronic Journal of Statistics, 5, 1015–1053.
[5] Chang, L., Roberts, S., & Welsh, A. (2018). Robust Lasso Regression Using Tukey's Biweight Criterion. Technometrics, 60(1), 36–47.
[6] Maronna, R. A., Martin, R. D., Yohai, V. J., & Salibián-Barrera, M. (2019). Robust Statistics: Theory and Methods (2nd ed.). Wiley.
[7] Li, C., & Li, H. (2008). Network-Constrained Regularization and Variable Selection for Analysis of Genomic Data. Bioinformatics, 24(9), 1175–1182.

Poster Session

Poster Session

Piotr Waz (Medical University of Gdańsk, Poland)
A Novel Computational Tool for AMD Diagnosis

Gianfranco Piscopo (University of Naples Federico II, Italy)
Robust Assessment of Benford's Law under Skewed Generalized Error Distributions via Nonparametric Combination

Igor Böhm (Charles University, Czech Republic)
Spherical Projection Depth

Sandra Esperanza Melo (The National University of Colombia, Colombia)
Methodology for the Longitudinal Analysis of thrips Counts in Chrysanthemum Cultivation

Angel Lopez Oriona (King Abdullah University of Science and Technology, Saudi Arabia)
Amortized Neural Clustering of Time Series based on Statistical Features

Dorota Bielinska Waz (Medical University of Gdańsk, Poland)
Novel alignment-free bioinformatics methods

A Novel Computational Tool for AMD Diagnosis

Piotr Wąż[1]*, Dorota Bielińska-Wąż[2], Katarzyna Zorena[3], Anna Murawska[3]
[1]Department of Nuclear Medicine, Medical University of Gdańsk, 80-210 Gdańsk, Poland
[2]Department of Radiological Informatics and Statistics, Medical University of Gdańsk, 80-210
Gdańsk, Poland
[3]Department of Immunobiology and Environment Microbiology, Medical University of Gdańsk, 80-210 Gdańsk, Poland
*Corresponding Author Email: phwaz@gumed.edu.pl

Vision loss in the central visual field resulting from Age-Related Macular Degeneration (AMD) represents a major health burden worldwide, especially in individuals over 60. AMD is responsible for nearly half of all cases of blindness [1]. This chronic, progressive disease affects the outer retinal layers and the choroid in the central macula. The number of AMD cases is expected to rise due to increasing life expectancy and greater exposure to risk factors that promote degenerative changes in the macula [2].

Our study analyzed 132 eyes from 66 patients [3], classified according to AMD progression using the four-point Age-Related Eye Disease Scale (AREDS) [4]. As the most widely utilized system for categorizing AMD, the AREDS scale divides AMD progression into the following stages:

1. AREDS 1 (control group): Absence of AMD or presence of only a few small drusen (< 63 µm in diameter).
2. AREDS 2: Early AMD, characterized by numerous small drusen (> 15 µm), several intermediate-sized drusen (63–125 µm), or Retinal Pigment Epithelium (RPE) abnormalities, such as increased pigmentation or depigmentation.
3. AREDS 3: Intermediate AMD, including numerous medium-sized drusen, at least one large druse (> 125 µm), or geographic atrophy not involving the central macula.
4. AREDS 4: Advanced AMD, involving geographic atrophy of the RPE affecting the macula or neo-vascular maculopathy. This includes Choroidal Neovascularization, serous or hemorrhagic retinal or RPE detachment, exudates and hard fibrovascular proliferations beneath the retina and RPE, and discoid scars (choroidal fibrosis).

