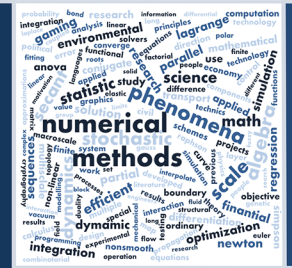


October  
13-15 | 2022

Instituto Politécnico de Tomar  
Unidade Departamental de Matemática e Física



VIII Workshop  
*on*  
**COMPUTATIONAL DATA ANALYSIS  
AND NUMERICAL METHODS**

**BOOK OF ABSTRACTS**

## WELCOME TO THE VIII WCDANM | 2022

Dear participants, colleagues and friends,

on behalf of the executive and organizing committees of the VIII Workshop on Computational Data Analysis and Numerical Methods (VIII WCDANM) it is a great honour and a privilege to warmly welcome you to this event, which is organized by the Polytechnic Institute de Tomar (located in the centre of Portugal in the beautiful city of Tomar), with the support of some Portuguese research centres. We hope that the final result can exceed the expectations of participants, sponsors and organizers.

This meeting will be broadcasted by videoconference (webinar), with important contributions from Portuguese researchers: Professor Carlos A. Coelho (NOVA University of Lisbon), Professor José Augusto Ferreira (University of Coimbra), Professor M. Ivette Gomes (CEAUL and DEIO, University of Lisbon) and Professor Ana Leonor Silvestre (IST, University of Lisbon) as Plenary Speakers. The high scientific level of the oral and poster presentations and an active audience will certainly contribute to the success of the meeting. Some of the works accepted (theoretical and applied) by the VIII WCDAM involve data science, big data, data mining and machine learning, in different areas of research. A very special thanks to this scientific community, because only with their important contribution it is possible to hold this event.

This year, the VIII WCDANM also has two courses available, one on “Neural Computing” coordinated by Professor Padhu Seshaiyer and Mr. Alonso Gabriel Ogeda (George Mason University, USA) and another in Statistical Quality Control coordinated by Professor Marien Graham (University of Pretoria, South Africa), who kindly and promptly accepted our invitation and to whom we are very grateful.

A special thanks is also due to the Members of the Executive, Scientific and Organizing Committees. In particular, Professor Fernando Carapau (University of Évora, Portugal), Professor Anuj Mubayi (Illionois State University, USA), Professor Milan Stehlík (Johannes Kepler University, Austria) of the Executive Committee and the tireless colleagues of the local Organizing Committee, Manuela Fernandes, Helena Grilo, Ana Nata (hosts of the Polytechnic Institute of Tomar, Portugal). For the dissemination of the scientific results achieved at the meeting, participants have available the “Journal of Applied Statistics” (Taylor & Francis) and “Neural Computing and Applications” (Springer), and this year also the new journal “Research in Statistics” (Taylor & Francis).

We hope that VIII WCDANM can provide an intellectual stimulus and an opportunity for the respective scientific community to work together, as well as to disseminate their scientific research. We wish you all a pleasant workshop!

Tomar, October 13–15, 2022.

Chairman of the Executive Committee of VIII WCDANM,



Luís Miguel Grilo

Polytechnic Institute of Tomar

Open University, Portugal

NOVAMath (Center for Mathematics and Applications), FCT NOVA, NOVA University of Lisbon, Portugal

## Committees

### Organizing Committee

Luís Miguel Grilo, Instituto Politécnico de Tomar, Portugal (Chairman of the Workshop)  
Ana Nata, Instituto Politécnico de Tomar, Portugal  
Helena Luzia Grilo, Instituto Politécnico de Tomar, Portugal  
Maria Manuela Fernandes, Instituto Politécnico de Tomar, Portugal

### Executive Committee

Anuj Mubayi, Illinois State University, USA  
Fernando Carapau, Universidade de Évora, Portugal  
Luís M. Grilo, Instituto Politécnico de Tomar, Portugal & Universidade Aberta, Portugal  
Milan Stehlík, Johannes Kepler University, Austria

### Scientific Committee

Adélia Sequeira, IST, Universidade de Lisboa, Portugal  
Alberto Pinto, Universidade do Porto, Portugal  
Ana Catarina Carapito, Universidade da Beira Interior, Portugal  
Ana Cristina Nata, Instituto Politécnico de Tomar, Portugal  
Ana Isabel Borges, Instituto Politécnico do Porto, Portugal  
Ashwin Vaidya, Montclair State University, USA  
Anuj Mubayi, Illinois State University, USA  
Arminda Manuela Gonçalves, Universidade do Minho, Portugal  
Amílcar Oliveira, Universidade Aberta, Portugal  
Alexandra Moura, Instituto Superior de Economia e Gestão, Portugal  
Alberto Simões, Universidade da Beira Interior, Portugal  
Adelaide Figueiredo, Universidade do Porto, Portugal  
Aldina Correia, Instituto Politécnico do Porto, Portugal  
Catarina Nunes, Universidade Aberta, Portugal  
Carlos Braumann, Universidade de Évora, Portugal  
Carla Santos, Instituto Politécnico de Beja, Portugal  
Carlos Ramos, Universidade de Évora, Portugal  
Carlos Agra Coelho, Universidade Nova de Lisboa, Portugal  
Célia Nunes, Universidade da Beira Interior, Portugal  
Cristina Dias, Instituto Politécnico de Portalegre, Portugal  
Delfim F. M. Torres, Universidade de Aveiro, Portugal  
Dora Gomes, Universidade Nova de Lisboa, Portugal  
Eliana Costa e Silva, Instituto Politécnico do Porto, Portugal  
Fátima de Almeida Ferreira, Instituto Politécnico do Porto, Portugal  
Feliz Minhós, Universidade de Évora, Portugal  
Fernanda Figueiredo, Universidade do Porto, Portugal  
Filipe Marques, Universidade Nova de Lisboa, Portugal  
Fernando Carapau, Universidade de Évora, Portugal  
Henrique Pinho, Instituto Politécnico de Tomar, Portugal  
Ilda Inácio, Universidade da Beira Interior, Portugal  
Isabel Silva Magalhães, Universidade do Porto, Portugal  
Irene Rodrigues, Universidade de Évora, Portugal



João Tiago Mexia, Universidade Nova de Lisboa, Portugal  
João Janela, Instituto Superior de Economia e Gestão, Portugal  
José Gonçalves Dias, Instituto Universitário de Lisboa, Portugal  
José António Pereira, Universidade do Porto, Portugal  
José Augusto Ferreira, Universidade de Coimbra, Portugal  
José Carrilho, Universidade de Évora, Portugal  
Leonid Hanin, Idaho State University, USA  
Luís Castro, Universidade de Aveiro, Portugal  
Luís M. Grilo, Instituto Politécnico de Tomar, Portugal & Universidade Aberta, Portugal  
Malay Banerjee, Indian Institute of Technology Kampur, India  
Manuel L. Esquível, Universidade Nova de Lisboa, Portugal  
Manuel Branco, Universidade de Évora, Portugal  
Manuela Oliveira, Universidade de Évora, Portugal  
Maria Luísa Morgado, Universidade de Trás-os-Montes e Alto Douro, Portugal  
Maria do Rosário Ramos, Universidade Aberta, Portugal  
Marien Graham, University of Pretoria, Pretoria, South of Africa  
Milan Stehlík, Johannes Kepler University, Austria  
Padmanabhan Seshaiyer, George Mason University, USA  
Paulo Correia, Universidade de Évora, Portugal  
Pedro Oliveira, Universidade do Porto, Portugal  
Philippe Fraunié, Université de Toulon, France  
Russell Alpizar-Jara, Universidade de Évora, Portugal  
Susana Faria, Universidade do Minho, Portugal  
Tanuja Srivastava, Indian Institute of Technology, India  
Teresa Oliveira, Universidade Aberta, Portugal

## Sponsored by



# Technical Specifications

**Title**

VIII Workshop on Computational Data Analysis and Numerical Methods

**Web-page**

<http://www.wcdanm.ipt.pt/2022>

**Editor**

Instituto Politécnico de Tomar  
Quinta do Contador – Estrada da Serra  
2300–313 Tomar

**Editors**

Luís M. Grilo (Instituto Politécnico de Tomar, Portugal & Universidade Aberta, Portugal)  
Ana Nata (Instituto Politécnico de Tomar, Portugal)  
Helena Luzia Grilo (Instituto Politécnico de Tomar, Portugal)  
Maria Manuela Fernandes (Instituto Politécnico de Tomar, Portugal)

**Authors**

Many authors.

**Published in a PDF format by:**

Instituto Politécnico de Tomar  
Copyright © 2022 left to the authors of individual papers  
All rights reserved.

**ISBN:** 978-989-8840-70-7

# Contents

|  |     |
|--|-----|
| Welcome to the VIII WCDANM   2022..... | i   |
| Committees .....                       | iii |
| Sponsors.....                          | v   |
| Technical Specifications.....          | vi  |

---

## Invited Speakers

---

|   |   |
|---|---|
| <b>Carlos A. Coelho</b><br><i>A Likelihood Ratio Test for one-way MANOVA Which works for both the low and high-dimensional settings .....</i>   | 2 |
| <b>José Augusto Ferreira</b><br><i>Exogenous stimuli responsive drug delivery systems: mathematical modeling and numerical simulation .....</i> | 3 |
| <b>M. Ivette Gomes</b><br><i>Further Tales of Heavy Tails and Generalized Means .....</i>   | 4 |
| <b>Ana Leonor Silvestre</b><br><i>The method of fundamental solutions applied to incompressible viscous flows entrepreneurship .....</i>        | 6 |

---

## Organized Sessions

---

|   |    |
|---|----|
| <b>Organized Session 1</b><br><b>Statistics and Modelling</b><br><b>Organizer: Milan Stehlík.....</b>   | 7  |
| <b>Arif S. Bagwan</b><br><i>Applications of Fixed point theorems in Fractional Calculus .....</i>   | 9  |
| <b>Zdeněk Fabián</b><br><i>Parametric estimation based on scalar-valued score functions .....</i>   | 11 |
| <b>Mukhtar Ahmad Khanday</b><br><i>Modeling human physiology with special reference to local Hyperthermic treatment of solid tumors .....</i> | 12 |
| <b>Jozef Kiselák</b><br><i>On Particle-Size Distribution of Three-dimensional Convex Similar Bodies .....</i>                                 | 13 |
| <b>Milan Stehlík</b><br><i>Statistical and Learning Functions .....</i>   | 14 |

---

---

|  |    |
|--|----|
| <b>Organized Session 2</b>   |    |
| <b>Statistical Methods and Applications</b>  |    |
| <b>Organizer:</b> <i>Fernanda Figueiredo</i> .....   | 15 |
| <hr/>  |    |
| <b>Frederico Caeiro</b>  |    |
| <i>A Comparison of Several Generalized and Extended Hill Estimators</i> .....  | 16 |
| <b>Adelaide Figueiredo</b>   |    |
| <i>Analysing the indicators of Europe 2020 agenda using the STATIS methodology</i> .....   | 17 |
| <b>Fernanda Otilia Figueiredo</b>  |    |
| <i>Digital technologies use and training: a multivariate data analysis</i> .....   | 19 |
| <b>Lígia Henriques-Rodrigues</b>   |    |
| <i>Lehmer's mean-of-order-p in the estimation of the Weibull tail coefficient</i> .....  | 21 |
| <hr/>  |    |
| <b>Organized Session 3</b>   |    |
| <b>Contributions of Statistics for professional and personal decision making</b>   |    |
| <b>Organizer:</b> <i>Carla Santos</i> .....  | 23 |
| <hr/>  |    |
| <b>Ana Cantarina</b>   |    |
| <i>Multivariate collective risk models and a special case</i> .....  | 24 |
| <b>Ana Dias</b>  |    |
| <i>Characterization of the dissertations of a master's degree in occupational health and safety</i> .....  | 25 |
| <b>Cristina Dias</b>   |    |
| <i>A new approach for analyzing multi-environment trials</i> .....   | 27 |
| <b>Carla Santos</b>  |    |
| <i>On estimation in an extended mixed model with balanced design</i> .....   | 28 |
| <hr/>  |    |
| <b>Organized Session 4</b>   |    |
| <b>Computational Data Analysis and Numerical Methods for Predicting the Dynamics of Infectious Diseases</b>  |    |
| <b>Organizers:</b> <i>Viswanathan Arunachalam and Padmanabhan Seshaiyer</i> .....  | 30 |
| <hr/>  |    |
| <b>Alvan Caleb Arulandu</b>  |    |
| <i>Physics-Informed Neural Networks for Informed Vaccine Distribution in Heterogeneously Mixed Populations</i> .....   | 31 |
| <b>Sohini Banerjee &amp; Long Nguyen</b>   |    |
| <i>Using Physics-Informed Neural Networks to Model and Simulate the Dynamics of Infectious Diseases in Enclosed Environments</i> .....                         | 33 |
| <b>Alonso Gabriel Ogueda &amp; Erika Johanna Martínez Salinas</b>  |    |
| <i>A Review and Application of Disease Informed Neural Network</i> .....   | 35 |
| <b>Viswanathan Arunachalam</b>   |    |
| <i>A stochastic model for epidemics with a diagnostic test for the spread of the SARS-Cov-2 virus: A case study for the Nariñense region in Colombia</i> ..... | 36 |

