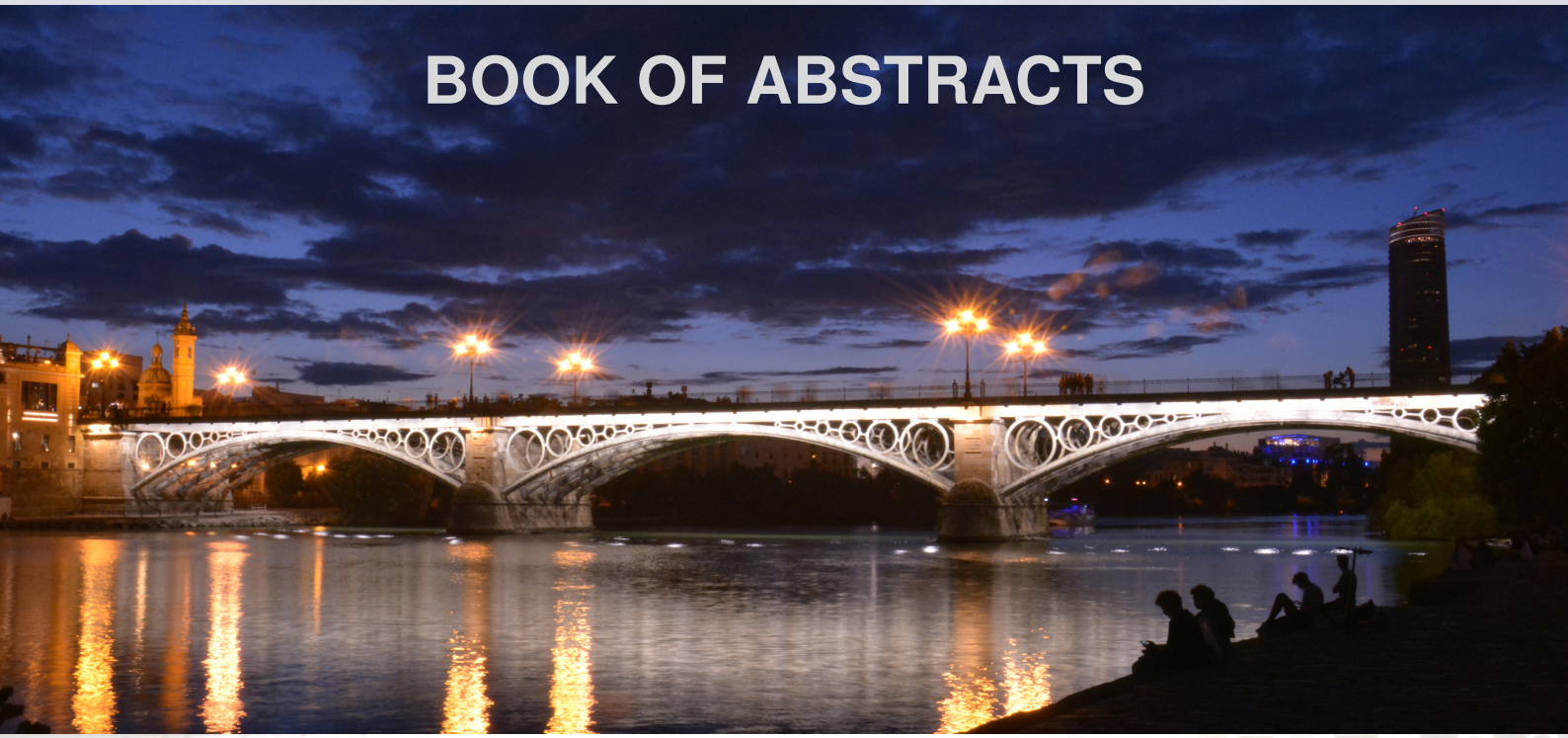
SYSØR 2025

5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING



BOOK OF ABSTRACTS

SEVILLA, 5 - 7 NOVEMBER 2025

ORGANIZERS

imus instituto de matemáticas universidad de sevilla

Seio Sociedad de Estadística, Investigación Operativa y Ciencia de Datos

SPONSORS

EURO THE ASSOCIATION OF EUROPEAN OPERATIONAL RESEARCH SOCIETIES

Junta de Andalucía
Consejería de Economía, Hacienda, Fondos Europeos y Diálogo Social
Instituto de Estadística y Cartografía de Andalucía

decide

uc3m Instituto Flores de Lemus

uc3m | Universidad Carlos III de Madrid

BOOK OF ABSTRACTS

5th Spanish Young Statisticians and Operational Researchers Meeting

5 - 7 November 2025

Seville, Spain

Editors: Organizing Committee

Conference website: https://www.imus.us.es/congresos/5SYSORM/

Welcome

Dear Participant,

As the President of SEIO, it is my great pleasure to welcome you to SYSORM, our youngest and most vibrant gathering of emerging researchers. Now in its fifth edition, the meeting continues to foster a collaborative community of talented young researchers in Statistics and Operations Research, building on the success of previous editions held in Granada (2017), El Escorial (2019), Elche (2022), and Santiago de Compostela (2024).

SEIO is the Spanish society dedicated to the advancement of Statistics and Operations Research, and to fostering the development of their methods and applications in the service to the wider community. With more than 900 members in Spain and around the world, SEIO actively participates in world-wide international activities, being a member of leading organizations such as EURO, FENStatS, ISI, EMS, ALIO, IFORS, and CIMPA, and maintaining collaborations with the IMS, as well as sister societies in Portugal and Latin America.

SEIO disseminates knowledge through its official journals TEST and TOP, its magazine BEIO, co-edited book series, conferences and workshops. SEIO promotes social outreach and statistical literacy, managing a YouTube channel featuring short videos for the public alongside research and thematic seminars. The society's Statistical Literacy Manifesto, launched in 2024, encourages informed decision-making and critical thinking, and has been endorsed by more than 50 national and international entities.

SEIO supports education at all levels, from secondary school competitions to international initiatives, inspiring students' curiosity and interest in Statistics and Operations Research. SEIO fosters equity, diversity, and inclusion, particularly through its Women's Committee, which enhances the visibility and career progression of women in Statistics, Operations Research, and Data Science. In addition, 22 Working Groups organise specialised activities and promote international collaboration and knowledge sharing across these disciplines.

This year has been particularly important for SEIO, as we have updated our statutes to reflect the evolving times, adopting inclusive and contemporary language, and explicitly embracing Data Science as one of our core disciplines. Starting from June 2025, SEIO stands for the Spanish Society of Statistics, Operations Research and Data Science.

Statistics and Operations Research develop and thrive by addressing real-world challenges. Today, the greatest opportunities lie in collaboration across disciplines to tackle the complex problems of our time. By combining expertise from Statistics, Operations Research, and Data Science, we can push the boundaries of data-driven learning and discovery, while respecting the unique identity and core principles of each field.

This meeting celebrates this spirit of innovation and collaboration. It brings together talented young researchers to explore, learn, and shape the future of these disciplines. We are excited

to see the ideas, connections, and collaborations that will emerge—so be curious, be bold, and enjoy this experience!

Warm regards,
Lola Martínez Miranda
President of SEIO

Committees

Organizing Committee

Guerrero, Vanesa (Chair)	Carlos III University of Madrid
Sinova, Beatriz (Chair)	University of Oviedo
Fischetti, Martina	University of Seville
García-García, José	University of Oviedo
Gómez-Vargas, Nuria	University of Seville
Pérez-Fernández, Sonia	University of Oviedo
Sillero-Denamiel, M. Remedios	University of Seville
Temprano, Francisco	Carlos III University of Madrid
Torrejón, Alberto	University of Seville

Scientific Committee

Castilla, Elena (Chair)	Rey Juan Carlos University
Minuesa, Carmen (Chair)	University of Extremadura
Alonso-Pena, María	University of Santiago de Compostela
Baldomero-Naranjo, Marta	University of Cádiz

Organizers and sponsors

Organizers

Sponsors

Program

The program of the 5th Spanish Young Statisticians and Operational Researchers Meeting (SYSORM 2025) comprises both scientific and social activities distributed over the three days of the conference (November 5-7, 2025).

Scientific program

The scientific program of the 5th SYSORM aims to support the training of young researchers through the insights of distinguished experts in the fields of Statistics and Operations Research, while promoting interaction and knowledge exchange across these disciplines. The meeting will only have a continuous session including the following activities:

- 4 plenary talks.
- 2 sponsor sessions.
- 32 contributed talks.

Plenary sessions last one hour, sponsor sessions thirty minutes, and contributed talks are arranged in eight sessions of twenty-five minutes per speaker. To encourage thematic diversity and foster interaction between Statistics and Operations Research, the contributed sessions have been designed to balance presentations from both areas or their intersection.

All plenary talks, sponsor presentations and contributed sessions, as well as coffee breaks, will take place at IMUS facilities.

Social program

The social program of the 5th SYSORM is intended to encourage networking and informal exchange among participants, and includes the following events:

- Welcome cocktail at “El Cabildo” restaurant.
- Guided tour of the Real Alcázar of Seville.
- Gala dinner at “El 29” restaurant.

Schedule

The detailed schedule for the three-day conference is presented in the following tables, where the color code below is used to indicate the thematic area of each contributed presentation:

Operational Research
Statistics
Statistics and Operational Research

Wednesday 5

08:15 - 09:00	Registration		
09:00 - 09:30	Opening		
09:30 - 10:30	Plenary I (Rosa E. Lillo)	Chair: Minuesa, Carmen	
10:30 - 11:00	Coffee break		
11:00 - 11:25	Session I (5 talks)	Chair: Morala, Pablo	Camacho Moro, Jesús
11:25 - 11:50			Martín-Chávez, Pedro
11:50 - 12:15			Segura, Paula
12:15 - 12:40			Cabello García, Esteban
12:40 - 13:05			Cruz Pérez, Lidia
13:05 - 14:30	Lunch		
14:30 - 14:55	Session II (4 talks)	Chair: Camacho Moro, Jesús	Navas Orozco, Antonio
14:55 - 15:20			De Souza, Marcelo
15:20 - 15:45			Pulido Bravo, Belén
15:45 - 16:10			Corberán, Teresa
16:10 - 16:45	Coffee break		
16:45 - 17:45	Plenary II (Christian P. Robert)	Chair: Alonso, María	
17:45 - 18:10	Session III (3 talks)	Chair: De Souza, Marcelo	Tobar Fernández, Cristina
18:10 - 18:35			García-García, José
18:35 - 19:00			Gómez-Vargas, Nuria

20:30 -	Welcome cocktail El Cabildo

Thursday 6

09:30 - 09:55	Session IV (5 talks)	Chair: Pulido Bravo, Belén	Cuesta Santa Teresa, Marina
09:55 - 10:20			Olivares, Adam
10:20 - 10:45			Hernández, Aitor
10:45 - 11:10			Santos Pascual, Miguel
11:10 - 11:35			Bernárdez Ferradás, Alejandro
11:35 - 12:00	Coffee break		
12:00 - 13:00	Plenary III (Francisco Saldahna)	Chair: Baldomero, Marta	
13:00 - 13:30	Sponsor I (IECA)		Figueras Téllez, Cecilia
13:30 - 15:00	Lunch		
15:00 - 15:25	Session V (5 talks)	Chair: Cuesta Santa Teresa, Marina	Terán Viadero, Paula
15:25 - 15:50			Serrano Ortega, Diego
15:50 - 16:15			Nácher, Lorena
16:15 - 16:40			González-Vázquez, Héctor
16:40 - 17:05			Algendi, Abdalrahman

19:00 -	Visit to the Real Alcázar

Friday 7

09:15 - 09:40	Session VI (4 talks)	Chair: Martín-Chávez, Pedro	Corrales, Daniel
09:40 - 10:05			Álvarez, Begoña
10:05 - 10:30			García Arce, Pablo
10:30 - 10:55			Rodríguez-Ballesteros, Sofía
10:55 - 11:25	Coffee break		
11:25 - 12:25	Plenary IV (María Merino)	Chair: Castilla, Elena	
12:25 - 12:55	Sponsor II (Decide)		León, Javier
12:55 - 15:00	Lunch		
15:00 - 15:25	Session VII (3 talks)	Chair: Terán Viadero, Paula	Romero Madroñal, Marcos
15:25 - 15:50			Torrejón, Alberto
15:50 - 16:15			Saavedra Martínez, Samuel
16:15 - 16:45	Coffee break		
16:45 - 17:10	Session VIII (3 talks)	Chair: Segura, Paula	Guillén, María
17:10 - 17:35			Morala, Pablo
17:35 - 18:00			Castro Gomez, José Carlos
18:00 - 18:30	Closing		

21:00 -	Gala dinner El 29

Contents

Plenary talks

SYSØR 2025
5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Our stochastic journey with the network scale-up method

Rosa E. Lillo[1,2] (✉), Sergio Díaz-Aranda[3], Juan M. Ramírez[3], José Aguilar[3] and Antonio Fernández-Anta[4]

[1]Department of Statistics, Universidad Carlos III de Madrid, Spain; [2]UC3M-Santander Big Data Institute, Universidad Carlos III de Madrid, Spain; [3]IMDEA Networks Institute, Madrid, Spain; [4]IMDEA Software Institute, Madrid, Spain

rosaelvira.lillo@uc3m.es

Rosa E. Lillo is a Full Professor of Statistics and Operations Research at Universidad Carlos III de Madrid (UC3M), director of the UC3M–Santander Big Data Institute, and head of the ENIA Chair: AImpulsa.

Network scale-up methods (NSUM) estimate the prevalence or size of population groups from aggregated relational data (ARD) counting of how many contacts a respondent reports in specified categories during a survey (see [4]). By relying on indirect reports rather than personal disclosure, NSUM has become a key tool for studying sensitive or hard-to-reach populations, including sex workers, people who use drugs, those who have had abortions, people with HIV or COVID-19, and even illicit networks.

Since our initial encounter with NSUM through the CoronaSurveys initiative (coronasurveys.org), we have used it as a cost-effective instrument for real-time inference in pandemic monitoring, voting intentions, and caregiving dynamics. Beyond these practical uses, our contributions, pivoting between stochastic processes, probability, and statistics, are threefold: (i) probabilistic bounds for prediction errors that apply broadly across NSUM estimators are derived [3]; (ii) the performance of well-known estimators in realistic simulation studies that incorporate plausible network topologies is evaluated [1]; and (iii) robust NSUM estimators that reduce bias in indirect surveys are proposed [2]. The goal of the talk is to present the practical–theoretical view of NSUM, highlighting its versatility and its potential to address complex societal questions while outlining stimulating challenges for the statistical community.

Keywords: aggregated relational data; hidden population; network scale-up method; robust estimators; stochastic bounds.