Inclusion criteria for the AMD and control groups were an age of over 55 years and an AMD diagnosis aligned with the AREDS scale and Polish Society of Ophthalmology guidelines. Statistical analyses were conducted using the R programming language. Quantitative variables were summarized using median, minimum, and maximum values. The Kruskal-Wallis test was applied to compare variables across AMD advancement groups, with post-hoc tests following significant results. Ordinal regression, implemented via a generalized linear model, was utilized to predict ordinal variables, where only relative ordering is important. The dependent variable was the AMD advancement level. Independent variables included Central Retinal Thickness (CRT), average Ganglion Cell Complex (GCC) thickness, Macular Pigment Optical Density

(MPOD), Early Treatment Diabetic Retinopathy Study (ETDRS), Snellen visual acuity, and patient age. Measurements were performed using the Zeiss Cirrus HD-OCT model 400. Logit functions were used in the analysis. Since, in this case, the maximization of probability (or the logarithm of probability) does not have an analytical solution, the Iteratively Reweighted Least Squares (IRLS) technique was employed to estimate the regression coefficients. The models included all the collected values for each variable. The quality of the models was assessed by evaluating the statistical significance of the coefficients, the -2 log likelihood value, and the frequency of correctly predicted categories based on the values of the independent variables. This modeling resulted in a tool for estimating AMD progression. Due to the limited sample size, models were restricted to pairs of independent variables. The article highlights pairs of variables with statistically significant regression coefficients, alongside their Odds Ratios (OR), Confidence Intervals (CI), and thresholds for ordinal categories. The statistical significance level was set at $\alpha = 0.05$. Using the proposed model, classification maps were generated, serving as a graphical tool to support the diagnosis of AMD [3]. These maps enable the classification of patients' eyes into specific groups (control group, AREDS 2, AREDS 3, or AREDS 4) based on the values of variables represented on their axes. Classification maps, based on these models, visually represent the predicted probabilities for each AMD stage, using a color-coded scheme. This alternative computational approach facilitates the accurate diagnosis of all stages of AMD with high or good precision.

In summary, this work describes a diagnostic graphical tool (classification maps) recently developed and published by the authors to support the detection of AMD [3]. These maps, constructed using an ordinal regression model, visually represent the progression of AMD. In this model, the degree of AMD advancement serves as the ordinal dependent variable. Independent variables, such as CRT, GCC, MPOD, ETDRS scores, Snellen visual acuity, and patient age are incorporated into the analysis and represented on the axes of the maps.

Keywords: *Data analysis, medical informatics, biostatistics, mathematical modeling, Age-Related Macular Degeneration (AMD)*

References

[1] Fleckenstein, M. et al. (2021). Age-related macular degeneration. Nat. Rev. Dis. Primers., 7, 31.
[2] Pennington, K., DeAngelis, M. (2016). Epidemiology of age-related macular degeneration (AMD): asssociations with cardiovascular disease phenotypes and lipid factors. Eye and Vision, 3, 34.
[3] Wąż, P., Zorena, K., Murawska, A., Bielińska-Wąż, D. (2023). Classification Maps: A New Mathematical Tool Supporting the Diagnosis of Age-Related Macular Degeneration. J. Pers. Med., 13, 1074.
[4] Ferris, F.L. et al. (2005). Age-Related Eye Disease Study (AREDS) Research Group. A simplified severity scale for age-related macular degeneration: AREDS Report No. 18. Arch Ophthalmol., 123, 1570–1574.

Robust Assessment of Benford's Law under Skewed Generalized Error Distributions via Nonparametric Combination

Gianfranco Piscopo[1], Valerio Ficcadenti[2], Maria Longobardi[1], Massimiliano Giacalone[3]
[1] Department of Mathematics and Applications "Renato Caccioppoli", University of Naples Federico II, Naples, Italy
[2] London South Bank University, London, United Kingdom
[3] Department of Economics, University of Campania "Luigi Vanvitelli", Italy

Benford's Law describes the non-uniform distribution of leading digits in many real-world datasets and is widely used in data analysis, fraud detection, and model validation. However, its validity may be affected by structural properties of the underlying data-generating process, such as skewness and tail behavior, which are central concerns in robust statistics.