---

---

|   |    |
|---|----|
| <b>Organized Session 5</b>  |    |
| <b>Statistical Modeling and Applications</b>  |    |
| <b>Organizer: José A. Pereira</b> . . . . .   | 37 |
| <hr/>   |    |
| <b>Patrícia Antunes</b>   |    |
| <i>Estimation in Gamma distributed Bi-additive models</i> . . . . .   | 38 |
| <b>Kwaku Opoku-Ameyaw</b>   |    |
| <i>A novel test for grouping levels of a factor: An application to cocoa breeding experiments in acidic soils</i> . . . . .   | 39 |
| <b>José Pereira</b>   |    |
| <i>Estimation of the Best Interval Between Radiography in Early Bone Loss Detection Using Logistic Regression Classifiers</i> . . . . .                                     | 41 |
| <hr/>   |    |
| <b>Organized Session 6</b>  |    |
| <b>Mathematical models</b>  |    |
| <b>Organizer: Sílvia Barbeiro</b> . . . . .   | 43 |
| <hr/>   |    |
| <b>Rafael Henriques</b>   |    |
| <i>Mathematical model to reconstruct the mechanical properties of an elastic medium</i> . . . . .   | 44 |
| <b>Afonso Costa</b>   |    |
| <i>Wave propagation in viscoelastic materials: analytical and numerical study</i> . . . . .   | 45 |
| <b>Milene Santos</b>  |    |
| <i>Modelling light propagation in the cornea: treatment of curved boundary domains</i> . . . . .  | 46 |
| <b>Sandra Vaz</b>   |    |
| <i>Local analysis of Lotka-Volterra type model</i> . . . . .  | 47 |
| <hr/>   |    |
| <b>Organized Session 7</b>  |    |
| <b>Threshold Behavior in Epidemiological and Health Sciences</b>  |    |
| <b>Organizer: Swati DebRoy</b> . . . . .  | 48 |
| <hr/>   |    |
| <b>Swati DebRoy</b>   |    |
| <i>Identifying the mode of action of drug from viral load data</i> . . . . .  | 49 |
| <b>Isaac Fung</b>   |    |
| <i>Is the epidemic growing or declining? Applying the instantaneous reproduction number method to COVID-19 data: case studies in the United States and Canada</i> . . . . . | 50 |
| <b>Anuj Mubayi</b>  |    |
| <i>Multidimensional evidence generation for health technology assessment</i> . . . . .  | 51 |
| <b>Omar Saucedo</b>   |    |
| <i>Incorporating human mobility data into epidemiological models</i> . . . . .  | 52 |
| <hr/>   |    |
| <b>Organized Session 8</b>  |    |
| <b>Statistical Modeling and Forecasting</b>   |    |
| <b>Organizer: A. Manuela Gonçalves</b> . . . . .  | 53 |

---



---

|   |    |
|---|----|
| <b>A. Manuela Gonçalves</b>   |    |
| <i>Short-term forecast by state-space model approach: the case study of maximum temperature</i> ..... | 54 |
| <b>F. Catarina Pereira</b>  |    |
| <i>Simulation study of time series with outliers in the context of state space models</i> .....       | 55 |
| <b>M. Rosário Ramos</b>   |    |
| <i>Projections of amputations in diabetics in Portugal with count regression model</i> ..             | 57 |

---

|  |    |
|--|----|
| <b>Organized Session 9</b>                                   |    |
| <b>Statistical and computational methods in applications</b> |    |
| <b>Organizer: Manuela Neves</b> .....                        | 59 |

---

|   |    |
|---|----|
| <b>Luís M. Grilo</b>  |    |
| <i>Optimism and stress as predictors of burnout in college students: PLSc-SEM approach</i> .....            | 60 |
| <b>Manuela Neves</b>  |    |
| <i>Revisiting computational procedures in extreme value parameter estimation</i> .....                      | 62 |
| <b>Nelson T. Jamba</b>  |    |
| <i>Approximation methods for the likelihood function in parameter estimation for SDE mixed models</i> ..... | 64 |

---

|                |  |
|----------------|--|
| <b>Posters</b> |  |
|----------------|--|

---

|   |    |
|---|----|
| <b>João M. Alves</b>  |    |
| <i>Big Data Analytics in Water Consumption</i> .....  | 67 |
| <b>Susana Faria</b>   |    |
| <i>Fitting mixtures of linear mixed models: a case study</i> .....  | 69 |
| <b>Susana Faria</b>   |    |
| <i>Mixtures of linear mixed models: variable selection using CEM algorithm</i> .....  | 71 |
| <b>Marta Ferreira</b>   |    |
| <i>Extremal Index: estimation and resampling</i> .....  | 72 |
| <b>Jaime Jerónimo</b>   |    |
| <i>Survival analysis of women diagnosed with breast cancer in Angola</i> .....  | 73 |
| <b>Miguel C. Lobo</b>   |    |
| <i>Retail store object detection using Yolo</i> .....   | 75 |
| <b>Inês J. Sequeira</b>   |    |
| <i>An application of logistic regression model to the study of constitutional imbalances in human chromosomal fragile sites</i> ..... | 77 |
| <b>Index of authors</b> .....   | 79 |

## **Invited Speakers**

# A Likelihood Ratio Test for one-way MANOVA Which works for both the low and high-dimensional settings

Carlos A. Coelho<sup>1,2</sup>

<sup>1</sup>Mathematics Department and NOVA Math-NOVA School of Science and Technology

<sup>2</sup>NOVA University of Lisbon, Caparica, Portugal

**E-mail address:** *cmac@fct.unl.pt*

## Keywords

Asymptotic Normal distribution, Exact distribution, Generalized Integer Gamma distribution, Generalized Near-Integer Gamma distribution, Product of Beta random variables.

---

A Likelihood Ratio Test is developed for the one-way high-dimensional MANOVA. In the high-dimensional setting this test is able to outperform in terms of power most existing tests in most situations, displaying an extraordinary behavior even for extremely skewed distributions as well as heavy tailed distributions, including those with no expected value, in which case it becomes a test for location. Still, it shows a better Type I error control than existing tests and non-inflated power values. Furthermore, the test presented is able to work with samples of size just 1, for all samples, except one of them, and its statistic has a very nice and simple asymptotic Normal distribution, which does not require any restrictions on sample sizes in order to hold. Furthermore, it can also be used as a common Likelihood Ratio Test in the low-dimensional setting. Extended simulation results are presented.

---

## Exogenous stimuli responsive drug delivery systems: mathematical modeling and numerical simulation

José Augusto Ferreira

University Coimbra, CMUC, Department of Mathematics, Coimbra, Portugal

**E-mail address:** *ferreira@mat.uc.pt*

---

Stimuli responsive drug delivery systems are medical devices that are able to deliver drugs locally in the target tissue in a controlled way and without affecting the healthy tissues. The drugs are encapsulated in nanostructures that are able to be safely transported to the target tissue where the drug release takes place enhanced by a stimulus.

This talk aims to discuss the mathematical modelling and the numerical simulation of the drug delivery driven by exogenous stimuli.

### Keywords

Drug delivery systems, Mathematical modelling, Numerical simulation.

---

Chemotherapy is the traditional cancer treatment often combined with radiation therapy and surgical resection. However, the delivery of chemotherapeutic agents leads to severe side effects due to the non-selective toxicity of the drugs. To avoid this serious limitation, an intensive research has been carried out in chemotherapy delivery devices that are able to locally deliver the drugs in a controlled way without affecting the healthy tissue.

Stimuli responsive nano-scale drug delivery systems such as tunable prodrugs, polymeric micelles, inorganic nanoparticles, nanotubes, nanorods, dendrimers, lipid-based drug delivery systems and carrier-based drug delivery systems have been used in several cancer scenarios to safely transport chemotherapeutic cocktails to the target tissue and release the therapeutic agents there. In this talk we will be mainly focused on exterior stimuli -light, temperature, ultrasound, electric fields, magnetic fields, etc.

From the mathematical point of view, the description of the drug delivery in the target tissue enhanced by a stimulus leads to several challenging problems: the mathematical description of the drug release, its transport and absorption by the target should be combined with the description of the effect of the stimulus. From the numerical point of view, the development of stable and accurate numerical methods that are able to simulate the entire drug delivery process also leads to several interesting challenges. The dependence of the drug release and drug transport on the stimulus requires the use of convenient numerical methods for the drug enhancer that does not degrade the properties of the numerical approximations of the drug concentration.

This talk will be focused on the mathematical modeling and the numerical simulation of drug delivery driven by exterior stimuli in different scenarios. Mathematical models and numerical methods will be discussed. Numerical results illustrating the behaviour of the models and the obtained theoretical results will be presented.

**Acknowledgements** This work was partially supported by the Center for Mathematics of the University of Coimbra (CMUC) - UIDB/00324/2020, funded by the Portuguese Government through FCT/MCTES.

## Further Tales of Heavy Tails and Generalized Means

M. Ivette Gomes<sup>1,2</sup>, Frederico Caeiro<sup>3,4</sup>, Fernanda Figueiredo<sup>2,5</sup> and Lúgia Henriques-Rodrigues<sup>6,7</sup>

<sup>1</sup>Universidade de Lisboa (UL), Portugal

<sup>2</sup>Centro de Estatística e Aplicações (CEAUL), Portugal

<sup>3</sup>Universidade Nova de Lisboa, Portugal

<sup>4</sup>Centro de Matemática e Aplicações (CMA), Portugal

<sup>5</sup>Universidade do Porto, Portugal

<sup>6</sup>Universidade de Évora, Portugal

<sup>7</sup>Centro de Investigação em Matemática e Aplicações (CIMA), Portugal

**E-mail addresses:** *ivette.gomes@fc.ul.pt; fac@fct.unl.pt; otilia@fep.up.pt; ligiahr@uevora.pt*

Statistics of extremes have been faced with many challenges, essentially due to the complexity of extreme events, high relevant to society in the most diverse fields. Generalised means (GMs) have been used with success for heavy right-tails, in the estimation of a positive extreme value index, among other parameters of extreme events. We shall now make use of GMs in the estimation of the conditional tail expectation, also crucial for an adequate risk assessment.

### Keywords

Conditional tail expectation, Generalised means, Semi-parametric estimation.

In the field of statistical *extreme value theory* (EVT), a great variety of alternative methodologies are available to deal with the management of risks of extreme events, commonly associated with a heavy right-tail and a positive *extreme value index* (EVI)  $\xi$ . The risk is generally expressed either by the *value-at-risk* ( $\text{VaR}_q$ ), the maximum size of the loss that is exceeded with probability  $1 - q$ ,

$$\text{VaR}_q := Q(q), \quad \text{with } Q(q) := \inf\{x \geq 0 : F(x) \geq q\}, \quad q \in (0, 1),$$

or by the *conditional tail expectation* (CTE), defined as

$$\text{CTE}_q \equiv \text{CTE}_q(X) = E(X|X > Q(q)) = \frac{1}{1-q} \int_q^1 Q(s) ds, \quad q \in (0, 1).$$

Contrarily to VaR, and as can be seen in [1], the CTE provides information about the potential loss beyond the VaR level and is a coherent risk measure. Paying attention to the right tail, with  $X_{i:n}$ ,  $1 \leq i \leq n$ , denoting the  $i$ -th ascending order statistic associated with a random sample  $X_i$ ,  $1 \leq i \leq n$ , and with  $s_+ := \max(s, 0)$ , the CTE-estimator,

$$\widetilde{\text{CTE}}_{k,n,q}(X; H_{k,n}) = \frac{1}{1-q} \sum_{j=1}^{n-k} \left( \binom{j}{n} - q \right)_+ - \left( \binom{j-1}{n} - q \right)_+ \Big) X_{j:n} + \frac{kX_{n-k:n}}{n(1-q)(1-H_{k,n})},$$

is proposed in [4], where  $\xi$  is estimated through the Hill EVI-estimator in [5], given by

$$H_{k,n} := \frac{1}{k} \sum_{i=1}^k V_{i,k}, \quad V_{i,k} := \ln X_{n-i+1:n} - \ln X_{n-k:n}, \quad k = 1, 2, \dots, n-1. \quad (1)$$

The interesting performance of EVI-estimators based on *generalized means* (GMs), leads us to consider a simple generalization of the Hill EVI-estimators, in (1), studied in [2], among others, under a second-order framework. Indeed, since the H EVI-estimators are the logarithm of the *geometric mean* (or *mean-of-order-0*) of  $U_{ik} := \exp(V_{i,k}) = X_{n-i+1:n}/X_{n-k:n}$ ,  $1 \leq i \leq k$ , with  $V_{i,k}$  given in (1), we can more generally consider the mean-of-order- $p$  of  $U_{i,k}$ , being led to the *mean-of-order- $p$*  ( $MO_p$ ) EVI-estimators (see also [3], where these estimators are dealt with under a third-order framework). More generally than  $\widehat{CTE}_{k,n,q}(X; H_{k,n})$ , we can base the CTE-estimation on the  $MO_p$  EVI-estimators, or even on other GM EVI-estimators, like the ones in [6–8]. Consistency and asymptotic normality of the estimators under study is proved under regular frameworks. The performance of the new estimators for finite samples is illustrated through a Monte-Carlo simulation.