References

[1] S. Díaz-Aranda, J. Aguilar, J.M. Ramírez, D. Rabanedo, A. Fernández-Anta, and R.E. Lillo. Performance analysis of NSUM estimators in social-network topologies. *The American Statistician*, 79(2):247–264, 2024. doi:10.1080/00031305.2024.2421361.

[2] S. Díaz-Aranda, J.M. Ramírez, J. Aguilar, R.E. Lillo, and A. Fernández-Anta. Robust network scale-up method estimators. *Social Networks*, 84:46–61, 2026. doi:10.1016/j.socnet.2025.08.002.

[3] S. Díaz-Aranda, J.M Ramírez, M. Daga, J.P. Champati, J Aguilar, R.E. Lillo, and A. Fernández Anta. Error bounds for the network scale-up method. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '25)*, Toronto, ON, Canada, 2025. Association for Computing Machinery. doi:10.1145/3711896.3736940.

[4] I. Laga, L. Bao, and X. Niu. Thirty years of the network scale-up method. *Journal of the American Statistical Association*, 116(535):1548–1559, 2021. doi:10.1080/01621459.2021.1935267.

SYSØR 2025
5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

A Bayesian decision-theoretic framework for privacy

Christian P. Robert[1,2] (✉)

[1]CEREMADE, Université Paris Dauphine, PSL, France; [2]Department of Statistics, University of Warwick, United Kingdom

xian@ceremade.dauphine.fr

Christian P. Robert is Professor at CEREMADE, Université Paris-Dauphine and at the University of Warwick, Department of Statistics, UK. He is also a 2023-2029 ERC Synergy awardee member for the OCEAN project and a Paris School of AI chair since 2019. He has been the Editor of the Journal of the Royal Statistical Society (Series B, Statistical Methodology) and a Deputy-Editor for the journal Biometrika. He was President of the International Society for Bayesian Analysis (ISBA) in 2008 and is a Fellow of the RSS, the IMS, the ISBA and the ASA. His reseach interests cover Bayesian statistics (decision theory, model choice, foundations, objective Bayesian methodology), Computational statistics (Monte Carlo methodology, MCMC methods, sequential importance sampling, approximate Bayesian computation (ABC), convergence diagnoses) and Latent variable models (mixtures, hidden Markov models). He has written a dozen books and more than 200 articles in these domains.

While several results in the literature (e.g., [1], [4]) demonstrate that Bayesian inference approximated by MCMC output can achieve differential privacy ([2]) with zero or limited impact on the ensuing posterior, we reassess this perspective via an alternate "exact" MCMC perturbation inspired from [3] within a federated learning setting. Since the ensuing privacy criterion is mostly related to a slowing-down of MCMC convergence rather than a generic gain in protecting data privacy, we propose an alternative decision-theoretic framework that accommodates more realistic privacy constraints.

This is a joint and on-going work with Joshua Bon, Judith Rousseau, and Julien Stoehr.

Keywords: Bayesian decision theory; Bayesian inference; computational methods; differential privacy; MCMC methods.

References

[1] C. Dimitrakakis, B. Nelson, Z. Zhang, A. Mitrokotsa, and B.I.P. Rubinstein. Differential privacy for Bayesian inference through posterior sampling. *Journal of Machine Learning Research*, 18(11):1–39, 2017. URL: http://jmlr.org/papers/v18/15-257.html.

[2] C. Dwork. Differential privacy. In *33rd International conference on Automata, Languages and Programming*, pages 1–12. Springer, 2006. URL: https://link.springer.com/chapter/10.1007/11787006_1.

[3] G.K. Nicholls, C. Fox, and A.M. Watt. Coupled MCMC with a randomized acceptance probability, 2012. URL: https://arxiv.org/abs/1205.6857, arXiv:1205.6857.

[4] W. Zhang and R. Zhang. DP-fast MH: Private, fast, and accurate Metropolis-Hastings for large-scale Bayesian inference. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 41847–41860. PMLR, 23–29 Jul 2023. URL: https://proceedings.mlr.press/v202/zhang23aw.html.

5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Robust and stochastic facility location: an overview

Francisco Saldanha-da-Gama[1] (✉)

[1]Sheffield University Management School, United Kingdom

francisco.saldanha-da-gama@sheffield.ac.uk

Francisco Saldanha-da-Gama is Chair of Supply Chain Management at Sheffield University Management School (UK). His research interests include Operations Research, Supply Chain Management, Logistics, Decision-Making under Uncertainty, Facility Location, and Project Scheduling. He is the Editor-in-Chief of Computers & Operations Research, and a member of the Editorial Advisory Board of several other journals.

A discrete facility location problem is a classical example of a network design and optimization problem. This is a problem that consists of determining the best location for one or several facilities to serve a set of demand points. The exact meaning of "best" depends on the particular application considered (objectives, constraints, subsidiary decisions, etc). These problems are often found at the core of more comprehensive problems arising in areas such as Logistics, Telecommunications, Supply Chain Network Design and Transportation. Although the network structure underlying a discrete facility location problem can be looked at as being simple, the problems easily become computationally challenging.

Many facility location problems involve strategic decisions that must hold for some considerable time after being implemented. During this time, unpredictable changes may occur in the underlying data. Sources of uncertainty in facility location include demand levels, travel time, cost for supplying the customers, location of the customers, presence or absence of the customers, and price for the commodities, to mention a few [1]. In many cases it is desirable to make decisions that hedge against such uncertainty.

In this talk we use some classical discrete facility location problems to show how uncertainty can be embedded in this type of problems. Particular emphasis is put on robust optimization and stochastic programming techniques.

Keywords: discrete facility location; robust optimization; stochastic programming.

References

[1] F. Saldanha-da-Gama and S. Wang. *Facility Location under Uncertainty: Models, Algorithms and Applications*. Springer, 2024. doi:10.1007/978-3-031-55927-3.

5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Stochastic optimization in location and dispatching for emergency medical services

María Merino Maestre[1,2] (✉)

[1]Department of Mathematics, University of the Basque Country-UPV/EHU, Spain; [2]Basque Center for Applied Mathematics-BCAM, Spain

maria.merino@ehu.eus, mmerino@bcamath.org

María Merino Maestre is Full Professor (under contract) in Statistics and Operations Research at UPV/EHU, where she earned her Mathematics degree (1999) and Ph.D. (2005). Her research covers stochastic and combinatorial optimization, risk management, and applications in industrial, health, and humanitarian contexts. She is researcher at BCAM, a member of SEIO's Women Commission, the EURO WISDOM Forum, coordinates the Ph.D. Program in Mathematics and Statistics at UPV/EHU and leads collaborations with the Basque Public Health Service.

Efficient management of Emergency Medical Services is vital, particularly when resources are limited and rapid response is critical. Strategic ambulance location and dispatch planning are key components, but these are challenged by uncertainties related to emergency timing, location, and response conditions such as traffic or ambulance availability. Various methodological approaches have been explored, notably queuing theory, simulation techniques, and mathematical optimization. For comprehensive reviews, see [3, 1, 2].

[4] proposes a novel approach to the multi-objective nature of location and dispatch problems by formulating a two-level hierarchical compromise model. It prioritizes emergency coverage while simultaneously balancing secondary objectives, such as minimizing average response time and Conditional Value-at-Risk to account for worst-case scenarios. The resulting stochastic problems with cross-scenario constraints are solved via a matheuristic scenario-decomposition algorithm. The framework is validated using real-world data from the Basque Country, Spain, demonstrating its effectiveness and practical relevance.

Keywords: branch-and-fix coordination; conditional value-at-risk; hierarchical compromise; OR in health services; stochastic programming.

References

[1] R. Aringhieri, M.E. Bruni, S. Khodaparasti, and J.T. van Essen. Emergency medical services and beyond: Addressing new challenges through a wide literature review. *Computers & Operations Research*, 78:349–368, 2017. doi:10.1016/j.cor.2016.09.016.

[2] V. Bélanger, A. Ruiz, and P. Soriano. Recent optimization models and trends in location, relocation, and dispatching of emergency medical vehicles. *European Journal of Operational Research*, 272(1):1–23, 2019. doi:10.1016/j.ejor.2018.02.055.

[3] L. Brotcorne, G. Laporte, and F. Semet. Ambulance location and relocation models. *European Journal of Operational Research*, 147(3):451–463, 2003. doi:10.1016/S0377-2217(02)00364-8.

[4] I. Gago-Carro, U. Aldasoro, D.-J. Lee, and M. Merino. Hierarchical compromise optimization of ambulance locations under stochastic travel times. *Computers & Operations Research*, 184:107208, 2025. doi:10.1016/j.cor.2025.107208.

Sponsor talks

5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

New information needs in the workplace. Preparation of longitudinal indicators based on census data on social security affiliation. The Andalusian case

Cecilia Figueras-Téllez[1] (✉) and Joaquín Planelles-Romero[1]

[1]Institute of Statistics and Cartography of Andalusia (IECA), Spain

cecilia.figueras@juntadeandalucia.es

Cecilia Figueras-Téllez has a degree in Statistics from the University of Seville. Recently incorporated into the Institute of Statistics and Cartography of Andalusia, specifically in the Demographic and Social Statistics service. She works on a project about the longitudinal study of affiliations and the labor market. Previously, she worked at Ingebau as a data statistician, and she has a specialization in Data Science and Big Data through Datahack.

The labour market is undergoing an accelerated transformation that affects all its areas: the workplace, the composition of the workforce and even organisational models. This ongoing evolution responds to the need to adapt to the changes driven by technological progress [3], globalization and recurrent social and economic crises. Given this dynamic context, new challenges arise for statistical production. It is essential to incorporate indicators that reflect this new reality of employment. Some studies about the work life have been carried out regarding the paradigm shift from the transversal to the longitudinal, allowing us to focus on specific cases such as that of youth labour tracks, comparing two cohorts in the city of Barcelona in 2008 and 2016 [1]. However, the objective of this study goes further, dealing with the continuous sample of working lives, MCVL [4], prepared by Ministerio de inclusión, Seguridad Social y Migraciones (MISM), INE and Agencia Estatal de Administración Tributaria (AEAT), after considering the loss of sample and possible biases [2].

Thus, the objective of this work is to enrich statistical production through indicators that move from cross-sectional to longitudinal study. Based on the monthly files that the regional statistical offices get from the Social Security, until now exploited from a cross-sectional perspective, research has been carried out aimed at building a longitudinal database throughout 2024 in Andalusia. This new structure makes it possible to offer more relevant information that is adjusted to current demands for decision-making, such as the transformation of affiliation over time, or changes in its sequence. This paper presents the work carried out with a broad description of the methodology applied for the construction of the database, an analysis of the main challenges that have been faced in the process and the solutions finally implemented, based on this innovative approach.

Keywords: affiliation to Social Security; labor market; longitudinal analysis.

References

[1] F. Antón-Alonso, S. Porcel, and I. Cruz-Gómez. La precarización creciente de las trayectorias laborales juveniles en la ciudad de barcelona. un análisis integrando las perspectivas de curso vital y generacional. *Papers-Revista de Sociología*, 108(1):1–26, 2023. doi:10.5565/rev/papers.30151-26.

[2] J.M. Arranz, C. García-Serrano, and V. Hernanz. How do we pursue "labormetrics"? An application using the MCVL. *Estadística Española*, 55(181):231–254, 2013. URL: https://ebuah.uah.es/dspace/handle/10017/9661.

[3] J.J. Dolado, F. Felgueroso, and J.F. Jimeno. Past, present and future of the spanish labour market: when the pandemic meets the megatrends. *Applied Economic Analysis*, 29(85):21–41, 2021. doi:10.1108/AEA-11-2020-0154.

[4] Ministerio de Inclusión, Seguridad Social y Migraciones. *MCVL Muestra continua de vidas laborales, guía de contenido*, 2024. URL: https://portaldatos.seg-social.gob.es/mcvl.

SYSØR 2025
5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

A multi-stage optimization framework for urban rail crew scheduling

Manuel Navarro García[1] (✉) and Javier León Caballero[1]

[1]Decide4AI, Spain

manuel.navarro@decidesoluciones.es

Manuel Navarro García is an Optimization Consultant at Decide4AI, where he develops mathematical models to support strategic planning and improve operational efficiency. He holds degrees in Mathematics and Physics and a PhD in Mathematical Engineering from the Carlos III University of Madrid. His work bridges mathematical optimization and data science, with experience in both academic research and industrial applications.

The crew scheduling problem in public transport systems, particularly in urban rail networks, consists in assigning operational duties such as train driving and station supervision to a set of available workers. The objective is to comply with operational and labor regulations while maximizing coverage and improving aspects like reducing unnecessary transfers and balancing workload among workers. From a methodological perspective, recent research has largely shifted toward metaheuristics and decomposition strategies due to their flexibility and computational efficiency [1, 2, 3]. Nevertheless, real-world applications still pose significant challenges, as labor regulations and organizational responsibilities evolve over time. This motivates the need for adaptable, modular, and maintainable solution frameworks.

We present a multi-stage optimization framework developed in an industrial context to address crew scheduling in a metropolitan subway system. The approach begins by segmenting long activity blocks into feasible work units using a mixed-integer formulation. Driving services are first constructed using constraint programming to satisfy scheduling constraints, and a column generation approach is applied to select the subset that maximizes coverage while improving secondary operational criteria. A local search phase is then applied to mitigate the effects of the initial segmentation by introducing small modifications that substantially reduce transfers between stations. Finally, station supervision tasks—considered lower priority—are assigned via a mixed-integer program to complete the schedules. The framework is currently in production, delivering robust and high-quality schedules across a range of planning scenarios.

Keywords: constraint programming; decomposition techniques; railway crew scheduling.

References

[1] E.J.W. Abbink, L. Albino, T. Dollevoet, D. Huisman, J. Roussado, and R.L. Saldanha. Solving large scale crew scheduling problems in practice. *Public Transport*, 3:149–164, 2011. doi:10.1007/s12469-011-0045-x.

[2] D. Huisman. A column generation approach for the rail crew re-scheduling problem. *European Journal of Operational Research*, 180(1):163–173, 2007. doi:10.1016/j.ejor.2006.04.026.

[3] F. Leutwiler and F. Corman. A review of principles and methods to decompose large-scale railway scheduling problems. *EURO Journal on Transportation and Logistics*, 12:100107, 2023. doi:10.1016/j.ejtl.2023.100107.

Contributed talks

5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Upper Lipschitz rates in linear programming

Jesús Camacho[1] (✉), María Josefa Cánovas[1], Helmut Gfrerer[2] and Juan Parra[1]

[1] Center of Operation Research, University Miguel Hernandez of Elche, Spain; [2] Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Austria

j.camacho@umh.es

Jesús Camacho is an Assistant Professor at the University Miguel Hernández of Elche. His research activity is focused in optimization and, particularly, variational rates of linear and convex problems. Before enrolling his PhD in the University Miguel Hernández of Elche, he did a MSc in and BSc in Mathematics at University of Seville.