In this work, we provide a robust assessment of Benford's Law within the framework of the Skewed Generalized Error Distribution (SGED), a flexible family that simultaneously captures asymmetry and tail thickness, encompassing classical distributions such as the normal and Laplace as special cases.

We conduct a Monte Carlo study over a grid of SGED parameters, varying both skewness and shape. Deviations from Benford's Law are evaluated using classical goodness-of-fit measures and a permutation-based Nonparametric Combination (NPC) framework, enabling global and distribution-free inference.

The results show that conformity to Benford's Law is not universal within the SGED family. Heavy tails and strong skewness induce systematic departures, while lighter tails and near-symmetric configurations exhibit closer agreement. These findings suggest that deviations from Benford's Law may arise from intrinsic distributional features rather than anomalous behavior.

Our study highlights the importance of incorporating skewness and tail effects into Benford-based diagnostics and supports the use of robust, nonparametric methods for reliable assessment in complex data settings.

Keywords: *Benford's Law, Skewed Generalized Error Distribution, Robust Statistics, Monte Carlo Simulation, Nonparametric Combination*

Spherical Projection Depth

Igor Böhm[1]*, Filip Bočinec[1], Stanislav Nagy[1], Hyemin Yeon[2]
[1]Department of Probability and Mathematical Statistics, Charles University, Sokolovská 83, Prague, 186 75, Czech Republic
[2]Department of Mathematical Sciences, Kent State University, Summit Street, Kent, 44242, Ohio, USA
*Corresponding Author Email: bohm@karlin.mff.cuni.cz

Depth functions are a well-established tool in nonparametric statistics, providing a way to quantify centrality and to perform robust inference. Numerous depth notions have been developed for Euclidean spaces. One notable example is the projection depth introduced by Zuo and Serfling [1], which measures outlyingness via directional projections combined with robust summary statistics.

In contrast, depth functions for non-Euclidean spaces, such as spheres in $\mathbb{R}^d$, are less developed than their Euclidean counterparts. Several well-established constructions extend Euclidean depth notions to spherical settings, including angular halfspace depth [2] and angular simplicial depth [3]. In this contribution, we explore a notion of projection depth on the sphere by adapting the projection-based construction to its intrinsic geometry.

The proposed approach replaces linear projections with projections onto great circles passing through the point of interest. For each such circle, the data are projected and an outlyingness measure is computed using robust statistics. The overall outlyingness is then obtained by taking the worst-case behavior over all directions.

We study theoretical properties of the proposed depth, including consistency and robustness, and examine its behavior on selected examples.

Keywords: *Data depth, Projection depth, Spherical data, Directional statistics, Robust statistics*

References

[1] Zuo, Y., Serfling, R. (2000). General notions of statistical depth functions. *The Annals of Statistics* 28, 461–482.
[2] Small, C.G. (1987). Measures of centrality for multivariate and directional distributions. *Canadian Journal of Statistics* 15, 31–39.
[3] Liu RY, Singh K (1992) Ordering directional data: concepts of data depth on circles and spheres. *The Annals of Statistics* 20(3):1468–1484

Methodology for the Longitudinal Analysis of thrips Counts in Chrysanthemum Cultivation

Sandra E Melo Martínez[1]*, John E Puerto Pinzón[2], Oscar O Melo Martínez[2]
[1] Department of Agronomy, Universidad Nacional de Colombia, Email: semelom@unal.edu.co
[2] Department of Statistics, Universidad Nacional de Colombia
Email: jepuertop@unal.edu.co, Email: oomelom@unal.edu.co

This research presents a methodology for data modeling in agronomic experiments where the response variable is a count. In such studies, problems of overdispersion often arise, leading to biased inferences due to underestimated standard errors, inflated significance tests, and misleading conclusions if not properly accounted for. For this study, 200 chrysanthemum plants were selected within a completely randomized design (CRD) The plants were arranged across five plots, each containing 40 plants, in plots of 4 m² with an approximate planting density of 25 plants/m².