**Acknowledgements** Research partially supported by National Funds through **FCT**—Fundação para a Ciência e a Tecnologia, projects UIDB/MAT/0006/2020 (CEA/UL), UIDB/MAT/0297/2020 (CMA/UNL) and UIDB/MAT/04674/2020 (CIMA).

## References

- [1] P. Artzner, F. Delbaen, J-M. Eber and D. Heath. Coherent measures of risk. *Math. Finance* **9**: 203–228, 1999.
- [2] M. F. Brilhante, M. I. Gomes and D. Pestana. A simple generalisation of the Hill estimator. *Comput. Statist. and Data Analysis* **57(1)**: 518–535, 2013.
- [3] F. Caeiro, M. I. Gomes, J. Beirlant and T. de Wet. Mean-of-order  $p$  reduced-bias extreme value index estimation under a third-order framework. *Extremes* **19(4)**: 561–589, 2016.
- [4] E. H. Deme, S. Girard and A. Guillou. Reduced-Bias Estimator of the Conditional Tail Expectation of Heavy-Tailed Distributions. In: Hallin, M., Mason, D., Pfeifer, D., Steinebach, J. (eds.) *Mathematical Statistics and Limit Theorems*. Springer, Cham, 2015.
- [5] B. M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* **3**: 1163–1174, 1975.
- [6] H. Penalva, F. Caeiro, M. I. Gomes and M. M. Neves. *An Efficient Naive Generalisation of the Hill Estimator: Discrepancy between Asymptotic and Finite Sample Behaviour*. Notas e Comunicações CEAUL 02/2016, 2016.
- [7] H. Penalva, M. I. Gomes, F. Caeiro and M. M. Neves. A couple of non reduced bias generalized means in extreme value theory: an asymptotic comparison. *Revstat-Statist. J.* **18(3)**: 281–298, 2020.
- [8] H. Penalva, M. I. Gomes, F. Caeiro and M. M. Neves. Lehmer’s mean-of-order- $p$  extreme value index estimation: a simulation study and applications. *J. Applied Statistics* **47(13–15)** (*Advances in Computational Data Analysis*): 2825–2845, 2020.



---

## The method of fundamental solutions applied to incompressible viscous flows

Ana Leonor Silvestre, Carlos Alves, Rodrigo Serrão and  
Nuno F.M. Martins

Department of Mathematics and CEMAT, Instituto Superior Técnico, Universidade de Lisboa

**E-mail address:** *ana.silvestre@math.tecnico.ulisboa.pt*

### Keywords

International Entrepreneurship, Perceptual variables, Country-level.

---

We begin by presenting a numerical study of the Method of Fundamental Solutions (MFS) for Stokes boundary value problems in 2 and 3 dimensions. The basis of the implementation of the MFS are Stokes fundamental solutions (Stokeslets and Stresslets) in the primal velocity-pressure formulation of the problems.

Then, we will consider the so-called domain method of fundamental solutions (MFS-D) for a Stokes problem containing a nonzero body force, and propose two methods for the numerical construction of a particular solution: one method is based on the Neuber-Papkovich potentials and the second method relies on a Helmholtz-type decomposition for the body force, enabling the construction of divergence-free basis functions.

The accuracy of the MFS and MFS-D will be illustrated through a series of numerical tests. The application of the MFS-D to the Navier-Stokes equations will also be addressed.

### References

- [1] C. J. S. Alves and A. L. Silvestre. On the application of the method of fundamental solutions to nonlinear partial differential equations, *Eng. Anal. Bound. Elem.* **92**: 267–274, 2018.
- [2] C. J. S. Alves, R. Serrão and A. L. Silvestre. Fundamental solutions for the Stokes equations: Numerical applications for 2D and 3D flows, *Appl. Numer. Math.* **170**: 55–73, 2021.
- [3] C. J. S. Alves, N. F. M. Martins and A. L. Silvestre. Numerical methods with particular solutions for nonhomogeneous Stokes and Brinkman systems, *Adv. Comput. Math.* **48**: 44, 2022.

## **Organized Sessions**

## Organized Session 1

### Statistics and Modelling

Organizer: Milan Stehlik

---

# Applications of Fixed point theorems in Fractional Calculus

Arif S. Bagwan<sup>1</sup>

<sup>1</sup>Pimpri Chinchwad College of Engineering, Pune, Maharashtra State, India - 411044

**E-mail addresses:** *arif.bagwan@gmail.com; pachpatte@gmail.com*

---

In this talk, I will give you the brief introduction to fractional calculus and will be presenting one of our recent work on the existence and uniqueness results for the solutions to implicit fractional differential equations involving generalized Katugampola derivative with nonlocal initial condition. By means of some classical fixed point theorem techniques such as Krasnosel'skii fixed point theorem and Banach contraction principle the main results have been carried out.

## Keywords

Generalized Katugampola derivative, Implicit differential equation; Existence; Uniqueness, Fixed point.

---

The concept of Fractional Calculus is the most popular concept of modern mathematics in past few decades which was introduced at the same time as that of classical calculus. The concept of classical calculus fails to describe many complex phenomena which gives a birth to fractional calculus [6]. It has opened a new era of mathematical research which in turn, provides an impetus to the development of numerous mathematical disciplines such as theory of fractional difference and differential equations, Control theory, Fluid flow, Viscoelasticity, Diffusion, Probability, Statistics, Electrical circuits, Electro-analytical chemistry, Neurology, Optics, Acoustics, Rheology, Bioengineering and many more.

The fractional order initial and boundary value problems plays vital role in many fields of science and engineering applications. Many industries like Mechanical, Automobile, Aerospace, Electronics and communication, Pharmaceutical, Chemical, Petroleum, Nanotechnology and Biotechnology rely on the study of initial and boundary value problems. They are also useful to describe many complex phenomena in designing and manufacturing high technological products in aforementioned fields. The existence, uniqueness and stability theory of such initial and boundary value problems remain the area of interest for many research scholars. See [1–8] and references therein. One of the most important tools in the study of existence of solutions to initial as well as boundary value problems are the fixed point theorems and their generalizations.

## References

- [1] S. Abbas, M. Benchohra, J. Lagreg, A. Alsaedi and Y. Zhou. Existence and Ulam stability for fractional differential equations of Hilfer-Hadamard type. *Adv. Difference Equ.* **180**, DOI 10.1186/s13662-017-1231-1, 2017.
- [2] B. Ahmad and S. K. Ntouyas. Initial value problem of fractional order Hadamard-type functional differential equations. *Electron. J. Differential Equations* **2015**, 1–9, 2015.

- 
- [3] A.S. Bagwan and D. B. Pachpatte. Existence and Stability of nonlocal initial value problems involving generalized Katugampola derivative, *Kragujevac J. Math.* **46**, (3), 443–460, 2022.
  - [4] M. Benchohra and J. E. Lazreg. Existence and Ulam stability for nonlinear implicit fractional differential equations with Hadamard derivative, *Stud. Univ. Babeş-Bolyai Math.*, **62**, 27–38, 2017.
  - [5] K. M. Furati, M. D. Kassim and N.e-. Tatar. Existence and uniqueness for a problem involving Hilfer fractional derivative, *Comput. Math. Appl.* **64**: 1616–1626, 2012.
  - [6] K. S. Miller and B. Ross. A Introduction to the Fractional Calculus and Differential Equations, *John Wiley, New York* 1993.
  - [7] D. S. Oliveira and E. C. de Oliveira. Hilfer-Katugampola fractional derivative, *Comput. Appl. Math.* **37**: 3672–3690, 2018.
  - [8] D. Vivek, K. Kanagrajan and E. M. Elsayed. Nonlocal initial value problems for implicit differential equations with Hilfer-Hadamard fractional derivative, *Nonlinear Anal. Model. Control.* **23**: 341–360, 2018.

---

# Parametric estimation based on scalar-valued score functions

Zdeněk Fabián<sup>1</sup>

<sup>1</sup>Institute of Computer Science, Czech Academy of Sciences, Prague

**E-mail address:** `zdenek@cs.cas.cz`

---

We outline mathematical background to an alternative theory of statistics: Define scalar-valued score functions of continuous random variables, discuss some of their properties, generalize them for parametric families of distributions and show how to use them in parametric estimation.

## Keywords

Score function, Score mean, Score moment method.

---

Let  $Y$  be random variable with distribution  $G$  on the entire real line with location parameter that is, with density  $g(y - \mu)$ . The score

$$U(y - \mu) = \frac{\partial}{\partial \mu} \log g(y - \mu) \quad (1)$$

is undoubtedly the best inference function for location distributions.

The question was how to generalize it for a general  $\theta$ .

Hundred years ago R. A. Fisher had introduced the vector-valued score (Fisher score function) with components consisting of derivatives of  $\log f(x, \theta)$  according components of  $\theta$ . It have led to maximum likelihood (ML) estimates of parameters having minimal possible asymptotic variance, and considered due to this property to be the best ones. This line has been followed last hundred years. It appeared, however, that ML estimates are the best for data generated from the considered model only. The finding have given rise to robust statistics initiated by Huber. Now it is known a large number of scalar-valued, bounded and easily manipulated inference functions, having, unfortunately, often a loose relation to the assumed model.

We noticed that there is an undetected crossroad: a possibility to generalize (1) by another way.

**Acknowledgements:** The work was supported by Institutional Subsidy No. RVO: 67985807.

## References

- [1] Z. Fabián, Z. Score function of distribution and revival of the moment method. *Commun. Stat. Theory Meth.* **45(4)** 2016.
- [2] Z. Fabián. Mean, mode or median? The score mean. *Commun. Stat. Theory Meth.* **50(10)**, 2021.



## Modeling human physiology with special reference to local Hyperthermic treatment of solid tumors

Mukhtar Ahmad Khanday<sup>1</sup>

<sup>1</sup>Department of Mathematics University of Kashmir, Srinagar-190006 Jammu and Kashmir, India

**E-mail address:** *khanday@uok.edu.in*

### Keywords

Mathematical modeling, Partial differential equation, Heat and mass transfer, Human physiology, Malignant tumor.

---

Applied mathematics is playing a pivotal role in all disciplines of natural sciences, humanities and social sciences. The tools and techniques available in mathematics help researchers, scientists and academicians to identify the optimal and best possible solution of a problems arising from surroundings with efficient and meaningful results. The complex nature of human physiology can be handled using appropriate concepts of mathematical formulae and computer software's. The plan of the present talk is to discuss the position and relevance of mathematics in human physiology. The partial differential equations are enormously used to understand the dynamics of heat and mass transfer in biological tissues. A brief and concise information regarding mathematical modeling, partial differential equations, human physiology, heat and mass transfer in living tissues etc. can be discussed in the proposed talk. In addition to this, treatment of malignant tissues through local Hyperthermic approach can be presented.

# On Particle-Size Distribution of Three-dimensional Convex Similar Bodies

Jozef Kiselák<sup>1</sup>

<sup>1</sup>Institute of Mathematics, Pavol Jozef Šafárik University in Košice, Slovak Republic

**E-mail address:** *jozef.kiselak@upjs.sk*

---

In the talk, we will present the solution to an old problem posed by Santaló of determining the size distribution of particles derived from the size distribution of their sections. We give an explicit form of particle-size distributions of convex similar bodies for random planes and random lines, which naturally generalize famous Wicksell's corpuscle problem. We achieve the results by applying the method of model solutions for solving well-known Santaló's integral equations. We give a partial result related to the question of the existence and uniqueness of these solutions. We also emphasize that the original form of a solution to Wicksell's problem is insufficient. We finally illustrate our approach in several examples and possible numerical approaches

## Keywords

Stereology, Particle-size distribution, Integral equation, Mellin transform.