During the last century, variational analysis has provided qualitative and quantitative properties to ensure robustness or stability of the mathematical program we are dealing with. In this general matter, the seminal contributions of Hoffman (see [3]), stand out. In particular, the rising interest in the field of upper Lipschitz constants associated to linear problems is illustrated by the number of articles around it; we specially mention here the stating points of our research [4, 5].

In this presentation, we study the so-called sharp Hoffman constant (a global measure of stability) for linear problems subject to canonical perturbations (affecting both the objective function and the right-hand side). This quantifier reflects the ratio between parameter perturbations and the displacement of solutions. As an analytical definition, it intrinsically involves infinitely many parameters and points to check. Our contribution, detailed in [1, 2], is to provide a computable formula for this constant in terms of the nominal problem data. In other words, our approach reduces the problem to considering a finite number of parameters and points.

Keywords: calmness; Hoffman constants; linear programming; Lipschitz upper semicontinuity; variational analysis.

References

[1] J. Camacho, M.J. Cánovas, H. Gfrerer, and J. Parra. Hoffman Constant of the Argmin Mapping in Linear Optimization, 2025. arXiv:10.48550/arXiv.2307.01034.

[2] J. Camacho, M.J. Cánovas, and J. Parra. From calmness to Hoffman constants for linear semi-infinite inequality systems. *SIAM Journal on Optimization*, 32(4):2859–2878, 2022. doi:10.1137/21M1418228.

[3] A.J. Hoffman. On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards*, 49(4):263–265, 1952. doi:10.6028/jres.049.027.

[4] M.H. Li, K.W. Meng, and X.Q. Yang. On error bound moduli for locally lipschitz and regular functions. *Mathematical Programming Series A*, 171:463–487, 2018. doi:10.1007/s10107-017-1200-1.

[5] J. Peña, J.C. Vera, and L.F. Zuluaga. New characterizations of hoffman constants for systems of linear constraints. *Mathematical Programming*, 187:79–109, 2021. doi:10.1007/s10107-020-01473-6.

SYSØR 2025
5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Estimating the eigenmeasure of the stochastic neutron transport equation

Pedro Martín-Chávez[1] (✉), Emma Horton[1] and Andreas E. Kyprianou[1]

[1]Department of Statistics, University of Warwick, United Kingdom

Pedro.Martin-Chavez.1@warwick.ac.uk

Pedro Martín-Chávez is a Postdoctoral Research Fellow at University of Warwick funded by MaThRad (EPSRC grant EP/W026899/2). His research activity is focused on stochastic modeling through branching processes, with applications to nuclear engineering. He obtained his PhD from University of Extremadura (UEx). Previously, he worked at UEx as Substitute Lecturer and Research Assistant, and did a MSc in Mathematics at UEx and a BSc in Mathematics and Physics at University of Sevilla.

The neutron transport equation (NTE) models the evolution of particle populations in fissile media and plays a central role in nuclear engineering. Its solution can be interpreted probabilistically as the mean behavior (expectation semigroup) of a neutron branching process (NBP), where individual neutrons move, scatter, and produce fission events randomly. This semigroup is a family of operators describing how the expected value of a test function of the process evolves over time, and its long-term behavior admits a Perron–Frobenius decomposition into a leading eigenvalue, an associated eigenfunction, and an eigenmeasure (see [2]). Estimating this eigentriple is crucial for understanding criticality and particle distribution in the system. Recent Monte Carlo methods (cf. [1]) provide efficient estimators for the eigenvalue and the eigenmeasure—the latter corresponding to the quasi-stationary distribution of the NBP—but the eigenfunction remains more computationally demanding to estimate, as it requires simulations initialized at every point in the phase space.

In this work, we propose a new method to estimate the eigenfunction by leveraging a time-reversal approach. Starting from a many-to-one representation of the expectation semigroup, we reinterpret the mean behavior of the NBP in terms of a single (forward) neutron random walk (NRW), where particles propagate in the direction of their velocity between scattering events (there is no fission). We then introduce a backward NRW, in which particles propagate in the opposite direction, using the same physical parameters but guided by the estimated eigenmeasure. By performing Monte Carlo simulations of this backward process, we recover the quasi-stationary distribution associated with the reversed dynamics. This in turn allows us to construct an estimate of the eigenfunction in a more computationally tractable way. The proposed method offers a practical route to complete the estimation of the eigentriple associated with the stochastic formulation of the NTE.

Keywords: Monte Carlo simulation; neutron branching process; neutron transport equation.

References

[1] A.M.G. Cox, S.C. Harris, A.E. Kyprianou, and M. Wang. Monte Carlo methods for the neutron transport equation. *SIAM/ASA Journal on Uncertainty Quantification*, 10(2):775–825, 2022. doi:10.1137/21M1390578.

[2] E. Horton and A.E. Kyprianou. *Stochastic Neutron Transport*. Birkhäuser Cham, 2023. doi:10.1007/978-3-031-39546-8.

5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Interpreting clusters with mathematical optimization

Paula Segura[1] (✉) and Alfredo Marín[2]

[1]Department of Mathematics for Economics and Business, University of Valencia, Spain; [2]Department of Statistics and Operations Research, University of Murcia, Spain

paula.segura-martinez@uv.es

Paula Segura holds a PhD in Statistics and Optimization from the University of Valencia. She previously completed a MSc in Advanced Mathematics at the University of Murcia and a BSc in Mathematics at the University of Alicante. Her research focuses on the study of mixed integer linear programming problems, particularly in transportation and logistics, arc routing, and location.

One of the most popular unsupervised learning methods is cluster analysis [2]. The growing popularity of machine learning methods in data driven decision making in recent years has led to an increase in their complexity, and the interpretability of these methods has also been directly affected. One way to improve interpretability in cluster analysis is through the so-called post-hoc models, in which a set of predefined clusters is available and the goal is to find an explanation that characterizes each cluster. In [1], the authors propose a post-hoc methodology for interpreting clusters, based on defining a prototype for each cluster that characterizes it as precisely as possible. The explanation of each cluster should apply to as many individuals as possible within the cluster, and to as few individuals as possible in the remaining clusters.

As an extension of the work developed in [1], we present in this talk a new approach for improving the interpretability of the results of cluster analysis through distance-based explanations. Let us consider a set of clusters, whose individuals present K different measures or characteristics, a dissimilarity between each pair of individuals associated with each of these measures, and an integer parameter $q \leq K$. The goal is to find an explanation that characterizes each cluster by choosing for each of them a minimum set of representative individuals or prototypes. This chosen set must guarantee that those individuals that are close to one prototype in at least q out of the K measures are allocated to its same cluster. We propose two new mathematical optimization models based on a different definition of closeness between individuals, inspired by classic location analysis problems.

Keywords: clusters; combinatorial optimization; interpretability; location.

References

[1] E. Carrizosa, K. Kurishchenko, A. Marín, and D. Romero–Morales. Interpreting clusters via prototype optimization. *Omega*, 107:102543, 2022. doi:10.1016/j.omega.2021.102543.

[2] G. Gan, C. Ma, and J. Wu. *Data Clustering: Theory, Algorithms, and Applications*. ASA-SIAM Series on Statistics and Applied Probability, 2007. doi:10.1137/1.9780898718348.

SYSOR 2025
5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Dirichlet mixed models for compositional small area estimation

Esteban Cabello[1] (✉), María Dolores Esteban[1], Tomàš Hobza[2], Domingo Morales[1] and Agustín Pérez[3]

[1]Center of Operations Research, Miguel Hernández University of Elche, Spain; [2]Department of Mathematics, Czech Technical University in Prague, Czech Republic; [3]Department of Economic and Financial Studies, Miguel Hernández University of Elche, Spain.

ecabello@umh.es

Esteban Cabello is a third-year PhD student in the PhD in Statistics, Optimization and Applied Mathematics at Miguel Hérnandez University of Elche. His research activity is focused on temporal and spatial linear mixed models applicable to the estimation of small-area multivariable indicators. Before enrolling his PhD, he did a MSc in Applied Statistics at Granada University and a BSc in Mathematics at Málaga University.

Small area estimation (SAE) provides model-based alternatives to direct estimators when producing reliable predictors for domains with limited sample sizes. A central challenge in SAE arises when the target variables are compositional, such as labour market indicators, which are proportions constrained to sum to one. Traditional area-level SAE models are linear and univariate or multivariate, but not designed to account directly compositional constraints. Recent works have addressed compositional constrains modeling log-ratio transformations [1, 2] or under Beta and Dirichlet distributions in fixed-effects frameworks [3].

We propose a novel nonlinear multivariate area-level Dirichlet mixed model to estimate labour force compositions while preserving the compositional structure. The model introduces domain-level random effects through a logit link to jointly estimate proportions constrained to the unit simplex. A Laplace approximation is used for parameter estimation. Predictors of proportions, totals and rates of small areas are obtained, and their mean square errors are estimated by parametric bootstrap. Several simulation experiments are carried out to analyze the behaviour of the fitting algorithm, the small area predictors and the bootstrap procedure. Finally, an application to real data from the Spanish Labour Force Survey, in the last quarter of 2022, is given. The target is the estimation of domain proportions of employed, unemployed and inactive people and unemployment rates by province and sex.

Keywords: area-level models; compositional data; Dirichlet mixed models; labour force survey; small area estimation.

References

[1] M.D. Esteban, M.J. Lombardía, E. López-Vizcaíno, D. Morales, and A. Pérez. Small area estimation of proportions under area-level compositional mixed models. *Test*, 29:793–818, 2020. doi:10.1007/s11749-021-00767-x.

[2] J. Krause, J.P. Burgard, and D. Morales. Robust prediction of domain compositions from uncertain data using isometric logratio transformations in a penalized multivariate Fay–Herriot model. *Statistica Neerlandica*, 76(1):65–96, 2022. doi:10.1111/stan.12253.

[3] Z.-Z. Tang and G. Chen. Zero-inflated generalized dirichlet multinomial regression model for microbiome compositional data analysis. *Biostatistics*, 20(4):698–713, 2019. doi:10.1093/biostatistics/kxy025.

5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

The total distance dominating set problem

Lidia Cruz[1] (✉), Ana D. López-Sánchez [1] and Eva Barrena[1]

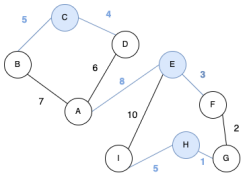
[1]Department of Economics, Quantitative Methods, and Economic History, University Pablo de Olavide, Spain

lcruper@upo.es

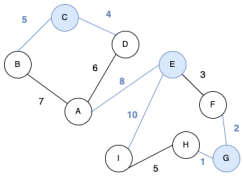
Lidia Cruz is a first-year PhD student in the PhD in Analysis, Modelling, and Applications of Domination Problems in Graphs at University Pablo de Olavide. Her research activity is focused on the design and implementation of metaheuristic algorithms for various domination problems in graphs. Before enrolling her PhD, she completed a Double Degree in Computer Science and Mathematics at University of Seville and a Master in Intelligent Systems in Energy and Transport at University of Seville.

Domination problems in graphs have been extensively studied due to their broad applicability in real-world scenarios. Specifically, the Minimum Dominating Set Problem (MDSP) [1] aims to identify the smallest subset of nodes such that every node not included in the subset is adjacent to at least one node within it. Practical applications of this problem include service location, communication network design, and social network analysis. However, the MDSP does not consider graph with distances neither the distances between dominating and dominated nodes, leading to inefficient connections in terms of cost, time, or quality.

The Total Distance Dominating Set Problem (TDDSP) focuses on achieving a trade-off between minimizing the size of the dominating set and minimizing the total distance between the dominating nodes and the nodes they dominate. Figure 3.1 presents an example of two optimal solutions for the MDSP. However, with respect to the TDDSP, the dominating set depicted in Figure 3.1b is preferable, as it yields a smaller total distance.



(a) Dominating set with total distance equal to 30.



(b) Dominating set with total distance equal to 26.

Figure 3.1: Example of dominating sets with the same cardinality but different total distance.

The TDDSP is a bi-objective problem and, since the MDSP is NP-hard, the TDDSP is also NP-hard. To solve it, a metaheuristic based on Variable Neighborhood Search (VNS) [2] has been implemented. This procedure is designed to overcome the issue of getting stuck in a local minima by automatically changing neighbourhoods during the search.

Keywords: dominating set; multiobjective optimization; variable neighborhood search; weighted graphs.

References

[1] E.J. Cockayne and S.T. Hedetniemi. Towards a theory of domination in graphs. *Networks*, 7:247–261, 1977. doi:10.1002/net.3230070305.

[2] P. Hansen, N. Mladenovic, and J. Moreno-Pérez. Variable neighbourhood search: Methods and applications. *4OR*, 175:367–407, 2010. doi:10.1007/s10479-009-0657-6.

SYSOR 2025
5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Robustness guarantees for counterfactual explanations

Emilio Carrizosa[1] and Antonio Navas-Orozco[1] (✉)

[1] Instituto de Matemáticas de la Universidad de Sevilla, Universidad de Sevilla, Spain

anavas1@us.es

Antonio Navas-Orozco is a first-year PhD student in the Doctorate Programme in Mathematics at Universidad de Sevilla. His research activity is focused on mathematical optimization applied to machine learning and, in particular, the study of robustness in counterfactual explanations. Before enrolling his PhD, he worked as a researcher at the SANEVEC Project, and did a MSc in Physics and Mathematics at Universidad de Granada and a double BSc in Physics and Mathematics at Universidad de Sevilla.

Generalized Linear Models (GLMs) [1] are well-established Data Science models (including linear and logistic regression) which are widely-used for regression and classification tasks. Counterfactual Explanations (CEs) are tools to make the outcome of Machine Learning models interpretable to the end-user and eventually provide actionable recourse to obtain more favourable outcomes. Specifically, they provide individual explanations of decisions, specifying the smallest feasible change to a given input to switch the model's output into a more desirable one. One of the drawbacks of CEs is the lack of robustness to changes in the model [2] and, in particular, the data used to train it. In the case of GLMs, the model is trained by computing the Maximum Likelihood Estimator (MLE) β for the given data. However, this β is not necessarily the real β_0 of the true underlying data distribution.

In this work, we propose a mathematical optimization model to approach robustness to model misspecification in CEs applied to GLMs. Robustness is understood as guaranteeing that the outcome of the real model, parametrized by the unknown β_0, is sufficiently high, using only the available information: the MLE β and well-known properties of GLMs. To address this problem, we propose a statistical test on the outcome of the model, so that the outcome of the real model is high enough with controlled, high probability. The resulting optimization problem is a SOCP (Second-Order Cone Program) readily solvable with off-the-shelf solvers such as Gurobi. Our approach can include models not belonging to the GLM family, fitting a local (in the given record) GLM surrogate, and favour sparsity, understood as minimizing the l_0 norm of the proposed changes, at the expense of an increased computational cost for higher number of features.