Five treatments were evaluated, including a control and a traditional management involving weekly rotation of agrochemicals. Foliar applications of nematodes were performed weekly for four weeks, using a volume of five liter of water per treatment in each application. The research was conducted in the municipality of San Antonio del Tequendama Cundinamarca, Colombia, in two chrysanthemums nurseries with plants in pots.

The measurements were taken over four weeks on the same plants, and during each observation period the number of larvae, adults, and the total thrips population were recorded from ten flowers per treatment. The objective of this experiment was to determine whether the biological control program affected the abundance of Frankliniella occidentalis thrips individuals on chrysanthemum plants throughout the implementation period, thus evaluating the effectiveness of the program in reducing these pest populations.

Thrips represent one of the main threats to ornamental plants, causing direct damage by feeding and transmitting viral diseases, which negatively affect crop productivity and overall quality. In many agronomic experiments, repeated measurements are taken on the same individuals, generating temporal correlation among observations that complicates both data analysis and interpretation. Assuming independence of observations over time is inappropriate, as it ignores within plant correlations across different time points.

To address this study, a zero inflated longitudinal Negative Binomial model was considered for larval counts, accounting for temporal correlations and providing a good fit to the data. Specifically, the study proposes the use of Negative Binomial Mixed Models (NBMM) for longitudinal agronomic data, following the methodology outlined by Zhang (2020). The extended NBMM can incorporate different types of fixed and random effects, as well as various correlations structures among within subject observations, thereby fully addressing the specific properties of longitudinal count data. In addition, generalized linear mixed models are suggested when observations are correlated in a way that requires random effects. The statistical analysis revealed that applying the highest doses of the entomopathogenic nematodes Heterorhabditis bacteriophora provides effective control of the pest across its different life stages.

Keywords: *Zero Inflated Negative Binomial Model, overdispersion, longitudinal analysis, count data, mixed effects.*

References

[1] Barrera, J., Meza, C., and Arribas-Gil, A. (2025). A stochastic method to estimate a zero-inflated two-part mixed model for human microbiome data. arXiv preprint arXiv:2504.15411.
[2] Zhang, Y., Han, S., Chen, Y., Chen, L., and Li, F. (2020). Zero-inflated gaussian mixed models for differential abundance analysis in microbiome studies. Journal of the American Statistical Association, 115(531):1021–1035

Amortized Neural Clustering of Time Series based on Statistical Features

Ángel López-Oriona*, Ying Sun

King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
*Corresponding Author Email: angel.lopezoriona@kaust.edu.sa

This work introduces an algorithm-agnostic approach to feature-based time series clustering via amortized neural inference. By training neural networks to approximate the optimal partitioning rule from simulated data, the proposed framework reduces reliance on conventional clustering methods, such as K-means, K-medoids, or hierarchical clustering, and their associated objective functions and heuristics. Leveraging statistical features, such as autocorrelations and quantile autocorrelations, the approach learns a data-driven affinity structure from which clustering partitions can be recovered, without requiring explicit prior specification of cluster shapes or structures. In addition, one version of the method can automatically determine the number of clusters, avoiding ad-hoc selection procedures. Comprehensive empirical studies show that the proposed framework achieves competitive or superior clustering accuracy relative to traditional methods, even in challenging scenarios where competing techniques are provided with the true number of clusters. An application to financial time series of stock returns illustrates its practical utility. By reducing the need for algorithm selection and calibration, the proposed framework opens new possibilities for automated, adaptive, and data-driven clustering of temporal data across scientific and industrial domains.

