



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

CONFERENCE PROCEEDINGS

Santiago de Compostela, 14th – 17th July 2026



DEPARTAMENTO DE ESTATÍSTICA,
ANÁLISE MATEMÁTICA E OPTIMIZACIÓN



Seio
Rede de Estatística
e Investigação
Operativa

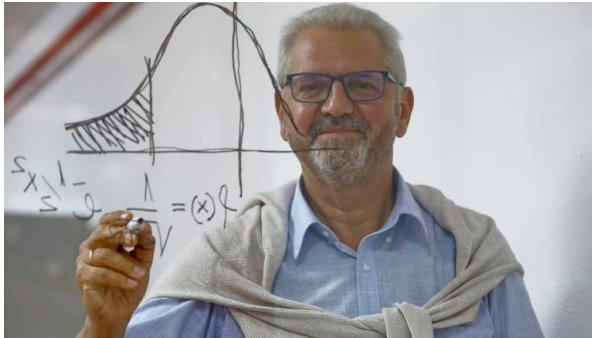


uc3m | Universidad Carlos III de Madrid
Instituto Flores de Lemus

Contents

About this workshop	2
Committees	4
Scientific Committee	4
Organizing Committee	4
Schedule	5
Summary	5
Tuesday, 14 July	6
Wednesday, 15 July	6
Thursday, 16 July	7
Friday, 17 July	7
Invited talks	9
Contributed talks	22
Posters	53
Index	69

About this workshop



On the occasion of Prof. González-Manteiga's retirement, the Working Groups on Functional Data Analysis, Nonparametric and Spatio-Temporal Statistics of the Sociedad de Estadística, Investigación Operativa y Ciencia de Datos (SEIO) have organized this workshop in his honour, bringing together colleagues, former students and friends to celebrate a distinguished academic career that has had a profound and lasting impact on modern Statistics.

Wenceslao González-Manteiga (São Paulo, Brazil, 1956) graduated in Mathematics from the Universidade de Santiago de Compostela in 1979. He received his Ph.D. from the same university in 1983 with the dissertation *Construcción axiomática, consistencias y distribuciones asintóticas de estimadores no paramétricos para funciones de densidad y de regresión*, supervised by José Antonio Cristóbal Cristóbal. He subsequently joined the Department of Statistics and Operations Research, where he developed his entire academic career, becoming Full Professor in 1992.

One of Professor González-Manteiga's most enduring achievements has been the creation and consolidation of the school of nonparametric statistics in Galicia. Under his leadership, this group has evolved into one of Spain's leading research schools in the field, training several generations of researchers and making a significant contribution to the development of nonparametric statistics. Nonparametric inference has been the central focus of his research, while his contributions have also encompassed goodness-of-fit testing, resampling methods, censored and incomplete data models, spatial point processes, ROC curve analysis, robust estimation, and small area estimation, among many other topics. These methodological developments have been always closely linked to applications in biomedicine, environmental sciences, economics, energy, economy and official statistics.

Throughout his career, he has supervised more than thirty Ph.D. dissertations and led numerous competitive research projects, contributing to the training of new generations of researchers and establishing an extensive network of national and international collaborations. His commitment to the advancement of Statistics has extended well beyond research and teaching, playing a leading role in the creation and consolidation of scientific institutions such as the *Sociedade Galega para a Promoción da Estatística e da Investigación de Operacións* (SGAPEIO) and the *Instituto Tecnolóxico de Matemática Industrial* (ITMATI), now integrated into the *Centro de Investigación e Tecnoloxía Matemática de Galicia* (CITMAga). Since 2009, he has also been a member of the *Consello Galego de Estatística*.

His scientific achievements have been recognized through numerous national and international distinctions. In 2017, he was elected an IMS Fellow of the Institute of Mathematical Statistics and became a Full Member of the *Real Academia Galega de Ciencias*. In 2020, he received the Medal of the Sociedad de Estadística e Investigación Operativa (SEIO) and the

SEIO–Fundación BBVA Award for the Best Methodological Contribution to Statistics. The following year, he was awarded the Spanish National Statistics Award in recognition of his contributions to the development of nonparametric statistics. In 2024, he was elected a Corresponding Member of the *Real Academia de Ciencias Exactas, Físicas y Naturales de España*.

Beyond his scientific contributions and the many distinctions he has received, Professor González-Manteiga's greatest legacy lies in the people and the scientific community he has helped to build. His intellectual rigor, academic generosity, and ability to inspire new ideas and foster collaborations have influenced several generations of researchers, many of whom continue to develop and expand the lines of research he helped establish. This conference is therefore not only a tribute to an outstanding academic career, but also a celebration of a way of understanding research founded on excellence, collaboration, and a lasting commitment to the advancement of Statistics.

As the Organizing Committee, we are delighted to bring together colleagues, former students and friends in celebration of Professor Wenceslao González-Manteiga's remarkable career. We hope that this workshop and the contributions gathered in this volume stand as a fitting tribute to his outstanding scientific achievements and to the lasting influence he continues to have on the field of Statistics.

Committees

Scientific Committee

- Ana M. Aguilera (Universidad de Granada)
- José M. Angulo (Universidad de Granada)
- Mario Francisco Fernández (Universidade da Coruña)
- Beatriz Pateiro López (Universidade de Santiago de Compostela)
- José Luis Torrecilla (Universidad Autónoma de Madrid)
- Lola Ugarte (Universidad Pública de Navarra)

Organizing Committee

- María Alonso Pena (Universidade de Santiago de Compostela)
- Maribel Borrajo García (Universidade de Santiago de Compostela)
- Mercedes Conde Amboage (Universidade de Santiago de Compostela)
- Isabel Fuentes Santos (Universidad de La Laguna)
- María José Ginzo Villamayor (Universidade de Santiago de Compostela)
- Andrea Meilán Vila (Universidad Carlos III de Madrid)
- Silvia Novo Díaz (Universidad Carlos III de Madrid)
- Belén Pulido Bravo (Universidad Nacional de Educación a Distancia)

Editors: Organizing Committee

Conference website: <https://sites.google.com/view/fnpsts/>

Schedule

Summary

Tuesday, 14th July	Wednesday, 15th July	Thursday, 16th July	Friday, 17th July
09:30 Registration 09:45 Opening 10:15 Invited Talk I	09:30 Invited Talk III 10:30 Contributed Session IV	09:30 Invited Talk V 10:30 Contributed Session VI	10:00 Invited Talk VII 11:00 Contributed Session VIII
11:15 Coffee Break	11:20 Coffee Break	11:20 Coffee Break	11:50 Coffee Break
11:45 Contributed Session I	11:50 Invited Talk IV 12:50 Contributed Session V	11:50 Contributed Session VII	12:20 Invited Talk VIII 13:20 Contributed Session IX
13:25–15:00 Lunch	13:25–15:00 Lunch	13:25–15:00 Lunch	14:10–15:30 Lunch
15:00 Invited Talk II 16:00 Contributed Session II	15:00 Social Activity	15:00 Invited Talk VI 16:00 Poster Session	15:30 Contributed Session X
16:50 Coffee Break 17:20 Contributed Session III 19:30 Welcome Reception		17:00 Coffee Break 21:00 Conference Dinner	17:10 Closing

Tuesday, 14th July

- 09:30–09:45 Registration
- 09:45–10:15 Opening
- 10:15–11:15 **Invited talk I** *Chair: Silvia Novo*
Some hypothesis testing on functional time series (**Manuel Febrero-Bande**)
- 11:15–11:45 **Coffee Break**
- 11:45–13:25 **Contributed Session I** *Chair: Luis A. Rodríguez*
 - Independent component analysis approaches for enhancing functional linear regression estimation (**Ana M. Aguilera**)
 - Dimensionality reduction and beta functional glucose modeling (**Pedro Delicado**)
 - Fertility rate estimation: combining functional data and small area models (**Esther López-Vizcaíno**)
 - Goodness-of-fit asymptotics under weak and strong correlated functional data (**María D. Ruiz-Medina**)
- 13:25–15:00 **Lunch**
- 15:00–16:00 **Invited talk II** *Chair: Beatriz Pateiro*
Imbalance classification under capacity constraints (**Ricardo Fraiman**)
- 16:00–16:50 **Contributed Session II** *Chair: Isabel Fuentes*
 - Kernel-based goodness-of-fit tests in functional regression (**Daniel Diz-Castro**)
 - Structural multichannel frequency-modulated Möbius models for electroencephalographic signal analysis (**Christian Canedo**)
- 16:50–17:20 **Coffee Break**
- 17:20–18:10 **Contributed Session III** *Chair: Silvia Novo*
 - Tissue Doppler imaging as a predictor of cardiotoxicity in breast cancer patients: functional data analysis approach (**Beatriz Piñeiro-Lamas**)
 - Uniform confidence bands for conditional colocalization curves (**Luis A. Rodríguez-Ramírez**)
- 19:30 Welcome Reception (Pazo de Fonseca)

Wednesday, 15th July

- 9:30–10:30 **Invited talk III** *Chair: Isabel Fuentes*
Local summary statistics and Dirichlet processes for spatial intensity estimation and automatic cluster detection (**Jorge Mateu**)
- 10:30–11:20 **Contributed Session IV** *Chair: Isabel Fuentes*
 - Nonlinear panel model with unobserved stochastic trends: an application to the environmental Kuznets curve (**Cheyenne Amoroso**)
 - Spatio-temporal air quality risk modelling (**María Bugallo**)
- 11:20–11:50 **Coffee Break**

- 11:50–12:50 **Invited talk IV** *Chair: Maribel Borrajo*
Nonparametric inference with low quality data (**María D. Martínez-Miranda**)
- 12:50–13:40 **Contributed Session V** *Chair: Maribel Borrajo*
 - Nonparametric test for the dependence of spatial point processes on covariates (**Isabel Fuentes-Santos**)
 - Prediction intervals for linear and non-linear quantile autoregression (**César Sánchez-Sellero**)
- 13:40–15:00 **Lunch**
- 15:00 Social activity (visita Illa de Ons)

Thursday, 16th July

- 9:30–10:30 **Invited talk V** *Chair: Andrea Meilán*
On an extension of the Cox model for time-varying covariates under dependent censoring with unknown association (**Ingrid Van Keilegom**)
- 10:30–11:20 **Contributed Session VI** *Chair: Andrea Meilán*
 - Defying the curse of dimensionality with higher-order kernels (**José E. Chacón**)
 - Dependence structures for random vectors (**Juan A. Cuesta-Albertos**)
- 11:20–11:50 **Coffee Break**
- 11:50–13:30 **Contributed Session VII** *Chair: José E. Chacón*
 - Goodness-of-fit tests under dependent censoring (**Adrián Lago**)
 - Optimal convex M-estimation for censored linear regression (**Paula Soto-Rodríguez**)
 - Minimax manifold estimation for hypersurfaces under convexity-type restrictions (**Héctor González-Vázquez**)
 - Covariate-dependent multivariate reference regions: modelling and inference via distributional regression (**Óscar Lado-Baleato**)
- 13:25–15:00 **Lunch**
- 15:00–16:00 **Invited talk VI** *Chair: María J. Ginzo*
Bagging cross-validated bandwidth selection in nonparametric curve estimation (**Ricardo Cao**)
- 16:00–17:00 **Poster Session** *Chair: María Alonso*
- 17:00–17:30 **Coffee Break**
- 21:00 Conference dinner (Pazo de San Lorenzo)

Friday, 17th July

- 10:00–11:00 **Invited talk VII** *Chair: Belén Pulido*
On variable selection in functional data analysis (**Antonio Cuevas**)
- 11:00–11:50 **Contributed Session VIII** *Chair: Belén Pulido*

- From trajectories to spatio-temporal fields under preferential sampling (**Raquel Menezes**)
- Neyman orthogonal model checking in generalized partially linear models (**Rui Costa-Miranda**)
- 11:50–12:20 **Coffee Break**
- 12:20–13:20 **Invited talk VIII** *Chair: María Alonso*
ROC curves, robustness and functional data (**Ana M. Bianco**)
- 13:20–14:10 **Contributed Session IX** *Chair: María Alonso*
 - On the spherical cardioid distribution and its goodness-of-fit (**Eduardo García-Portugués**)
 - A pivotal goodness-of-fit test for the spherical Cauchy family (**Diego Bolón**)
- 14:10–15:30 **Lunch**
- 15:30–17:00 **Contributed Session X** *Chair: Jose Ameijeiras*
 - An RKHS-based method for testing regression model specification (**María Vidal-García**)
 - Generalized distance covariance for linearity and independence in scalar-on-function regression with missing at random responses (**Pedro Galeano**)
 - Quadratic specification tests with high-dimensional nuisance parameters (**Juan C. Escanciano**)
 - Tests for regression models based on the error distribution (**Juan C. Pardo-Fernández**)
- 17:10–17:30 Closing

Invited talks

Tuesday 14th July



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

A review of some tools for Functional Time Series

14th July
10:15–11:15
Invited talk I

Manuel Febrero-Bande¹ (✉), Wenceslao González-Manteiga¹, María Dolores Ruiz-Medina²,
Morteza Amini³ and Mohamad Darbalaei^{1,3}

¹ Department of Statistics, Mathematical Analysis and Optimization, University of Santiago de Compostela, Spain; ² Department of Statistics and Op. Research, University of Granada, Spain; ³ University of Tehran, Iran
`manuel.febrero@usc.es`

Functional Time Series (FTS) has evolved significantly since the foundational functional autoregressive (FAR) model introduced by [1] to extensions considering different lagged structures of temporal dependence. As it happens in a multivariate context, there is a need on developing tools for analyzing and testing temporal dependence that may consider the two dimensions of complexity: functional nature of the objects and possibly nonlinear dependence structure. Two main tools are revised in this work. A test based on random projections for checking FTS models (using similar techniques as in [2]) and a diagnostic tool for determining the lag dependence of a FTS based on distance correlation procedures as in the work of [5] or on computing the norms of the lagged autocovariance operators (see, for instance, [3] or [4]).

Keywords: Functional Time Series; Functional Autoregressive Process; Distance Correlation; Testing.

Acknowledgements This research has been partially supported by PID2024-158017NB-I00, financed by MCIU/AEI/10.13039/501100011033 and by FEDER UE, and by grant ED431C 2021/24 (Grupos competitivos) financed by Xunta de Galicia through European Regional Development Funds (ERDF).

Bibliography

- [1] D. Bosq. *Linear Processes in Function Spaces. Theory and Applications*. Lecture Notes in Statistics, Springer-Verlag, (2000).
- [2] J.A. Cuesta-Albertos, E. García-Portugués, M. Febrero-Bande and W. González-Manteiga. *Goodness-of-fit tests for the functional linear model based on randomly projected empirical processes*. The Annals of Statistics, **47**(1) (2019) 439–467. doi:10.1214/18-AOS1693
- [3] P. Kokoszka and M. Reimherr. *Determining the order of the functional autoregressive model*. Journal of Time Series Analysis, **34**(1) (2012) 116–129. doi:10.1111/j.1467-9892.2012.00816.x
- [4] G. Mestre, J. Portela, G. Rice, A. Muñoz San Roque and E. Alonso. *Functional time series model identification and diagnosis by means of auto and partial autocorrelation analysis*. Computational Statistics & Data Analysis, **155**:107108 (2021). doi:10.1016/j.csda.2020.107108
- [5] Z. Zhou *Measuring nonlinear dependence in time-series, a distance correlation approach*. Journal of Time Series Analysis, **33**(3) (2012) 438–457. doi:10.1111/j.1467-9892.2011.00780.x



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Inbalanced Classification under Capacity Constrains

Daniel Fraiman¹ (✉), Ricardo Fraiman¹

¹Departamento de Matemática y Ciencias, Universidad de San Andrés, Argentina;
fraimanricardo@gmail.com

14th July
15:00–16:00
Invited talk II

Detecting observations from a minority class under severe class imbalance is a central challenge in applications such as fraud detection, medical screening, and industrial quality control. In these settings, each positive prediction triggers a costly follow-up action, an MRI scan, a transaction audit, whose execution is subject to real operational constraints. This paper proposes a formal classification framework under capacity constraints: given a user-defined bound limit b on the proportion of observations that can be labeled as belonging to the minority class, the goal is to find the classifier that maximizes sensitivity on that class. We characterize the optimal classifier under this constraint and establish its equivalence with the classical Bayes classifier under a reweighting of the prior probabilities. We also introduce a capacity-adjusted performance metric M that accounts for the effective detection rate when the capacity constraint is binding. The framework is implemented on top of standard learning methods, k-NN, SVM, random forests, and neural networks, and statistical consistency is established for each. We further introduce a capacity-aware support vector machine that exploits the constraint during training and achieves the strongest empirical performance. Experiments on the Taiwanese credit card default dataset confirm that capacity-constrained classifiers substantially outperform classical approaches, *Keywords*: keyword1; keyword2; keyword3; keyword4; keyword5.