Keywords: counterfactual explanations; data-driven optimization; machine learning; multi-objective decision-making; robustness.

References

[1] A. Agresti. *Foundations of Linear and Generalized Linear Models*. John Wiley & Sons, 2015.

[2] T. Laugel, M.J. Lesot, C. Marsala, X. Renard, and M. Detyniecki. Unjustified classification regions and counterfactual explanations in machine learning. In *Machine Learning and Knowledge Discovery in Databases*, pages 37–54, Cham, 2020. Springer. doi:10.1007/978-3-030-46147-8_3

5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Optimizing drone delivery of medical supplies in disaster scenarios

Marcelo de Souza[1] (✉) and Clara dos Santos Becker[1]

[1]Department of Software Engineering, Santa Catarina State University, Brazil

marcelo.desouza@udesc.br

Marcelo de Souza is a Brazilian researcher at Santa Catarina State University. He holds a PhD in Computer Science from the Federal University of Rio Grande do Sul (2022). His research focuses on artificial intelligence and optimization techniques, with applications across diverse domains, including industry, social systems, and healthcare.

This work addresses the challenge of medical supply delivery in disaster scenarios, where traditional ground transportation is often prevented. We model this problem as a Vehicle Routing Problem (VRP) [2], specifically tailored for drone operations. The model incorporates real-world constraints, including drone limitations and autonomy, translated into route time limit and payload capacity. Our objective is to optimize drone routes to efficiently deliver medical supplies to multiple affected locations. Several works explore the potential of drones in (post-)disaster respose [3]. In particular, drones enable safer, faster, and more cost-effective medical supply delivery and facilitate remote patient assessment prior to the arrival of on-site medical personnel.

To solve the VRP, we employ a random-key optimization framework [1], exploring a suite of metaheuristic algorithms (e.g. GRASP, iterated local search, large neighborhood search, among others). We leverage automatic configuration techniques to obtain high-performing parameter settings. A comprehensive performance analysis is presented, studying the behavior and effectiveness of each metaheuristic under diverse parameter configurations. This approach allows us to determine the most robust and efficient solution strategies for drone-based medical supply delivery. We validate our model and solution methodology using a case study based on recent flooding in southern Brazil. This real-world application demonstrates the practical benefits of drone-based delivery, particularly in situations where land access is severely restricted or impossible. We quantify the cost of drone deployment and compare it to alternative delivery methods, highlighting the potential for significant improvements in response time and accessibility. The findings provide evidence for the efficacy of drones as a valuable technology in disaster response, enabling rapid and optimized medical supply distribution to affected populations.

Keywords: metaheuristics; random-key optimization; vehicle routing problem.

References

[1] A.A. Chaves, M.G.C. Resende, M.J.A. Schuetz, J.K. Brubaker, H.G. Katzgraber, E.F. Arruda, and R. Silva. A random-key optimizer for combinatorial optimization. 2024. arXiv:2411.04293.

[2] G.B. Dantzig and J.H. Ramser. The truck dispatching problem. *Management Science*, 6(1):80–91, 1959. doi:10.1287/mnsc.6.1.80

[3] S.M.S.M. Daud, M.Y.P.M. Yusof, C.C. Heo, L.S. Khoo, M.K.C. Singh, M.S. Mahmood, and H. Nawawi. Applications of drone in disaster management: A scoping review. *Science & Justice*, 62(1):30–42, 2022. doi:10.1016/j.scijus.2021.11.002

Addressing dependence and enhancing robustness in network scale-up method estimators

Antía Enríquez[1], Rosa E. Lillo[1,2] and Belén Pulido[3] (✉)

[1]UC3M-Santander Big Data Institute (IBiDat), Universidad Carlos III de Madrid, Spain; [2]Department of Statistics, Universidad Carlos III de Madrid, Spain; [3]Department of Statistics, Operational Research and Numeric Calculus, Universidad Nacional de Educación a Distancia, Spain

belen.pulido@ccia.uned.es

Belén Pulido is an Assistant Professor at Universidad Nacional de Educación a Distancia. She received her PhD in Mathematical Engineering (Statistics) from Universidad Carlos III de Madrid (UC3M) in 2024. Her thesis was focused on the extension of classical unsupervised classification methods to the functional data analysis context. Her postdoctoral work includes studying network scale-up method. Prior to her PhD, she did a MSc in Big Data at UC3M, and a BSc in Mathematics at Universidad de Málaga.

The Network Scale-up Method (NSUM) is an indirect survey-based technique used to estimate the size of hard-to-reach or hidden populations. These are groups that are difficult to count directly through traditional surveys, usually due to factors like stigma, illegality, or simply being a small fraction of the overall population. Instead of trying to count every individual in the hidden group, NSUM leverages the social networks of a sample from the general population. The core idea is: "How many people do you know?" followed by "Out of those, how many belong to X?", where X is the population of interest. The NSUM has been used to estimate the size of a variety of subpopulations, including female sex workers, drug users, and even children who have been hospitalized for choking. A comprehensive overview of these methods can be found in [1].

While NSUM has proven valuable across various applications, conventional estimators are often applied to estimate the size of a single hidden population at a time. When the goal is to simultaneously estimate the sizes of multiple distinct hidden populations using data from the same general population survey, existing methods typically overlook the dependencies that arise between the estimators for these different populations. This research addresses these limitations by proposing two different types of estimators. Firstly, we propose novel estimators that consider copula functions to explicitly model and incorporate the dependency structures inherent in multivariate survey data. Secondly, we develop robust estimators based on statistical depth concepts, designed to be less sensitive to outliers and deviations from assumed data distributions. These methodologies are validated using both synthetic and real datasets. Furthermore, we will present preliminary insights from applying these enhanced NSUM techniques to a study, funded by the Instituto de las Mujeres (Ministerio de Igualdad), aimed at examining societal perceptions of gender equality and the distribution of care, highlighting their practical utility.

Keywords: dependency estimation; indirect surveys; network scale-up method; robust estimators.

References

[1] I. Laga, L. Bao, and X. Niu. Thirty years of the network scale-up method. *Journal of the American Statistical Association*, 116(535):1548–1559, 2021. doi:10.1080/01621459.2021.1935267.

SYSØR 2025
5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

The periodic drone arc routing problem with irregular services and maximum benefits

Teresa Corberán[1] (✉), Renata Mansini[2], Isaac Plana[3] and José María Sanchis[4]

[1]Departament d'Estadística i Investigació Operativa, Universitat de València, Spain; [2]Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Brescia, Italy; [3]Departament de Matemàtiques per a l'Economia i l'Empresa, Universitat de València, Spain; [4]Departament de Matemàtica Aplicada, Universitat Politècnica de València, Spain

tecorfa@alumni.uv.es

Teresa Corberán is a fourth-year PhD student in the PhD in Statistics and Operations Research at the University of Valencia. Her research focuses on arc routing problems involving drones to perform services along the arcs or edges of a network. She develops both exact and heuristic algorithms to solve these problems. Also, she works in a company where she implements heuristic algorithms to address real-world problems in routing, demand planning, and forecasting.

Arc routing problems involving drones have gained interest due to their applicability in inspection, maintenance, and monitoring tasks. This work addresses an extension of the Periodic Rural Postman Problem with Irregular Services [2], also considering the use of drones. Given that drones can travel off the network and fly directly between any two points (not necessarily the endpoints of the required lines) this becomes a continuous optimization problem where the shape of the lines to be serviced must be taken into account.

We study the Periodic drone Arc Routing Problem with Irregular Services and Maximum Benefits (PdARP–IS–MB) over a finite, discrete time horizon divided into periods. Each required edge has a service frequency per period and must be traversed by a single drone at least once over the entire horizon. However, a required edge can be traversed in each period a number of times greater than the given frequency if the maximum distance the drone can travel allows it (but not more than once a day). We consider that these additional traversals of the required edges are beneficial, so each required edge has a benefit associated. The objective is to plan a daily route satisfying all service constraints while maximizing the total benefit from the additional traversal of the required edges. We present a mixed-integer programming formulation and develop both an exact branch-and-cut algorithm and a Kernel Search heuristic [1] to solve the problem. Computational experiments demonstrate the effectiveness of our approach and highlight the balance between solution quality and computation time.

Keywords: arc routing; branch and cut; drones; irregular services; kernel search.

References

[1] E. Angelelli, R. Mansini, and M.G. Speranza. Kernel Search: a new heuristic framework for portfolio selection. *Computational Optimization and Applications*, 51:345–361, 2012. doi:10.1007/s10589-010-9326-6.

[2] E. Benavent, Á. Corberán, D. Laganà, and F. Vocaturo. The periodic rural postman problem with irregular services on mixed graphs. *European Journal of Operational Research*, 276(3):826–839, 2019. doi:10.1016/j.ejor.2019.01.056.

5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Artificial intelligence and simheuristics for the stochastic prize-collecting traveling salesman problem

C. Tobar-Fernández[1] (✉), Ana D. López-Sánchez[2] and Jesús Sánchez-Oro[1]

[1]Department of Informatics and Statistics, University Rey Juan Carlos, Spain; [2]Department of Economics, University Pablo de Olavide, Spain

c.tobar.2024@alumnos.urjc.es

Cristina Tobar-Fernández is a second-year PhD student in the PhD Programme in Information and Communication Technologies at Rey Juan Carlos University, within the line of research in Artificial Intelligence. Her research activity focuses on the design and implementation of algorithms to minimize uncertainty in operations research problems. She is pursuing an Industrial PhD in collaboration with the company OGA.ai where she works as Optimization Scientist. She holds a Double Bachelor's Degree in Mathematics and Statistics from the University of Seville and a Master's Degree in Big Data from Carlos III University of Madrid.

In recent years, the integration of stochastic elements into classical optimization problems has gained increasing relevance. Among these, the Prize-Collecting Traveling Salesman Problem (PCTSP) [1] stands out as a flexible variant of the classical Traveling Salesman Problem (TSP), where not all nodes must be visited. While exact and heuristic methods have been widely studied for its deterministic version, real-world applications often involve uncertainty in key parameters, making purely deterministic approaches insufficient. To address this, hybrid methodologies combining simulation and optimization—commonly referred to as simheuristics—have emerged as a promising line of research in Operations Research [2]. Simultaneously, advancing Artificial Intelligence (AI) techniques offers new opportunities to improve the solutions under uncertainty.

This work proposes a novel AI driven simheuristic framework to solve the PCTSP under demand uncertainty, integrating predictive Machine Learning models and data-driven scenario generation into a Greedy Randomized Adaptive Search Procedure (GRASP) based metaheuristic. Instead of relying only on expected values or random sampling, our method incorporates supervised learning techniques to generate approximations of demand, and unsupervised learning via clustering to produce scenarios. A two-phase simulation evaluates solutions: a fast initial screening followed by intensive analysis of top candidates. Validation on a real enriched dataset shows the hybrid model significantly outperforms classical methods.

Keywords: simheuristics; stochastic optimization; uncertainty; vehicle routing.

References

[1] E. Balas. The prize collecting traveling salesman problem. *Networks*, 19(6):621–636, 1989. doi:10.1002/net.3230190602.

[2] A.A. Juan, J. Faulin, S.E. Grasman, M. Rabe, and G. Figueira. A review of simheuristics: Extending metaheuristics to deal with stochastic combinatorial optimization problems. *Operations Research Perspectives*, 2:62–72, 2015. doi:10.1016/j.orp.2015.03.001.

SYSOR 2025
5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Confidence intervals for Cronbach's α coefficient with interval-valued data based on log-transformed estimates

José García-García[1] (✉) and M. Asunción Lubiano[1]

[1]Department of Statistics, and Operational Research, and Didactics of Mathematics, University of Oviedo, Spain

garciagarjose@uniovi.es

José García-García is a PhD candidate in Mathematics and Statistics at the University of Oviedo (Spain) and in Digital and Analytic Sciences at the University of Salzburg (Austria) under joint supervision. He did a MSc in Data Analysis for Business Intelligence and a BSc in Mathematics (end of degree award), both at the University of Oviedo. Founded by a Severo Ochoa fellowship from the Principality of Asturias, his research is focused on the development of statistical techniques for meta-analysis with imprecise random elements.

Internal consistency reliability assessment is a key step in designing and analyzing data from questionnaires aimed to evaluate intrinsically imprecise constructs since it reflects the extent to which the items and the rating scales allow to collect statistically coherent responses. Cronbach's α coefficient is probably the most common index for this purpose when traditional numerically-encoded scale-based data is considered. The emergence of more complex but also richer and more informative measurement instruments such as the so-called interval-valued rating scales –where responses are given as ranges rather than as single-point values– motivates the extension of the definition of this coefficient to different frameworks. In the particular case of interval-valued data, such an extension could be done, as suggested in [1], by following a distance-based approach which models the mechanism that generates such data through the concept of random interval.

Synthesizing evidence on internal consistency across multiple studies from a meta-analytical perspective requires the use of confidence intervals for the reported reliability estimates. In [2], asymptotic and bootstrap confidence intervals for the extended Cronbach α coefficient have been obtained. Since applying transformations can sometimes improve convergence rates, the goal of this work is to study the performance of asymptotic and bootstrap confidence intervals for such an index based on log-transformed estimates as well as comparing the obtained results with those from the non-transformed scenario. To this end, the necessary statistical background on interval-valued data is to be recalled, the extended Cronbach α coefficient and its estimators are to be presented, and the log-transformed confidence intervals are to be analyzed, proving their consistency via limit distribution.

Keywords: bootstrap; confidence interval; extended Cronbach's α coefficient; interval-valued rating scale; random interval.

References

[1] J. García-García, M.Á. Gil, and M.A. Lubiano. On some properties of Cronbach's α coefficient for interval-valued data in questionnaires. *Advances in Data Analysis and Classification*, 19(3):831–854, 2025. doi:10.1007/s11634-024-00601-w.

[2] J. García-García and M.A. Lubiano. Confidence intervals and one-sample-based hypothesis testing for Cronbach's α coefficient for interval-valued data. *Statistics and Computing*, 35(5):148, 2025. doi:10.1007/s11222-025-10676-w.