Keywords: *time series, clustering, simulation-based learning, amortized inference, neural networks, stock returns*

Novel alignment-free bioinformatics methods

Dorota Bielińska-Wąż[1]*, Piotr Wąż[2]
[1]Department of Radiological Informatics and Statistics, Medical University of Gdańsk, 80-210
Gdańsk, Poland
[2]Department of Nuclear Medicine, Medical University of Gdańsk, 80-210 Gdańsk, Poland
*Corresponding Author Email: djwaz@gumed.edu.pl

To address the challenges posed by rapidly expanding biomedical datasets, the field has recently introduced numerous novel mathematical frameworks for large-scale biological data analysis [1, 2]. A prominent example are alignment-free bioinformatics methods - efficient computational techniques for comparing DNA, RNA, and protein sequences without relying on traditional alignment (reviewed in [3]). These approaches, drawing on mathematics, computer science, and biology, deliver high computational efficiency and scalability [4]. In this work, we showcase our contributions to the field, which include innovative alignment-free methods [5–16]. Specifically, we have developed a series of novel alignment-free bioinformatics methods, such as:

- 2D-Dynamic Representation of DNA Sequences [5],
- Four-Component Spectral Representation of DNA Sequences [6],
- 3D-Dynamic Representation of DNA Sequences [7],
- 20D-Dynamic Representation of Protein Sequences [8],
- Spectral-Dynamic Representation of DNA Sequences [9],
- 4D-Dynamic Representation of DNA/RNA Sequences [10].

Some of these methods we combined with supervised machine learning algorithms (C5.0 and random forest), an unsupervised machine learning algorithm (*K*-means clustering), and a statistical method (Principal Component Analysis). In particular we proposed:

- 2D-Dynamic Representation of DNA/RNA Sequences combined with C5.0 [11],
- 3D-Dynamic Representation of DNA/RNA Sequences combined with random forest algorithm [12],
- 20D-Dynamic Representation of Protein Sequences combined with Principal Component Analysis [13],
- 20D-Dynamic Representation of Protein Sequences combined with *K*-means clustering [14],

- 4D-Dynamic Representation of DNA/RNA Sequences combined with *K*-means clustering [15].

We have successfully applied our methods in various studies, for example as shown in this presentation 4D-Dynamic Representation of DNA/RNA Sequences - to studies on coronaviruses [10], time evolution analysis of Zika virus genome sequences [10], and genetic diversity studies of *Echinococcus multilocularis* in red foxes in Poland [15,16].

Keywords: *Data analysis, bioinformatics, alignment-free methods, machine learning, Principal Component Analysis*