Bibliography

- [1] Chawla, N. V.; Bowyer, K. W.; Hall, L. O.; Kegelmeyer, W. P. *SMOTE: Synthetic Minority Over-sampling Technique*. Journal of Artificial Intelligence Research. **16** (2011) 321-357.
- [2] Devroye, L., Györfi, L. and Lugosi, G. *A Probabilistic Theory of Pattern Recognition*. Springer. (2013)
- [3] He, H. and Garcia, E. A. *Learning from imbalanced data*. IEEE Transactions on Knowledge and Data Engineering **21** 1263-1284 (2009).
- [4] Steinwart, I. *Consistency of support vector machines and other regularized kernel classifiers*. **51** 128-142 (2009).

Wednesday 15th July



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Local summary statistics and Dirichlet processes for spatial intensity estimation and automatic cluster detection

15th July
09:30–10:30
Invited talk III

Jorge Mateu

Department of Mathematics, Universitat Jaume I, Castellon, Spain

`mateu@uji.es`

Common practice in spatial point process modelling dictates that formal analysis begins with intensity estimation, which is carried out by exploiting external covariates, when available. Parametric intensity modelling often uses the Poisson likelihood function, but this underperforms when the data come from a more complex model, with some kind of interaction among points. To address this shortcoming, we propose a method that incorporates local second-order characteristics to account for spatial dependencies in the model fitting procedure. Our method relies on a locally weighted Poisson log-likelihood, which avoids making explicit assumptions about the type and degree of spatial interaction. We further propose a non-parametric test for the detection of interaction between points. In addition, when the intensity outlines the existence of a clustering behaviour, we propose a spatio-temporal Dirichlet process mixture model on Euclidean and linear network windows to automatically detect space-time clusters of the events. We consider a Bayesian hierarchical model and adopt a convolution kernel estimator to account for the network structure in the city. Several applications will be illustrated.

Keywords: Dirichlet process; intensity estimation; linear network, local characteristics; Markov chain Monte Carlo; spatial point patterns

Acknowledgements This work has been funded by grant PID2022-141555OB-I00 from Ministry of Science and Innovation.

Bibliography

- [1] N. D'Angelo, G. Adelfio, J. Mateu, O. Cronie. *Local summary statistics for spatial point process intensity estimation*. Spatial Statistics (2026). doi:10.1016/j.spasta.2026.100986.
- [2] S. Lee, W. Chang, J. Mateu, H. Lee, J. Park. *A spatio-temporal Dirichlet process mixture model on linear networks for crime data*. Journal of the Royal Statistical Society: Series A (2026). doi:10.1093/jrssa/qnag052.
- [3] J. Park, S. Yi, W. Chang, J. Mateu. *A spatio-temporal dirichlet process mixture model for Coronavirus disease-19*. Statistics in Medicine. **42(30)** (2023) 5555-5576.



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Nonparametric inference and forecasting with low quality data

15th July
11:50–12:50
Invited talk IV

María Dolores Martínez Miranda¹ (✉)

¹Department of Statistics and Operations Research, University of Granada, Spain

`mmiranda@ugr.es`

It is not uncommon that we would like to forecast certain events before they happen. These events may include the number of crimes of a particular type, cases of insurance fraud, new hospital admissions during a pandemic, or the numbers of births and deaths in a country or worldwide. In all these situations, the information available for statistical inference and forecasting is often of low quality: data are typically incomplete or inconsistently reported; only part of the relevant information is collected; available data are often aggregated through occurrence and exposure measures; and the definition of exposure may change over time. In this talk, I will review some recent work and ongoing research motivated by applications in survival analysis, demography and insurance. Despite their different contexts, these problems share a common challenge: developing consistent and reliable inference and forecasts from available but low-quality data. The talk will also revisit several projects developed jointly with Professor González-Manteiga, whose influence has shaped much of the research presented here.

Keywords: age-period-cohort models; counting processes; local linear estimation; survival analysis.

Acknowledgements This work was supported by Project PID2024-158017NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by FEDER, UE.

Bibliography

- [1] M.L. Gámiz, E. Mammen, M.D. Martínez-Miranda, J.P. Nielsen. *Double one-sided cross-validation of local linear hazards*. Journal of the Royal Statistical Society Series B. **78** 755–779.
- [2] M.L. Gámiz, E. Mammen, M.D. Martínez-Miranda, J.P. Nielsen. *Monitoring a developing pandemic with available data*. (2023) [arXiv:2308.09919](#).
- [3] M.L. Gámiz, E. Mammen, M.D. Martínez-Miranda, J.P. Nielsen. *Low quality exposure and point processes with a view to the first phase of a pandemic*. (2023) [arXiv:2308.09918](#).
- [4] W. González-Manteiga, M.D. Martínez-Miranda, A. Pérez-González. *The choice of smoothing parameter in nonparametric regression through wild bootstrap*. Computational Statistics & Data Analysis. **47** (2004), 487–515.
- [5] E. Mammen, M.D. Martínez-Miranda, J.P. Nielsen, M. Vogt. *Calendar effect and in-sample forecasting*. Insurance: Mathematics and Economics. **96** (2021) 31–52.

Thursday 16th July



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

When censoring is not innocent: Challenges and solutions for dependent censoring in survival analysis

16th July
09:30–10:30
Invited talk V

Ingrid Van Keilegom¹ (✉)

¹ORSTAT, KU Leuven, Belgium

ingrid.vankeilegom@kuleuven.be

Survival analysis relies on a deceptively simple assumption: censoring is independent of the event of interest. In practice, however, this assumption is often violated. Patients may drop out because their health deteriorates, treatments may be discontinued in response to emerging risk, and follow-up may depend on factors that are themselves related to survival. Such dependent censoring can introduce substantial bias, undermine the validity of standard estimators, and lead to misleading scientific conclusions. In this talk, I will explore why dependent censoring is one of the most challenging and frequently overlooked threats to reliable time-to-event analysis. Using examples from clinical and observational studies, I will illustrate how dependent censoring arises, how it distorts inference, and why conventional survival methods can fail. I will then discuss methodological approaches for addressing dependent censoring, with particular emphasis on copula-based models that explicitly characterize the dependence between survival and censoring times. These models provide a flexible framework for sensitivity analysis and for quantifying the impact of departures from the independent censoring assumption. In addition, I will present recent developments on the partial identification of survival quantities under dependent censoring, as well as new results on identifiability and verifiability in enriched data settings. Dependent censoring will serve as a motivating example of a broader question at the heart of statistical inference: what can and cannot be learned from incomplete data?

Keywords: keyword1; keyword2; keyword3; keyword4; keyword5.

Acknowledgements This work was supported by

Bibliography

- [1] W. Stute, W. González-Manteiga, M.A. Presedo-Quindimil. *Bootstrap approximations in model checks for regression*. Journal of the American Statistical Association. **93(441)** (2011) 141–149. doi:10.1080/01621459.1998.10474096
- [2] J. Ramsay and B. Silverman. *Functional Data Analysis*. Springer Series in Statistics, 3rd Edition. (2020)
- [3] Latex – A document preparation system <http://www.latex-project.org>
- [4] L. Freijeiro-González, M. Febrero-Bande, W. González-Manteiga. *Guidelines for LASSO and derivatives use under different dependence and scale structures*. (2025) arXiv:2506.08582



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Bagging cross-validated bandwidth selection for nonparametric regression

16th July
15:00–16:00
Invited talk VI

Ricardo Cao¹ (✉), Daniel Barreiro² and Mario Francisco¹

¹Group MODES, CITIC, Department of Mathematics, University of A Coruña, Spain; ²NEXT DIGITAL, Spain

`ricardo.cao@udc.es`

Bandwidth selection is a key issue in nonparametric density and regression estimation, as it controls the bias–variance trade-off of kernel estimators. Classical cross-validation methods provide reliable bandwidth choices but suffer from high computational cost and large variability, especially for large sample sizes, since they typically require $O(n^2)$ operations.

This work proposes the use of bagging and subbagging techniques to construct efficient cross-validated bandwidth selectors for the Parzen–Rosenblatt density estimator and the Nadaraya–Watson regression estimator. The approach is based on computing bandwidths over subsamples, appropriately rescaling them, and aggregating the results. The resulting estimators achieve substantial reductions in variability while significantly decreasing computational burden.

Asymptotic properties of the proposed bagging bandwidths are derived, showing improved convergence rates under suitable conditions. Simulation studies confirm that the method outperforms standard cross-validation both in statistical accuracy and computational efficiency. The methodology is further illustrated with an application to COVID-19 data, highlighting its practical usefulness for large datasets.

Keywords: bagging; cross-validation; kernel regression; nonparametric estimation; subsampling.

Acknowledgements This research is part of the Grants PID2020-113578RB-I00 and PID2023-147127OB-I00 "ERDF/EU", funded by MCIN/AEI/10.13039/501100011033/. It has also been supported by the Xunta de Galicia (Grupos de Referencia Competitiva ED431C-2024/14) and by CITIC as a center accredited for excellence within the Galician University System and a member of the CIGUS Network, receives subsidies from the Department of Education, Science, Universities, and Vocational Training of the Xunta de Galicia. Additionally, it is co-financed by the EU through the FEDER Galicia 2021-27 operational program (Ref. ED431G 2023/01).

Friday 17th July



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

On variable selection in functional data analysis

17th July
10:00–11:00
Invited talk VII

Antonio Cuevas¹ (✉)

¹Department of Mathematics, Universidad Autónoma de Madrid, Spain

antonio.cuevas@uam.es

In statistical problems where the available data are functions $x(t)$, with $t \in [0, 1]$, a natural strategy is to apply dimension reduction methods aimed at replacing the functional data with multivariate “projections”, defined on different ways.

Several popular dimension reduction methods (principal components, partial least squares,...) are based on linear continuous projections of type $x \in L^2[0, 1] \mapsto (\langle v_1, x \rangle, \dots, \langle v_p, x \rangle)$, where $v_j \in L^2[0, 1]$ and $\langle \cdot, \cdot \rangle$ denotes the standard L^2 inner product.

In this talk, we will review some recent contributions on the alternative approach of using variable selection (VS), i.e., transformations of type $x = x(t) \mapsto (x(t_1), \dots, x(t_p))$, as a competing paradigm.

We argue that, in addition to the obvious advantages in terms of interpretability, there is now some solid support for the VS-based methodologies, including a suitable theoretical framework and asymptotic results.

Keywords: functional data analysis, impact points, variable selection.

Acknowledgements This work was partially supported by research grant PID2023-148081NB-I00, from the Spanish Government.

Bibliography

- [1] J.R. Berrendero, A. Cuevas, J.L. Torrecilla. *On the use of reproducing kernel Hilbert spaces in functional classification*. Journal of the American Statistical Association **113**(523) (2018) 1210–1218. doi:10.1080/01621459.2017.1320287
- [2] J.R. Berrendero, B. Bueno-Larraz, A. Cuevas. *An RKHS model for variable selection in functional linear regression*. Journal of Multivariate Analysis **170** (2019) 25–45. doi:10.1016/j.jmva.2018.04.008
- [3] J.R. Berrendero, B. Bueno-Larraz, A. Cuevas. *On functional logistic regression: some conceptual issues*. Test **32** (2023) 321–349. doi:10.1007/s11749-022-00836-9
- [4] J.R. Berrendero, A. Cholaquidis, A. Cuevas. *On the functional regression model and its finite-dimensional approximations*. Statistical Papers **56** (2024) 5167–5201. doi:10.1007/s00362-024-01567-9
- [5] J.R. Berrendero, A. Coín, A. Cuevas. *A Bayesian approach to functional regression: theory and computation*. Bayesian Analysis, to appear. doi:10.1214/25-BA1549



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

ROC curves, robustness and functional data

17th July
12:20–13:20
Invited talk VIII

Ana M. Bianco¹ (✉)

¹Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

anambianco@gmail.com

ROC curves are a useful graphical tool for assessing the accuracy of a classification rule, providing an overall view of a classifier's ability to discriminate between two groups.

Outliers may affect the assessment of diagnostic accuracy, motivating the development of robust alternatives. In this talk, we review several results illustrating the impact that outlying observations can have on ROC curve estimation and discuss some robust approaches proposed in the literature.

Our focus is on conditional (covariate-specific) ROC curves in the presence of functional covariates, with particular emphasis on the robustness of the corresponding estimation procedures. Since regression models play a central role in both the direct and induced approaches, atypical observations in either the responses or the covariates may severely affect the resulting estimators. This issue becomes even more challenging in the context of functional data. Our methodology is semiparametric and combines robust regression estimators with weighted empirical distribution estimators based on an adaptive procedure designed to reduce the influence of outlying observations.

Through a numerical study, we illustrate the behavior of robust estimators of the conditional ROC curve in the presence of functional covariates, demonstrating their improved performance relative to classical estimators when the data are contaminated.

Keywords: conditional ROC curve; functional covariates; robustness; induced method; direct method.

Contributed talks

Tuesday 14th July



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Independent component analysis approaches for enhancing functional linear regression estimation

Ana M. Aguilera¹ (✉), Helena Ortiz¹, Christian Acal¹, and Marc Vidal²

¹Department of Statistics and O. R., Institute of Mathematics, University of Granada, Spain;

²Department of Data Analysis, Ghent University, Belgium

aaguiler@ugr.es

14th July
11:45–13:25
Contributed
Session I

Estimating functional linear regression often introduces computational difficulties due to the infinite-dimensional nature and high correlation of functional predictors [1]. While conventional methodology circumvents these issues using dimensionality reduction framework like Principal Component Analysis (PCA) or Partial Least Squares (PLS) [2], these techniques are limited to maximizing variance or response covariance. This study introduces an alternative estimation framework centered on Functional Independent Component Analysis (FICA) [3, 4]. Rather than relying on second-order statistics, the proposed method isolates latent source signals by maximizing kurtosis, offering a deeper look into the underlying data structure. By substituting the original functional curves with a strategic selection of independent components, this approach optimizes the regression framework. We evaluate this methodology using a comprehensive simulation study designed to establish robust criteria for optimal component and basis expansion selection. Ultimately, this framework offers a promising approach for fields handling complex multi-signal compositions, such as chemometrics and neuroscience.

Keywords: functional data; kurtosis operator; independent components; linear regression.

Acknowledgements This work was supported by project PID2023-149087NB-I00 and IMAG María de Maeztu grant CEX2020-001105-M, funded by MICIU/AEI/10.13039/501100011033 and by FEDER, UE.