---

Stereology is originally concerned with the determination of three-dimensional structures from two-dimensional or one-dimensional observational data. It provides practical techniques for extracting quantitative information about this structure and is based on fundamental principles of geometry and statistics, see e.g. [1]. In [2] we give an explicit solution of integral equations related to the reconstruction of particles from cross sections in  $\mathbb{R}^3$ . We focus on Santaló's formulation [3], i.e. on the estimation problems concerned with ascertaining the size distribution of similarly shaped convex particles, capable of complete size specification by one size parameter and randomly distributed in a convex opaque field. It also involves a famous original Wicksell's corpuscle problem, described figuratively as the 'tomato salad problem', dealing with the determination of the distribution of spherical particles from planar sections, see [4].

**Acknowledgements:** This work was supported by the grant VEGA MŠ SR 1/0526/20.

## References

- [1] A. Baddeley and E. B. V. Jensen. *Stereology for Statisticians. Chapman & Hall/CRC Monographs on Statistics & Applied Probability*. Taylor & Francis, 2004.
- [2] J. Kiselák and G. Baluchová. On Particle-Size Distribution of Convex Similar Bodies in  $\mathbb{R}^3$ . *J. Math. Imaging Vis.* **63**(1): 108–119, 2021.
- [3] Luis Santaló. *Integral geometry and geometric probability*. With a foreword by Mark Kac. Addison-Wesley Publishing Company, 1. edition, 1976.
- [4] S. D. Wicksell. The corpuscle problem. A mathematical study of a biometric problem. *Biometrika* **17**(1-2): 84–99, 06 1925.

---

# Statistical and Learning Functions

Milan Stehlík<sup>1,2,3,4</sup>

<sup>1</sup>Department of Applied Statistics, Johannes Kepler University in Linz, Austria

<sup>2</sup>Linz Institute of Technology, Johannes Kepler University in Linz, Austria

<sup>3</sup>Instituto de Estadística, Universidad de Valparaíso, Valparaíso, Chile

<sup>4</sup>Universidad Andrés Bello, Facultad de Ingeniería, Viña del Mar, Chile

**E-mail address:** *milan.stehlik@jku.at*

---

During the talk we speak on learning mechanisms of data transformation and aggregation. We also define statistical functions. We also provide transfer functions for neural networks, including SPOCU transfer function. Statistical learning will be also discussed.

## Keywords

Statistical learning, Aggregation, Transfer functions.

---

During the talk we speak on learning mechanisms of data transformation and aggregation. This will be connected to the information theory approach to Machine learning, Neural Computing and Artificial Intelligence, which may provide us with a new perspectives of methodological and applied research for statistical inference and optimal design. Several questions will be addressed, e.g. What is optimal learning of complex data? Can we learn negatively?

Importance of Statistics and Optimal designs will be acknowledged. Some advanced statistical techniques, like from algebraic statistics and nonparametric statistics will be advertised. I will also introduce topological and semi-topological data analysis (TDA and STDA) for evolving complex data systems. Topological aggregation for standard aggregation functions have been studied in [2]. Special transfer function for complex data has been developed in [1]. Also applications to image analysis, finance [3], ecology, machine learning, neural networks, neural computing and neuroscience will be given.

**Acknowledgements:** This work has received funding from FCT - Fundação para a Ciência e a Tecnologia, I.P., under the scope of the projects UIDB/00297/2020 and UIDP/00297/2020 (Center for Mathematics and Applications).

## References

- [1] J. Kiseřák, Y. Lu, J. Švihra, P. Szepe, and M. Stehlík. "SPOCU": scaled polynomial constant unit activation function, *Neural Computing and Applications* **33**: 3385–3401, 2021.
- [2] M. Stehlík. On convergence of topological aggregation functions *Fuzzy Sets and Systems* **287**: 48–56, 2016.
- [3] M. Stehlík, Ch. Helpersdorfer, P. Hermann, J. Šupina, L. M. Grilo, J. P. Maidana, F. Fuders, and S. Stehlíková. Financial and risk modelling with semicontinuous covariances, *Information Sciences*, **394-395**: 246–272, 2017.

## Organized Session 2

### Statistical Methods and Applications

Organizer: Fernanda Figueiredo

# A Comparison of Several Generalized and Extended Hill Estimators

Frederico Caeiro<sup>1,2</sup>, Ivanilda Cabral<sup>2,3</sup> and M. Ivette Gomes<sup>4,5</sup>

<sup>1</sup>Universidade Nova de Lisboa, Portugal

<sup>2</sup>Centro de Matemática e Aplicações (CMA), Portugal

<sup>3</sup>Faculdade de Ciências e Tecnologia, Universidade de Cabo Verde, Cape Verde

<sup>4</sup>Universidade de Lisboa (UL), Portugal

<sup>5</sup>Centro de Estatística e Aplicações (CEAUL), Portugal

**E-mail addresses:** *fac@fct.unl.pt; ivanilda.cabral@docente.unicv.edu.cv; ivette.gomes@fc.ul.pt*

## Keywords

Comparison of estimators, Extreme value index, Mean squared error, Statistics of extremes.

For heavy tails, classical extreme value index estimators, such as the Hill estimator (Hill, 1975), usually have a strong bias. Consequently, new interesting alternative estimators have appeared in the literature. In this work we compare several generalized and extended Hill estimators of the extreme value index. The comparison is performed not only with the Hill, but also with an asymptotically unbiased Hill estimator. The comparison study is performed asymptotically, under a third-order framework, and for finite samples, through a Monte Carlo simulation study.

**Acknowledgements:** Research supported by National Funds through FCT-Fundação para a Ciência e a Tecnologia, projects UIDB/00006/2020 (CEA/UL) and UIDB/00297/2020 (CMA/UNL) and by Fundação Calouste Gulbenkian.

## References

- [1] J. Beirlant, P. Vynckier and J. Teugels. Excess functions and estimation of the extreme-value. *Bernoulli* **2**(4): 293–318, 1996. <https://doi.org/10.2307/3318416>
- [2] I. Cabral, F. Caeiro and M. I. Gomes. On the Comparison of Several Classical Estimators of the Extreme Value Index. *Communications in Statistics - Theory and Methods* **51**(1): 179–196, 2022. <https://doi.org/10.1080/03610926.2020.1746970>
- [3] M. I. Fraga Alves, M. I., Gomes, L. de Haan and C. Neves. The mixed moment estimator and location invariant alternatives. *Extremes* **12**: 149–185, 2009. <https://doi.org/10.1007/s10687-008-0073-3>
- [4] H. Penalva, M. I. Gomes, F. Caeiro and M. M. Neves. A couple of non Reduced Bias Generalized Means in Extreme Value Theory: an Asymptotic Comparison. *Revstat-Statistical Journal* **18**(3): 281–298, 2020. <https://doi.org/10.57805/revstat.v18i3.301>

## Analysing the indicators of Europe 2020 agenda using the STATIS methodology

Adelaide Figueiredo<sup>1,2</sup> and Fernanda Otília Figueiredo<sup>1,3</sup>

<sup>1</sup>Universidade do Porto, Faculdade de Economia, Portugal

<sup>2</sup>LIAAD - INESC-TEC, Universidade do Porto, Portugal

<sup>3</sup>CEAUL, Centro de Estatística e Aplicações da Universidade de Lisboa, Portugal

**E-mail addresses:** *adelaide@fep.up.pt; otilia@fep.up.pt*

---

In this study we consider the headline indicators of Europe 2020 agenda for the EU countries in recent years. The indicators belong to the following areas: employment, education, research & development (R&D) and innovation, poverty and social exclusion and climate change and energy. Our aim is to study the evolution of the EU countries and of the indicators along the years. For analysing the data we applied a three-way data method: the STATIS methodology.

### Keywords

Europe 2020 strategy, EU countries, Principal component analysis and three-way data.

---

The Europe 2020 strategy was adopted by the European Council in 2010. It points out three key priorities to strengthen the EU economy: smart growth, sustainable growth and inclusive growth (see [1]). First, a smart growth can be achieved with the development of an economy based on knowledge, research and innovation. Second, a sustainable growth is reached with the reduction of greenhouse gas emissions, increase in energy efficiency and increasing the share of renewable energy in final energy consumption. Third, a inclusive growth is obtained with policies to fostering job creation and poverty reduction. Based on these three priorities, the EU adopted targets for the following indicators: employment rate; early leavers from education and training rate; population, aged 30 to 34, with higher education; expenditure on R&D as % of GDP; population at risk of poverty; greenhouse gas emissions; renewable energy consumption; primary energy consumption and final energy consumption. These indicators are strongly related as referred in [1]. For example, higher levels of education are associated with better jobs, and increasing the employment rate helps to reduce poverty. A greater investment in R&D and innovation in the economy would improve competitiveness and foster job creation.

We collected data from Pordata of these indicators for the EU countries in several years of the period 2010-2019. As the indicators are correlated, we applied a multivariate data method: the Statist methodology (see [2]) for analysing these data. Then, we identified which countries are more similar or distant regarding all the indicators and we analysed the evolution of the EU countries and of the indicators along the period of time 2010-2019.

**Acknowledgements:** This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within the following projects: LA/P/0063/2020 (INESC TEC) and UIDB/00006/2020 (CEAUL).

## References

- [1] Eurostat. *Smarter, greener, more inclusive? Indicators to support the Europe 2020 Strategy*. European Union, 2018.
- [2] C. Lavit, Y. Escoufier, R. Sabatier and P. Traissac. The act (statis method). *Comput. Statist. and Data Analysis* **18**: 87–119, 1994.

---

## Digital technologies use and training: a multivariate data analysis

Fernanda Otília Figueiredo<sup>1,2</sup> and Adelaide Figueiredo<sup>1,3</sup>

<sup>1</sup>Universidade do Porto, Faculdade de Economia, Portugal

<sup>2</sup>CEAUL, Centro de Estatística e Aplicações da Universidade de Lisboa, Portugal

<sup>3</sup>LIAAD - INESC-TEC, Universidade do Porto, Portugal

**E-mail addresses:** *otilia@fep.up.pt; adelaide@fep.up.pt*

---

This study aims to compare the European countries in which concerns to Information and Communications Technology education and the use of digital in people's daily lives. Several indicators associated with digital skills, training and usage were selected to implement this study, and the corresponding data to the last years were collected from Eurostat database. To perform the statistical multivariate data analysis we applied a Double Principal Component Analysis.

### Keywords

Digital, DPCA, ICT education, Multivariate analysis.

---

Digital literacy and Information and Communications Technology (ICT) education are hot issues nowadays, not only at schools but also for the society in general. At schools, ICT education can have a great impact on student learning, when teachers are digitally literate and understand how to integrate it into curriculum. In the era of Industry 4.0, digitalization have transformed the society and the economy, with a great impact on everyday life. With the increased application of digital technologies into a broad range of economic sectors such as manufacturing, retail, transport, energy, finance, healthcare and education, the type of skills that are required by employers across the EU are changing. The ICT sector itself has had a rapid growth in the demand for ICT specialists.

The main objective of this study is to understand how the different countries of the European Union are prepared to deal with the new challenges of the digital era and ICT education. In order to analyze the evolution of several indicators associated with digital skills, training and use of internet, we collected data from Eurostat database, along of different years. We have several data tables, one for each of the years of the study, that must be analyzed jointly. Thus, we applied a Double Principal Component Analysis (DPCA), a method of multivariate data analysis introduced in Bouroche [1] to analyze three-way data with quantitative variables.

Some general conclusions can be taken from the results of this study. Some results confirm what our intuition was, but others are interesting findings. For instance, the countries of the north and central of Europe have already financed a lot of money for providing training in ICT skills to their populations, and everyone feels comfortable in using the Internet for communicating, purchase, selling, finding things, learning, working and so on. On the contrary, the countries of the East and South of Europe have made little investment in ICT education and, in most of them, a huge part of the population did not have digital skills, or even has never used the Internet. The same scenario is observed in an enterprise environment: the enterprises have a small percentage of employed persons with ICT education, but even though, they do not provide training to their personnel to develop such skills.



**Acknowledgements:** This work has received funding from National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia under the projects UIDB/00006/2020 (CEAUL) and LA/P/0063/2020 (INESC TEC).