References

[1] Du, X., Zhu, R., Li, Y., Anjum, A. (2019). Language model-based automatic prefix abbreviation expansion method for biomedical big data analysis. Future Gener. Comput. Syst., 98, 238–251.
[2] Lötsch, J. Malkusch, S., Ultsch, A. (2021). Optimal distribution- preserving downsampling of large biomedical data sets (opdis- Downsampling). PLoS One, 16, e0255838.
[3] Ramanathan, N., Ramamurthy, J., Natarajan, G. (2022). Numerical characterization of DNA sequences for alignment-free sequence comparison - A review. Comb. Chem. High T. Scr., 25, 365–380.
[4] Wang, T., Yu, Z.G., Li, J. (2024). CGRWDL: Alignment-free phylogeny reconstruction method for viruses based on chaos game representation weighted by dynamical language model. Front. Microbiol., 15, 1339156.
[5] Bielińska-Wąż, D., Clark, T., Wąż, P., Nowak, W., Nandy, A. (2007). 2D-dynamic representation of DNA sequences. Chem. Phys. Lett., 442, 140–144.
[6] Bielińska-Wąż, D. (2014). Four-component spectral representation of DNA sequences. J. Math. Chem., 47, 41–51.
[7] Wąż, P., Bielińska-Wąż, D. (2014). 3D-dynamic representation of DNA sequences. J. Mol. Model., 20, 2141.
[8] Czerniecka, A., Bielińska-Wąż, D., Wąż, P., Clark, T. (2016). 20D-dynamic representation of protein sequences. Genomics, 107, 16–23.
[9] Bielińska-Wąż, D., Wąż, P. (2017). Spectral-dynamic representation of DNA sequences, J. Biomed. Inform., 72 (2017) 1–7.
[10] Bielińska-Wąż, D., Wąż, P. (2021). Non-standard bioinformatics characterization of SARS-CoV-2. Comput. Biol. Med., 131, 104247.
[11] Panas, D., Wąż,P., Bielińska-Wąż, D., Nandy, A., Basak. S.C. (2018). An Application of the 2D-Dynamic Representation of DNA/RNA Sequences to the Prediction of Influenza A Virus Subtypes. MATCH Commun. Math. Comput. Chem., 80, 295–310.
[12] Bielińska-Wąż, D., Wąż, P., Panas, D. (2022). Applications of 2D and 3D-Dynamic Representations of DNA/RNA Sequences for a Description of Genome Sequences of Viruses. Comb. Chem. High T. Scr., 25, 429–438.
[13] Bielińska-Wąż, D., Wąż, P., Błaczkowska, A., Mandrysz, J., Lass, A., Gładysz, P., Karamon, J. (2024). Mathematical Modeling in Bioinformatics: Application of an Alignment-Free Method Combined with Principal Component Analysis. Symmetry, 16, 967.
[14] Bielińska-Wąż, D., Wąż, P., Błaczkowska, A. (2026). 20D-Dynamic Representation of Protein Sequences Combined with K-means Clustering. Comb. Chem. High T. Scr., 29, 929–942.
[15] Gładysz, P., Bielińska-Wąż, D., Wąż, P., Samorek-Pieróg, M., Karamon, J., Lass, A. (2026). Euro-Asian hybrids of Echinococcus multilocularis from red foxes in northern and northeastern Poland result from secondary contact between long-isolated populations. Scientific Reports, 16, 9986.
[16] Bielińska-Wąż, D., Wąż, P. Lass, A. Karamon, J. (2022). 4D-Dynamic Representation of DNA/RNA Sequences: Studies on Genetic Diversity of Echinococcus Multilocularis in Red Foxes in Poland. Life, 12, 877

Author Index

A

A. Fırat Özdemir, 101
Abdullah Yalçınkaya, xxvi, 163
Agustin Mayo-Iscar, x, xiii, xxv, 8, 74
Agustín Mayo-Iscar, 35, 37
Alan A. Cohen, 95
Alberto González-Sanz, xiv, 7, 29, 31
Aldo Corbellini, xiii, 33, 36
Ana M. Bianco, 28
Andrea Cerioli, x, xvii, 35, 72, 74
Angel Lopez Oriona, xx, 166
Ángel López-Oriona, 122, 173
Anthony C. Atkinson, 36
Anthony Christidis, 26
Antonio Panico, xxi, 125, 128
Arthur Daisomont, xvii, 72, 76
Aylin Alın, x, xiv, 12

B

Badr-Eddine Chérief-Abdellatif, xiv, xxii, 68, 69, 96, 99
Bent Nielsen, 106
Birdal Şenoğlu, 126, 127
Bohan Wu, 31
Bruno A. Pimentel, 61
Burak Dilber, xv, 100, 101
Burcu Koca Guler, xxii, 77, 80

C

Can Hakan Dağıdır, xvii, 72, 73
Ceylan Yozgatlıgil, 80, 81
Christian Hennig, xiii, 33, 34, 35
Christophe Croux, x, 124
Christopher G. Small, 52
Claudio Agostinelli, x, xxii, 96
Columbia SOH Group, 95
Conceição Amado, xxv, 24, 28
Cynthia Rush, xiv, 29, 32

D

Daniel Malinsky, 95
David Bounie, 131
David E. Tyler, x, xvi, 1, 4
David Iseri Inouye, xxii, 77, 78
Diogo Pinheiro, 57
Domenico Perrotta, x, xvii, 36, 72
Dorota Bielinska Waz, xx, 166
Dorota Bielińska-Wąż, 167, 174
Dušan Pokorný, 120