Bibliography

- [1] J. Ramsay and B. Silverman. *Functional data analysis*. Springer Series in Statistics, 2nd Edition. (2005)
- [2] M. Febrero-Bande, P. Galeano, W. González-Manteiga. *Functional principal component regression and functional partial least-squares regression: An overview and a comparative study*. International Statistical Review. **85** (2017) 61–83. doi:10.1111/insr.12116
- [3] M. Vidal, A. M. Aguilera. *Novel whitening approaches in functional settings*. Stat. **12** (2023) e516. doi:10.1002/sta4.516
- [4] M. Vidal, M. Leman, A. M. Aguilera. *Functional independent component analysis by choice of norm: A framework for near-perfect classification*. Advanced Data Analysis and Classification. (2025) 1–29. doi:10.1007/s11634-024-00622-5



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Dimensionality reduction and Beta functional glucose modeling

14th July
11:45–13:25
Contributed
Session I

Nihan Acar-Denizli¹ and Pedro Delicado¹ (✉)

¹Department of Statistics and Operations Research,
Universitat Politècnica de Catalunya - BarcelonaTECH, Spain

pedro.delicado@upc.edu

We present a functional data analysis approach for studying time-dependent, continuous glucose monitoring data with repeated measures for each individual in a longitudinal study. After scaling the glucose concentration curves to the interval $[0, 1]$, we model them by using a Beta distribution with two time-varying parameters. In this context, we develop a local linear maximum likelihood smoothing procedure that is valid when more than one parameter depends on time. Specifically, we extend local likelihood estimation from one to two time-varying parameters by developing theoretical expressions for estimation and for approximating the leave-one-out cross-validation. Once a Beta distribution with time-varying parameters is estimated for each individual in the sample, we summarize the information contained in the individual estimates by means of multidimensional scaling based on the Integrated Aitchison distance.

We evaluate the performance of our proposal in terms of computation time and model fit using a synthetic dataset as well as a real clinical trial dataset. We compare our method with existing methods in the literature [1, 2, 3, 4] and we conclude that our approach competes with spline-based estimation methods, the dominant line of functional regression models today.

Keywords: Beta functional model; Continuous glucose monitoring; Generalized nonparametric models; Local linear likelihood estimation; Time-varying parameters.

Acknowledgements This work was supported by the project PID2023-148158OB-I00 (MICIU/AEI/10.13039/501100011033 and European Union Next Generation EU/PRTR).

Bibliography

- [1] S. Wood. *Generalized Additive Models: An Introduction with R*, 2nd ed., Chapman and Hall/CRC, New York, (2017).
- [2] I. Gaynanova, N. Punjabi, C. Crainiceanu. *Modeling continuous glucose monitoring (CGM) data during sleep*, Biostatistics (Oxford, England) **23** (2022) 223–239. doi:10.1093/biostatistics/kxaa023
- [3] R. A. Rigby, D. M. Stasinopoulos. *Generalized additive models for location, scale and shape, (with discussion)*, Journal of the Royal Statistical Society Series C: Applied Statistics **54** (2005) 507–554. doi:10.1111/j.1467-9876.2005.00510.x
- [4] N. Acar-Denizli, P. Delicado. *Local constant likelihood estimation for Beta distribution with time varying parameters.*, in: G. Aneiros, E. G. Bongiorno, A. Goia, M. Hušková (Eds.), New Trends in Functional Statistics and Related Fields, Springer Nature Switzerland, Cham, (2025) 7–14.



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Fertility Rate Estimation: Combining Functional Data and Small Area Models

Esther López-Vizcaíno¹ (✉), M José Lombardía² and Cristina Rueda³

¹Instituto Galego de Estatística, Spain; ²Department of Mathematics, University of Coruña, Spain;

³Department of Statistics and Operations Research, University of Valladolid, Spain

`esther.lopez.vizcaino@xunta.gal`

14th July
11:45–13:25
Contributed
Session I

Accurate fertility rate estimation at the municipal level is essential for demographic planning and policy-making, yet small-area projections remain methodologically challenging due to data sparsity and variability. This research addresses these challenges by integrating Functional Data Analysis (FDA) and Small Area Estimation (SAE) techniques to improve fertility rate estimates in Galician municipalities from 2002 to 2022.

We introduce a novel Frequency Modulated Mobius (FMM) model [2], which provides a flexible and interpretable approach for modeling fertility trends. The proposed methodology is evaluated against existing estimation techniques, including splines, and the Dokumentov et al. (2018) model [1]. Simulation experiments and real data analysis demonstrate that FMM models, outperform alternative methods, offering smoother fertility curves and lower estimation errors (RMSE).

To assess the forecasting accuracy of our model, we apply k-fold cross-validation, using data from 2002-2022 to validate predictions. The FMM-based models perform best for fertility prediction (2023-2032), revealing a continuous decline in fertility rates and a shift towards later childbearing in most of the municipalities.

This research contributes a new methodological framework for fertility estimation, providing a powerful tool for demographers, policymakers, and regional planners.

Keywords: small area estimation; Frequency Modulated Mobius model; FDA; fertility rates.

Acknowledgements This research/work is part of the grants PID2020-113578RB-I00 and PID2023-147127OB-I00 "ERDF/EU", funded by MCIN/AEI/10.13039/501100011033/. It has also been supported by the Xunta de Galicia (Grupos de Referencia Competitiva ED431C-2024/14) and by CITIC as a center accredited for excellence within the Galician University System and a member of the CIGUS Network, receives subsidies from the Department of Education, Science, Universities, and Vocational Training of the Xunta de Galicia. Additionally, it is co-financed by the EU through the FEDER Galicia 2021-27 operational program (Ref. ED431G 2023/01)

Bibliography

- [1] A. Dokumentov and R. J. Hyndman (2017). *Smoothing of Two-Dimensional Demographic Data, Optionally Taking into Account Period and Cohort Effects*. R package version 0.2. <http://cran.r-project.org/package=smoothAPC>
- [2] C. Rueda, Y. Larriba and S. D. Peddada (2019). *Frequency Modulated Mobius Model Accurately Predicts Rhythmic Signals in Biological and Physical Sciences*. Scientific Reports. **9**, 18701 (2019). doi:10.1038/s41598-019-54569-1



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Goodness-of-fit asymptotics under weak and strong correlated functional data

María D Ruiz-Medina ¹ (✉)

¹Department of Statistics and Operations Research, University of Granada, Spain

mruiz@ugr.es

14th July
11:45–13:25
Contributed
Session I

This work presents recent advances in the asymptotic analysis of GOF from weak and strong correlated functional data. In the weak-dependent case, under an $AR\mathbb{H}(1)$ process framework, the suitability of a specific bounded linear operator, or of a family of operators under model misspecification is tested. The test statistic involves an empirical process with functional marks, whose parameter (index set), and state space is a separable Hilbert space $\mathbb{H}$. Its convergence to time-changed $\mathbb{H}$ -valued Brownian motion is proved under the null, as well as its consistency under the alternative. The second approach allows fitting Long Range Dependence (LRD). The test statistic is formulated in the spectral domain assuming stationarity. The limit distribution under the null, as well as the consistency of the test are derived, when functional data are fully observed. Under discretely observed functional data, the bandwidth parameter decreasing rate, and the spatial sampling frequency and truncation parameter increasing rates are jointly fitted to the functional sample size increasing, ensuring similar asymptotic GOF results. This work is based on the results obtained in collaboration with Wenceslao González-Manteiga, Rosa M. Crujeiras and Manuel Febrero-Bande (see [1], [2], [3]).

Keywords: Correlated functional data; functional empirical process; functional spectral domain; GOF test asymptotics; weighted periodogram operator.

Acknowledgements This work has been partially supported by grants PID2022-142900NB-I00 funded by MICIU / AEI/10.13039/501100011033 / ERDF, EU, and CEX2020-001105-M, funded by MICIU / AEI/10.13039/501100011033.

Bibliography

- [1] W. González-Manteiga, M.D. Ruiz-Medina and M. Febrero-Bande. *Testing the goodness-of-fit of a functional autoregressive model*. (2023) [arXiv:2303.09644](#)
- [2] M.D. Ruiz-Medina and R.M. Crujeiras. *Testing LRD in the spectral domain for functional time series in manifolds*. *Electronic Journal of Statistics*. **19** (2025) 3601–3642. doi:10.1214/25-EJS2421
- [3] M.D. Ruiz-Medina and R.M. Crujeiras. *An LRD spectral test for irregularly discretely observed contaminated functional time series in manifolds*. (2025) [arXiv:2511.00518](#)



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Kernel-based goodness-of-fit tests in functional regression

Daniel Diz-Castro¹ (✉), Manuel Febrero-Bande¹ and Wenceslao González Manteiga¹

¹Department of Statistics, Mathematical Analysis and Optimization, University of Santiago de Compostela, Spain

daniel diz.castro@usc.com

14th July
16:00–16:50
Contributed
Session II

The literature on goodness-of-fit (GoF) tests for regression models involving functional data and infinite-dimensional parameters is scarce. Testing the linearity of the regression function has been previously discussed in [1], [2] and the references therein. Moreover, the proposal in [3] allows for testing the structure of regression models with compactly supported covariates and reproducing kernel Hilbert space-based estimators. However, a general framework for GoF tests in functional regression without restrictions on the parameter estimators or the domain of the covariates is still missing. In this contribution, we explore a kernel-based approach to GoF tests in regression and show that the theoretical developments introduced in [4] for the Euclidean case can be extended to more general settings, including functional regression models. We present an easily computable test statistic and show that, under suitable assumptions on the regression function, the model errors and the convergence rate of parameter estimators, its asymptotic behavior is the same as in the finite-dimensional setting. The test can also be reliably calibrated using a multiplier bootstrap scheme. A succinct simulation study involving functional data illustrates the finite-sample performance of the proposed method.

Keywords: goodness-of-fit; functional data; kernels.

Acknowledgements This work was supported by the GRC grant ED431C 2025/03, funded by the Xunta de Galicia, and also from projects PID2020-116587GB-I00 and PID2024-158017NB-I00, funded by MICIU/AEI/10.13039/501100011033 and by FEDER, UE.

Bibliography

- [1] J. A. Cuesta-Albertos, E. García-Portugués, M. Febrero-Bande, W. González-Manteiga. *Goodness-of-fit tests for the functional linear model based on randomly projected empirical processes*. The Annals of Statistics, **47**(1) (2019) 439-467. doi:10.1214/18-AOS1693
- [2] V. Patilea and C. Sánchez-Sellero. *Testing for lack-of-fit in functional regression models against general alternatives*. Journal of Statistical Planning and Inference, **209** (2020) 229–251. doi:10.1016/j.jspi.2020.04.002
- [3] A. Sancetta. *Testing subspace restrictions in the presence of high dimensional nuisance parameters*. Electronic journal of statistics, **16**(2) (2022) 5277-5320. doi:10.1214/22-EJS2058
- [4] J.C. Escanciano. *A Gaussian process approach to model checks*. The Annals of Statistics, **52**(5) (2024) 2456-2481. doi:10.1214/24-AOS2443



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Structural multichannel Frequency-Modulated Möbius models for electroencephalographic signal analysis

Christian Canedo¹ (✉) and Cristina Rueda¹

¹Department of Statistics and Operations Research, University of Valladolid, Spain

christian.canedo@uva.com

14th July
16:00–16:50
Contributed
Session II

Electroencephalographic signals provide a non-invasive window into brain activity and play a central role in the assessment of neurological disorders and brain dysfunction. Their multichannel nature reflects the spatial propagation of neural activity across the scalp, creating important challenges for statistical modeling and signal interpretation. Traditional approaches often analyze channels separately or rely on black-box machine learning techniques with limited physiological interpretability.

In this work, we explore the use of the Frequency-Modulated Möbius framework for the analysis of multichannel electroencephalographic recordings [1]. The model provides an interpretable representation of oscillatory patterns through a small number of parameters describing amplitude, phase, shape, and temporal dynamics. Its multichannel extension enables the joint characterization of temporal and spatial information while preserving physiological meaning.

The research is motivated by an ongoing interdisciplinary collaboration involving statisticians, neurologists, and neurophysiologists focused on post-anoxic patients. The main objective is to develop quantitative electroencephalographic biomarkers to improve neurological outcome prediction and support clinical decision-making. Methodological developments include structured spatial representations, parsimonious parameterizations, and source-oriented formulations exploiting dependence between channels.

Keywords: electroencephalography; frequency-modulated möbius models; multichannel signals; statistical signal processing; structural constraints.

Acknowledgements This work was supported by the Spanish Ministry of Science, Innovation and Universities under Grant PID2023-147839OB-I00.

Bibliography

- [1] C. Canedo, C. Rueda. *Physiologically Guided Modeling for EEG Multichannel Signals*. *Bioengineering*. **13**(2) (2026) 131. doi:10.3390/bioengineering13020131



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Tissue Doppler Imaging as a predictor of cardiotoxicity in breast cancer patients: a functional data analysis approach

Beatriz Piñeiro-Lamas¹ (✉), Ana López-Cheda¹ and Ricardo Cao¹

¹Grupo de Modelización, Optimización e Inferencia Estadística, Departamento de Matemáticas, Facultade de Informática, Universidade da Coruña, CITIC, Campus de Elviña, 15071, A Coruña, Spain
b.pineiro.lamas@udc.es

14th July
17:20–18:10
Contributed
Session III

Cardiotoxicity is a side effect of some oncological treatments whose early detection remains a clinical challenge. In this work, a functional data analysis approach is proposed to evaluate the predictive role of baseline Tissue Doppler Imaging using data from 270 breast cancer patients. Using an image preprocessing algorithm [1], the myocardial contraction and relaxation velocity function is obtained for each patient. Through a permutation test [2], the early diastolic filling wave (E') is found to differ significantly between groups: at baseline, patients who later developed cardiotoxicity exhibited a smoother, rightward-displaced wave, consistent with slower and delayed myocardial relaxation. Based on these findings, a scalar parameter related to this wave, defined as the ratio between the timing of the E' peak within the cardiac cycle and its peak velocity, is proposed as a practical tool for baseline risk stratification [2].

Keywords: functional data; image preprocessing; permutation test; risk stratification.

Acknowledgements This work, integrated into the framework of PERTE for Vanguard Health, has been co-financed by the Spanish Ministry of Science, Innovation and Universities with funds from the European Union NextGenerationEU, from the Recovery, Transformation and Resilience Plan (PRTR-C17.I1) and from the Autonomous Community of Galicia within the framework of the Biotechnology Plan Applied to Health. The authors acknowledge partial support by the MICINN Grant PID2023-147127OB-I00, partial support of Xunta de Galicia (Grupos de Referencia Competitiva ED431C 2024/14) and partial support from the Centro de Investigación de Galicia “CITIC”, funded by Xunta de Galicia and the European Union European Regional Development Fund (ERDF) - Galicia 2021–2027 Program, Grant ED431G 2023/01.

Bibliography

- [1] Piñeiro-Lamas B, López-Cheda A, Cao R, Ramos-Alonso L, González-Barbeito G, Barbeito-Caamaño C, Bouzas-Mosquera A (2023). A cardiotoxicity dataset for breast cancer patients. *Scientific Data*, 10, 527. doi:10.1038/s41597-023-02419-1
- [2] Piñeiro-Lamas B, López-Cheda A, Cao R, Lesta-Mellid R, Bouzas-Mosquera A, Barbeito-Caamaño C (2025). Tissue Doppler Imaging as a predictor of therapy-related cardiac dysfunction in breast cancer patients. *International Journal of Cardiovascular Imaging*, 41(8), 1505–1512. doi:10.1007/s10554-025-03439-1



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Uniform confidence bands for conditional colocalization curves

14th July
17:20–18:10
Contributed
Session III

Santiago Arenas Velilla¹, Axel Munk^{1, 2} and Luis Alberto Rodríguez Ramírez¹ (✉)

¹Institute for Mathematical Stochastics, Georg August-Universität Göttingen, Germany; ²Cluster of Excellence “Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells” (MBExC), University Medical Center, Robert-Koch-Straße 40, 37075 Göttingen, Germany

`luis.rodriguez@uni-goettingen.de`

Entropic optimal transport (EOT) has emerged as a powerful tool for the analysis of high-dimensional and structured data due to its computational tractability and statistical robustness. Beyond its established use in machine learning and imaging sciences, EOT-based methodologies have recently been introduced in cell biology to quantify spatial interactions between protein assemblies through colocalization analysis. In this work, we extend the framework developed in [1] to the multimarginal setting and introduce a conditioned multimarginal colocalization process based on entropic transport couplings. We establish asymptotic weak convergence results for this process and derive corresponding inferential procedures, including uniform confidence bands and bootstrap consistency. The proposed methodology is validated through simulation experiments and applied to mitochondrial protein colocalization data.