## References

- [1] J. M. Bouroche. Analyse des données ternaires: la double analyse en composantes principales. Thèse de 3ème cycle. Université de Paris VI, 1975.

## Lehmer's mean-of-order-p in the estimation of the Weibull tail coefficient

Lígia Henriques-Rodrigues<sup>1,2</sup>, Frederico Caeiro<sup>3,4</sup> and  
M. Ivette Gomes<sup>5,6</sup>

<sup>1</sup>Universidade de Évora, Portugal

<sup>2</sup>Centro de Investigação em Matemática e Aplicações (CIMA), Portugal

<sup>3</sup>Universidade Nova de Lisboa, Portugal

<sup>4</sup>Centro de Matemática e Aplicações (CMA), Portugal

<sup>5</sup>Universidade de Lisboa (UL), Portugal

<sup>6</sup>Centro de Estatística e Aplicações (CEAUL), Portugal

**E-mail addresses:** *ligiahr@uevora.pt; fac@fct.unl.pt; ivette.gomes@fc.ul.pt*

The Lehmer's mean-of-order-p ( $L_p$ ), generalizes de arithmetic mean and has enabled amendments in the estimation of a positive extreme value index (EVI), the main parameter in a Pareto type-tail. Most estimators of the Weibull tail coefficient (WTC) are strongly related to the Hill estimator, the classical EVI-estimator. We will use the  $L_p$  EVI-estimators and study the improvements on the WTC estimation.

### Keywords

Weibull tail coefficient, Generalized means, Semi-parametric estimation.

The *Weibull tail-coefficient* (WTC) is a relevant parameter in statistics of extremes when dealing with Weibull type tails. Among the several models that have this type of tails we mention the normal, the gamma, the Weibull, and the logistic distributions. The WTC is the parameter  $\theta$  in a right-tail function of the type

$$\bar{F}(x) = 1 - F(x) = e^{-\mathcal{H}(x)}, \quad \text{with} \quad \mathcal{H} \in \mathcal{R}_{1/\theta}, \quad \theta \in \mathbb{R}^+,$$

with  $\mathcal{H}(x) := -\ln(1 - F(x))$ , a regularly varying cumulative hazard function with index  $1/\theta$ . Among the several classes of WTC-estimators available in the literature we will consider the class introduced in [1] and [4],

$$\hat{\theta}^G(k) := \frac{H(k)}{\frac{1}{k} \sum_{i=1}^k \ln \ln((n+1)/i) - \ln \ln((n+1)/(k+1))}, \quad (2)$$

and the Hill type WTC-estimator introduced in [3]

$$\hat{\theta}^{GG}(k) := \ln(n/k)H(k), \quad (3)$$

with  $H(k)$  the classical Hill (H) estimator ([5]) of a positive extreme value index (EVI),  $\xi$ ,

$$H(k) := \frac{1}{k} \sum_{i=1}^k V_{ik}, \quad V_{ik} := \ln \frac{X_{n-i+1:n}}{X_{n-k:n}}, \quad 1 \leq i \leq k < n, \quad k = 1, 2, \dots, n-1. \quad (4)$$

The Lehmer's mean-of-order-p ( $L_p$ ) EVI-estimators were introduced and studied, under a second order framework, in [6,?,8], being defined as

$$L_p(k) := \frac{M_{k,n}^{(p)}}{p M_{k,n}^{(p-1)}} \quad [L_1(k) \equiv H(k)], \quad (5)$$

consistent for  $\xi > 0$  and real  $p > 0$ , and where  $M_{k,n}^{(p)}$  is the  $p$ -moment of the log-excesses,  $V_{ik}$ , in (4),  $M_{k,n}^{(p)} := \frac{1}{k} \sum_{i=1}^k V_{ik}^p$ ,  $p \geq 1$ .

Generalized means have been successfully used in the estimation of  $\xi$  allowing, for adequate values of the threshold, a reduction in both bias and root mean square error. These facts led us to use them in the estimation of WTC, thence, generalizations of the estimators in (2) and (3) based on the  $L_p$  EVI-estimators, in (5), will be considered and studied under a second order setup (see also ([2]), where a generalization of the estimator in (3) was used in the estimation of the WTC). Their performance for finite samples will be assessed through a small scale Monte-Carlo simulation study.

**Acknowledgements:** Research partially supported by National Funds through FCT - Fundação para a Ciência e a Tecnologia, projects UIDB/MAT/04674/2020 (CIMA), UIDB/MAT/0297/2020 (CMA/UNL) and UIDB/MAT/0006/2020 (CEA/UL).

## References

- [1] J. Beirlant, J. L. Teugels and P. Vynckier. *Practical Analysis of Extreme Values*. Leuven University Press, 1996.
- [2] F. Caeiro, L. Henriques-Rodrigues and M. I. Gomes. The Use of Generalized Means in the Estimation of the Weibull Tail Coefficient. *Computational and Mathematical Methods* **2022**, Article ID 7290822, 12 pages, 2022.
- [3] L. Gardes and S. Girard. Comparison of Weibull tail-coefficient estimators. *Revstat-Statistical Journal* **4(2)**: 163–188, 2006.
- [4] S. Girard. A Hill type estimator of the Weibull tail-coefficient. *Communications in Statistics-Theory and Methods* **33(2)**: 205–234, 2004.
- [5] B. M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* **3**: 1163–1174, 1975.
- [6] H. Penalva, F. Caeiro, M. I. Gomes and M. M. Neves. *An Efficient Naive Generalisation of the Hill Estimator: Discrepancy between Asymptotic and Finite Sample Behaviour*. Notas e Comunicações CEAUL 02/2016.
- [7] H. Penalva, M. I. Gomes, F. Caeiro and M. M. Neves. A couple of non reduced bias generalized means in extreme value theory: an asymptotic comparison. *Revstat-Statist. J.* **18(3)**: 281–298, 2020a.
- [8] H. Penalva, M. I. Gomes, F. Caeiro and M. M. Neves. Lehmer’s mean-of-order- $p$  extreme value index estimation: a simulation study and applications. *J. Applied Statistics* **47**: 13–15 (*Advances in Computational Data Analysis*), 2825–2845, 2020b.

## Organized Session 3

### Contributions of Statistics for professional and personal decision making

Organizer: Carla Santos

## Multivariate collective risk models and a special case

Ana Cantarinha<sup>1</sup>, Elsa Moreira<sup>2,3</sup> and João T. Mexia<sup>2,3</sup>

<sup>1</sup>Polytechnic Institute of Beja, Portugal

<sup>2</sup>Nova University of Lisbon, Portugal

<sup>3</sup>Center for Mathematics and Applications - NOVAMATH, Portugal

**E-mail addresses:** *aicantarinha@gmail.com; efnm@fct.unl.pt*

### Keywords

Collective models, Asymptotic distributions, Confidence intervals, Risk theory.

---

Univariate collective models have played an important role in Actuarial Mathematics. The inference about these models is usually made for the totals of claims.

We now present a multivariate version of these models that may be of interest, as a special case, for application in forest fires. The inference in this case is now made for the total burnt area and the number of fires.

### References

- [1] N. L. Bowers, H. U. Gerber, J. C. Hickman, D. A. Jones and C. J. Nesbet. *Actuarial Mathematics*. The Society of Actuaries, 1986.
- [2] A. Cantarinha, E. Moreira, M. Oliveira and J. T. Mexia. A risk model for forest fires based on asymptotic results for multivariate collective models. Single models and structured families of *Communications in Statistics - Theory and Methods*, 2022.
- [3] A. Khuri. *Advanced calculus with applications to statistics*. John Wiley & Sons, 2003.
- [4] S. Marques, J. G. Borges, J. Garcia-Gonzalo, F. Moreira, J. M. B. Carreiras, M. M. Oliveira, A. Cantarinha, B. Botequim and J. M. C. Pereira. Characterization of wild-fires in Portugal. *European Journal of Forest Research* **130(5)**: 775–784, 2011.
- [5] Mexia, J. T. *Variance free models*, Trabalhos de Investigação, Departamento de Matemática, Faculdade de Ciências e Tecnologia. Universidade Nova de Lisboa, 1990.

## Characterization of the dissertations of a master's degree in occupational health and safety

Ana Dias<sup>1,2</sup> and Carla Santos<sup>1,3</sup>

<sup>1</sup>Polytechnic Institute of Beja, Portugal

<sup>2</sup>CBPST: Centro de Boas Práticas em Segurança no Trabalho, Polytechnic Institute of Beja, Portugal

<sup>3</sup>NOVAMATH: Center for Mathematics and Applications -FCT- New University of Lisbon, Portugal

**E-mail addresses:** *ana.dias@ipbeja.pt; carla.santos@ipbeja.pt*

---

With the aim of presenting suggestions for new topics for future students' dissertations within the scope of the Master's degree in Workplace Safety and Hygiene, in order to diversify the topics covered and framing them in the priorities established in the Strategic Framework of the European Commission for Health and Safety at Work for 2021-2027, which aims to improve the health and safety of workers, and in the research priorities in the field of Occupational Safety and Health, presented in the report of the European Agency for Safety and Health at Work (EU-OSHA), a study was developed with the aim of characterizing the 57 dissertations already completed in this master's degree.

### Keywords

EU-OSHA, Master's degree dissertations, OSH.

---

The need to ensure occupational safety and health (OSH) has been recognized for a long time, however scientific research on OSH is a relatively new area that needs expanding. In 2012, the European Agency for Safety and Health at Work (EU-OSHA) produced a report with a view to defining research priorities in the field of OSH for the following years, based on four main themes: "Demographic evolution - sustainable work for longer and healthier working lives", "Globalization and the evolution of working conditions", "Research in the field of OSH in favour of new safe technologies" and, "The new or increasing professional exposure to chemical and biological agents". These themes reflect the global economic, social, and technological challenges the EU faces and establish the link between the OSH research priorities identified in the report and the targets set by the Europe 2020 Strategy. From a perspective more focused on reinforcing the promotion of workers' well-being, the priorities mentioned above should be added to that established in the Strategic Framework of the European Commission for Health and Safety at Work for 2021-2027, which defines the main priorities and actions to improve the health and safety of workers. This framework is based on three main priorities, "Anticipating and managing change in the context of green, digital and demographic transitions"; "Improve the prevention of work-related accidents and illnesses and pursue a Vision Zero approach to work-related deaths" and, "Increase preparedness to respond to current and future health crises". The master's degree in the scientific area of OSH taught at the Polytechnic Institute of Beja, named master's degree in Health and Safety (HS), aims to offer advanced training for the exercise of the profession of Occupational Safety Senior Technician, but also to provide opportunities for applied research related to OSH and access to advanced 3rd cycle studies. As part of a strategy of continuous improvement of the master's degree in HS, and of

diversification of the topics covered in the future students' final dissertations, a descriptive-exploratory study was carried out with the objective of characterizing the dissertations carried out from the beginning of this course, in the academic year of 2015/2016, to the academic year that has now ended, 2021/2022. In this study, the 57 master's degree in HS dissertations contained in the scientific repository of the Polytechnic Institute of Beja and in the master's course coordination file were analysed, and their themes were classified by areas, temporal distribution, relationship with the students' professional activity, and according to their framing in the priorities outlined by EU-OSHA for research in the field of OSH, and compliance with the European Commission's Strategic Framework for Health and Safety at Work.

## References

- [1] Lei n. 42/2012 de 28 de agosto. Diário da República, 1.<sup>a</sup> série N. 166. Lisboa: Assembleia da República.
- [2] EU-OSHA. Prioridades da investigação no domínio da segurança e saúde no trabalho na Europa: 2013–2020 Agência Europeia para a Segurança e Saúde no Trabalho. Luxemburgo. 2014. DOI: 10.2802/92348.
- [3] European Commission (EC). EU Strategic Framework on Health and Safety at Work 2021–2027. In Occupational Safety and Health in a Changing World of Work; European Commission: Brussels, Belgium, 2021.