Preference estimation in inverse multiobjective optimization

Emilio Carrizosa[1], Nuria Gómez-Vargas[1] (✉) and Veronica Piccialli[2]

[1]Department of Statistics and Operations Research, University of Seville, Spain; [2]Department of Computer, Control and Management Engineering, Sapienza University of Rome, Italy

ngvargas@us.es

Nuria Gómez-Vargas is a fourth-year PhD student in the PhD in Mathematics in the department of Statistics and Operations Research at University of Seville (US). Her research activity falls into Contextual Decision-Making, where Supervised Machine Learning is used to address the uncertainty in Operational Research problems. Before enrolling and during her PhD, she has worked in research projects in collaboration with business, and did a MSc in Mathematics and a BSc in Mathematics at US.

Inverse Optimization (IO) [3] aims to infer the hidden structure of a decision-making process from observed decisions. In many real-world scenarios, decisions are driven by unobserved preferences across conflicting objectives. Existing works inferring the preference structures either rely on geometric proximity of decisions [5], which may misrepresent similarity when feasible sets differ, or assume fixed preferences [1]. Moreover, ensuring that inferred preferences are interpretable [4] and aligned with domain knowledge remains an open challenge.

We propose in [2] a novel framework that integrates clustering into inverse multiobjective optimization to group decision-makers by latent preferences. By minimizing optimality gaps —differences between observed and optimal values under inferred preferences— we obtain Mixed-Integer Quadratic Programs (MIQPs). We design an alternating optimization heuristic to warm-start global solvers and include constraints that promote sparse, interpretable preferences. Our method is validated on a real-world diet recommendation problem, successfully uncovering robust and meaningful decision patterns under noisy data.

Keywords: clustering; explainable decision-making; inverse optimization; multiple objective programming; preference estimation.

References

[1] R. Blanquero, E. Carrizosa, and N. Gómez-Vargas. On contextual inverse multiobjective problems, 2025. URL: https://www.researchgate.net/publication/387788641_On_Contextual_Inverse_Multiobjective_Problems.

[2] E. Carrizosa, N. Gómez-Vargas, and V. Piccialli. Preference estimation in inverse multiobjective optimization, 2025. URL: https://www.researchgate.net/publication/391692902_Preference_Estimation_in_Inverse_Multiobjective_Optimization.

[3] T.C.Y. Chan, R. Mahmood, and I.Y. Zhu. Inverse optimization: Theory and applications. *Operations Research*, 73(2):1046–1074, 2023. doi:10.1287/opre.2022.0382.

[4] M. Goerigk and M. Hartisch. A framework for inherently interpretable optimization models. *European Journal of Operational Research*, 310(3):1312–1324, 2023. doi:10.1016/j.ejor.2023.04.013.

[5] Z. Shahmoradi and T. Lee. Optimality-based clustering: An inverse optimization approach. *Operations Research Letters*, 50(2):205–212, 2022. doi:10.1016/j.orl.2021.12.012.

Feature selection for shape-constrained smooth additive models

Marina Cuesta[1] (✉), María Durbán[1] and Vanesa Guerrero[1]

[1]Department of Statistics, Carlos III University of Madrid, Spain

marincue@est-econ.uc3m.es

Marina Cuesta holds a PhD in Information and Communication Technologies from Rey Juan Carlos University, where her research focused on explainable machine learning with data visualization. Since November 2024, she has been a postdoctoral researcher at Carlos III University, working on smooth regression models. She earned an MSc in Statistical and Computational Information Processing from Complutense and Politechnic Universities of Madrid, and a BSc in Mathematics and Statistics from Complutense University.

Shape-constrained smooth additive regression models stand out as a flexible and interpretable tool for modeling data. Their estimation can include shape requirements, such as monotonicity or non-negativity, to reflect domain-specific knowledge [2]. In an increasingly data-driven world, modeling a response variable often involves many potential covariates. If redundant or noisy variables are present, their inclusion can compromise interpretability and increase the risk of overfitting. Thus, even inherently interpretable models can become difficult to understand in high-dimensional problems. Feature selection has become essential in modern regression analysis to overcome these challenges [1]. By identifying the most informative covariates, sparsity is used as a proxy with the aim of not compromising predictive performance.

In this work, we address the feature selection problem in the context of shape-constrained smooth additive regression models using a mathematical optimization perspective, generalizing the work in [3]. The problem is approached by imposing a constraint within the estimation of shape-constrained smooth additive regression models that limits the number of covariates with nonzero coefficients to a fixed value. As a result, a conic optimization problem with binary variables is obtained. The proposed approach provides a structured way to perform feature selection while respecting the shape constraints imposed on the covariates. The method has been evaluated on several simulated datasets with different numbers of covariates, showing promising results.

Keywords: B-splines; feature selection; mathematical optimization; shape-constrained regression.

References

[1] J. Li, K. Cheng, S. Wang, F. Morstatter, R.P. Trevino, J. Tang, and H. Liu. Feature selection: A data perspective. *ACM Computing Surveys*, 50(6):1–45, 2017. doi:10.1145/3136625.

[2] M. Navarro-García, V. Guerrero, and M. Durbán. A mathematical optimization approach to shape-constrained generalized additive models. *Expert Systems with Applications*, 255:124654, 2024. doi:10.1016/j.eswa.2024.124654.

[3] M. Navarro-García, V. Guerrero, M. Durbán, and A. del Cerro. Feature and functional form selection in additive models via mixed-integer optimization. *Computers & Operations Research*, 176:106945, 2025. doi:10.1016/j.cor.2024.106945.

SYSØR 2025
5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Applied Bayesian nonparametric modeling under likelihood ratio order constraints

Adam Olivares[1] (✉), Víctor Peña[1], Michael Jauch[2] and Andrés F. Barrientos[2]

[1]Department of Statistics and Operations Research, Universitat Politècnica de Catalunya, Spain;
[2]Department of Statistics, Florida State University, United States

adam.olivares@upc.edu

Adam Olivares is a first-year PhD student in the PhD program in Statistics and Operations Reserach at Universitat Politècnica de Catalunya. His research activity is focused on Bayesian nonparametric inference, Interpretable Machine Learning and their intersection. Adam holds a MSc in Data Science from Universitat Pompeu Fabra and a BSc in Economics from Universitat Autònoma de Barcelona.

In many applications, domain knowledge often suggests that two distributions should satisfy a stochastic order. Such orderings arise naturally in fields such as economics and finance, as well as biology, medicine, and public health [3]. Jauch et al. [1] introduced a mixture representation for univariate distribution functions F and G so that $F \leq G$ with respect to the likelihood ratio order. Distributions F and G are ordered in the likelihood ratio order when the ratio of their probability density functions, $f(x)/g(x)$, is a non-increasing function in x. To model this, Jauch et al. [1] showed that the ratio of two probability density functions is monotone if and only if one can be expressed as a mixture of one-sided truncations of the other. They proposed a Bayesian nonparametric model for density estimation using Dirichlet process mixtures [2] to enforce this constraint and applied it to medical data.

We consider two extensions of the representation introduced in [1]: (1) An extension to multivariate F and G that are ordered according to the weak likelihood ratio order, and (2) An extension to K univariate stochastically ordered distributions $F_1 \leq \ldots \leq F_K$ with respect to the likelihood ratio order. These mixture representations can be applied to hidden Markov models and logistic and ordinal regression models.

Keywords: Bayesian methods; mixture representations; monotone likelihood ratio order; regression models; stochastic order.

References

[1] M. Jauch, A.F. Barrientos, V. Peña, and D.S. Matteson. Mixture representations and Bayesian nonparametric inference for likelihood ratio ordered distributions. *Bayesian Analysis*, pages 1–24, 2025. doi:10.1214/25-BA1519.

[2] A.Y. Lo. On a class of Bayesian nonparametric estimates: I. density estimates. *The Annals of Statistics*, 12(1):351–357, 1984. doi:10.1214/aos/1176346412.

[3] M. Shaked and J.G. Shanthikumar. *Stochastic Orders*. Springer Series in Statistics. Springer, New York, NY, 2007. doi:10.1007/978-0-387-34675-5.

An iterated greedy algorithm for the rank pricing problem

Herminia I. Calvete[1,2], Carmen Galé[1,2], Aitor Hernández[1,2] (✉) and José A. Iranzo[1,2]

[1]Instituto Universitario de Investigación de Matemáticas y Aplicaciones, University of Zaragoza, Spain;
[2]Departamento de Métodos Estadísticos, University of Zaragoza, Spain

aitor.hernandez@unizar.es

Aitor Hernández is a third-year PhD student in the PhD programme in Mathematics and Statistics at University of Zaragoza. His research activity is focused on bilevel and multi-objective optimization. Before enrolling his PhD, he did a MSc in Mathematical modelling, Statistics and Computing and a BSc in Mathematics at University of Zaragoza.

The rank pricing problem (RPP) consists in determining the optimal prices for a set of products offered by a company to a set of customers interested in some of them, taking into account both the customers' budgets and personal preferences [1]. The RPP admits a bilevel formulation in which the company acts as the leader setting the prices and the customers act as the followers who react to those prices by deciding whether or not to purchase a product. Calvete et al. [1] solved the RPP exactly by formulating and analyzing several mathematical models, while Calvete et al. [2] proposed an evolutionary algorithm to address larger instances.

This research presents several important contributions. It introduces the first iterated greedy algorithm specifically designed to solve the RPP. Two dedicated procedures are defined to partially destroy incumbent solutions throughout the execution of the algorithm. Several local search methods are also proposed to further enhance the performance of the algorithm. Finally, extensive computational experiments are conducted to show that the algorithm provides high-quality solutions within very short computational times. These experiments confirm the effectiveness of the proposed algorithm by comparing it against exact and heuristic methods from the literature.

Furthermore, both Domínguez et al. [4] and Calvete et al. [3] have studied the variant of the problem where customers' preferences may include ties. The algorithm proposed in this study is easily extendable to also handle this variant.

Keywords: bilevel optimization; iterated greedy; metaheuristics; rank pricing problem.

References

[1] H.I. Calvete, C. Domínguez, C. Galé, M. Labbé, and A. Marín. The rank pricing problem: Models and branch-and-cut algorithms. *Computers and Operations Research*, 105:12–31, 2019. doi:10.1016/j.cor.2018.12.011.

[2] H.I. Calvete, C. Galé, A. Hernández, and J.A. Iranzo. An evolutionary algorithm for the rank pricing problem. In *Metaheuristics, MIC 2024*, pages 360–366, 2024. doi:10.1007/978-3-031-62922-8_28.

[3] H.I. Calvete, C. Galé, A. Hernández, and J.A. Iranzo. A novel approach to pessimistic bilevel problems. an application to the rank pricing problem with ties. *Optimization*, 2024. doi:10.1080/02331934.2024.2388204.

[4] C. Domínguez, M. Labbé, and A. Marín. The rank pricing problem with ties. *European Journal of Operational Research*, 294(2):492–506, 2021. doi:10.1016/j.ejor.2021.02.017.

SYSØR 2025
5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

A two-level Plackett-Luce model for preference modelling in route choice

Miguel Santos-Pascual[1] (✉) and David Ríos-Insua[1]

[1]Department of Statistics and Operations Research, Instituto de Ciencias Matemáticas, Spain

migsanpas@gmail.com

Miguel Santos-Pascual is a master's student and future PhD student in statistics at ICMAT. His first steps on the research activity are focused on the implementation of Bayesian inference techniques so as to forecasting customers' behaviour. Before starting his PhD, he completed an MSc in statistics at Universidad Carlos III de Madrid while working at ICMAT, and earned a BSc in mathematics and physics from the University of Valladolid.

This work combines the Plackett–Luce (PL) model, a generalization of the Bradley–Terry model used to represent how individuals rank a set of alternatives [3], with a Bayesian inference framework based on Markov Chain Monte Carlo (MCMC) methods. The implementation relies on the No-U-Turn Sampler (NUTS), a gradient-based variant of MCMC that adapts to the geometry of the parameter space by automatically determining the number of steps to take during sampling [2]. This combination enables efficient inference in high-dimensional parameter spaces, as demonstrated in [4]. The integration of NUTS with probabilistic ranking models such as PL allows for a more flexible and accurate representation of decision-making processes, extending previous work on efficient Bayesian approaches for modeling ranked data [1].

The work focuses on the development of the algorithms of a mobility platform, contributing to three main components: the creation of a synthetic data generation process to simulate user inputs in the absence of real data; the design and implementation of a Bayesian inference framework using MCMC methods for learning preference structures based on a variant of the PL model; and the application of the inferred model to predict user choices and extract relevant insights. These contributions support the personalization and adaptability of the system, forming the foundation for intelligent decision-making in mobility contexts.

Keywords: Bradley-Terry; Markov chain Monte Carlo; No-U-Turn sampler; Plackett-Luce; route choice.

References

[1] F. Caron and A. Doucet. Efficient bayesian inference for generalized bradley–terry models. *Journal of Computational and Graphical Statistics*, 21(1):174–196, 2012. doi:10.1080/10618600.2012.638220.

[2] M.D. Hoffman and A. Gelman. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014. URL: https://www.jmlr.org/papers/volume15/hoffman14a/hoffman14a.pdf.

[3] R.L. Plackett. The analysis of permutations. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 24(2):193–202, 1975. URL: https://academic.oup.com/jrsssc/article/24/2/193/6953554?login=false.

[4] M. Plummer. Simulation-based Bayesian analysis. *Annual Review of Statistics and Its Application*, 10(1):401–425, 2023. URL: https://www.annualreviews.org/content/journals/10.1146/annurev-statistics-122121-040905.

SYSØR 2025
5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Dirichlet values for balanced games

Alejandro Bernárdez Ferradás[1] (✉), Miguel Ángel Mirás Calvo[2] and Estela Sánchez-Rodríguez[1]

[1]Department of Statistics and Operations Research, University of Vigo, Spain; [2]Department of Mathematics, University of Vigo, Spain

alejandro.bernardez@uvigo.es

Alejandro Bernárdez Ferradás is a first-year PhD student in Statistics and Operations Research at the University of Vigo. His research focuses on game theory, specifically on analyzing the influence of weights and clones in coalitional games, with concrete applications to economic problems. Before starting his PhD, he earned a Bachelor's degree in Economics and a Master's degree in Statistics and Operations Research. He also worked as a Data Analyst at Hijos de Rivera, S.A.U.

This contribution falls within the field of game theory, specifically within cooperative games. In this work, we introduce a new family of values, the Dirichlet values, for balanced positive coalitional games. These values are defined by means of the Dirichlet distribution, a multivariate distribution over the $(n-1)$-regular simplex, parameterized by an n-dimensional vector of positive reals, which can be interpreted as a probability distribution over all the possible divisions of one unit among n players [3]. The Dirichlet value for a positive balanced game is given by the expectation of the conditional distribution of a Dirichlet distribution over the (scaled) core of the game. A notable member of this family is the core-center solution [2], which corresponds to the flat Dirichlet distribution (the uniform distribution).