Keywords: delta method; empirical process; Hadamard differentiability; non parametric inference.

Acknowledgements This work was supported by the Deutsche Forschungsgemeinschaft through DFG RU 5381.

Bibliography

- [1] Arenas-Velilla, S., Munk, A., & Rodríguez, L. A. *Distributional Convergence of Empirical Entropic Optimal Transport and Statistical Applications*. (2026). [arXiv:2604.12940](https://arxiv.org/abs/2604.12940)

Wednesday 15th July



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Nonlinear panel model with unobserved stochastic trends: an application to the environmental Kuznets curve

Cheyenne Amoroso¹ (✉), Carolina García-Martos² and Siem Jan Koopman³

¹Department of Mathematics, Universidade da Coruña, Spain;

²Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid, Spain;

³Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, The Netherlands.

`cheyenne.amoroso@udc.es`

15th July
10:30–11:20
Contributed
Session IV

This paper introduces a dynamic nonlinear panel data model with unobserved stochastic trends to overcome econometric challenges in estimating the Environmental Kuznets Curve (EKC). Conventional EKC literature [1] relies on polynomial specifications, which risk spurious relationships when applied to integrated processes [2]. To address this, we replace polynomials with a bounded, nonlinear difference-of-sigmoids function that separates the emissions pressure of economic growth from the offsetting force of the green transition. Furthermore, unobserved country-specific stochastic trends are incorporated to account for time-varying features, such as technological progress and policy shifts. By formulating the model in state-space form [3], we evaluate the likelihood and extract unobserved trends using the Kalman filter.

Applying this methodology to a panel of CO₂ emissions and GDP for 15 major economies, we identify a global EKC pattern while avoiding spurious regression pitfalls. Notably, we find that European countries began their green transition at an earlier stage of economic development, reaching their CO₂ peak at a lower level than comparable non-European economies. Introducing a regional parameter shift captures this structural difference while leveraging the full panel to robustly identify the global EKC.

Keywords: environmental Kuznets curve; non-linear panel data; state-space models.

Acknowledgements This work was supported by INDITEX–UDC 2024 grant for predoctoral research stays, during which part of this work was carried out, Xunta de Galicia (Competitive Reference Groups, ED431C–2024/14) and CITIC (CIGUS Network), subsidised by the Xunta de Galicia (Regional Ministry of Education, Science, Universities and Vocational Training) and co-funded by the ERDF (ERDF Galicia 2021–27 Operational Programme, ED431G 2023/01).

Bibliography

- [1] G.M. Grossman and A.B. Krueger. *Economic growth and the environment*. The Quarterly Journal of Economics. **110**(2) (1995) 353–377. doi:10.2307/2118443
- [2] M. Wagner. *The environmental Kuznets curve, cointegration and nonlinearity*. Journal of Applied Econometrics. **30**(6) (2015) 948–967. doi:10.1002/jae.2406
- [3] J. Durbin and S.J. Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press, 2nd Edition. (2012)



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Spatio-temporal air quality risk modelling

15th July
10:30–11:20
Contributed
Session IV

María Bugallo¹ (✉), Domingo Morales², Nicola Salvati³ and Francesco Schirripa Spagnolo³

¹Department of Mathematics, University of A Coruña, Spain; ²Center of Operations Research, Miguel Hernández University of Elche; ³Department of Economics and Management, University of Pisa, Italy.

maria.bugallo@udc.es

This study introduces a robust spatio-temporal M-quantile framework for modeling air quality risk indicators under complex dependence structures and incomplete spatial coverage [1]. By integrating Geographically Weighted Regression [2] with dual spatio-temporal weighting schemes, the proposed methodology captures local distributional characteristics and heterogeneous dynamics without relying on restrictive random-effects assumptions. Tailored for Small Area Estimation, the framework delivers reliable out-of-sample predictions and accurate risk mapping for unmonitored domains, backed by analytical estimators of the mean squared error. Validated through extensive simulation studies, the model is applied to U.S. county-level air quality data from 2016 to 2023. The empirical findings reveal persistent pollution hotspots exceeding warning thresholds in parts of California and southern Florida, underscoring the framework's potential for environmental monitoring and public health risk assessment.

Keywords: M-quantile regression; spatio-temporal modelling; out-of-sample prediction; geographically weighted regression; mean squared error estimation.

Acknowledgements

This research is part of the Juan de la Cierva fellowship JDC2024-053513-I funded by the Ministerio de Ciencia, Innovación y Universidades (MICIU/AEI/10.13039/501100011033) and co-funded by the European Social Fund Plus (FSE+). It has also been supported by the Xunta de Galicia (Competitive Reference Groups ED431C-2024/14) and by MODES and CITIC, as a center accredited for excellence within the Galician University System and a member of the CIGUS Network, receives subsidies from the Department of Education, Science, Universities, and Vocational Training of the Xunta de Galicia. Additionally, it is co-financed by the EU through the FEDER Galicia 2021-27 operational program (Ref. ED431G 2023/01).

Bibliography

- [1] M. Bugallo, D. Morales, N. Salvati, F. Schirripa Spagnolo. *Small Area Estimation using Spatio-Temporal M-Quantile Models*. Journal of the Royal Statistical Society Series A. (2026). doi:<https://doi.org/10.1093/jrsssa/qnag055>
- [2] A. S. Fotheringham, C. Brunsdon, M. Charlton. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley and Sons, New York (2002).



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Nonparametric test for the dependence of spatial point processes on covariates.

Isabel Fuentes-Santos¹ (✉), Wenceslao González-Manteiga² and Jorge Mateu²

¹Department of Mathematics, Statistics and Operational Research, Universidad de La Laguna, Spain

²Department of Statistics, Mathematical Analysis and Optimisation, Universidade de Santiago de Compostela, Spain, ³Department of Mathematics, Universitat Jaume I, Spain

ifuentesa@ull.edu.es

15th July
12:50–13:40
Contributed
Session V

We propose an extension of the no-effect regression test [1] to the spatial point process framework to test for the effect of covariates prior to their inclusion in a kernel intensity estimator [3].

Let $\{X_i\}_{i=1}^N$ be a point pattern in a bounded region, $W \subset \mathbb{R}^2$, and Z a real-valued covariate. We consider the normalized residual field $s_0(u) = \lambda_0(u) - 1/|W|$, where $\lambda_0(u)$ is the density of event locations [2]. Under the independence hypothesis, $s_0(u)$ and Z are independent. This leads to a regression problem where we test the dependence of residuals $Y = \{y_i = s_0(u_i)\}_{i=1}^n$ on $Z = \{Z(u_i)\}_{i=1}^n$. We propose an F -type statistic to assess the covariate effect by comparing the residual sum of squares under a zero-mean null ($\mathcal{H}_0 : E[y_i|z_i] = 0$) against a non-parametric alternative ($\mathcal{H}_1 : E[y_i|z_i] = m(z_i)$), where m is estimated by kernel regression.

Since $\lambda_0(\cdot)$ and $s_0(\cdot)$ are unknown, the empirical statistic $\hat{F}$ is derived from the smoothed residuals $\hat{s}_{0,h}(u) = \hat{\lambda}_{0,h}(u) - 1/|W|$. The test is implemented over a random sample $\{u_i\}_{i=1}^n$ to avoid preferential sampling, and calibrated by permutation-based bootstrap. We evaluate the performance of the test in a simulation study and through its application to test the effect of environmental conditions on cetacean distribution in European Atlantic waters.

Keywords: density of event locations; no-effect test; permutation test; smoothed residual field.

Acknowledgements This work was supported by Project PID2024-158017NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by FEDER, UE.

Bibliography

- [1] A.W. Bowman, A. Azzalini. *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations* (Vol. 18). OUP Oxford. (1997)
- [2] M.I. Borrajo-García, W. González-Manteiga, M.D. Martínez-Miranda. *Bootstrapping kernel intensity estimation for inhomogeneous point processes with spatial covariates*. Computational Statistics and Data Analysis, **144** (2020) 106875 doi:<https://doi.org/10.1016/j.csda.2019.106875>
- [3] I. Fuentes-Santos, W. González-Manteiga, J. Mateu. *Consistent smooth bootstrap kernel intensity estimation for inhomogeneous spatial Poisson point processes*. Scandinavian Journal of Statistics. **43**(2) (2016) 416-435. doi: <https://doi.org/10.1111/sjos.12183>



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Prediction intervals for linear and non-linear quantile autoregression

Silvia Novo¹ (✉) and César Sánchez-Sellero²

¹Department of Statistics, Universidad Carlos III de Madrid, Spain; ²Department of Statistics,
Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, Spain

snovo@est-econ.uc3m.es

15th July
12:50–13:40
Contributed
Session V

This work introduces new methods for constructing prediction intervals using quantile-based techniques. The procedures are developed for classical (homoscedastic and linear) autoregressive models and for linear and non-linear quantile autoregressive models (see [1], [2] and [3], respectively). The proposed methods combine quantile estimation with multiplier bootstrap schemes to approximate the sampling variability of parameter estimates, together with bootstrap replications of future observations. We consider both percentile-based and predictive-root-based constructions. Theoretical results establish the validity and pertinence of the proposed methods. Simulation experiments evaluate their finite-sample performance and show that the proposed methods yield improved coverage properties and computational efficiency relative to existing approaches in the literature. The empirical usefulness of the methods is illustrated through applications to U.S. unemployment rate data and retail gasoline prices. The contributions for linear models are already available in [4], while the new developments for non-linear quantile models are in progress.

Keywords: quantile autoregression; prediction intervals; bootstrap.

Acknowledgements The first author acknowledges funding from Universidad Carlos III de Madrid through the Grants for Young PhDs program of the University's Own Research and Transfer Program 2024. The second author acknowledges support from Project PID2024-158017NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by FEDER, UE, and from the Xunta de Galicia (Consellería de Educación, Ciencia, Universidades e FP) under the GRC grant ED431C 2025/03.

Bibliography

- [1] G.E.P. Box, G.M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day. (1976)
- [2] R. Koenker, Z. Xiao. *Quantile autoregression*. Journal of the American Statistical Association. **101**(475) (2006) 980-990. doi:10.1198/016214506000000672
- [3] A.F. Galvao Jr, G. Montes-Rojas, J. Olmo. *Threshold quantile autoregressive models*. Journal of Time Series Analysis. **32**(3) (2011) 253-267. doi:10.1111/j.1467-9892.2010.00696.x
- [4] S. Novo, C. Sánchez-Sellero. *Prediction intervals for quantile autoregression*. (2025) arXiv:2512.22018

Thursday 16th July



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Defying the curse of dimensionality with higher-order kernels

José E. Chacón¹ (✉) and Tarn Duong²

¹Departamento de Matemáticas, Universidad de Extremadura, Spain; ²Independent researcher, France
jechacon@unex.es

16th July
10:30–11:20
Contributed
Session VI

In non-parametric kernel smoothing, higher-order kernels offer improved statistical properties, such as reduced bias, compared to standard second-order kernels [1]. This bias reduction is a key component in our proposed approach to attenuate the ‘curse of dimensionality’ for kernel estimators. This attenuation relies on an unexpected result that a judicious choice of the kernel order, dependent on the data dimension, leads to a mean square convergence rate that remains constant as the data dimension increases. To make this theoretical approach practically feasible it is necessary to construct multivariate kernels of arbitrary order. By employing a novel vectorized differential analysis framework [2], we derive recursive and closed-form expressions for a maximally general family of spherically symmetric multivariate kernels, with particular attention to the widely used Gaussian-based kernels [3].

Keywords: bias reduction; curse of dimensionality; Gaussian-based kernel; spherically symmetric; vectorized derivative.

Acknowledgements This work was partially supported by grant PID2023-148081NB-I00 from the Spanish Ministerio de Ciencia, Innovación y Universidades and grant GR24055 from the Junta de Extremadura.

Bibliography

- [1] J.E. Chacón, T. Duong. *Multivariate Kernel Smoothing and Its Applications*. Chapman and Hall/CRC. (2018)
- [2] J.E. Chacón, T. Duong. *Higher order differential analysis with vectorized derivatives*. (2021) [arXiv:2011.01833](#)
- [3] M.P. Wand, W.R. Schucany. *Gaussian-based kernels*. Canadian Journal of Statistics. **18(3)** (1990) 197–204. [doi:10.2307/3315450](#)



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Depence structures for random vectors

16th July
10:30–11:20
Contributed
Session VI

Juan A. Cuesta-Albertos¹ (✉), Eustasio del Barrio², Carlos Matrán² and David Rodríguez²

¹Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria, Spain;

²Departamento de Estadística e Investigación Operativa and IMUVA, Universidad de Valladolid, Spain

juan.cuesta@gmx.com

As is well known, every d -dimensional probability distribution can be characterized using its one-dimensional marginals and its copula, where the copula describes the dependence structure between the marginals of the given distribution (see [1]).

An interesting question is whether two probabilities share the same copula. However, an important point not usually mentioned in the literature is that these probabilities may have this property in one basis but not in others. Take, for example, the uniform distribution on the unit square and the two-dimensional normal distribution with a covariance matrix equal to the identity matrix. The Gaussian distribution has independent marginal distributions in every basis, but this only occurs in the canonical basis and its permutations for the uniform distribution on the square.

In this talk, we elaborate on this problem and propose a procedure based on properties of the Wasserstein metric to identify the bases in which copula equality between two distributions may occur. In developing this research, we discovered a, in our opinion, striking limitation in defining multidimensional location-scatter families properly.

We also provide examples to show the interest of this question from the Economics perspective (see [2]).

Keywords: copulas; optimal transport; L_2 -Wasserstein distance; location-scatter families; elliptical model; component-wise related distributions

Acknowledgements This research has been partially supported by grants No PID2021-128314NB-I00 and PID2022-139237NB-I00 funded by MCIN/AEI/10.13039/501100011033/FEDER.

Bibliography

- [1] R.B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. (1999)
- [2] D. Dupuis, E. Jacquier, N. Papageorgiou, and B. Remillard. *Empirical evidence on the dependence of credit default swaps and equity prices*. Journal of Future Markets, **29** (8) (2009) 695–712 doi:10.1002/fut.20382



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Goodness-of-fit tests under dependent censoring

Adrián Lago¹ (✉), Juan Carlos Pardo-Fernández¹ and Ingrid Van Keilegom²

¹Department of Statistics and Operations Research, University of Vigo, Spain; ²Department of Decision Sciences and Information Management, KU Leuven, Belgium

adrian.lago@uvigo.gal

16th July
11:50–13:30
Contributed
Session VII

In survival analysis, the variable of interest is frequently only partially observed due to censoring, which biases the information provided by the sample. This has yielded extensive methodological developments for inference under these conditions. A common feature of many statistical methods is the assumption that the censoring mechanism is independent of the target variable. However, this assumption can be unrealistic in practical applications. When this occurs, classical techniques designed for right-censored data may no longer be valid and may produce misleading results.

In this talk, we first introduce the issues caused by dependent censoring when estimating the distribution function, both parametrically and nonparametrically. Following the estimation methodology proposed recently in [1], we propose Kolmogorov–Smirnov and Cramér–von Mises tests for model assessment when data are subject to dependent censoring and the dependency structure is modelled by an Archimedean copula. The asymptotic behaviour of the new tests is addressed theoretically. Following the difficult application of the results derived, we propose a bootstrap resampling plan to approximate the null distribution of the proposed tests. Their finite-sample performance under the null and alternative hypotheses is investigated in a simulation study. A real data example is discussed.