## A new approach for analyzing multi-environment trials

Cristina Dias<sup>1,4</sup>, Carla Santos<sup>2,4</sup> and João T. Mexia<sup>3,4</sup>

<sup>1</sup>Polytechnic Institute of Portalegre, Portugal

<sup>2</sup>Polytechnic Institute of Beja, Portugal

<sup>3</sup>Faculty of Science and Technology, Department of Mathematics

<sup>4</sup> Center of Mathematics and Its Applications (CMA), New University of Lisbon, Lisbon, Portugal

**E-mail addresses:** *cpsd@ippportalegre.pt; carla.santos@ipbeja.pt; jtm@fct.unl.pt*

Trials in a network of trials are usually carried out using uniform agricultural techniques. Thus, from trial to trial, the main differences will be related to location and year. It should be noted that Calinski et al. (1987a), (1987b), (1995a) and (1995b) admit that these two factors are additive in the model they presented for a network of trials. In this paper, we will consider jointly delineated trial networks for cultivar comparison. It is our goal to extend the technique of Conjoint Regression Analysis (CRA) to encompass networks of trials with incomplete blocks. Trials often use a-designs to compare cultivars, i.e. trials with incomplete blocks. We conclude that the use of  $L_2$  environmental indices, allowed us to build a CRA theoretical foundation, integrating it into statistical inference for normal models. Thus, (i) it was possible to study the action of local and year factors on the environmental indices; (ii) to apply the technique to connected networks of incomplete blocks, thus increasing the number of cultivars to be compared and enabling the application of the technique to networks of  $\alpha$ -designs.

### Keywords

Randomized block designs, Networks, Environmental indices.

**Acknowledgements:** This work is funded by national funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P., under the scope of the project UIDB/00297/2020 (Center for Mathematics and Applications).

### References

- [1] T. Calinski and S. Kageyama. *Block Designs: A Randomization Approach*. Vol. I. Analysis, Lecture Note in Statistics vol. 150, Springer-Verlag, New York, 2000.
- [2] C. Dias, C. Santos, M. Varadinov and J. T. Mexia. ANOVA like analysis for structured families of stochastic matrices, *12th International Conference of Computational Methods in Sciences and Engineering, AIP Conf. Proc.* **1790**: 140005–140009, 2016.
- [3] M. Fonseca. Design of Experiments: An Introduction Based on Linear Models by Max D. Morris, *International Statistical Review*, **79(1)**: 139–140, 2011.
- [4] J. A Nelder. *The analysis of randomized experiments with orthogonal block structure, II. Treatment structure and the general analysis of variance, Proc. Roy. Soc. Ser. A* **283**: 163–178, 1965.



## On estimation in an extended mixed model with balanced design

Carla Santos<sup>1,5</sup>, Cristina Dias<sup>2,5</sup>, Célia Nunes<sup>3</sup> and João T. Mexia<sup>4,5</sup>

<sup>1</sup>Polytechnic Institute of Beja, Portugal

<sup>5</sup>NOVAMATH: Center for Mathematics and Applications -FCT- New University of Lisbon, Portugal

<sup>2</sup>Polytechnic Institute of Portalegre, Portugal

<sup>3</sup>Department of Mathematics and Center of Mathematics and its Applications -FCT- New University of Lisbon, Portugal

<sup>4</sup> Department of Mathematics -FCT- New University of Lisbon, Portugal

**E-mail addresses:** *carla.santos@ipbeja.pt; cpsd@ipportalegre.pt; celian@ubi.pt; jtm@fct.unl.pt*

---

Important results for estimation can be achieved when considering a class of mixed models, named models with orthogonal block structure, which is characterized by having a covariance matrix that is a linear combination of known pairwise orthogonal projection matrices, that add up to the identity matrix. In particular, it is possible to obtain least square estimators that are best linear unbiased estimators in the subclass of models with commutative orthogonal block structure. Adopting an approach that is based on the algebraic structure of the models, we studied the possibility of extending a balanced mixed model, which could lead to a model with commutative orthogonal block structure.

### Keywords

Best linear unbiased estimator, Estimation, Models with commutative orthogonal block structure, Polynomial regression.

---

In areas such as agriculture, biological and medical research, genetics, economics, or industry, the design of experiments is often based on mixed models, taking advantage of their versatility and power. In the framework of the design of experiments in agricultural trials a special class of mixed models emerged, called models with orthogonal block structure (OBS), which is characterized by its covariance matrix being a linear combination of known pairwise orthogonal projection matrices that add up to the identity matrix. OBS allow optimal estimation for variance components of blocks and contrasts of treatments, but it is possible to get Least Square Estimators that are Best Linear Unbiased Estimators if we consider models with commutative orthogonal block structure (COBS), that is a more restrict class of OBS whose covariance matrix commute with the orthogonal projection matrix on the space spanned by the mean vector. In this work we study the possibility of extending a mixed model with balanced design, using an approach based on the algebraic structure of the models, which can constitute a starting point for obtaining COBS from a balanced mixed model by increasing the degree of polynomial regression.

**Acknowledgements:** This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the project UIDB/00297/2020 (Centro de Matemática e Aplicações)

## References

- [1] J. A. Nelder. The analysis of randomized experiments with orthogonal block structure I, Block structure and the null analysis of variance. *Proc. R. Soc. Lond. Ser. A Math., Phys. Eng. Sci.* **283**: 147–162, 1965.
- [2] C. Nunes, C. Santos and J. T. Mexia. Relevant statistics for models with commutative orthogonal block structure and unbiased estimator for variance components, *J. Interdiscip. Math.* **11**: 553–564, 2008.
- [3] C. Santos, C. Nunes and J. T. Mexia OBS, COBS and mixed models associated to commutative Jordan Algebra. *Bulletin of the ISI, LXII, Proceedings of 56th session of the International Statistical Institute, Lisbon* **11**: 3271–3274, 2008.
- [4] C. Santos, C. Nunes, C. Dias and J. T. Mexia Models with commutative orthogonal block structure: a general condition for commutativity. *Journal of Applied Statistics* **47**:13–15, 2421–2430, 2020.

## Organized Session 4

### Computational Data Analysis and Numerical Methods for Predicting the Dynamics of Infectious Diseases

Organizers: Viswanathan Arunachalam and Padmanabhan Seshaiyer

# Physics-Informed Neural Networks for Informed Vaccine Distribution in Heterogeneously Mixed Populations

Alvan Caleb Arulandu<sup>1</sup> and Padmanabhan Seshaiyer<sup>2</sup>

<sup>1</sup>Thomas Jefferson High School for Science and Technology, USA

<sup>2</sup>George Mason University, USA

**E-mail addresses:** *calebarulandu@gmail.com; pseshaiy@gmu.edu*

---

In this work, we propose a data-driven framework for modeling meta-populations containing preferentially-mixed subgroups of different activity levels. Specifically, we use a physics-informed neural networks (PINNs) approach to both estimate the SIR compartmentalized differential equation model and approximate disease parameters from training data. These parameters are then used in a negative gradient approach to calculate an optimal vaccine distribution plan for informed policy decisions.

## Keywords

PINNs, Heterogeneity, Vaccine Distribution, Epidemiology, COVID-19.

---

Accurate numerical and physical models play an important role in modeling the spread of infectious disease as well as informing policy decisions. Vaccination programs rely on the estimation of disease parameters from limited, error-prone reported data. Using physics-informed neural networks (PINNs) [1,4] as universal function approximators of the SIR compartmentalized differential equation model [3], we create a data-driven framework that uses reported data to estimate disease spread and approximate corresponding disease parameters. We apply this to data from a London boarding school, demonstrating the framework's ability to produce accurate disease and parameter estimations despite noisy data. However, real-world populations contain sub-populations, each exhibiting different levels of risk and activity. Thus, we expand our framework to model meta-populations of preferentially-mixed subgroups with various contact rates [4]. Optimal parameters are estimated through PINNs which are then used in a negative gradient approach to calculate an optimal vaccine distribution plan for informed policy decisions. Together, our work creates a data-driven tool for future infectious disease vaccination efforts in heterogeneously mixed populations.

## References

- [1] F. Brauer, C. Castillo-Chavez and Z. Feng. *Mathematical Models in Epidemiology*. Springer, NY., 2019. <https://doi.org/10.1007/978-1-4939-9828-9>
- [2] Z. Feng, A. N. Hill, P. J. Smith and J. W. Glasser. An elaboration of theory about preventing outbreaks in homogeneous populations to include heterogeneity or preferential mixing. *Journal of Theoretical Biology* **386**: 177–187, 2015. <https://doi.org/10.1016/j.jtbi.2015.09.006>

- [3] M. Raissi, P. Perdikaris and G. E. Karniadakis. Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations. *Journal of Computational Physics* **378**: 686–707, 2019. <https://doi.org/10.1016/j.jcp.2018.10.045>
- [4] S. Shaier, M. Raissi and P. Seshaiyer. Data-driven approaches for predicting spread of infectious diseases through DINNs: Disease Informed Neural Networks. *Letters in Biomathematics* **9(1)**: 71–105, 2022.

---

# Using Physics-Informed Neural Networks to Model and Simulate the Dynamics of Infectious Diseases in Enclosed Environments

Sohini Banerjee<sup>1</sup>, Long Nguyen<sup>2</sup>, Rohan Singh<sup>3</sup> and Padmanabhan Seshaiyer<sup>2</sup>

<sup>1</sup>University of Chicago, Chicago, IL

<sup>2</sup>George Mason University, Fairfax, VA

<sup>3</sup>Birla Institute of Technology and Science, Pilani, India

**E-mail addresses:** *sohinib@uchicago.edu; lnguye33@gmu.edu; f20180102@pilani.bits-pilani.ac.in; pshaiyer@gmu.edu*

---

Infectious diseases often have high rates of spread in enclosed spaces. In this study, we modeled the dynamics of an infectious disease such COVID in an enclosed space through a compartmental model that combines an epidemiology model coupled with scalar transport and computational fluid dynamics with the effects of the environment. We used physics-informed neural networks (PINNs) as tool for implementing the proposed model both as a forward and inverse methodology applied to benchmark problems.

## Keywords

Compartmental Model, Enclosed Spaces, Neural Networks.

---

Infectious diseases often have high rates of spread in enclosed spaces. Scientists are creating mathematical models to represent the transmission of the disease to better understand how to prevent individuals from being infected. In this study, we modeled the dynamics of an infectious disease such COVID in an enclosed space through a compartmental model [3,3] that combines an epidemiology model coupled with scalar transport and computational fluid dynamics with the effects of the environment. We used physics-informed neural networks (PINNs) as tool for doing analysis on the model [1,4]. Specifically, we modeled the dynamics of an infectious disease such COVID in an enclosed space through a compartmental model consisting of five sub-populations including Susceptible, Exposed, Infected, Recovered for the human population and the Concentration of the contaminant. We used PINNs as an inverse parameter estimation technique to identify optimal transmission and recovery rates in this model both in the absence and presence of contaminants. We studied the impact of the peak number of infected individuals as a function of increased concentration effects. We also calculated the basic reproduction number for the coupled differential equation model that helps us to provide insights into the spread of the epidemic. Our computations suggest that the model created and the PINNs approach employed can be reliable candidates for understanding spread of diseases in enclosed spaces.

**Acknowledgements:** This work is partially supported by the National Science Foundation DMS 2031029.

## References

- [1] F. Brauer, C. Castillo-Chavez and Z. Feng. *Mathematical Models in Epidemiology*. Springer, NY., 2019 <https://doi.org/10.1007/978-1-4939-9828-9>
- [2] L. Nguyen, M. Raissi and P. Seshaiyer. Modeling, Analysis and Physics Informed Neural Network approaches for studying the dynamics of COVID-19 involving human-human and human-pathogen interaction. *Computational and Mathematical Biophysics* **10(1)**: 1–17, 2022.
- [3] M. Raissi, P. Perdikaris, G. E. Karniadakis. Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Non-linear Partial Differential Equations. *Journal of Computational Physics* **378**: 686–707, 2019 <https://doi.org/10.1016/j.jcp.2018.10.045>
- [4] M. Raissi, N. Ramezani, and P. Seshaiyer. On parameter estimation approaches for predicting disease transmission through optimization, deep learning and statistical inference methods. *Letters in biomathematics* **6(2)**: 1–26, 2019.

# A Review and Application of Disease Informed Neural Network

Alonso Gabriel Ogueda<sup>1</sup>, Erika Johanna Martínez Salinas<sup>2</sup>,  
Viswanathan Arunachalam<sup>2</sup> and Padmanabhan Seshaiyer<sup>1</sup>

<sup>1</sup>George Mason University, USA.

<sup>2</sup>Universidad Nacional de Colombia, Bogotá Colombia.

**E-mail addresses:** *aogueda@gmu.edu; ejmartinezsa@unal.edu.co; varunachalam@unal.edu.co; pshaiy@gmu.edu*

## Keywords

Compartmental Models, Epidemiology, Neural Networks, Transport, Deep Learning.

---

In this work, we present an application of Disease Informed Neural Networks [1] through the use of deep learning and its application to real data. We show how these approaches are capable of predicting the behavior of a disease described by modified compartmental models [4] that include parameters and variables associated with the movement of the population between neighboring cities. We show that our model validates real-data and also how such Physics Informed Neural Network based methods [4]- [2] predicts optimal parameters for a given dataset.