We analyze whether or not the Dirichlet values satisfy some basic properties. In particular, we study how the Dirichlet values treat symmetric players (or clones). We identify a subclass of coalitional games for which the payoff recommended by the core-center solution for each group of symmetric players coincides with the allocation selected by the Dirichlet value, with parameters given by the sizes of each group of clones, when applied to the reduced game without clones. In particular, this subclass includes the coalitional games associated with airport problems [1]. Remarkably, for these types of games, the calculation of the core-center solution can be performed on a lower-dimensional space by computing the suitable Dirichlet value for the reduced game.

Keywords: clones; coalitional games; core-center; Dirichlet distribution; weighted values.

References

[1] A. Bernárdez Ferradás, M.Á Mirás Calvo, C. Quinteiro Sandomingo, and E. Sánchez-Rodríguez. Airport problems with cloned agents, 2025. Preprint, submitted for publication.

[2] J. González Díaz and E. Sánchez-Rodríguez. A natural selection from the core of a TU game: The core-center. *International Journal of Game Theory*, 36:27–46, 2007. doi:10.1007/s00182-007-0074-5.

[3] K.W. Ng, G.-L. Tian, and M.-L. Tang. *Dirichlet and Related distributions. Theory, Method and Applications*. Wiley, 2011. doi:10.1002/9781119995784.

SYSØR 2025
5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Management of shared photovoltaic systems on apartment buildings: a two-stage stochastic optimization model

Paula Terán-Viadero[1] (✉), Antonio Alonso-Ayuso[2], F. Javier Martín-Campo[3] and Elisenda Molina[3]

[1]Dept. of Financial and Actuarial Economics & Statistics, Complutense University of Madrid, Spain; [2]DSLAB-CETINIA, University Rey Juan Carlos, Spain; [3]Dept. of Statistics and Operational Research, Interdisciplinary Mathematics Institute, Complutense University of Madrid, Spain

pauteran@ucm.es

Paula Terán-Viadero is a PhD in Operations Research, particularly dealing with Cutting and Packing problems. She received her PhD in June 2024 by Complutense University of Madrid (UCM). She has worked in the private sector, developing integer linear mathematical optimisation models to solve problems arising from real-world applications. Since January 2025 she is an assistant professor at Complutense University of Madrid.

This work addresses the problem of determining how to distribute the energy generated by a photovoltaic system in residential buildings. Distribution coefficients specify how the solar energy is allocated to each consumption point (either a neighbour or a community facility, as elevators, lighting, pool ...). According to Spanish regulations, all generated energy must be fed into the grid, and the distribution company calculates how much solar energy each consumption point receives each hour, based on these distribution coefficients. The coefficients can be updated every four months and cannot be modified during this period. The energy assigned to each consumer reduces their electricity bill, and any surplus may be compensated, usually at a lower rate. Our proposal is to define an optimization model that allows to obtain the distribution coefficients that maximize the economic performance of the photovoltaic installation.

In previous works ([1, 2]), the problem is studied under a deterministic assumption: production, demand and prices are considered known in advance. Our proposal incorporates uncertainty in the model by developing a two-stage stochastic optimization approach, where uncertainty is represented through a scenario tree, allowing for the addition of risk aversion measures. Additionally, this work also highlights the complexity of dealing with common areas, which typically have higher and more seasonal consumption patterns and whose cost-sharing rules cannot be altered.

Keywords: energy allocation coefficients; photovoltaic energy; stochastic optimisation.

References

[1] F. Lazzari, G. Mor, J. Cipriano, F. Solsona, D. Chemisana, and D. Guericke. Optimizing planning and operation of renewable energy communities with genetic algorithms. *Applied Energy*, 338:120906, 2023. doi:10.1016/j.apenergy.2023.120906.

[2] Á. Manso-Burgos, D. Ribó-Pérez, T. Gómez-Navarro, and M. Alcázar-Ortega. Local energy communities in Spain: Economic implications of the new tariff and variable coefficients. *Sustainability*, 13(10):10555, 2021. doi:10.3390/su1310555.

Prediction regions for random forests in metric spaces

Diego Serrano[1] (✉) and Eduardo García-Portugués[1]

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

dieserra@est-econ.uc3m.es

Diego Serrano is a first-year PhD student in the PhD in Statistics for Data Science at Universidad Carlos III de Madrid (UC3M). His research involves statistical methods for analyzing complex data valued in metric spaces, particularly the development and application of random forests for metric data. He held a research position in the Department of Mathematics at Universidad Autónoma de Madrid (UAM), and he completed a MSc in Statistics for Data Science at UC3M and a BSc in Mathematics at UAM.

Random forests have been recently generalized for variables taking values in general metric spaces. The most relevant methods in this sense are Fréchet Random Forests (FRF) [1] and Random Forest Weighted Local Constant Fréchet Regression (RFWLCFR) [3]. Recent works [2, 5] enable uncertainty estimation in metric spaces but inherit the inefficiencies of split-conformal inference derived from splitting the data into two subsamples, as they are not tailored to random forests. In contrast, [4] constructs prediction regions using out-of-bag errors from a single forest, thus leveraging the full dataset for both prediction and uncertainty quantification. However, these prediction regions are only applicable with Euclidean data.

We present a new methodology to quantify the uncertainty in the prediction with FRF/RFWLCFR through the consideration of confidence regions for the metric-valued response. Asymptotic coverage theory is presented in four different scenarios depending on the type of coverage. The proposed prediction regions are compared and illustrated through simulations using RFWLCFR in different metric spaces, such as the Euclidean space, unit sphere, and the space of symmetric positive definite matrices. More precise and computationally efficient estimates are obtained when compared to generic split-conformal approaches. The prediction regions are also applied to forecast the estimated death location of sunspot groups using their birth locations.

Keywords: confidence regions; Fréchet regression; metric space data analysis; random forests.

References

[1] L. Capitaine, J. Bigot, R. Thiébaut, and R. Genuer. Fréchet random forests for metric space valued regression with non Euclidean predictors. *Journal of Machine Learning Research*, 25(355):1–41, 2024. URL: http://jmlr.org/papers/v25/20-1173.html.

[2] G. Lugosi and M. Matabuena. Uncertainty quantification in metric spaces, 2024. arXiv:2405.05110.

[3] R. Qiu, Z. Yu, and R. Zhu. Random forest weighted local Fréchet regression with random objects. *Journal of Machine Learning Research*, 25(107):1–69, 2024. URL: http://jmlr.org/papers/v25/23-0811.html.

[4] H. Zhang, J. Zimmerman, D. Nettleton, and D.J. Nordman. Random forest prediction intervals. *The American Statistician*, 74(4):392–406, 2020. doi:10.1080/00031305.2019.1585288.

[5] H. Zhou and H.-G. Müller. Conformal inference for random objects, 2024. arXiv:2405.00294.

5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

The capacitated dispersion problem with upgradings and downgradings

Lorena Nácher[1] (✉), Mercedes Landete[1], Marina Leal[1] and Juanjo Peiró[2]

[1]Department of Statistics, Mathematics and Computer Science, Center of Operations Research (CIO), Miguel Hernández University of Elche, Spain; [2]Department of Statistics and Operations Research, University of Valencia, Spain

lnacher@umh.es

Lorena Nácher is a second-year PhD student in Statistics, Optimization, and Applied Mathematics at Miguel Hernández University of Elche. Her research focuses on new optimization models for hierarchical classification based on spanning trees, as well as on location problems. Before enrolling her PhD, she completed a Master's degree in Computational Statistics and Data Science for Decision-Making and a Bachelor's degree in Business Statistics, both at the same university, receiving the best academic record award in the Master's.

The Capacitated Dispersion Problem (CDP) addresses the selection of a subset of facilities whose total capacity satisfies a minimum demand, while maximizing the minimum distance between any pair of them. This problem is found in real-world situations where it is important to cover a certain area and ensure that the selected facilities are well separated, such as in emergency planning, environmental monitoring, data center placement, or the allocation of evacuation shelters. The CDP is an extension of the classic p-dispersion problem introduced by [1], adding capacity constraints that make the problem more complex but also more realistic. Recent works, such as [2], proposes improved mathematical models and valid inequalities, while [3] presents heuristic methods useful for solving large instances.

In this work, we extend the classical capacitated dispersion problem by introducing modification strategies for facilities and/or their connections. Specifically, we adapt the Kuby, Edge, and Telescopic formulations to incorporate capacity upgrades at facilities and distance downgrades in selected connections—that is, increasing connection lengths to promote dispersion. We also consider modifications that affect all connections incident to an upgraded facility. Each extension preserves the original max–min dispersion structure while allowing a limited number of changes. We apply all three formulations to each proposed scenario and conduct a comprehensive computational study to compare their performance in terms of solution quality, computational effort, and robustness across different instance types and parameter settings.

Keywords: capacitated dispersion problem; connection downgrading; facility upgrading; mixed integer linear programming.

References

[1] M.J. Kuby. Programming models for facility dispersion: The p-dispersion and maxisum dispersion problems. *Geographical Analysis*, 19(4):315–329, 1987. doi:10.1111/j.1538-4632.1987.tb00133.x.

[2] M. Landete, J. Peiró, and H. Yaman. Formulations and valid inequalities for the capacitated dispersion problem. *Networks*, 81(2):294–315, 2023. doi:10.1002/net.22132.

[3] J. Peiró, I. Jiménez, J. Laguardia, and R. Martí. Heuristics for the capacitated dispersion problem. *International transactions in operational research*, 28(1):119–141, 2021. doi:10.1111/itor.12799.

A minimax optimal filament estimator

Héctor González-Vázquez[1,2] (✉), Beatriz Pateiro-López[1,2] and Alberto Rodríguez-Casal[1,2]

[1]Department of Statistics, Mathematical Analysis and Optimization, University of Santiago de Compostela, Spain; [2]Galician Centre for Mathematical Research and Technology (CITMAga), Spain

hector.gonzalez@usc.es

Héctor González-Vázquez is a first-year PhD student in the Statistics and Operations Research PhD program at the University of Santiago de Compostela. His research activity is focused on the development of set estimation tools in manifold statistics. Before enrolling his PhD, he did a MSc in Statistical Techniques and a BSc in Mathematics, both at the University of Santiago de Compostela.

Many situations involve dealing with high-dimensional data that actually exhibit a lower-dimensional structure. Principal components analysis (PCA) is a classical linear method to perform dimension reduction by projecting the data onto a linear subspace. In contrast to this linearity, the so called manifold hypothesis assumes that data lie around a lower-dimensional submanifold embedded in the Euclidean ambient space, leading to a variety of non-linear dimension reduction techniques known as manifold learning. These methods aim to reduce the dimension by recovering this underlying manifold, which also helps understand the true structure of the data. When the manifold is a one-dimensional curve, the problem is known as filament estimation.

This work presents a new filament estimator. It follows the ideas of the Euclidean distance transform (EDT) estimator proposed by [1] and improves it by making use of a shape condition, the r-convexity, which generalizes convexity. In particular, instead of the more general Devroye-Wise support estimator, we propose to use the r-convex hull of the sample to estimate the support of the sampling distribution, see [3]. Our estimator achieves the optimal minimax rate in Hausdorff distance (up to logarithmic factor) in the noise model studied by [2] when the ambient space is the bidimensional Euclidean plane. We also illustrate this new estimator with a real data application.

Keywords: filament estimation; manifold learning; minimax rate; r-convex hull.

References

[1] C. Genovese, M. Perone-Pacifico, I. Verdinelli, and L. Wasserman. The geometry of nonparametric filament estimation. *Journal of the American Statistical Association*, 107(498):788–799, 2012. doi:10.1080/01621459.2012.682527.

[2] C. Genovese, M. Perone-Pacifico, I. Verdinelli, and L. Wasserman. Minimax manifold estimation. *Journal of Machine Learning Research*, 13(43):1263–1291, 2012. URL: https://jmlr.csail.mit.edu/papers/v13/genovese12a.html.

[3] A. Rodríguez-Casal. Set estimation under convexity type assumptions. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*, 43(6):763–774, 2007. doi:10.1016/j.anihpb.2006.11.001.

SYSØR 2025
5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

A home healthcare routing and scheduling problem with ferry-dependent travel times

Abdalrahman Algendi[1] (✉), Sebastián Urrutia[1] and Lars Magnus Hvattum[1]

[1]Logistics College, Molde University, Norway

abal@himolde.no

Abdalrahman Algendi is a third-year PhD student in Quantitative Logistics at Molde University College. His research activity is focused on developing optimization-based decision-support tools for complex real-world systems, with applications in health care, petroleum, and defense logistics. Before enrolling in his PhD, he earned a MSc in Logistics Analytics from Molde University and a BSc in Economics and Mathematics from the Norwegian University of Life Sciences.

Home healthcare routing and scheduling problems have attracted increasing attention due to the growing demand for in-home medical services and the logistical complexity of delivering them efficiently [1]. These problems often involve rich constraints, such as time windows, skill requirements, and synchronization between multiple caregivers. When ferry-dependent regions are considered, additional challenges arise due to the need to align caregiver schedules with infrequent and fixed ferry timetables. While several exact and heuristic methods have been proposed for related vehicle routing problems, few address the integration of maritime transport constraints into caregiver routing models [2].

This study investigates a home healthcare routing and scheduling problem in a setting where caregivers must utilize ferry services to reach certain patients, which is the case in many areas along the Norwegian coast. The objective is to minimize the sum of route durations across all caregivers, thereby maximizing the time available for other potential tasks. To tackle the problem, we formulate a mixed-integer linear programming model and develop a matheuristic based on weighted proximity search. The method initially constructs a feasible solution through a sequential graph expansion strategy and subsequently applies iterative refinements to improve solution quality. Computational tests on 20 realistic instances, derived from two Norwegian healthcare centers, show that the proposed weighted proximity search outperforms both direct model solving using a commercial solver and the classical proximity search in terms of feasibility, optimality, and solution quality. In particular, the method is able to solve instances involving up to 80 patients (135 visits) within a 3600-second time limit.

Keywords: mixed-integer programming; skill matching; synchronization; time-dependent travel time.

References

[1] A. Algendi, S. Urrutia, L.M. Hvattum, and B.I. Helgheim. Home healthcare staffing, routing, and scheduling problem with multiple shifts and emergency considerations. *Networks*, 85(2):223–242, 2025. doi:10.1002/net.22260.

[2] M. Masmoudi, J. Euchi, and P. Siarry. Home healthcare routing and scheduling: operations research approaches and contemporary challenges. *Annals of Operations Research*, 343:701–751, 2024. doi:10.1007/s10479-024-06244-6.