Keywords: Dependent Censoring; Copulas; Goodness-of-Fit; Survival Analysis.

Acknowledgements This work was supported by the Grant PID2023-148811NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU.

Bibliography

- [1] Czado, C., Van Keilegom, I. *Dependent censoring based on parametric copulas*. *Biometrika*. **110**(3) (2023) 721–738. doi:10.1093/biomet/asac067



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Optimal convex M-estimation for censored linear regression

16th July
11:50–13:30
Contributed
Session VII

Paula Soto-Rodríguez¹ (✉), Jacobo de Uña-Álvarez¹ and Juan Carlos Pardo-Fernández¹

¹SIDOR Research Group, Department of Statistics and Operations Research, CITMAga, University of Vigo, Spain

p.soto@uvigo.gal

Ordinary least squares is the most widely used estimation method in linear regression, yet its robustness and efficiency can be compromised when the errors depart from normality. M-estimators based on convex loss functions provide an interesting alternative, as convexity simplifies both the theoretical analysis and the computation of the estimator. Building on this idea, [1] introduce a semiparametric procedure that adaptively constructs the optimal convex loss function, yielding an estimator with the smallest asymptotic variance within the class of convex M-estimators. In this work, we extend the framework of [1] to censored linear regression, a setting commonly encountered in survival analysis. We adapt their estimation algorithm to this setting and propose several estimators specifically designed to correct the bias introduced by censoring. We further establish their consistency and asymptotic normality, and assess their finite-sample performance through simulation studies across a range of censoring scenarios.

Keywords: asymptotic theory; censored data; semiparametric estimation; survival analysis.

Acknowledgements This work is part of the R+D+I project grant PID2023-148811NB-I00, funded by MCIN/AEI/10.13039/501100011033/, Spain and by “ERDF A way of making Europe”/EU.

Bibliography

- [1] O.Y. Feng, Y.-C. Kao, M. Xu, R.J. Samworth. *Optimal convex M-estimation via score matching*. *Annals of Statistics*. **54**(1) (2026) 408–441. doi:10.1214/25-A0S2572



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Minimax manifold estimation for hypersurfaces under convexity-type restrictions

16th July
11:50–13:30
Contributed
Session VII

Héctor González-Vázquez¹ (✉), Beatriz Pateiro-López¹ and Alberto Rodríguez-Casal¹

¹Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, Spain; Galician Centre for Mathematical Research and Technology (CITMAga), Santiago de Compostela, Spain

hector.gonzalez@usc.es

Many situations involve dealing with Euclidean data that actually tend to lie near a lower-dimensional manifold embedded in the ambient space, a setting referred to as manifold hypothesis. The objective of manifold estimation is to approximate the underlying manifold based on the sample of points. In this work we propose a manifold estimator in a tubular noise model. Our estimator builds upon the Euclidean Distance Transform (EDT) estimator, originally proposed by [1]. First, we generalize the EDT estimator allowing for the use of any pilot estimator of the support. Second, we propose to estimate the support using the r -convex hull of the sample, which generalizes the convex hull and provides greater flexibility [2]. Our estimator achieves the optimal minimax rate of convergence in Hausdorff distance (up to logarithmic factor), as shown by [1], to estimate codimension-1 manifolds, known as hypersurfaces. We also present a real data illustration involving the estimation of filaments in the two-dimensional plane.

Keywords: hypersurface; manifold estimation; minimax rate; r -convex hull; set estimation.

Acknowledgements The first author acknowledges financial support from *axuda de apoio á etapa predoutoral da Xunta de Galicia* ED481A-2025-173, partially co-funded by the European Union within the FSE+ Galicia 2021-2027 programme. This work is part of the project PID2024-158017NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by FEDER, UE, and the project ED431C 2025/03, funded by *Consellería de Educación, Ciencia, Universidades e Formación Profesional*.

Bibliography

- [1] C. Genovese, M. Perone-Pacifico, I. Verdini, L. Wasserman. *Minimax manifold estimation*. Journal of Machine Learning Research. **13(43)** (2012) 1263–1291.
- [2] A. Rodríguez-Casal. *Set estimation under convexity type assumptions*. Annales de l’Institut Henri Poincaré - Probabilités et Statistiques. **43(6)** (2007) 763–774. doi:10.1016/j.anihpb.2006.11.001



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Covariate-dependent multivariate reference regions: modelling and inference via distributional regression

Óscar Lado-Baleato¹ (✉), Giampiero Marra²

¹Department of Statistics, Operations Research, and Mathematics Education, Asturias, Spain;

²Department of Statistical Science, University College London, London, UK

ladooscar@uniovi.es

16th July
11:50–13:30
Contributed
Session VII

Reference regions are essential diagnostic tools, yet classical multivariate proposals [3] assume a homogeneous population and ignore covariate effects. Multivariate distributional regression overcomes this by modelling all parameters of the joint response distribution conditional on covariates. Copula-based approaches [2] provide flexible marginal specifications and parametric copulas that capture complex dependence patterns. When both marginal and copula parameters vary with covariates, the entire multivariate distribution adapts to patient characteristics. This work integrates conditional copula models to estimate covariate-dependent multivariate reference regions. The conditional joint density is estimated via flexible additive predictors, and regions are defined as highest predictive density sets. Bayesian posterior simulation yields an inference method for the estimated region as a tolerance region [1] that ensure nominal coverage across the covariate space. Simulation studies demonstrate correct conditional coverage, while an analysis of glycemic markers shows clinical gains over univariate reference intervals that ignore markers' correlation.

Keywords: conditional reference regions; copula; distributional regression; penalized splines.

Acknowledgements The first author acknowledges the Biostatistics and Biomedical Data Science Group at the University of Santiago de Compostela (GRID-BDS) for their generous financial support.

Bibliography

- [1] Ó. Lado-Baleato, C. Cadarso-Suárez, T. Kneib, F. Gude. *Multivariate reference and tolerance regions based on conditional transformation models: Application to glycemic markers*. Biometrical Journal. **65**(8) (2023) 2200229. doi:10.1002/bimj.202200229
- [2] G. Marra, R. Radice. *Copula Additive Distributional Regression Using R*. CRC Press. (2025).
- [3] J.C. Boyd. *Reference regions of two or more dimensions*. Clinical Chemistry & Laboratory Medicine. **42**(7) (2004) 739–746. doi:10.1515/CCLM.2004.125

Friday 17th July



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

From trajectories to spatio-temporal fields under preferential sampling

Daniela Silva^{1,3} and Raquel Menezes^{2,3}(✉)

¹Division of Modelling and Management of Fishery Resources, Portuguese Institute for the Sea and Atmosphere (IPMA), Portugal; ²Department of Mathematics, University of Minho, Portugal; ³Centre of Mathematics, University of Minho, Portugal

`rmenezes@math.uminho.pt`

17th July
11:00–11:50
Contributed
Session VIII

Integrating high-frequency movement data into spatio-temporal models poses important challenges, particularly regarding data alignment and preferential sampling. This work proposes a framework to transform Automatic Identification System (AIS) tracking data into geostatistical inputs for species distribution modeling.

First, an unsupervised multivariate Hidden Markov Model (HMM) is applied to vessel trajectories, using speed and course data to distinguish navigation from fishing activity [1]. The resulting fishing locations are used as a proxy for species distribution. To address preferential sampling caused by the non-random distribution of fishing effort, we adopt a shared latent process model in which both sampling intensity and observations depend on a common latent Gaussian field [2]. These data are then incorporated into a hierarchical spatio-temporal model using the SPDE approach with a Barrier model to represent coastal boundaries and complex spatial dependence [3].

The framework is illustrated through a case study of European sardine (*Sardina pilchardus*) along the Portuguese coast, investigating relationships with time-lagged environmental drivers. This approach connects movement ecology and geostatistics, enabling spatial inference from trajectory data while accounting for preferential sampling effects.

Keywords: AIS data; geostatistics; preferential sampling; spatio-temporal modelling; fish data.

Acknowledgements This work was partially supported by Portuguese Funds through FCT (Foundation for Science and Technology) within the Project UID/00013/2025 - Centre of Mathematics of the University of Minho ([doi:10.54499/UID/00013/2025](https://doi.org/10.54499/UID/00013/2025)) and the Project 2024.15617.PEX ([doi:10.54499/2024.15617.PEX](https://doi.org/10.54499/2024.15617.PEX)).

Bibliography

- [1] S. Lebedev, A. Lee, G. Varoquaux, C. Farrow. *hmmlearn: Unsupervised learning and inference of Hidden Markov Models*. (2021) <https://hmmlearn.com>
- [2] P. Diggle, R. Menezes, T. Su. *Geostatistical inference under preferential sampling*. Journal of the Royal Statistical Series C. **59** (2) 191-232. (2010) [doi:10.1111/j.1467-9876.2009.00701.x](https://doi.org/10.1111/j.1467-9876.2009.00701.x)
- [3] D. Silva, R. Menezes, G. Araújo, R. Rosa, A. Moreno, A. Silva, S. Garrido. *Joint model for zero-inflated data combining fishery-dependent and fishery-independent sources*. Spatial Statistics, page 100930. (2025) [doi:10.1016/j.spasta.2025.100930](https://doi.org/10.1016/j.spasta.2025.100930)



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Neyman orthogonal model checking in generalized partially linear models

17th July
11:00–11:50
Contributed
Session VIII

Rui Costa-Miranda¹ (✉), Rita Gaio¹, Christian Heumann², Wenceslao González-Manteiga³

¹Department of Mathematics of the Faculty of Sciences of the University of Porto and Centre of Mathematics of the University of Porto, Portugal; ²Ludwig-Maximilians University of Munich, Germany; ³Faculty of Mathematics of the University of Santiago de Compostela, Spain

up201804962@up.pt

We extend recent Neyman orthogonal integrated conditional moments tests for parametric models to the setting of generalized partially linear models, where the parameter space is infinite-dimensional. The proposed approach relies on local polynomial approximations to accommodate the nonparametric component in a finite-dimensional framework. The null hypothesis is formulated through conditional moment restrictions of the model errors, which are shown to be equivalent to a continuum of projected unconditional moment restrictions [1]. From that, the test statistic is constructed from the model residuals and the estimated covariance kernel of a Gaussian process, which is projected onto the orthocomplement of the conditional moment score subspace [2]. This ensures robustness with respect to the estimation, and allows for fast bootstrap procedures where there is no need to refit the model at each iteration, thus resulting in more efficient computations. To account for the heteroscedasticity in generalized models, we consider crude, standardized, and working residuals for separate goodness-of-fit tests. Additionally, a dimension reduction improvement for the conditional covariate space is employed using a random projections approach [3]. We evaluate the performance of the proposed tests across multiple scenarios, focusing on computational complexity, empirical size, and power.

Keywords: conditional moment restrictions; Gaussian processes; generalized partially linear models; goodness-of-fit; Neyman orthogonality.

Acknowledgements Rui Costa-Miranda was granted a doctoral research fellowship financed by FCT - Fundação para a Ciência e Tecnologia, I.P., with reference 2024.03100.BD. Rita Gaio and Rui Costa-Miranda were partially supported by CMUP, member of LASI, which is financed by national funds through FCT, for the project with reference UID/00144.

Bibliography

- [1] X. Li, H. Liang, W. Härdle, H. Liang. *Model checking for generalized partially linear models*. *Test*. **33**(2) (2024) 361–378.
- [2] J.C. Escanciano. *A Gaussian process approach to model checks*. *The Annals of Statistics*. **52**(5) (2024) 2456–2481.
- [3] J.A. Cuesta-Albertos, E. García-Portugués, M. Febrero-Bande, W. González-Manteiga. *Goodness-of-fit tests for the functional linear model based on randomly projected empirical processes*. *The Annals of Statistics*. **47**(1) (2019) 439–467.



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

On the spherical cardioid distribution and its goodness-of-fit

17th July
13:20–14:10
Contributed
Session IX

Eduardo García-Portugués¹ (✉)

¹Department of Statistics, Universidad Carlos III de Madrid, Spain

edgarcia@est-econ.uc3m.es

In this talk, we study the spherical cardioid distribution [1], a higher-dimensional and arbitrary-order generalization of the circular cardioid distribution. This distribution is rotationally symmetric and generates unimodal, multimodal, axial, and girdle-like densities. We identify various properties of the spherical cardioid that make it highly tractable, including explicit expressions for vectorized moments and efficient simulation. The moments of the spherical cardioid of order k up to order $k - 1$ coincide with those of the uniform distribution on the sphere, highlighting its closeness to the latter. We derive estimators by the method of moments and maximum likelihood, and their asymptotic distributions. We give the machinery for a bootstrap goodness-of-fit test based on the projected empirical cumulative distribution function approach, including the projected distribution and closed-form expressions for test statistics. An application to modeling the orbits of long-period comets shows the usefulness of the spherical cardioid distribution in real data analyses.

Keywords: Chebyshev polynomials; directional statistics; Gegenbauer polynomials; goodness-of-fit; projections.

Acknowledgements This work has been partially supported by grants PID2021-124051NB-I00 and PID2024-158399NB-I00, funded by MICIU/AEI/10.13039/501100011033 and ERDF/EU.

Bibliography

- [1] E. García-Portugués. *On the spherical cardioid distribution and its goodness-of-fit*. (2026) [arXiv:2601.16095](#)



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

A pivotal goodness-of-fit test for the Spherical Cauchy family

17th July
13:20–14:10
Contributed
Session IX

Diego Bolón ¹ (✉), Davy Paindaveine¹

¹ECARES and Department of Mathematics, Université Libre de Bruxelles, Belgium

diego.bolon.rodriguez@ulb.be

The Spherical Cauchy family is a parametric family of distributions for directional data that was recently introduced by [1]. One of its main properties is that it is a transformation model under the group of Möbius transformations in the sphere: given any two distributions belonging to the Spherical Cauchy family, there exists a Möbius transformation that maps one distribution onto the other. We use this property to introduce a general approach for testing whether the distribution of a given sample belongs to this family. We introduce a class of test statistics that are pivotal, i.e. their distribution under the null hypothesis does not depend on the true value of the parameters. We then derive the limiting distribution of the proposed test statistics and show their consistency against general alternatives. Finally, we illustrate the behaviour in practice of this procedure with a simulation study.

Keywords: directional data; Goodness-of-Fit test; pivotal statistic.

Acknowledgements The authors are supported by two “Projet de Recherche” grants from the FNRS (Fonds National pour la Recherche Scientifique), Communauté Française de Belgique.

Bibliography

- [1] Kato, S. and McCullagh, P. *Some properties of a Cauchy family on the sphere derived from the Möbius transformations.* Bernoulli **26**(4) (2020) 3224–3248. doi:10.3150/20-BEJ1222



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

An RKHS-based method for testing regression model specification

María Vidal-García¹ (✉), Ingrid Van Keilegom², Rosa M. Crujeiras¹ and Wenceslao González-Manteiga¹

¹CITMaga, Universidade de Santiago de Compostela, Spain; ²Katholieke Universiteit Leuven, Belgium
mariavidal.garcia@usc.es

17th July
15:30–17:00
Contributed
Session X

In the last decade, there have been considerable advances in RKHS procedures [1] (Machine Learning) and energy distance methods [2] (Statistics) in the context of hypothesis testing. This talk focuses on a proposal to extend these methodologies for specification testing of the regression function.

First, the methodology will be briefly introduced within a unified framework that encompasses the different perspectives in the literature [3]. Then, the problem of specification of the regression function is introduced, along with the classical residual-based process approach [4]. The adaptation of the RKHS methodology to this context yields a new test. This proposal will be studied both theoretically and computationally under the null hypothesis, fixed and local alternatives. The problem of choosing the tuning parameter will be discussed in detail. Finally, the performance of the proposed methodology is compared with previously proposed tests [5], identifying scenarios where it is advantageous in terms of power and local sensitivity, and those where the extra computational cost may not be justified.