## References

- [1] S. Cuomo, V. S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi and F. Piccialli. Scientific Machine Learning through Physics-Informed Neural Networks: Where we are and What's next. arXiv preprint arXiv:2201.05624, 2022.
- [2] H. Gao, M. J. Zahr and J. X. Wang. Physics-informed graph neural Galerkin networks: A unified framework for solving PDE-governed forward and *inverse problems*. *Computer Methods in Applied Mechanics and Engineering* **390**: 114502, 2022.
- [3] M. Raissi, N. Ramezani and P. Seshaiyer. On parameter estimation approaches for predicting disease transmission through optimization, deep learning and statistical inference methods. *Letters in biomathematics* **6(2)**: 1–26, 2019.
- [4] S. Shaier, M. Raissi and P. Seshaiyer. Data-driven approaches for predicting spread of infectious diseases through DINNs: Disease Informed Neural Networks. *Letters in Biomathematics* **9(1)**: 71–105, 2022.



## A stochastic model for epidemics with a diagnostic test for the spread of the SARS-Cov-2 virus: A case study for the Nariñense region in Colombia

Andres Felipe Rincón-Prieto<sup>1</sup>, Andres Ríos-Gutiérrez<sup>1</sup> and  
Viswanathan Arunachalam<sup>1</sup>

<sup>1</sup>Department of Statistics, Universidad Nacional de Colombia, Bogotá, Colombia.

**E-mail addresses:** *varunachalam@unal.edu.co; afrinconp@unal.edu.co; asriosg@unal.edu.co*

### Keywords

Stochastic perturbations, False negative, Covid-19.

---

SARS-Cov-2 is a virus that has spread worldwide and generated a global health crisis. In this work, we present the modified SEIR model with a diagnostic test to describe the dynamics of the spread of COVID-19 in the Pacific Nariñense region in Colombia based on the data registered by the National Institute of Health (INS)[1,2]. We introduce the stochastic perturbations in the transmission rates to describe the behavior of the proposed model considering the susceptible, exposed, false negative tests that have undergone the diagnostic test, untested, tested positive, recovered, and deceased. The statistical data were analyzed at two different times: before and after vaccination. We also include the effect of vaccination as a control measure and compare the basic reproduction number for the proposed model with and without vaccination.

**Acknowledgements:** This work has received funding from Directorate-Bogotá (DIB), Universidad Nacional de Colombia under the project No. 51145

### References

- [1] Instituto Nacional de Salud de Colombia. Coronavirus (COVID-19) en Colombia. <https://www.ins.gov.co/Noticias/Paginas/Coronaviruss.aspx>.
- [2] Niño-Torres, David, Andrés Ríos-Gutiérrez, Viswanathan Arunachalam, Comfort Ohajunwa, and Padmanabhan Seshaiyer. "Stochastic modeling, analysis, and simulation of the COVID-19 pandemic with explicit behavioral changes in Bogotá: A case study." *Infectious Disease Modelling* 7 **no. 1**: 199–211, 2022.

## Organized Session 5

### Statistical Modeling and Applications

Organizer: José A. Pereira

## Estimation in Gamma distributed Bi-additive models

Patrícia Antunes<sup>1</sup>, Sandra S. Ferreira<sup>1,2</sup> and Dário Ferreira<sup>1,2</sup>

<sup>1</sup>Center of Mathematics and Applications, University of Beira Interior, Covilhã, Portugal

<sup>2</sup>Department of Mathematics University of Beira Interior, Covilhã, Portugal

**E-mail addresses:** *antunes.patricia03@gmail.com; sandraf@ubi.pt; dario@ubi.pt*

---

In our work, we present an application in the field of mixed models, where we introduce bi-additive models. For that, we consider the case in which the components of the vector of the random part of the models are distributed as Gamma distribution. We simulate the vectors from the random part and their estimates achieve good precision.

### Keywords

Cumulants, Gamma distribution, Mixed models.

---

### References

- [1] P. Antunes, S. S. Ferreira, D. Ferreira, C. Nunes and J. T. Mexia. Estimation in additive models and ANOVA-like applications, *Journal of Applied Statistics* **47**(13–15): 2374–2383, 2020.
- [2] N. Balakrishnan, N. L. Johnson and S. Kotz. A note on relationships between moments, central moments and cumulants from multivariate distributions. *Statistics & Probability Letters* **39**(1): 49–54, Elsevier, 1998.
- [3] E. Demidenko. *Mixed Models: Theory and Applications*. New Jersey: John Wiley & Sons, 2004.
- [4] N. L. Johnson, S. Kotz and N. Balakrishnan. *Continuous Univariate Distributions* 2<sup>nd</sup> edition, Vol. 1. New Jersey: John Wiley, 1994
- [5] I. A. Koutrouvelis, G. C. Canavos and A. G. Kallioras. Cumulants Plots for Assessing the Gamma Distribution. *Communications in Statistics-Theory and Methods* **39**(4): 626–641, 2010.
- [6] E. L. Lehmann. *Testing Statistical Hypotheses*, 2<sup>nd</sup> edition. New York: John Wiley, 1986.
- [7] Q. Zheng. Computing relations between moments and cumulants. *Computational Statistics* **17**(4): 507–515, 2002.

## A novel test for grouping levels of a factor: An application to cocoa breeding experiments in acidic soils

Kwaku Opoku-Ameyaw<sup>1,2</sup>, Célia Nunes<sup>1,3</sup>, Manuel L. Esquível<sup>4,5</sup> and João T. Mexia<sup>5</sup>

<sup>1</sup>Center of Mathematics and Applications, University of Beira Interior, Covilhã, Portugal

<sup>2</sup>Cocoa Research Institute of Ghana, Ghana

<sup>3</sup>Mathematics Department, University of Beira Interior, Covilhã, Portugal

<sup>4</sup>Mathematics Department, Faculty of Sciences and Technology, Nova University of Lisbon, Portugal

<sup>5</sup>Center of Mathematics and its Applications, Faculty of Science and Technology, Nova University of Lisbon, Portugal

**E-mail addresses:** *koameyawjnr@yahoo.com; celian@ubi.pt; mle@fct.unl.pt; jtm@fct.unl.pt*

### Keywords

Nonparametric test, Grouping of factor levels, Edgeworth series expansion, Hypothesis intersection, Cocoa breeding database.

---

The complex nature of some biological datasets in the field of plant breeding has brought about the need to move to nonparametric tests since they hold under more general conditions than parametric tests (they do not rely on assumptions about the population's distribution from which the sample is drawn). In this work, a new nonparametric test is proposed for grouping factor levels. The proposed test is applied to a cocoa breeding experiment in Ghana, in which twelve cocoa varieties are compared. The varieties were mainly from three groups of varieties, two of which had a common ascendent, and we intended to test if this grouping is significant. The cocoa breeding experiment was carried out in four types of acidic soils, so resistance to low pH was an important goal of the plant breeding. The obtained results suggest that the grouping was significant for the most acidic soil.

**Acknowledgements:** This work was partially supported by national funds of FCT Foundation for Science and Technology under UIDB/00212/2020 and UIDB/00297/2020.

### References

- [1] Y. Ahenkorah. Influence of environment on growth and production of the cacao tree: soils and nutrition. Actes, Douala, Cameroun, 4–12 Nov 1979/7 Conference internationale sur la recherche cacaoyere (Proceedings, Douala, Cameroun, 4-12 Nov 1979/7 International Cocoa Research Conference). Lagos, Nigeria: Secretary General, Cocoa Producers' Alliance, 1981.
- [2] W. J. Conover, *Practical Nonparametric Statistics*, 3rd Edition, Wiley Series in Probability and Statistics, John Wiley & Sons, 1999.
- [3] D. E. Kornbrot. The rank difference test: A new and meaningful alternative to the Wilcoxon signed ranks test for ordinal data. *British Journal of Mathematical and Statistical Psychology* **43**(2): 241–264, 1990.

- [4] S. M. Mortazavian and S. Azizi-Nia, Nonparametric stability analysis in multienvironment trial of canola. *Turkish Journal of Field Crops* **19(1)**: 108–117, 2014.
- [5] F. S. Nahm. Nonparametric statistical tests for the continuous data: the basic concept and the practical use. *Korean Journal of Anesthesiology* **69(1)**: 8–14, 2016.
- [6] A. Stuart and J. K. Ord. *Kendall's Advanced Theory of Statistics*, Vol. 1, Distribution Theory, 6th Edition, John Wiley & Sons, Ltd, 1994.

# Estimation of the Best Interval Between Radiography in Early Bone Loss Detection Using Logistic Regression Classifiers

José Pereira<sup>1,2,3</sup>, Inês Oliveira<sup>1</sup>, Luzia Mendes<sup>1</sup> and Teresa Oliveira<sup>2,3</sup>

<sup>1</sup>Faculdade de Medicina Dentária, Universidade do Porto, Portugal

<sup>2</sup>Universidade Aberta, Portugal

<sup>3</sup>CEAUL | Centro de Estatística e Aplicações da Universidade de Lisboa, Portugal

**E-mail addresses:** *lobopereiramail@gmail.com; inesoliveira199624@gmail.com; luziamgoncalves@gmail.com; tpcaoliveira@gmail.com*

---

The early detection of bone loss progression is done by comparing two consecutive radiography and is of paramount importance to control periodontal disease progression. In this work we propose a supervised classification method through logistic regression models to address the estimation of the best interval between two radiography.

## Keywords

Logistic regression, Classifier and Periodontitis.

---

Periodontal disease (PD) is an infectious/inflammatory disease that affects the tooth supporting tissues and is characterized by alveolar bone loss. Being one of the most prevalent oral disorders, affecting about 20-50% of global population, [1] periodontitis, the destructive form of periodontal disease, is the leading cause of tooth loss among adults. The early diagnose of periodontitis and a good estimation of its progression rate allows an accurate diagnosis and prognosis and, consequently, an adequate treatment plan. The assessment of periodontal bone level changes is of paramount importance to estimate the rate of periodontitis progression, and therefore provide crucial information to achieve a more accurate prognosis.

Panoramic radiographs (PR) are part of the patient care protocol at the FMDUP clinic and the first radiological medium for periodontal bone level screening. The interval between two PR needs to be small enough to detect early changes of bone level, however must minimize the amount of radiation received.

Aiming to know the minimum interval between PR that allows detecting variations of bone level, we performed an observational study with convenience sample of 200 patients of the FMDUP clinic (400 PR, 2 per patient). The variation of interproximal bone level between two consecutive PR was assessed through a digital percentile ruler graded at 5% of tooth root and a binary classifier logistic regression was applied with R software [2]. Binomial generalized linear models were fitted to data, with the predictive variables age, gender, and interval between PR.

The results showed that the minimum interval between PR was 2.5 years with a sensitivity and specificity of 46.0% and 85.1%, respectively, and the area under the curve (AUC) was 69.6.

## References

- [1] M. Sanz, F. D'Aiuto, J. Deanfield and F. Fernandez-Avilésäk. European workshop in periodontal health and cardiovascular disease-scientific evidence on the association between periodontal and cardiovascular diseases: A review of the literature. *Eur Heart J Suppl B*:**B3-12**, 2010.
- [2] R Core Team. *A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. URL <https://www.R-project.org/>, 2019.

## Organized Session 6

### Mathematical models

Organizer: *Silvia Barbeiro*



---

## Mathematical model to reconstruct the mechanical properties of an elastic medium

Rafael Henriques<sup>1</sup>, José Santos<sup>1</sup> and Sílvia Barbeiro<sup>1</sup>

<sup>1</sup>CMUC, Department of Mathematics, University of Coimbra, Portugal

**E-mail addresses:** *rafael.henriques97@gmail.com; zeluis@mat.uc.pt; silvia@mat.uc.pt*

---

In this talk, a mathematical model to reconstruct the mechanical properties of an elastic medium is presented, in view of contributing to the resolution of the problem of optical coherence elastography [1].

The mathematical model for the mechanical deformations is based on the on time-harmonic equations of linear elasticity in multilayer domains. The numerical solution is obtained through finite elements in a three-dimensional domain. The performance of the method is illustrated with numerical examples.

The mathematical model for solving this direct problem is the computational basis for addressing the inverse problem of determining the set of parameters that characterize the mechanical properties of the medium, knowing the field of displacements for a given excitation. We consider different optimization methods to solve the inverse problem and discuss their performance. We report several computational results which illustrate their behavior in terms of accuracy and efficiency. We will give some lights for future work.