E

Elgiz Askeroğlu, xvii, 39, 44
Elvezio Ronchetti, xiii, xx, 1, 5
Erik Mendroš, xviii, 115, 118
Ersin YILMAZ, 161, 164
Eufrasio Lima Neto, xvi, 54

F

F. Sevinc Kurnaz, x, xviii, xxi, 85, 108
Fabio Centofanti, xxv, 10, 24, 25, 73, 76
Fatma Noyan Tekeli, 103
Fatma Zehra Doğru, xvii, 39, 41, 44
Filip Bočinec, xiv, 12, 18, 170
Francesca Greselin, 66
Fulya Gokalp Yavuz, xii, xv, xxii, 77, 80, 81, 100
Fulya Gokalp Yavuz[1], 110

G

Gabriela Cohen Freue, xxiii, xxv, 24, 26, 50
García-Escudero, xiv, 11, 33, 37, 38, 67, 157
Getoar Sopa, xiv, 68, 70
Gianfranco Piscopo, xx, 166, 169
Giorgia Zaccaria, xv, 63, 66
Giovanni C. Porzio, 88
Giulia Bertagnolli, xxii, 96, 97
Gokalp Tepekoylu, xxiii, 145
Gökalp TEPEKÖYLÜ, 152
Gulder Kemalbay, xv, 100, 105
Gunseli Aytac Cankurtaran, xxiv, 136
Günseli Aytaç Cankurtaran, 141

H

Han Ming Wu, xxiv, 156
Han-Ming Wu, 158
Hanna Jankowski, 132
Hasan Bulut, xxiv, 156
Hasan BULUT, 157
Hernando Ombao, 122
Hira L. Koul, xvi, 45, 47
Houyem Demni, xxi, 85, 88
Hüseyin Ünözkan, 152
Hyemin Yeon, 18, 170

I

Igor Böhm, xx, 166, 170
Ines Wilms, 65
Iris Aragon Mladosich, xix, 121
Iris Aragón Mladosich, 124
Isabel M. Rodrigues, 28

Ismail Yenilmez, xxiv, xxvi, 156, 163
İsmail YENİLMEZ, 161, 164

J

Jairo Diaz-Rodriguez, 132
Jan Picek, xvii, 45, 48
Jana Jureckova, x, xvi, 45, 46
Jens Klooster, xv, 100, 106
Jeremy Oguamalam, 64
Jeyshinee Pyneeandee, 26
Jingyu Cui, 93
John E Puerto Pinzón, 171
Joni Virta, 51
Juntong Chen, 71
Juraj Marusic, 32

K

Kamélia Daudel, 98
Kankanamge Madushanka, 109
Karen Kafadar, xvi, 1, 3
Kelly Ramsay, xxiii, 130, 132
Kihyun Han, 92

L

Leonardo Leone, xxiii, 130
Leonardo Leone[1], 131
Leopold Micheler, xviii, 115, 117
Liangyuan Hu, x, xviii, 20
Lorenzo Benzakour, 66
Luca Greco, 97
Lucia Trapote, 37
Luigi Grossi, 128
Luis Angel García Escudero, xiii, 33
Luis Angel Garcia-Escudero, 74
Luis Angel García-Escudero, 35

M

M. Rosário Oliveira, x, xvi, 54, 57
M. Tariqul Hasan, 93
Maral Karbaschi, xviii, 108, 114
Marcel Nutz, 31
Marco Avella Medina, xvii, 32, 115, 116
Marco Avella-Medina, xiv, 68
Marco Riani, 35, 36
Marcus Mayrhofer, xiii, 8, 9, 117
Maria Bianco, x, xxv, 24
Maria do Rosário Oliveira, xvi, 54
Maria Longobardi, 169
Maria-Pia Victoria Feser, 135
Maria-Pia Victoria-Feser, xi, xxiii, 130
Marko Voutilainen, 51
Martin Picard, 95
Massimiliano Giacalone, 169
Matias Salibian-Barrera, 27
Matthias Templ, xxi, 85, 86
Medhi Hirari, 10
Mehmet Yılmaz, xi
Meral Çetin, xxiv, 136, 137
Mia Hubert, xi, xiii, xv, xvi, 8, 10, 63, 73, 76, 88
Mouhamadou Lamine Ndao, 159
Müge Mutiş, xiv, 12, 14
Mustafa Gök, 107
Mutlu Altuntaş, xviii, 108, 112, 140