Keywords: energy distance; regression; RKHS; specification testing.

Acknowledgements This work is part of the R&D project PID2020-116587GB-I00 granted by MICIU/AEI/10.13039/501100011033 and it is also supported by the Xunta de Galicia (Consellería de Educación, Ciencia, Universidades e FP) under the GRC grant ED431C 2025/03.

Bibliography

- [1] B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, G. R. Lanckriet. *Hilbert space embeddings and metrics on probability measures*. The Journal of Machine Learning Research. **11** (2010) 1517–1561.
- [2] G. J. Székely, M. L. Rizzo. *The energy of data*. Annual Review of Statistics and Its Application **4**(1) (2017) 447–479. doi:10.1146/annurev-statistics-060116-054026
- [3] D. Sejdinovic, B. Sriperumbudur, A. Gretton, K. Fukumizu. *Equivalence of distance-based and RKHS-based statistics in hypothesis testing*. The annals of statistics. **41**(5) (2013) 2263–2291. doi:10.1214/13-AOS1140
- [4] Akritas, M.G. and Van Keilegom, I. *Non-parametric Estimation of the Residual Distribution*. Scandinavian Journal of Statistics. **28** (2001) 549–567. doi:10.1111/1467-9469.00254
- [5] I. Van Keilegom, W. González Manteiga, C. Sánchez Sellero. *Goodness-of-fit tests in parametric regression based on the estimation of the error distribution*. TEST. **17** (2008) 401–415. doi:10.1007/s11749-007-0044-z



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Generalized distance covariance for linearity and independence in scalar-on-function regression with missing at random responses

Manuel Febrero-Bande¹, Pedro Galeano² (✉) and Wenceslao González-Manteiga¹

¹Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, Spain; ²Department of Statistics, Universidad Carlos III de Madrid, Spain

pedro.galeano@uc3m.es

17th July
15:30–17:00
Contributed
Session X

A testing procedure is introduced to jointly assess linearity and independence between the covariate and the error term in the scalar-on-function regression model with responses missing at random (MAR). The test statistic is defined as the generalized distance covariance between the functional covariate and the residuals from a fitted linear model, see [1] and [2]. To handle MAR responses, three slope estimation approaches based on functional principal components (FPCs) are considered: (i) the simplified method, which discards observations with missing responses at the cost of potential information loss; (ii) the predicted value imputation method, which fills missing responses using a regression-based approach; and (iii) the noise-augmented multiple imputation method, which adds random error terms into the regression imputations to better preserve variability. In all cases, the optimal number of FPCs is selected via cross-validation. The null distribution of the test statistics is calibrated through wild bootstrap. Monte Carlo experiments indicate that the procedure achieves high power when the semimetric and its parameters associated with the generalized distance covariance are appropriately chosen, with statistics based on the predicted value imputation method yielding a moderate power gain. Finally, the proposed methodology is illustrated with an application modeling average number of sunny days observed at Spanish meteorological stations as a function of the average daily temperatures.

Keywords: Functional linear model; Functional principal components; Generalized distance covariance; Goodness-of-fit tests; Missing at random.

Acknowledgements The authors acknowledge financial support by MCIN/AEI/10.13039/501100011033: the first and third authors from grant PID2024-158017NB-I00, and the second author from grant PID2022-138114NB-I00.

Bibliography

- [1] R. Lyons. *Distance covariance in metric spaces*. Annals of Probability. **41**(5) (2013) 3284–3305. doi:10.1214/12-AOP803
- [2] D. Sejdinovic, B. Sriperumbudur, A. Gretton, K. Fukumizu. *Equivalence of distance-based and rkhs-based statistics in hypothesis testing*. Annals of Statistics. **41**(5) (2013) 2263–2291. doi:10.1214/13-AOS1140



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Kernel-based specification tests with high-dimensional nuisance parameters

Juan Carlos Escanciano¹, (✉)

¹Department of Economics, Universidad Carlos III de Madrid, Spain

jescanci@eco.uc3m.es

17th July
15:30–17:00
Contributed
Session X

We develop global kernel-based specification tests for semiparametric models defined by conditional moment restrictions in the presence of high-dimensional nuisance parameters. Early contributions to specification testing with nonparametric nuisance estimation, including [2], established rigorous asymptotic theory for conditional moment restrictions with infinite-dimensional nuisance parameters. Recent RKHS-based developments include [3, 4]. We extend this line of work to modern machine-learning settings through orthogonal moments, cross-fitting, and reproducing kernel Hilbert space methods. Specifically, we propose Kernel-Based Locally Robust (KLR) tests that combine locally robust moments [1] with RKHS-based specification testing, extending existing frameworks to high-dimensional nuisance parameters. The resulting procedures achieve inference that is first-order insensitive to nuisance estimation error. We establish oracle equivalence between the feasible process based on estimated nuisance functions and the corresponding infeasible process based on the true nuisance functions, yielding asymptotically valid inference under weak nuisance-rate conditions. A simple and fast multiplier bootstrap avoids nuisance re-estimation and consistently approximates the null distribution. We further investigate the information available for specification testing after nuisance projection. The spectrum of the orthogonalized covariance operator quantifies the remaining testable information and motivates an effective degree of overidentification that summarizes its magnitude. Applications include specification testing in high-dimensional linear and logistic regression models, significance testing with machine-learning regressions, and tests of constant conditional treatment effects.

Keywords: high dimensional models; kernel methods; model checks; multiplier bootstrap; orthogonal moments.

Bibliography

- [1] V. Chernozhukov, J.C. Escanciano, H. Ichimura, W.K. Newey and J.M. Robins. *Locally robust semiparametric estimation*. *Econometrica*, **90**, (2022), 1501–1535. doi:10.3982/ECTA18731.
- [2] M.A. Delgado, and W. González-Manteiga. *Significance testing in nonparametric regression based on the bootstrap*. *Annals of Statistics*, **29**, (2001), 1469–1507. doi:10.1214/aos/1013203466.
- [3] J.C. Escanciano. *Orthogonal specification tests for conditional moment restrictions in reproducing kernel Hilbert spaces*. *Annals of Statistics*, **52**, (2024), 2198–2226. doi:10.1214/24-AOS2458.
- [4] J.C. Escanciano and J. de Uña-Álvarez. *Goodness-of-Fit Tests for Censored and Truncated Data: Maximum Mean Discrepancy Over Regular Functionals*. (2026). arXiv:2602.08108



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Tests for regression models based on the error distribution

17th July
15:30–17:00
Contributed
Session X

Juan Carlos Pardo-Fernández¹ (✉)

¹Departamento de Estatística e IO and CITMaga, Universidade de Vigo, Spain
juancp@uvigo.gal

Over the last two decades, a wide variety of testing procedures in regression models based on the estimation of the error distribution have been developed. These include goodness-of-fit tests for the regression function and conditional variance function, tests for the equality of regression functions and tests for the equality of error distributions, among others. Broadly speaking, these procedures rely on comparing the distributions of two sets of regression residuals: one obtained through nonparametric estimation and the other computed under the restricted model specified under the null hypothesis. In this context, the contributions of Wenceslao González-Manteiga have been particularly influential; see, for example, [1], [2] and [3].

In this talk, we will review some of the most relevant developments in this line of research, including extensions to censored and dependent data, procedures based on the empirical characteristic function, and robust adaptations of the original methods. Special emphasis will be placed on Wenceslao's contributions to the development and consolidation of this area of research.

Keywords: goodness-of-fit tests; nonparametric regression; testing problems.

Acknowledgements This work was supported by Grant PID2023-148811NB-I00 (MICIU/AEI/10.13039/501100011033 and ERDF/EU).

Bibliography

- [1] I. Van Keilegom, W. González-Manteiga, C. Sánchez-Sellero. *Goodness-of-fit tests in parametric regression based on the estimation of the error distribution*. TEST **17** (2008) 401-415. doi:10.1007/s11749-007-0044-z
- [2] J. C. Pardo-Fernández, I. Van Keilegom, W. González-Manteiga. *Testing for the equality of k regression curves*. Statistica Sinica **17** (2007) 1115-1137.
- [3] W. González-Manteiga, M.D. Martínez-Miranda, I. Van Keilegom. *Goodness-of-fit test in parametric mixed effects models based on the estimation of the error distribution*. Biometrika **103** (2016) 133-146. doi:10.1093/biomet/asv061

Posters



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Spatio-temporal separability test for point processes on linear networks

Antía Castro-Díaz ¹ (✉), María Isabel Borrajo ^{1,2} and Wenceslao González-Manteiga ^{1,2}

¹ Centro de Investigación e Tecnoloxía Matemática de Galicia (CITMAga);

² Dpto. de Estatística, Análise Matemática e Optimización, U. de Santiago de Compostela, Spain

`antia.castro.diaz@rai.usc.es`

The statistical analysis of point processes has a long tradition in the literature, with a well-established methodological framework developed primarily for spatial settings on Euclidean domains. The growing availability of georeferenced data in fields such as epidemiology, criminology, or traffic safety has however revealed the limitations of purely spatial approaches, making spatiotemporal models increasingly necessary to capture the dynamic nature of real-world phenomena. Moreover, in many of these applications the events of interest do not occur freely in space but are constrained to follow a network structure, such as road or river networks, whose geometry departs fundamentally from the Euclidean setting and poses new methodological challenges that remain largely unaddressed.

In this context, testing the hypothesis of space-time separability constitutes a fundamental preliminary step, as it not only facilitates intensity estimation but also underpins the sound development of further methodology. This paper proposes a novel hypothesis test for separability in spatiotemporal point processes on linear networks, based on an adaptation of the LISTA (Local Indicators of Spatiotemporal Association) functions proposed by [1] for the inhomogeneous K -function. A simulation study is carried out to calibrate the test statistic and evaluate its performance under varying degrees of spatiotemporal interaction, analysing both Type I error control and statistical power. Finally, the methodology is illustrated through its application to the spatiotemporal distribution of traffic accidents on the road network of Madrid.

Keywords: linear networks; spatio-temporal point processes; separability; traffic accidents.

Acknowledgements María Isabel Borrajo and Wenceslao González-Manteiga acknowledge support from Project PID2024-158017NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by FEDER, UE and GRC grant ED431C 2025/03 funded by Xunta de Galicia (Consellería de Educación, Ciencia, Universidades e FP). Antía Castro-Díaz was supported by a funded internship granted by the *Consello Social da USC* in collaboration with CITMAga.

Bibliography

- [1] D'Angelo, N., Adelfio, G., Mateu, J. *Local inhomogeneous second-order characteristics for spatio-temporal point processes occurring on linear networks*, Stat Papers, **64**, (2023) 779–805. doi:10.1007/s00362-022-01338-4



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Assessing the adequacy of quantile regression models

Mercedes Conde-Amboage^{1,2} (✉), Wenceslao González-Manteiga^{1,2}, César A. Sánchez-Sellero^{1,2}
and Ingrid Van Keilegom³

¹Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, Spain;

² Galician Centre for Mathematical Research and Technology (CITMAga), Spain;

³ Katholieke Universiteit Leuven (KU Leuven), Belgium

mercedes.amboage@usc.es

Although mean regression achieved its greatest prominence in the twentieth century, quantile regression is used when the aim is to estimate different conditional positions of the response variable, namely its quantiles. This approach provides a more detailed description of the behaviour of the response variable, accommodates more general conditions on the error distribution, and does not require stringent assumptions such as homoscedasticity or normality. It also enjoys important robustness properties.

This poster reviews several lack-of-fit tests for quantile regression models that have been developed with the contribution of Wenceslao González-Manteiga as a co-author. Particular attention is given to their application in diverse settings, such as high-dimensional covariates (see [1]), right-censored response variable (see [2] or [3]), and cure rate models ([4]).

Keywords: bootstrap; lack-of-fit test; quantile regression.

Acknowledgements This work was supported by Project PID2024-158017NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by FEDER, UE and by the Xunta de Galicia (Consejería de Educación, Ciencia, Universidades e Formación Profesional) under the GRC grant ED431C 2024/07 and ED431C 2025/03.

Bibliography

- [1] M. Conde-Amboage, W. González-Manteiga and C. Sánchez-Sellero. *Nonparametric model check for cure rate quantile regression*. Preprint (2026).
- [2] M. Conde-Amboage, C. Sánchez-Sellero and W. González-Manteiga. *A lack-of-fit test for quantile regression models with high-dimensional covariates*. Computational Statistics & Data Analysis, **88**, (2015), 128-138. doi:10.1016/j.csda.2015.02.016.
- [3] M. Conde-Amboage, I. Van Keilegom and W. González-Manteiga. *A new lack-of-fit test for quantile regression with censored data*. Scandinavian Journal of Statistics, **48(2)**, (2021), 655-688. doi:10.1111/sjos.12512.
- [4] M. Conde-Amboage, I. Van Keilegom and W. González-Manteiga, W. *Application of quantile regression models for biomedical data*. In Statistical Methods at the Forefront of Biomedical Advances (pp. 83-113). Cham: Springer International Publishing. doi:10.1007/978-3-031-32729-2_5.



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

An integrated system for SO_2 prediction using Nonparametric, Semiparametric, Functional, and Spatio-Temporal models

Tomás R. Cotos Yáñez¹ (✉), Rubén Fernández Casal², Belén M. Fernández de Castro³, Javier Roca Pardiñas¹, José M. Prada Sánchez⁴, Manuel Febrero Bande⁴, Wenceslao González Manteiga⁴

¹Departamento de Estadística e Investigación Operativa, Universidade de Vigo (Spain); ²Departamento de Matemáticas, Universidade da Coruña (Spain); ³Zara Business Analytics, INDITEX (Spain);

⁴Departamento de Estadística, Análise Matemático e Optimización, Universidade de Santiago de Compostela (Spain)

cotos@uvigo.gal

In the 1990s and 2000s, the former Department of Statistics and Operations Research at the University of Santiago de Compostela, in collaboration with the Endesa-owned power plant in As Pontes (A Coruña, Spain), developed an integrated system, SIPEI (Sistema Integrado de Predicción Estadística de Inmisión), to predict ground-level SO_2 in the surrounding area.

At that time, the power plant operated an Air Quality Monitoring Network comprising 17 automatic stations located within a 30 km radius of the facility. These stations were equipped with monitoring systems that recorded data on various air pollution indicators, including SO_2 concentrations, at five-minute intervals.

SIPEI was based on several statistical models that generated different types of forecasts: a 30-minute and 1-hour point forecasts [1, 2], a spatial forecast [3], and a functional forecast [4].

Real data— SO_2 concentrations and meteorological information—together with model predictions, were integrated into an application through which different users at the power plant could monitor conditions and make decisions. Over the years, the application has evolved through several versions, from early implementations developed in Visual Basic to more recent versions built using Shiny.

Bibliography

- [1] J.M. Prada-Sánchez, M. Febrero-Bande, T.R. Cotos-Yáñez, W. González-Manteiga, J.L. Bermúdez-Cela, and T. Lucas-Domínguez. *Prediction of SO_2 pollution incidents near a power station using partially linear models and a historical matrix of predictor-response vectors*. *Environmetrics*, **11** (2000) 209–225.
- [2] J. Roca-Pardiñas, W. González-Manteiga, M. Febrero-Bande, J.M. Prada-Sánchez, and C.M. Cadarso-Suárez. *Predicting binary time series of SO_2 using generalized additive models with unknown link function*. *Environmetrics*, **15** (2004) 729–742.
- [3] R. Fernández-Casal, W. González Manteiga, and M. Febrero-Bande. *Flexible Spatio-Temporal Stationary Variogram Models*. *Statistics and Computing* **13**, (2003) 127–136.
- [4] B.M. Fernández de Castro, S. Guillas and W. González-Manteiga. *Functional Samples and Bootstrap for Predicting Sulfur Dioxide Levels*. *Technometrics*, **47** (2005) 212–222.