Designing incentives for colorectal cancer screening programs using adversarial risk analysis

Daniel Corrales[1, 2] (✉) and David Ríos-Insua[1]

[1]Mathematical Sciences Institute (ICMAT-CSIC), Spain; [2]Department of Mathematics, Autonomous University of Madrid, Spain

daniel.corrales@icmat.com

Daniel Corrales is a first-year PhD student at the Institute of Mathematical Sciences (ICMAT-CSIC). His research is focused on methodologies for securitising AI models, as well as on the development of statistical models for medical applications. Before joining ICMAT, he did a MSc in Machine Learning for Health at Universidad Carlos III de Madrid and a BSc in Mathematical Engineering at Universidad Complutense de Madrid.

Out of all invited susceptible EU citizens, only 14% participate in colorectal cancer (CRC) screening programs. This supposes a huge health and economic burden for a disease which makes up for 12% of the deaths due to cancer [3]. Importantly, early-stage cancer detection can result in more manageable, less expensive and more likely to be successful treatments. Thus, the importance of predictive and decision models in this field is crucial for characterising the influence of relevant risk factors on the development of the disease and the allocation of screening tests performed to the population. Furthermore, there is a pertinent need among policymakers regarding the design of incentive programs to improve the coverage, acceptability and adherence of high-risk citizens to screening.

This work addresses the challenge of promoting CRC screening from a policymaker's perspective, recognising that participating citizens may pursue objectives that diverge from public health goals. To manage this misalignment, we introduce a decision-making approach based on adversarial risk analysis [4], which allows for the design of optimal incentive schemes under uncertainty modelling on the participants' goals. Our methodology builds on established models of CRC risk and screening strategy optimisation [2, 1], and we demonstrate its application through illustrative use cases involving both individual and group-level optimal financial incentives.

Keywords: adversarial risk analysis; colorectal cancer; decision-making; incentive-scheme; screening.

References

[1] D. Corrales, D. Ríos-Insua, and M.J. González. A decision analysis model for colorectal cancer screening, 2025. arXiv:2502.21210.

[2] D. Corrales, A. Santos-Lozano, S. López-Ortiz, A. Lucia, and D. Ríos-Insua. Colorectal cancer risk mapping through bayesian networks. *Computer Methods and Programs in Biomedicine*, 257:108407, 2024. doi:10.1016/j.cmpb.2024.108407.

[3] E. Morgan, M. Arnold, A. Gini, V. Lorenzoni, C.J. Cabasag, M. Laversanne, J. Vignat, J. Ferlay, N. Murphy, and F. Bray. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from globocan. *Gut*, 72(2):338–344, 2023. doi:10.1136/gutjnl-2022-327736.

[4] D. Ríos-Insua, J. Ríos, and D. Banks. Adversarial risk analysis. *Journal of the American Statistical Association*, 104(486):841–854, 2009. doi:10.1198/jasa.2009.0155.

5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Fairness and equity in the physician scheduling problem

Begoña Álvarez[1,2] (✉), Marta Cildoz[3], Fermin Mallor[3] and Pedro M. Mateo[2]

[1]Department of Statistical Methods, University of Zaragoza, Spain; [2] Polytechnic University School of La Almunia, University of Zaragoza, Spain; [3] Department of Statistics, Computer Science and Mathematics, Public University of Navarra, Spain

balvarez@unizar.es

Begoña Álvarez is a PhD student in the PhD in Operation Research at the University of Zaragoza. Her research activity is focused on the development of matheuristics and metaheuristic algorithms for optimal scheduling in health services. Previously, she graduated in Mathematics at University of Zaragoza and obtained a Master in Statistics and Operations Research at Polytechnic University of Catalonia. Also she worked as a Data Analyst.

The physician scheduling problem has been studied widely [2]. However, few approaches consider the well-being of medical staff in them. In many cases, healthcare professionals work under unpleasant conditions due to long working hours and uneven shift distribution. As a result, problems arise that affect physicians' health and, consequently their efficiency and performance, two key aspects in the current context [3]. Therefore, achieving a fair and equitable workload distribution is essential. Incorporating fairness and equity requirements; temporal balance for heterogeneous staff; and ergonomics constraints, makes scheduling a challenging problem [1].

We introduce the concept of fairness and equity in physician scheduling. Then, we propose a methodology that incorporates these principles to try to ensure that all physicians are assigned shifts fairly, based on their individual circumstances, while also achieving an overall equitable distribution of the workload. We model this situation by defining the ideal number of shifts according to fairness and equity criteria. Since these values may conflict, we formulate a multi-objective linear programming problem to determine a compromise solution regarding the number and types of shifts to be assigned to each physician. To solve this model, we propose a variation of a lexicographic optimization procedure. The solution obtained from the previous model can subsequently be used as input parameters for designing more efficient algorithms for solving the physician scheduling problem.

Keywords: fairness and equity, matheuristics; optimization; physician scheduling problem.

References

[1] M. Cildoz, F. Mallor, and P. M. Mateo. A grasp-based algorithm for solving the emergency room physician scheduling problem. *Applied Soft Computing*, 103, 2021. doi:10.1016/j.asoc.2021.107151.

[2] M. Erhard, J. Schoenfelder, A. Fügener, and J.O. Brunner. State of the art in physician scheduling. *European Journal of Operational Research*, 265:1–18, 2018. doi:10.1016/j.ejor.2017.06.037.

[3] H.I. Koruca, M. Serdar, and E. Gulmez. Development of a new personalized staff-scheduling method with a work-life balance perspective: case of a hospital. *Annals of Operations Research*, 328:793–820, 2023. doi:10.1007/s10479-023-05244-2.

Inference-time robustness through flow-based input purification

Pablo García Arce[1, 2] (✉), Roi Naveiro[3] and David Ríos-Insua[1]

[1]ICMAT, Spain; [2]Universidad Autónoma de Madrid, Spain; [3]CUNEF Universidad, Spain

pablo.garcia@icmat.es

Pablo García Arce is a second-year PhD student in the PhD in Computer Science at Institute of Mathematical Sciences. His research activity is focused on the robustification of machine learning models against adversarial attacks. Before enrolling in his PhD, he worked at Predictia Intelligent Data Solutions as Data Scientist and did a MSc in Statistics and Computation at Universidad Complutense de Madrid and a BSc in Mathematics and Physics at Universidad de Cantabria.

While machine learning (ML) has brought significant benefits, it has also been subject to misuse, particularly through attempts by adversaries to manipulate models for their own gain. This has given rise to the field of adversarial machine learning (AML) [4, 5, 2]. Traditional ML assumes independently and identically distributed data, but in many real-world settings, adversaries may alter inputs, breaking this assumption. AML seeks to build models that are robust to such manipulations, focusing on understanding attacks, designing defenses, and developing frameworks to guide optimal protection strategies.

This paper introduces a pipeline to protect predictive models during deployment by purifying potentially manipulated inputs using normalizing flows [3]. Building on the framework in [1], the proposed method focuses on operational-time defenses against adversarial manipulation, without requiring changes to the original training process. The approach leverages normalizing flows to model the distribution of clean data and detect and correct adversarial perturbations at inference time. Additionally, the pipeline includes mechanisms to monitor for shifts in attack patterns, prompting adaptive purification strategies or triggering alerts for retraining if needed. Several examples illustrate the effectiveness of the method in preserving model performance under adversarial conditions.

Keywords: adversarial analysis; density estimation; machine learning.

References

[1] V. Gallego, R. Naveiro, A. Redondo, D. Ríos-Insua, and F. Ruggeri. Protecting classifiers from attacks: A Bayesian approach. 2024. arXiv:2004.08705.

[2] L. Huang, A. Joseph, B. Nelson, B. Rubinstein, and J. D. Tygar. Adversarial machine learning. In *Proceedings of the 4th ACM Workshop on Artificial Intelligence and Security*, pages 43–58, 2011. doi:10.1145/2046684.2046692.

[3] G. Papamakarios, E. Nalisnick, D. Jimenez Rezende, S. Mohamed, and B. Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(1):57, 2021. doi:10.5555/3546258.3546315.

[4] D. Ríos-Insua, R. Naveiro, R. Gallego, and J. Poulos. Adversarial machine learning: Bayesian perspectives. *Journal of the American Statistical Association*, 118(543):2195–2206, 2023. doi:10.1080/01621459.2023.2183129.

[5] Y. Vorobeychik and M. Kantarcioglu. *Adversarial Machine Learning*. Springer, 2018. doi:10.1007/978-3-031-01580-9.

5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Multi-mode resource-constrained project scheduling problem with time-dependent resource costs and capacities: a bi-objective approach

Sofía Rodríguez-Ballesteros[1] (✉), Javier Alcaraz[1] and Laura Anton-Sanchez[1]

[1]Center of Operations Research, Miguel Hernández University of Elche, Spain

sofia.rodriguezb@umh.es

Sofía Rodríguez-Ballesteros is a fourth-year PhD student in Statistics, Optimization and Applied Mathematics at Miguel Hernández University of Elche. Her research activity focuses on new extensions for project scheduling problems, and the development of solving methods adapted to specific contexts. Before enrolling her PhD, she worked on her Master's thesis at the Technological Institute for Industrial Mathematics, and did a MSc in Statistics and a BSc in Mathematics, both at University of Santiago de Compostela.

The resource-constrained project scheduling problem (RCPSP) is a classical NP-hard problem involving scheduling activities under precedence and resource limitations. A prominent variant, the multi-mode RCPSP (MRCPSP), allows executing activities in multiple modes, each characterized by distinct durations and resource demands. Additionally, introducing multiple objectives yields a set of trade-off solutions, known as the Pareto front (PF). Recently, [1] proposed a bi-objective RCPSP (RCPSP_TDRC), explicitly incorporating time-dependent resource costs to enhance realism.

Despite these promising advancements, integrating multiple execution modes and dynamic resource capacities into the bi-objective RCPSP_TDRC remains unexplored. This work introduces the bi-objective MRCPSP with time-dependent resource costs and capacities (MRCPSP_TDRCC), combining mode flexibility, variable resource costs, and fluctuating resource availability. We formulate the problem and propose an non-dominated sorting genetic algorithm II (NSGA-II)-based metaheuristic, inspired by [2], featuring a triple-list encoding, custom genetic operators, and penalization for infeasibility handling. Computational experiments on modified PSPLIB and MMLIB datasets validate the method's effectiveness, providing robust PF approximations and establishing a solid basis for further realistic scheduling research.

Keywords: metaheuristic; multi-mode resource-constrained project scheduling; multi-objective optimization; Pareto front; performance indicator.

References

[1] J. Alcaraz, L. Anton-Sanchez, and F. Saldanha-da-Gama. Bi-objective resource-constrained project scheduling problem with time-dependent resource costs. *Journal of Manufacturing Systems*, 63:506–523, 2022. doi:10.1016/j.jmsy.2022.05.002.

[2] J. Alcaraz, C. Maroto, and R. Ruiz. Solving the multi-mode resource-constrained project scheduling problem with genetic algorithms. *The Journal of the Operational Research Society*, 54(6):614–626, 2003. doi:http://www.jstor.org/stable/4101753.

Parameter equality testing in the many-populations regime

Marcos Romero-Madroñal[1,2] (✉), M. Remedios Sillero-Denamiel[1,2] and M. Dolores Jiménez-Gamero[1,2]

[1]Department of Statistics and Operations Reseach, Universidad de Sevilla, Spain; [2]Instituto de Matemáticas de la Universidad de Sevilla, Spain

mrmadronal@us.es

Marcos Romero-Madroñal is a second-year PhD student in Statistics at the University of Sevilla. His research focuses on the development of hypothesis testing procedures for comparing populations in asymptotic regimes. Before starting his PhD, he completed an MSc in Mathematics at the University of Sevilla, an MSc in Data Science and Computer Engineering at the University of Granada, and two BSc degrees in Mathematics and Physics at the University of Sevilla.

Testing the equality of a parameter across multiple populations is a classical problem in statistics, typically addressed under the assumption that the number of groups k is fixed and small relative to the within-group sample sizes $n_1, \ldots, n_k$. However, in many modern applications, k is large and may even exceed the group sizes. In this large-k, small-n regime, classical test statistics such as the ANOVA F-statistic and its rank-based analogues deviate from their usual reference distributions (see, e.g., [1]), motivating the development of new methods tailored to this setting. Several recent works have explored this regime in problems such as the k-sample problem (see, e.g., [4]), goodness-of-fit [2], and homoscedasticity [3].

This work introduces and studies a test for the comparison of an estimable parameter across k populations, when k is large and the sample sizes from each population are small when compared with k. The proposed test statistic is asymptotically distribution-free under the null hypothesis of parameter homogeneity, enabling asymptotically exact inference without parametric assumptions. Additionally, the behavior of the proposal is studied under alternatives. Simulations are conducted to evaluate its finite-sample performance, and a linear bootstrap method is implemented to improve its behavior for small to moderate k. Finally, an application to a real dataset is presented.

Keywords: consistency; many populations; testing; U-statistics.

References

[1] D.D. Boos and C. Brownie. ANOVA and rank tests when the number of treatments is large. *Statistics & Probability Letters*, 23(2):183–191, 1995. doi:10.1016/0167-7152(94)00112-L.

[2] M.D. Jiménez-Gamero. Testing normality of a large number of populations. *Statistical Papers*, 65:435–465, 2024. doi:10.1007/s00362-022-01384-y.

[3] M.D. Jiménez-Gamero, M. Valdora, and D. Rodríguez. Testing homoscedasticity of a large number of populations. *Statistical Papers*, 66(1):32, 2025. doi:10.1007/s00362-024-01650-1.

[4] D. Zhan and J.D. Hart. Testing equality of a large number of densities. *Biometrika*, 101(2):449–464, 2014. doi:10.1093/biomet/asu002.

Sorting is all you need

Víctor Blanco[1], Ivana Ljubic[2], Miguel A. Pozo[3], Justo Puerto[3] and Alberto Torrejón[3] (✉)

[1]Institute of Mathematics, University of Granada, Spain; [2]ESSEC Business School, Paris, France; [3]Institute of Mathematics, University of Seville, Spain

atorrejon@us.es

Alberto Torrejón is a PhD student in the Department of Statistics and Operations Research at the University of Seville and a member of the Institute of Mathematics of the University of Seville. His research focuses on mathematical modeling, algorithmic development, and statistical methodologies applied to data analysis and decision-making problems. In addition to his academic work, he is actively involved in organizing and participating in conferences, seminars, and outreach activities that promote the dissemination of mathematics, statistics, and operations research.

Sorting is a fundamental operation in mathematics and computer science, used to impose order and extract structure from data. Beyond its algorithmic importance, sorting underlies many optimization problems where ordered outcomes influence decision-making, whether through ranking, prioritization, or in the modelization of more abstract concepts such as equity, risk or envy.