N

Nabamallika Dehingia, xxiii, 145, 149
Ndeye Niang, xxiv, 156
Ndèye Niang, 159
Nuri Berk Ural, xv, 100, 107

O

Olcay Arslan, x, xi, xvi, xvii, 39, 40, 41, 45, 49, 53
Olcay ARSLAN, 84
Omer L. Gebizlioglu, 22
Omer Lütfi Gebizlioglu, xviii, 20
Onur Camli, xxiii, 145, 155
Onur Toka, xii, xviii, xxi, 85, 89, 121, 143
Oscar O Melo Martínez, 171
Özge Gürer, xx, 125, 126, 127
Özge Ustahüseyin, xxiv, 136
Özlem Kaymaz, xxii, 145
Özlem Türkşen, 141

P

Palaniappan Vellaisamy, 47
Pedro Duarte Silva, xvi, 54
Peter Filzmoser, x, xi, xiii, xv, xxii, 8, 63, 64, 117
Peter J. Rousseeuw, 1, 2, 73, 88
Peter Ruckdeschel, xxiii, 130, 133
Piotr Waz, xx, 166
Piotr Wąż, 167, 174

R

Renata M.C.R. Souza, 61
Renata Maria Cardoso Rodrigues Souza, xvi, 54
Renjun Ma, xxv, 91, 93

S

Sandra E Melo Martínez, 171
Sandra Esperanza Melo, xx, 166
Sara Suarez-Fernandez, 57
Sascha Desmettre, 133
Şenay Özdemir, xxii, 77
Şenay ÖZDEMİR, 84
Senol Celik, xxiv, 136
Şenol Çelik, 136, 140
Serenay Çakar, xxii, 77, 81
Serhan Tunçel, xxiv, 136, 137, 143
Shojaeddin Chenouri, xxiii, 50, 52
Shunan Sheng, xiv, 29, 31
Sophie Langer, xiv, 68, 71

Stanislav Nagy, xvii, 18, 115, 118, 120, 170
Stefan Van Aelst, xi, xv, xxv, 10, 24, 27, 63
Şükrü Acıtaş, xxiv, 156

T

Talha Arslan, xviii, 121, 123
Tanya Garcia, xxv, 91
Tianjun Ke, 116
Tobias Leopold, xxiii, 145, 146
Tomas Masak, 9, 64
Tony Sarre, 159
Tuğba Kapucu, xviii, 108, 110

U

Ufuk Beyaztas, xiv, 12, 13
Una Radojičić, 9, 64

V

Valentin Todorov, xii, xiii, 33, 36
Valerio Ficcadenti, 169
Varnan Sarangian, 52

X

Xiao Fang, xviii, 20, 21
Xin Dang, xviii, 108, 109

Y

Yangzhuoran Fin Yang, xv, 63, 65
Yanyuan Ma, xii, xxii, xxv, 1, 6, 91, 92
Yasemin Gunter, xv, 100
Yasemin Günter, 103
Yeşim Güney, xvi, xvii, 39, 45, 49
Yetkin Tuaç, xvii, xx, xxiii, 45, 49, 50, 53, 125
Ying Sun, 122, 173
Ying Wei, xxv, 91, 95
Yongli Sang, 109
Yuming Zhang, 135

Z

Zain Khan, 95
Zdeněk Hlávka, xiv, 12, 15
Ziling Ma, xviii, 121, 122
Zi-Yao Su, 21
Zoraida F. Rico, 30
Zoraida Rico, xiv, 29