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Semiparametric censorship and the Koziol-Green model

Jacobo de Uña-Álvarez¹ (✉)

¹Department of Statistics and Operations Research, Universidade de Vigo, Spain
jacobo@uvigo.gal

In this work I revisit the setting of censored data when the hazard function of the target event time, $\lambda_1(t)$ say, and that of the censoring time, $\lambda_2(t)$, are linked through a semiparametric model. That is, for a certain parametric model $\nu(t; \beta) > 0$ it holds $\lambda_2(t) = \nu(t; \beta)\lambda_1(t)$, the $\lambda_j(t)$'s being nonparametric [1]. The particular case in which $\nu(t; \beta)$ is free of t is called the Koziol-Green proportional hazards model. In general, the situation is referred to as semiparametric censorship, an informative censoring model under which Kaplan-Meier estimation is no longer efficient. Key results are reviewed; among these, I mention some relevant contributions from my PhD thesis [2], carried out under the supervision of Wenceslao González-Manteiga in the late nineties. I also discuss important applications of the semiparametric censorship model in fields like kernel smoothing [3], estimation with missing censoring indicators [3], multi-state models [4], and regression analysis [5], [7]. New simulation studies and real data illustrations are included.

Keywords: censored data; presmoothing; proportional hazards; Survival Analysis.

Acknowledgements Work supported by Grant PID2023-148811NB-I00, Spanish Ministry of Science, Innovation and Universities (MICIU/AEI/10.13039/501100011033 and ERDF/EU).

Bibliography

- [1] G. Dikta, J. Ghorai, C. Schmidt. *The central limit theorem under semiparametric random censorship models*. Journal of Statistical Planning and Inference **127**(1) (2005) 23–51. doi:10.1016/j.jspi.2003.09.013
- [2] J. de Uña-Álvarez. *Inferencia Estadística en Modelos de Censura Proporcional*. Tesis Doctoral, Universidad de Santiago de Compostela (1998)
- [3] J. de Uña-Álvarez, W. González-Manteiga, C. Cadarso-Suárez. *Kernel distribution function estimation under the Koziol-Green model*. Journal of Statistical Planning and Inference **87**(2) (2000) 199–219. doi:10.1016/S0378-3758(99)00192-5
- [4] J. Mendonça. *New Contributions in the Missing Censoring-Indicator Model*. Tesis Doctoral, Universidade de Vigo (2014)
- [5] A.P. Amorim, J. de Uña-Álvarez, L. Meira-Machado. *Presmoothing the transition probabilities in the illness-death model*. Statistics & Probability Letters **81**(7) (2011) 797–806. doi:10.1016/j.spl.2011.02.017
- [6] M.C. Iglesias-Pérez, J. de Uña-Álvarez. *Nonparametric estimation of the conditional distribution function in a semiparametric censorship model*. Journal of Statistical Planning and Inference **138**(10) (2008) 3044–3058. doi:10.1016/j.jspi.2007.11.011
- [7] G. Dikta, R. Külheim, J. Mendonça, J. de Uña-Álvarez. *Asymptotic representation of presmoothed Kaplan-Meier integrals with covariates in a semiparametric censorship model*. Journal of Statistical Planning and Inference **171** (2016) 10–37. doi:10.1016/j.jspi.2015.12.001



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Nonparametric tests for comparing ROC curves with covariates

Arís Fanjul-Hevia¹ (✉), Wenceslao González-Manteiga² and Juan Carlos Pardo-Fernández³

¹Departamento de Estadística e Investigación Operativa y Didáctica de la Matemática, Universidad de Oviedo, Spain; ² Departamento de Estadística, Análise Matemática e Optimización, Universidade de Santiago de Compostela, Spain; ³ Departamento de Estadística e Investigación Operativa, Universidade de Vigo, Spain

fanjularis@uniovi.com

The Receiver Operating Characteristic (ROC) curve is a statistical tool used to evaluate how well a binary classification method distinguishes between two categories. It is frequently applied, for example, in diagnostic problems. When multiple classification methods are available –or when a single method is applied to different populations– comparing their corresponding ROC curves is a standard approach to assessing their performance [1].

Whenever additional covariate information is available, it should be incorporated into the comparison of ROC curves, since covariates may affect the discriminatory capability of the diagnostic methods. This can be done using either covariate-specific or covariate-adjusted ROC curves. Our work first examines which type of curve is most appropriate for the different situations that may arise [4]. Based on these findings, we then propose several tests for comparing these curves in both paired and independent data settings [2, 3].

Keywords: bootstrap; covariates; hypothesis testing ; ROC curve.

Acknowledgements This work was supported by the Ministerio de Ciencia e Innovación (Grant Nos. PID2020-116587GB-I00 and PID2020-118101GB-I00) and the Ministerio de Ciencia, Innovación y Universidades (Grant No. PID2023-148811NB-I00).

Bibliography

- [1] A. Fanjul-Hevia, W. González-Manteiga. *A comparative study of methods for testing the equality of two or more ROC curves*. Computational Statistics. **33** (2018) 357–377. doi:10.1007/s00180-017-0783-6
- [2] A. Fanjul-Hevia, W. González-Manteiga, J.C. Pardo-Fernández. *A non-parametric test for comparing conditional ROC curves*. Computational Statistics & Data Analysis. **157** (2021) 1071469. doi:10.1016/j.csda.2020.107146
- [3] A. Fanjul-Hevia, J.C. Pardo-Fernández, I. Van Keilegom, W. González-Manteiga. *A test for comparing conditional ROC curves with multidimensional covariates*. Journal of Applied Statistics. **51** (1) (2022) 87–113. doi:10.1080/02664763.2022.2116409
- [4] A. Fanjul-Hevia, J.C. Pardo-Fernández, W. González-Manteiga. *A New Test for Assessing the Covariate Effect in ROC Curves*. Statistics in Medicine. **44** (23–24) (2025) e70284. doi:10.1002/sim.70284



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Two decades of SO₂ forecasting: A success story

Tomás R. Cotos-Yáñez¹, Rubén Fernández-Casal² (✉), Belén M. Fernández de Castro³, Javier Roca-Pardiñas¹, José M. Prada-Sánchez⁴, Manuel Febrero-Bande⁴, Wenceslao González-Manteiga⁴

¹Departamento de Estadística e Investigación Operativa, Universidade de Vigo (Spain); ²Departamento de Matemáticas, Universidade da Coruña (Spain); ³Zara Business Analytics, INDITEX (Spain);

⁴Departamento de Estadística, Análise Matemático e Optimización, Universidade de Santiago de Compostela (Spain)

ruben.fcasal@udc.es

This study presents several methods developed during the 1990s and 2000s to predict ground-level SO₂ concentrations based on observations collected at monitoring stations located near the Endesa-owned power plant in As Pontes (A Coruña, Spain).

For the 30-minute prediction horizon, the implemented forecasting model consists of a nonparametric component used to estimate the trend and a Box–Jenkins model applied to the residual series. Subsequently, changes in the operational context required forecasts to be generated one hour ahead. To improve the robustness of the model, it was extended by incorporating a linear combination of exogenous explanatory variables, whose coefficients are estimated as part of the prediction procedure, resulting in a partially linear model [1].

Alternative approaches included the use of generalized additive models with an unknown link function for predicting binary time series of SO₂ concentrations [2]. In addition, neural networks and nonparametric functional regression techniques were combined with boosting methods to improve prediction accuracy [4]. For spatio-temporal prediction, a generalization of the Shapiro and Botha approach was proposed, allowing the construction of flexible variogram models [3].

Keywords: nonparametric regression; partially linear models; spatial data; binary prediction; functional regression.

Bibliography

- [1] J.M. Prada-Sánchez, M. Febrero-Bande, T.R. Cotos-Yáñez, W. González-Manteiga, J.L. Bermúdez-Cela, and T. Lucas-Domínguez. *Prediction of SO₂ pollution incidents near a power station using partially linear models and a historical matrix of predictor-response vectors*. *Environmetrics*, **11** (2000) 209–225.
- [2] J. Roca-Pardiñas, W. González-Manteiga, M. Febrero-Bande, J.M. Prada-Sánchez, and C.M. Cadarso-Suárez. *Predicting binary time series of SO₂ using generalized additive models with unknown link function*. *Environmetrics*, **15** (2004) 729–742.
- [3] R. Fernández-Casal, W. González Manteiga, and M. Febrero-Bande. *Flexible Spatio-Temporal Stationary Variogram Models*. *Statistics and Computing* **13**, (2003) 127–136.
- [4] B.M. Fernández de Castro, S. Guillas and W. González-Manteiga. *Functional Samples and Bootstrap for Predicting Sulfur Dioxide Levels*. *Technometrics*, **47** (2005) 212–222.



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Covariate Selection Techniques for Big and Functional Data

Laura Freijeiro-González¹ (✉), Wenceslao González-Manteiga² and Manuel Febrero-Bande²

¹Department of Statistics and Operational Research, University of Vigo, Spain; ² CITMaga;
Department of Statistics, Mathematical Analysis and Optimization, University of Santiago de
Compostela, Spain

`laura.freijeiro@uvigo.gal`

In this work, we review several covariate selection techniques developed for high-dimensional and functional settings. In particular, we examine penalization methods tailored to Big Data contexts ([1] and [4]), as well as extensions of distance correlation applied to various functional frameworks ([2] and [3]). The methodological results will be presented, along with some applications to real datasets. All these studies were conducted jointly with Professor Wenceslao González Manteiga.

Keywords: big data; covariate selection; functional data.

Acknowledgements This work was supported by the Consellería de Cultura, Educación e Ordenación Universitaria along with the Consellería de Economía, Emprego e Industria of the Xunta de Galicia under Project ED481A-2018/264; and Project PID2024-158017NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by FEDER, UE. We also acknowledge to the Centro de Supercomputación de Galicia (CESGA) for computational resources.

Bibliography

- [1] L. Freijeiro-González, M. Febrero-Bande, W. González-Manteiga. *A Critical Review of LASSO and Its Derivatives for Variable Selection Under Dependence Among Covariates*. International Statistical Review. **90** (2021) 118-145. doi:10.1111/insr.12469
- [2] W. González-Manteiga. *A review on specification tests for models with functional data*. Spanish Journal of Statistics. **4**(1) (2022) 9–40. doi:10.37830/SJS.2022.1.02
- [3] L. Freijeiro-González, M. Febrero-Bande, W. González-Manteiga. *Novel specification tests for synchronous additive concurrent model formulation based on martingale difference divergence*. TEST. **32** (2023) 908-941. doi:10.1007/s11749-023-00857-y
- [4] L. Freijeiro-González, M. Febrero-Bande, W. González-Manteiga. *Guidelines for LASSO and derivatives use under different dependence and scale structures*. (2025) arXiv:2506.08582



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

From Gauss to Wences: tracing a mathematical genealogy

Eduardo García-Portugués¹ (✉)

¹Department of Statistics, Universidad Carlos III de Madrid, Spain

edgarcia@est-econ.uc3m.es

On the occasion of Prof. Wenceslao González-Manteiga's 70th birthday, we reconstruct his mathematical genealogy using curated information from the Mathematics Genealogy Project and Dialnet [1, 2]. Traced upwards, a chain of advisors connects Wenceslao to Carl Friedrich Gauss passing through Julio Rey Pastor. Traced downwards, his mentoring legacy is substantially richer than the 8 students that the Mathematics Genealogy Project listed in March 2026: the completed record, accepted in June 2026, now spans three generations of academic descendants. The poster tells both halves of this story, turning scattered academic records into the documented memory of a school of statistics. How many hops separate Gauss and other famous mathematicians from Wences, and how far his school has grown, is revealed at the poster.

Keywords: academic lineage; mathematical genealogy; mentoring; scientific legacy; statistics.

Acknowledgements This work was supported by grant PID2024-158399NB-I00 funded by MICIU/AEI/10.13039/501100011033 and ERDF/EU.

Bibliography

- [1] Mathematics Genealogy Project <https://www.mathgenealogy.org>
- [2] Dialnet, Universidad de La Rioja <https://dialnet.unirioja.es>



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Edgeworth expansions in nonparametric statistics

Pilar García-Soidán¹ (✉)

¹Department of Statistics and Operations Research, University of Vigo, Spain
pgarcia@uvigo.gal

The Edgeworth expansions provide a useful tool to refine the normal approximation derived from the Central Limit Theorem, in terms of the cumulants of the underlying distribution. This mechanism was originally applied to sums of independent and identically distributed random variables [1], by assuming Cramér's condition and the existence of moments up to a certain order. The validity of the Edgeworth expansions was further extended to triangular arrays, thus supporting their application to well-known problems in nonparametric statistics, related to density estimation [3] or distribution estimation [2]. For dependent data, the formal structure of the Edgeworth expansions can be preserved, under several conditions, although several problems may arise in this setting, such as the difficulties involved in the estimation of the cumulants, the slower rate of convergence of the resulting approximation or the fact that Cramér's condition typically fails. Among other uses, the Edgeworth expansions can be employed to construct confidence intervals with coverage closer to the nominal one than that achieved through the normal limit distribution, as well as to justify the accuracy of a Bootstrap approach.

Keywords: asymptotic expansion; Cramér's condition; cumulant; normal distribution.

Acknowledgements The author would like to thank Wenceslao González-Manteiga for the encouragement and support he provided, which motivated her to embark on her research journey.

Bibliography

- [1] R. N. Bhattacharya, R. Rao, *Normal approximation and asymptotic expansions*. Wiley. (1976)
- [2] P. García-Soidán, W. González-Manteiga, J. M. Prada-Sánchez. *Edgeworth expansions for nonparametric distribution estimation with applications*. Journal of Statistical Planning and Inference. **65** (1997) 213-231. doi:10.1016/S0378-3758(97)00059-1
- [3] P. Hall, *Edgeworth expansions for nonparametric density estimators, with applications*. Statistics. **22** (1992) 215-232. doi:10.1080/02331889108802305



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Comparing time series of mastitis pathogens using ARIMA models and clustering

María José Ginzo Villamayor¹ (✉), Gonzalo Fernández Rodríguez², María Luisa Barreal López³, María Belén Pombo Río³, Wenceslao González Manteiga¹, Alberto Prieto Lago², Noelia Lago Rodríguez² and J. González Palencia³

¹Department of Statistics, Mathematical Analysis and Optimization, University of Santiago de Compostela, Galician Centre for Mathematical Research and Technology (CITMaga), Spain;

²Department of Veterinary Pathology, University of Santiago de Compostela, Spain; ³Galician Interprofessional Milk Laboratory (LIGAL), Spain

mariajose.ginzo@usc.es

This study examines and compares the temporal trends of mastitis pathogens in dairy cattle using microbiological data from milk samples collected in Galicia (Spain) between 2005 and 2011. A total of 240,232 samples were analyzed to identify seasonal patterns and similarities in pathogen behavior. Time series analysis was performed using ARIMA models to account for stochastic dependence, while multivariate techniques such as cluster analysis were applied to group pathogens with similar epidemiological trends. The methodology allows decomposition of each series into trend and random components, enabling robust comparison between pathogens. Distances between series were computed based on trend differences, following a functional comparison approach inspired by [1], and clustering methods were used to identify homogeneous groups. In addition, the overall framework builds upon previous work on mastitis epidemiology using time series analysis [2]. Results reveal that traditional classifications of mastitis pathogens into contagious and environmental categories are often insufficient, as several pathogens exhibit mixed or distinct temporal behaviors. The findings highlight the complexity of mastitis epidemiology and demonstrate that time series and multivariate statistical approaches provide valuable tools for understanding pathogen dynamics.

Keywords: dairy cattle; epidemiology; mastitis; pathogens; time series; trends.