In this talk, we introduce a flexible modeling framework grounded in linear and mixed-integer programming that leverages sorting principles to address a broad class of combinatorial optimization problems. Our approach not only efficiently handles known data values but is also capable of managing scenarios where values must be inferred, such as assignment costs or route lengths determined during the solution process. In addition, by means of an ordering framework a wide range of objective functions can be modeled, from classical location and dispersion measures to more complex criteria involving fairness, robustness, or envy. We demonstrate its applicability to decision-making problems like facility location [1, 2] and vehicle routing, as well as for statistical problems such as robust regression [3]. The proposed models incorporate ordered and bilevel optimization techniques and are supported by comprehensive computational experiments. The results underscore the framework's versatility, its theoretical soundness, and its potential to produce high-quality, interpretable solutions across various application contexts.

Keywords: mathematical programming; optimization problems; ordered optimization.

References

[1] I. Ljubić, M.A. Pozo, J. Puerto, and A. Torrejón. Benders decomposition for the discrete ordered median problem. *European Journal of Operational Research*, 317(3):858–874, 2024. doi:10.1016/j.ejor.2024.04.030.

[2] M.A. Pozo, J. Puerto, and A. Torrejón. The ordered median tree location problem. *Computers & Operations Research*, 169:106746, 2024. doi:10.1016/j.cor.2024.106746.

[3] J. Puerto and A. Torrejón. A fresh view on least quantile of squares regression based on new optimization approaches. *Expert Systems with Applications*, 282:127705, 2025. doi:10.1016/j.eswa.2025.127705.

SYSØR 2025
5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Nonparametric cure rate estimation using presmoothing

Samuel Saavedra[1] (✉), Ana López-Cheda[1] and María Amalia Jácome[1]

[1]CITIC, MODES group, Department of Mathematics, Universidade da Coruña, Spain

samuel.saavedra@udc.gal

Samuel Saavedra is a first-year PhD student in Statistics and Operational Research at Universidade da Coruña. His research activity is focused on the development of presmoothed estimators for cure models. Before enrolling his PhD, he worked as substitute lecturer in Statistics and he completed both a MSc in Bioinformatics and a BSc in Biology, also at the Universidade da Coruña.

Survival analysis focuses on the time until an event of interest occurs, under the assumption that all individuals will eventually experience the event [3]. However, in some cases, a portion of the population may never encounter the event. This scenario leads to cure models, which account for two distinct groups: those who will eventually experience the event and those who will never be affected [1]. Accurately estimating the probability of the event occurring is essential in such contexts. To enhance the efficiency of nonparametric estimates in survival analysis —and, by extension, in cure models— presmoothing techniques can be employed. A key aspect of presmoothing is the selection of an appropriate smoothing parameter, which plays a critical role in ensuring accurate and reliable results [2].

In this context, we introduce a novel presmoothing-based approach to improve cure rate estimation using the Beran estimator within in mixture cure models (MCM) [4]. When the presmoothing bandwidth is optimally selected —a crucial step— this method either matches or outperforms the classical nonparametric estimator, demonstrating its clear advantage. Through simulation studies and applications to clinical datasets, including breast cancer, we validate the effectiveness of this approach. Furthermore, the methodology has been extended to incorporate both binary and continuous covariates, enabling more precise patient-specific estimations and enhancing the model's ability to extract meaningful insights from the data.

Keywords: cure models; nonparametric estimation; presmoothing; survival analysis.

References

[1] J.W. Boag. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(1):15–53, 1949. doi:10.1111/j.2517-6161.1949.tb00020.x.

[2] M.A. Jácome and R. Cao. Almost sure asymptotic representation for the presmoothed distribution and density estimators for censored data. *Statistics*, 41(6):517–534, 2007. doi:10.1080/02331880701529522.

[3] J.P. Klein and M.L. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science & Business Media, 1997. doi:10.1007/978-1-4757-2728-9.

[4] A. López-Cheda, R. Cao, M.A. Jácome, and I. Van Keilegom. Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. *Computational Statistics & Data Analysis*, 105:144–165, 2017. doi:10.1016/j.csda.2016.08.002.

SYSOR 2025
5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Influence of features with internal structure in multi-class classification problems

María D. Guillén[1] (✉), Juan Aparicio[1], Juan Carlos Gonçalves-Dosantos[1] and Joaquín Sánchez-Soriano[1]

[1]Center of Operations Research (CIO), Miguel Hernández University (UMH)

maria.guilleng@umh.com

María D. Guillén is an Assistant Professor in the Department of Statistics, Mathematics and Computing at the University Miguel Hernández of Elche. She holds a PhD in Statistics, Optimization and Applied Mathematics from the same university. Her research focuses on the intersection of Machine Learning and Data Envelopment Analysis. Before her PhD, she completed an MSc in Big Data at the University of Santiago de Compostela and the University of Murcia, and a BSc in Computer Engineering at the University of Murcia.

Understanding the contribution of individual features to the output of classification models is crucial to explainable artificial intelligence. Game-theoretic concepts such as the Shapley and Banzhaf values have gained popularity as model-agnostic approaches to feature influence due to their solid axiomatic foundations (e.g., [2]). Although many of these methods assume independence between features, real-world data often involve internal structures or natural groupings among predictors. In this sense, game-theoretic values like the Owen and Banzhaf-Owen values have been developed to handle coalitional structures, allowing to model a priori unions among players. Recent literature has begun to explore these ideas in the context of binary classification, providing new directions for model interpretation when dependencies between features defined over finite sets are present [1].

In this work, we propose a novel influence measure for multiclass classification problems based on the Owen value, which considers possible internal structures among features. We formally define the influence measure and provide an axiomatic characterization for it. Our approach is model-agnostic and supports both categorical and continuous variables (potentially with infinite range of values). Finally, we illustrate the practical utility of our method through simulations and an application to the PISA 2012 dataset, where features are naturally grouped into blocks reflecting students' socioeconomic background, school characteristics, and cognitive skills. Our results confirm the robustness and interpretability of the proposed measure and highlight its value for structured real-world classification tasks.

Keywords: classification problems; influence measure of features; machine learning; owen value.

References

[1] L. Davila-Pena, A. Saavedra-Nieves, and B. Casas-Méndez. On the influence of dependent features in classification problems: A game-theoretic perspective. *Expert Systems with Applications*, 280:127446, 2025. doi:10.1016/j.eswa.2025.127446.

[2] E. Strumbelj and I. Kononenko. An efficient explanation of individual classifications using game theory. *The Journal of Machine Learning Research*, 11:1–18, 2010. URL: http://jmlr.org/papers/v11/strumbelj10a.html.

SYSØR 2025
5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Polynomial benchmarks for feature-interaction explanations

Pablo Morala[1] (✉), J. Alexandra Cifuentes[2], Rosa E. Lillo[1,3] and Iñaki Úcar[1,3]

[1]UC3M–Santander Big Data Institute (IBiDat), Universidad Carlos III de Madrid; [2] Department of Quantitative Methods, Universidad Pontificia de Comillas; [3] Department of Statistics, Universidad Carlos III de Madrid, Spain

pablo.morala@uc3m.es

Pablo Morala is a postdoctoral researcher at the UC3M–Santander Big Data Institute, UC3M, where he earned a PhD in Mathematical Engineering (2024). His research focuses on eXplainable Artificial Intelligence, specifically for neural network models. He holds BSc degrees in Mathematics and Physics (Universidad de Oviedo, 2019) and an MSc in Statistics for Data Science (UC3M, 2020).

Due to the opaque nature of many machine learning and artificial intelligence models, especially neural networks, the field of eXplainable Artificial Intelligence (XAI) has given birth to many methods that rely on single-variable importances as explanations, such as the Shapley Additive exPlanations (SHAP). However, these methods often overlook the interactions between variables and how they may affect the final model predictions. Consequently, several SHAP extensions have been proposed, such as the Shapley-Taylor Interaction Index [3], FaithSHAP [4] or n-Shapley values [1]. Nevertheless, these advances lack a rigorous evaluation with controlled simulations, because the usage of real datasets does not provide an effective ground truth in interpretability.

This work has two main goals: (i) to close that gap through a comprehensive simulation benchmark based on data generated from controlled polynomials, where interactions can be modeled directly, and (ii) to compare these SHAP-based models with NN2Poly [2], a neural networks XAI method based on polynomial representations, that can explicitly capture these variable interactions. The simulation study uses rank-based metrics on the interactions detected by each method and also develops new local approaches with NN2Poly to provide local and global comparisons in all cases.

Keywords: explainability; interactions; interpretability; XAI.

References

[1] S. Bordt and U. von Luxburg. From Shapley values to generalized additive models and back. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, pages 709–745, 2023. URL: https://proceedings.mlr.press/v206/bordt23a.html.

[2] P. Morala, J.A. Cifuentes, R.E. Lillo, and I. Úcar. NN2Poly: a polynomial representation for deep feed-forward artificial neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 36(1):781–795, 2025. doi:10.1109/TNNLS.2023.3330328.

[3] M. Sundararajan, K. Dhamdhere, and A. Agarwal. The Shapley Taylor interaction index. In *Proceedings of the 37th International Conference on Machine Learning*, pages 9259–9268, 2020. URL: https://proceedings.mlr.press/v119/sundararajan20a.html.

[4] C.P. Tsai, C.K. Yeh, and P. Ravikumar. Faith-Shap: the faithful Shapley interaction index. *Journal of Machine Learning Research*, 24(94):1–42, 2023. URL: https://dl.acm.org/doi/abs/10.5555/3648699.3648793.

5TH SPANISH YOUNG STATISTICIANS AND OPERATIONAL RESEARCHERS MEETING

Prescriptive modality selection

Emilio Carrizosa[1], Vanesa Guerrero[2] and José Carlos Castro[1] (✉)

[1]Instituto de Matemáticas de la Universidad de Sevilla, Spain; [2]Department of Statistics, Universidad Carlos III de Madrid, Spain

joscasgom1@alum.us.es

José Carlos Castro is a first-year PhD student in the Department of Statistics and Operations Research at the Universidad de Sevilla (US). His research focuses on Prescriptive Analytics. He holds a double BSc in Mathematics and Physics and an MSc in Mathematics, both from the Universidad de Sevilla.

Accurate predictions for individuals often depend on multimodal data (e.g., genomics, imaging, clinical tests), but acquiring all modalities may be infeasible due to cost, time, or invasiveness. This raises the challenge of selecting which modalities to collect under budget constraints before making any prediction. Traditional methods focus on global feature subsets [3], often ignoring personalized decisions and acquisition costs. Recent advances (e.g., cost-sensitive feature selection [1], multimodal prediction [4], and prescriptive analytics [2]) highlight the need to integrate prediction and optimization.

We propose a framework that integrates a specialized Random Forest predictor with integer programming to optimize personalized data acquisition and prediction. It estimates prediction errors for each individual based on a fully observed training set and uses these estimates to determine which modalities to collect for each case, balancing predictive gain and acquisition cost. The same model is used to generate final predictions. The approach leverages model structure to reduce optimization complexity, avoids imputation, and ensures scalability for individualized prescriptions. This methodology has been successfully validated on both synthetic and real-world datasets.

Keywords: classification and regression trees; prescriptive model; random forest.

References

[1] S. Benítez-Peña, R. Blanquero, E. Carrizosa, and P. Ramírez-Cobo. Cost-sensitive feature selection for support vector machines. *Computers & Operations Research*, 106:169–178, 2019. doi:10.1016/j.cor.2018.03.005.

[2] D. Bertsimas and N. Kallus. From predictive to prescriptive analytics. *Management Science*, 66(3):1025–1044, 2020. doi:10.1287/mnsc.2018.3253.

[3] J. Li, K. Cheng, S. Wang, F. Morstatter, R.P. Trevino, J. Tang, and H. Liu. Feature selection: a data perspective. *ACM Computing Surveys (CSUR)*, 50(6):1–45, 2017. doi:10.1145/3136625.

[4] G. Yu, Q. Li, D. Shen, and Y. Liu. Optimal sparse linear prediction for block-missing multi-modality data without imputation. *Journal of the American Statistical Association*, 115(531):1406–1419, 2020. doi:10.1080/01621459.2019.1640012.

List of attendees

Algendi, Abdalrahman	Molde University
Alonso-Pena, María	University of Santiago de Compostela
Álvarez, Begoña	University of Zaragoza
Baldomero-Naranjo, Marta	University of Cádiz
Bernárdez Ferradás, Alejandro	University of Vigo
Cabello, Esteban	Miguel Hernández University of Elche
Camacho, Jesús	Miguel Hernández University of Elche
Castilla, Elena	Rey Juan Carlos University
Castro, José Carlos	University of Seville
Corberán, Teresa	University of Valencia
Corrales, Daniel	Mathematical Sciences Institute ICMAT-CSIC
Cruz, Lidia	University Pablo de Olavide
Cuesta, Marina	Carlos III University of Madrid
De Souza, Marcelo	Santa Catarina Estate University
Figueras-Téllez, Cecilia	Institute of Statistics and Cartography of Andalusia (IECA)
Fischetti, Martina	University of Seville
García Arce, Pablo	Mathematical Sciences Institute ICMAT-CSIC
García-García, José	University of Oviedo
Gómez-Vargas, Nuria	University of Seville
González-Vázquez, Héctor	University of Santiago de Compostela
Guerrero, Vanesa	Carlos III University of Madrid
Guillén, María D.	Miguel Hernández University of Elche
Hernández, Aitor	University of Zaragoza
León, Javier	Decide4AI
Lillo, Rosa E.	Carlos III University of Madrid
Martín-Chávez, Pedro	University of Warwick
Martínez Miranda, María Dolores	University of de Granada
Merino Maestre, María	University of the Basque Country UPV-EHU
Minuesa, Carmen	University of Extremadura
Morala, Pablo	Carlos III University of Madrid

Nácher, Lorena	Miguel Hernández University of Elche
Navarro García, Manuel	Decide4AI
Navas-Orozco, Antonio	University of Seville
Olivares, Adam	Polytechnic University of Catalonia
Pulido, Belén	National University of Distance Education
Robert, Christian P.	Paris-Dauphine University
Rodríguez-Ballesteros, Sofía	Miguel Hernández University of Elche
Romero-Madroñal, Marcos	University of Seville
Saavedra, Samuel	University of A Coruña
Saldanha-da-Gama, Francisco	Sheffield University Management School
Santos-Pascual, Miguel	Mathematical Sciences Institute ICMAT-CSIC
Segura, Paula	University of Valencia
Serrano, Diego	Carlos III University of Madrid
Sillero-Denamiel, M. Remedios	University of Seville
Sinova, Beatriz	University of Oviedo
Temprano, Francisco	Carlos III University of Madrid
Terán-Viadero, Paula	Complutense University of Madrid
Tobar-Fernández, Cristina	Rey Juan Carlos University
Torrejón, Alberto	University of Seville

List of authors