Acknowledgements This work was supported by Xunta de Galicia (ED431C 2025/03) and MCIN/AEI (10.13039/501100011033) through projects PID2024-158017NB-I00 and PID2020-116587GB-I00.

Bibliography

- [1] J.M. Vilar Fernández, W. González Manteiga. *Nonparametric comparison of curves with dependent errors*. Journal of Statistical Theory and Applications, **38**(2) (2004) 81–99. doi:10.1080/02331880310001634656
- [2] G. Fernández, M.L. Barreal, M.B. Pombo, et al. *Comparison of the epidemiological behavior of mastitis pathogens by applying time-series analysis in results of milk samples submitted for microbiological examination*. Veterinary Research Communications, **37** (2013) 259–267. doi:10.1007/s11259-013-9570-1



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Estimation of the conditional distribution function with censoring and truncation

María del Carmen Iglesias Pérez¹ (✉)

¹Department of Statistics and Operations Research, University of Vigo, Spain

mcigles@uvigo.es

This work is a tribute to Professor Wenceslao González Manteiga, my doctoral thesis advisor and the academic father, grandfather, and great-grandfather of the large family of Galician nonparametric statisticians. The objective is to review the papers we published together and analyze their impact through the citations received.

Two main papers resulted from my tesis (see [1] and [2]), whose titles are:

1. *Strong representation of a generalized product-limit estimator for truncated and censored data with some applications.*
2. *Bootstrap for the conditional distribution function with truncated and censored data.*

Article [1] has had a significant impact in the field, with 257 views and 53 citations in Taylor & Francis Online, 41 citations in Web of Science, and 86 in Google Scholar. Furthermore, its influence has been sustained over time and it is cited in studies on asymptotic convergence and empirical processes. The relevance of article [2] has been more limited, receiving 218 Accesses and 9 citations in Springer Nature Link, 6 citations in Web of Science and 25 in Google Scholar.

My gratitude and admiration to Wences for his enthusiasm, ability to learn, communicate, bring together ideas and people, for his intuition and discernment, his inspiring leadership, and his enormous vocation as a professor and researcher. It is an honor to learn from the greatest.

Keywords: conditional product-limit estimator; kernel estimator; almost sure representation; asymptotic normality; bootstrap.

Acknowledgements This work was supported by Grant PID2023-148811NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by "ERDF/EU".

Bibliography

- [1] C. Iglesias Pérez and W. González-Manteiga. *Strong representation of a generalized product-limit estimator for truncated and censored data with some applications*. Journal of Nonparametric Statistics. **10**(3) (1999) 213–244. doi:10.1080/10485259908832761
- [2] M. C. Iglesias Pérez and W. González-Manteiga. *Bootstrap for the conditional distribution function with truncated and censored data*. Annals of the Institute of Statistical Mathematics. **55** (2003) 331–357. doi:10.1007/BF02530503



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Study and comparison of Feret diameter in organic milks by density estimation

Aida Jover Ramos¹ (✉), Mara Fernanda Rodríguez Cao¹ and Luis Alberto Ramil Novo²

¹Departamento de Química Física, Universidade de Santiago de Compostela, Facultade de Ciencias, Lugo, Spain; ² MODESTYA and Departamento de Estatística, Análise Matemática e Optimización, Universidade de Santiago de Compostela, Facultade de Ciencias, Lugo, Spain

aida.jover@usc.es

The maximum diameter of milk fat globules, known as the Feret diameter, provides practical information for the dairy sector regarding various milk characteristics (see, e.g., [1] and [2]), including organoleptic properties, viscosity, processing methods, and homogenization intensity. A study is presented here on the distribution of Feret diameter, determined by optical microscopy, in various certified organic milks, from various origins and processing types. First, the Feret density function is estimated for each milk using the kernel method, selecting the smoothing parameter through cross-validation. Subsequently, the estimated densities are used to compare, in two phases, the Feret of the analyzed milks. In the first one, the differences between the different origins are examined, considering raw milk on one hand, and pasteurized milk on the other. In the second phase, for each origin, the differences between the different processing methods are examined. To carry out the comparisons, a k -sample test is used, which takes as a reference the common area [3] of the non-parametric estimates of the densities.

Keywords: kernel density estimation; common area test; organic milk; milk treatment; milk fat globule membrane.

Acknowledgements The authors thank P. Martínez Camblor for his support in the application of the test in [3]. AJR greatly acknowledges the support from the Spanish Ministry of Science, Innovation, and Universities under Grant No. PID2023-147148NB-I00. LARN is grateful for the support to this work by the Xunta de Galicia (Consellería de Educación, Ciencia, Universidades e FP) under the GRC grant ED431C 2025/03.

Bibliography

- [1] S.J. Lee, J.W. Sherbon. *Chemical changes in bovine milk fat globule membrane caused by heat treatment and homogenization of whole milk..* Journal of Dairy Research. **69**(4) (2002) 555-567. doi:10.1017/s002202990200571x
- [2] Y. Ma, L. Zhang, Y. Wu, P. Zhou. *Changes in milk fat globule membrane proteome after pasteurization in human, bovine and caprine species.* Food Chemistry. **279** (2019) 209-215. doi:10.1016/j.foodchem.2018.12.015
- [3] P. Martínez-Camblor, J. De Uña-Álvarez, N. Corral. *k-Sample test based on the common area of kernel density estimators.* Journal of Statistical Planning and Inference. **138**, **12** (2008) 4006-4020. doi:10.1016/j.jspi.2008.02.008



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

Nonparametric methods for functional data homogenization

Manuel Oviedo de la Fuente¹ (✉) and Rubén Fernández-Casal¹

¹Department of Mathematics, Universidade da Coruña, Spain

manuel.oviedo@udc.es

Functional data from environmental and climatological studies are often affected by point-wise anomalies, missing observations and outlying trajectories. These problems distort downstream functional data analyses and require a preliminary homogenization step. We propose a fully nonparametric procedure based on a heteroscedastic functional model: the trend, variance and dependence are estimated nonparametrically; point-wise anomalies are winsorized using leave-one-out residual kriging; incomplete curves are reconstructed by kriging prediction; and final functional outliers are identified through a depth-based bootstrap calibration. The last stage adapts the trimming approach of Febrero et al. [1], replacing the smoothed bootstrap by a nonparametric bootstrap from the fitted heteroscedastic model [2]. The method is illustrated with annual ground-level ozone curves recorded at the Yarner Wood monitoring station from 1988 to 2024, and supported by a Monte Carlo study comparing the proposed calibration with the original smoothed-bootstrap procedure. The methodology, implemented in the `npfda` package [3], reduces false positive detections while retaining competitive power for identifying functional outliers.

Keywords: functional data analysis; functional outliers; homogenization; kriging; nonparametric bootstrap.

Acknowledgements This work was supported by MCIN/AEI/10.13039/501100011033 through grant PID2020-113578RB-I00, by Xunta de Galicia (Grupos de Referencia Competitiva ED431C-2020/14), and by CITIC, supported by Xunta de Galicia through the CIGUS framework.

Bibliography

- [1] M. Febrero, P. Galeano, W. González-Manteiga. *Outlier detection in functional data by depth measures*. *Environmetrics*. **19** (2008) 331–345.
- [2] R. Fernández-Casal, S. Castillo, M. Flores. *A nonparametric bootstrap method for heteroscedastic functional data*. *Journal of Agricultural, Biological and Environmental Statistics*. **29** (2024) 169–184.
- [3] R. Fernández-Casal, S. Castillo, M. Flores, M. Oviedo. *npfda: Nonparametric functional data analysis*. R package version 0.2-1. (2025)



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

An overview of nonparametric missing data analysis in Wenceslao González Manteiga's research

Ana Pérez González¹ (✉)

¹Departamento de Estadística e Investigación Operativa, Universidad de Vigo

anapg@uvigo.gal

Missing data is a very common problem in statistical data analysis and has been the subject of numerous research papers, particularly in the field of nonparametric statistics. Professor González-Manteiga, in whose honor this workshop is held, has made significant contributions to this area. This poster provides a review of the principal contributions to nonparametric statistics with missing data developed by Professor González-Manteiga in collaboration with various co-authors. Among these authors, we may cite Boente, G.[2], Bianco, A. [1], Cotos-Yáñez, T. R.[3], Martínez Miranda, M.D. [4], Vilar-Fernández, J.M.[5] or Pérez-González, A.([1]-[5]) The synthesis encompasses several methodological frameworks, including regression function estimation, goodness-of-fit testing, dependent data or robust inference.

Keywords: Missing Data; Regression function; Goodness of fit Test; Robustness.

Acknowledgements This work was supported by the Grant PID2023-148811NB-I00 from Agencia Estatal de Investigación (Ministerio de Ciencia, Innovación y Universidades, Spanish Government) and PID2024-158017NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by FEDER, UE.

Bibliography

- [1] Bianco, A., Boente, G. González-Manteiga, W., Pérez-González, A. *Estimation of the marginal location under a partially linear model with missing responses*. Computational Statistics & Data Analysis. **54**(2) (2010) 546-564
- [2] G. Boente, W. González-Manteiga and A. Pérez-González. *Robust nonparametric estimation with missing data*. Journal of Statistical Planning and Inference, 139(2), 571–592. (2009)
- [3] T. R. Cotos-Yáñez, A. Pérez-González and W. González-Manteiga. *Model checks for nonparametric regression with missing data: a comparative study*. Journal of Statistical Computation and Simulation, 86(16), 3188–3204. (2016)
- [4] W. González-Manteiga, M. D. Martínez-Miranda and A. Pérez-González. *The choice of smoothing parameter in nonparametric regression through Wild Bootstrap*. Computational Statistics & Data Analysis, 47(3), 487–515. (2004)
- [5] A. Pérez-González, J. M. Vilar-Fernández and W. González-Manteiga. *Asymptotic properties of local polynomial regression with missing data and correlated errors*. Annals of the Institute of Statistical Mathematics, 61, 85–109. (2009)



WORKSHOP ON FUNCTIONAL NONPARAMETRIC AND SPATIO-TEMPORAL STATISTICS

An energy-distance control chart for functional data

Guillermo Portela¹, Salvador Naya² and Javier Tarrío-Saavedra² (✉)

¹Cátedra UDC-Inditex de IA en Algoritmos Verdes, UDC; ²MODES group, CITIC, UDC

javier.tarrio@udc.es

This work proposes a control chart for detecting perturbations in functional data generating processes through repeated two-sample tests between Hilbert-valued random elements. Functional observations are split into equally sized temporal blocks. Phase I compares each reference block with the remaining in-control blocks to estimate an empirical upper control limit, while Phase II sequentially compares new blocks with resampled Phase I data. The monitoring statistic follows the energy-distance approach of [1] and its connection with kernel and reproducing-kernel Hilbert space two-sample methods [2, 3]. The procedure is evaluated using bivariate functional data simulated by Karhunen–Loeve expansions, introducing gradual changes in mean, variance and interaction structure. Results show increasing sensitivity as perturbations become stronger.

Keywords: control charts; functional data analysis; energy distance; two-sample testing.

Acknowledgements This work is supported by Grants TSI-100925-2023-1, PID2023-147127OB-I00, ED431C 2024/014, and ED431G 2023/01.

Bibliography

- [1] G. González-Rodríguez, A. Colubi, W. González-Manteiga and M. Febrero-Bande. *A consistent test of equality of distributions for Hilbert-valued random elements*. Journal of Multivariate Analysis, **202**:105312, 2024. doi: 10.1016/j.jmva.2024.105312.
- [2] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf and A. Smola. *A kernel two-sample test*. Journal of Machine Learning Research, **13**(25):723–773, 2012. DOI not assigned. <https://www.jmlr.org/papers/v13/gretton12a.html>
- [3] D. Sejdinovic, B. Sriperumbudur, A. Gretton and K. Fukumizu. *Equivalence of distance-based and RKHS-based statistics in hypothesis testing*. The Annals of Statistics, **41**(5), 2263–2291, 2013. doi: 10.1214/13-AOS1140.

Index

Acal
Christian, 24

Acar-Denizli
Nihan, 25

Aguilera
Ana M., 24

Amini
Morteza, 11

Amoroso
Cheyenne, 33

Arenas Velilla
Santiago, 31

Barreal López
María Luisa, 63

Barreiro
Daniel, 18

Bianco
Ana M., 21

Bolón
Diego, 48

Borrajo
María Isabel, 54

Bugallo
María, 34

Canedo
Christian, 29

Cao
Ricardo, 18, 30

Castro-Díaz
Antía, 54

Chacón
José E., 38

Conde-Amboage
Mercedes, 55

Costa-Miranda
Rui, 46

Cotos-Yáñez
Tomás R., 56, 59

Crujeiras
Rosa M., 49

Cuesta-Albertos
Juan A., 39

Cuevas
Antonio, 20

Darbalaei
Mohamad, 11

de Uña-Álvarez
Jacobo, 41, 57

del Barrio
Eustasio, 39

Delicado
Pedro, 25

Diz-Castro
Daniel, 28

Duong
Tarn, 38

Escanciano
Juan Carlos, 51

Fanjul-Hevia
Arís, 58

Febrero-Bande
Manuel, 11, 28, 50, 56, 59, 60

Fernández de Castro
Belén M., 56, 59

Fernández Rodríguez
Gonzalo, 63

Fernández-Casal
Rubén, 56, 59, 66

Fraiman
Daniel, 12
Ricardo, 12

Francisco
Mario, 18

Freijeiro-González
Laura, 60

Fuentes-Santos
Isabel, 35

Gaio
Rita, 46

Galeano
 Pedro, 50

García-Portugués
 Eduardo, 47, 61

García-Martos
 Carolina, 33

García-Soidán
 Pilar, 62

Ginzo Villamayor
 María José, 63

González Palencia
 J., 63

González-Manteiga
 Wenceslao, 11, 28, 35, 46, 49, 50,
 54–56, 58–60, 63

González-Vázquez
 Héctor, 42

Heumann
 Christian, 46

Iglesias Pérez
 María del Carmen, 64

Jover Ramos
 Aida, 65

Koopman
 Siem Jan, 33

López-Cheda
 Ana, 30

López-Vizcaíno
 Esther, 26

Lado-Baleato
 Óscar, 43

Lago
 Adrián, 40

Lago Rodríguez
 Noelia, 63

Lombardía
 M. José, 26

Marra
 Giampiero, 43

Martínez-Miranda
 María Dolores, 15

Mateu
 Jorge, 14, 35

Matrán
 Carlos, 39

Menezes
 Raquel, 45

Morales
 Domingo, 34

Munk
 Axel, 31

Naya
 Salvador, 68

Novo
 Silvia, 36

Ortiz
 Helena, 24

Oviedo de la Fuente
 Manuel, 66

Pérez González
 Ana, 67

Paindaveine
 Davy, 48

Pardo-Fernández
 Juan Carlos, 40, 41, 52, 58

Pateiro-López
 Beatriz, 42

Piñeiro-Lamas
 Beatriz, 30

Pombo Río
 María Belén, 63

Portela
 Guillermo, 68

Prada-Sánchez
 José M., 56, 59

Prieto Lago
 Alberto, 63

Ramil Novo
 Luis Alberto, 65

Roca-Pardiñas
 Javier, 56, 59

Rodríguez-Casal
 Alberto, 42

Rodríguez
 David, 39

Rodríguez Cao
 Mara Fernanda, 65

Rodríguez Ramírez
 Luis Alberto, 31

Rueda
 Cristina, 26, 29

Ruiz-Medina
 María Dolores, 11, 27

Sánchez-Sellero
 César A., 36, 55

Salvati

Nicola, 34

Schirripa Spagnolo

Francesco, 34

Silva

Daniela, 45

Soto-Rodríguez

Paula, 41

Tarrío-Saavedra

Javier, 68

Van Keilegom

Ingrid, 17, 40, 49, 55

Vidal

Marc, 24

Vidal-García

María, 49