### Keywords

Optical Coherence Elastography, Linear Elasticity Equation, Finite Element Method, Inverse Problem.

---

**Acknowledgements:** This work was supported by FEDER Funds through the Operational Program for Competitiveness Factors - COMPETE and by Portuguese National Funds through FCT - Foundation for Science and Technology under the PTDC / EMD-EMD / 32162/2017 project “Optical Coherence Elastography, for imaging of the mechanical properties of the retina” and by the Centre for Mathematics of the University of Coimbra - UIDB/00324/2020, funded by the Portuguese Government through FCT/MCTES.

### References

- [1] R. Henriques. Mathematical model to reconstruct the mechanical properties of an elastic medium, master thesis <http://hdl.handle.net/10316/93596>, 2020.

# Wave propagation in viscoelastic materials: analytical and numerical study

Afonso Costa<sup>1</sup>

<sup>1</sup>University of Coimbra, Portugal

**Email addresses:** *uc2017270186@student.uc.pt; afonsofvnc@hotmail.com*

---

Here we present a study of a problem that describe the displacement of a wave in a viscoelastic medium. For the problem are established existence, uniqueness and stability results. Furthermore, a numerical method is proposed, a semi-discrete defined over a non-uniform partition and the fully discrete adding a temporal integration. It is proved that, both methods have quadratic spatial convergence and, in the case of the discrete method, the convergence in time is of 1st order.

## Keywords

Wave, Viscoelastic, Supraconvergence, Kelvin-Voigt, Finit differences method.

---

In this thesis, a boundary and initial conditions problem describing the displacement of a wave in a viscoelastic medium [1] obtained considering that the functional relationship between the stress and the strain of the medium is characterized by the Kelvin-Voigt model [2]. For the problem of initial and boundary conditions, existence results are established considering the method of separation of variables, uniqueness and stability using the energy method. From the numerical point of view, in this dissertation a numerical method belonging to the family of finite difference methods is proposed. The study of this method is done in two distinct phases: in a first phase we study the semi-discrete approximation defined over a non-uniform partition and, afterwards, we study the fully discrete method combining the spatial discretization already considered with a temporal integration defined from a uniform partition. In a proper functional context, it is proved that, for a discrete norm that can be seen as a discretization of the usual norm of  $H^1$ , both methods have quadratic spatial convergence and, in the case of the discrete method, the convergence in time is of 1st order.

**Acknowledgements:** This work was supervised by Doctor José Augusto Ferreira (CMUC).

## References

- [1] J. A. Ferreira, D. Jordão and L. Pinto. Drug delivery enhanced by ultrasound: Mathematical modeling and simulation. *Computers & Mathematics with Applications* **107**: 57–69, 2022.
- [2] W. Zhang, A. Capilnasiu and D. Nordsletten. Comparative analysis of nonlinear viscoelastic models across common biomechanical experiments. *Journal of Elasticity* **145.1**: 117–152, 2021.

## Modelling light propagation in the cornea: treatment of curved boundary domains

Milene Santos<sup>1</sup>, Adérito Araújo<sup>1,2</sup> and Sílvia Barbeiro<sup>1,2</sup>

<sup>1</sup>Department of Mathematics, University of Coimbra, Portugal

<sup>2</sup>CMUC, Department of Mathematics, University of Coimbra, Portugal

**E-mail address:** *milene-santos09@hotmail.com*

### Keywords

The Helmholtz's equation, Discontinuous Galerkin method, Curved boundary domains, Polynomial reconstruction.

To model the incidence and reflection of light in the cornea, we can use the Maxwell's equations, which describe the electromagnetic wave's propagation field. In this work, we focus on the Maxwell's equations in the time-harmonic form which translates in the Helmholtz's equation. We propose a numerical method based on nodal discontinuous Galerkin methods [3] combined with a strategy which is specially designed to deal with curved domains [4] which arise naturally in our domain of interest for the application. The treatment of boundary value problems in curved boundary domains has been a subject of growing interest in the numerical analysis community. The question that arises concerns the reduction of the order of convergence of numerical methods when considering the approximation of the domain by a polygonal mesh. The method proposed in this work, called DG-ROD (Reconstruction for Off-site Data), is based on a polynomial reconstruction of the boundary condition imposed on the computational domain that takes into account the boundary condition imposed on the physical domain. The numerical tests show a reduction of the error and an increase in the order of convergence of the method, in relation to the classical DG method.

**Acknowledgements:** This work has been supported by the Centre for Mathematics of the University of Coimbra - UIDB/00324/2020, funded by the Portuguese Government through FCT/MCTES, and by FCT/MCTES through the project reference PTDC/MAT-APL/28118/2017 and POCI-01-0145-FEDER-028118.

### References

- [1] A. Araújo, S. Barbeiro, R. Bernardes, M. Morgado and S. Sakic. A mathematical model for the corneal transparency problem. *Journal of Mathematics in Industry* **12**, 2022.
- [2] F. Bassi and S. Rebay. High-order accurate discontinuous finite element solution of the 2D Euler equations. *Journal of Computational Physics* **138**, 251–285, 1997.
- [3] R. Costa, S. Clain, R. Loubère and G. J. Machado. Very high-order accurate finite volume scheme on curved boundaries for the two dimensional steady-state convection-diffusion equation with Dirichlet condition. *Applied Mathematical Modelling* **54**: 752–767, 2018.
- [4] R. A. Farrell, D. E. Freund and R. L. Mccaly. Research on corneal structure. *Johns Hopkins APL Technical Digest* **11(1-2)**: 191–199, 1990.
- [5] J. Hesthaven and T. Warburton. *Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications*. Springer-Verlag, New York, 2008.

## Local analysis of Lotka-Volterra type model

Sandra Vaz<sup>1</sup>, Jorge Pinto<sup>2</sup> and Delfim F. M. Torres<sup>3</sup>

<sup>1</sup>CMA-UBI, Department of Mathematics of University of Beira Interior, Portugal

<sup>2</sup>Department of Mathematics of University of Beira Interior, Portugal, Portugal

<sup>3</sup>CIDMA, Department of Mathematics of University of Aveiro, Portugal

**E-mail addresses:** *svaz@ubi.pt; jorge.pinto@ubi.pt; delfim@ua.pt*

### Keywords

Lotka-Volterra Model, Stability, Mickens, Fractional Calculus.

---

In this work it is considered a modified Lotka-Volterra model applied to the predator-prey system but can be applied to the bank system [4]. The well-posedness of the model and dynamical consistency is shown by the Euler numerical scheme and Mickens NSFD scheme [2,3] and Caputo fractional derivative [1]. For the fractional model, we also prove the existence and unicity of solutions. Some numerical solutions are made to present our results.

**Acknowledgements:** The authors were partially supported by the Portuguese Foundation for Science and Technology (FCT): Sandra Vaz through the Center of Mathematics and Applications of *Universidade da Beira Interior* (CMA-UBI), project UIDB/00212/2020; Delfim F. M. Torres through the Center for Research and Development in Mathematics and Applications (CIDMA), project UIDB/04106/2020.

### References

- [1] M. Caputo, Linear Models whose Q is almost Frequency Independent-II, *Geophysical. J. R. astronomer. Soc* **13**: 529–539, 1967.
- [2] R. E. Mickens, Calculation of denominator functions for nonstandard finite difference schemes for differential equations satisfying a positivity condition, *Numer. Methods Partial Differential Equations* **23**: 672–691, 2007.
- [3] R. E. Mickens, NSFD discretization of interacting population models satisfying conservation laws, *Computers and Mathematics with Applications* **66**: 2307–2316, 2013.
- [4] N. Sumarti, A Dynamical System of deposits and loan volumes based on the Lotka - Volterra Model, *AIP Conference Proceedings*, 2004.

## Organized Session 7

### Threshold Behavior in Epidemiological and Health Sciences

Organizer: Swati DebRoy

---

## Identifying the mode of action of drug from viral load data

Swati DebRoy<sup>1</sup>

<sup>1</sup>University of South Carolina Beaufort, USA

**E-mail address** *SDEBROY@uscb.edu*

Mathematical models can be used in different ways to identify mode and action of a drug from viral load decline data upon administration of drug. Here we introduce a physiologically based mechanistic model along with parameter estimation techniques like generation of synthetic patient data. We discuss strategies used to simplify the model to suit limited quantity of time-series data. Lastly, we present the advantage of implementing a population-based parameter estimation approach which is especially useful for sparse data scenarios.

## Is the epidemic growing or declining? Applying the instantaneous reproduction number method to COVID-19 data: case studies in the United States and Canada

Isaac Fung<sup>1</sup>

Georgia Southern University, USA<sup>1</sup>

The effective reproduction number ( $R_t$ ) measures the transmission potential of an infectious disease in a population in the presence of population immunity, behavioral changes, and public health interventions. When  $R_t > 1$ , the epidemic grows; when  $R_t < 1$ , the epidemic declines. The instantaneous reproduction number method developed by Cori et al. (2013), as implemented in the R package EpiEstim, is frequently used to estimate the  $R_t$  of the coronavirus during the COVID-19 pandemic. In this presentation, Dr. Fung will describe with examples, the application of this method to the estimation of COVID-19 transmission potential at the state and local level in the United States. The limitations of public health surveillance data, the challenges of using these data to estimate  $R_t$ , and approaches that may be deployed to address such challenges will also be discussed.

## Multidimensional evidence generation for health technology assessment

Anuj Mubayi<sup>1</sup>

IQVIA, USA<sup>1</sup>

**E-mail address:** *anujmubayi@yahoo.com*

COVID-19 is having a profound impact on healthcare and society. The future of health care not merely will depend on convenience and less expensive technology but also will be multidimensional evidence-based while enhancing patients' personal health and needs in real time. This will require integrated evidence-generation rather siloed functioning of departments, disciplines, markets, and geographies. There will be radical ways to operate that will smoothly connect evidence generating process from randomized clinical trials to patients' medical and non-medical history, health economics and outcomes research, market access, and precision prevention strategies. That is, the paradigm shift incorporating partnership model will simultaneously enhance data engineering platforms, advanced analytics capabilities, modeling-based decision process, and cross-functional collaborations.



## Incorporating human mobility data into epidemiological models

Omar Saucedo<sup>1</sup>

<sup>1</sup>Virginia Tech, USA

In the past decade, human mobility data has become increasingly available with the introduction of smartphone devices. Not only did communication between acquaintances and access to information become easier; smartphones provide clues on the movement patterns of individuals throughout their day. Incorporating mobility data into an epidemiological model can offer valuable insight for implementing control strategies. In this talk, we will present analytical tools for approximating the basic reproduction number for a SIS-SI vector-borne disease network model. We will use cell phone data to estimate the movement patterns of individuals between different regions with the objective of understanding how the network structure influences vector-borne disease dynamics.

## Organized Session 8

### Statistical Modeling and Forecasting

Organizer: A. Manuela Gonçalves

## Short-term forecast by state-space model approach: the case study of maximum temperature

A. Manuela Gonçalves<sup>1</sup>, F. Catarina Pereira<sup>1</sup> and Marco Costa<sup>2,3</sup>

<sup>1</sup>University of Minho, Department of Mathematics and Centre of Mathematics, Portugal

<sup>2</sup>University of Aveiro, Águeda School of Technology and Management

<sup>3</sup>Centre for Research and Development in Mathematics and Applications, Portugal

**E-mail addresses:** *mneves@math.uminho.pt; id9976@alunos.uminho.pt; marco@ua.pt*

---

This work aims to improve the accuracy of short-term forecasts of a meteorological time series (maximum air temperature) obtained from an easily accessible online website. The proposed methodology is based on a state-space representation that incorporates the latent process, the state, which is estimated recursively using the Kalman filter. The proposed model linearly and stochastically relates the forecasts from the website (as a covariate) to the observations of the maximum temperature recorded during the study.

### Keywords

Forecasting, State space models, Kalman filter, Data assimilation, Meteorological time series.

---

**Acknowledgements:** A. Manuela Gonçalves was partially financed by Portuguese Funds through FCT (Fundação para a Ciência e a Tecnologia) within the Projects UIDB/00013/2020 and UIDP/00013/2020 of CMAT-UM. F. Catarina Pereira was financed by national funds through FCT through the individual PhD research grant UI/BD/150967/2021 of CMAT-UM. Marco Costa was partially supported by The Center for Research and Development in Mathematics and Applications (CIDMA) through the Portuguese Foundation for Science and Technology (FCT), references UIDB/04106/2020 and UIDP/04106/2020.

### References

- [1] A. M. Gonçalves and M. Costa. Predicting seasonal and hydrometeorological impact in environmental variables modelling via Kalman filtering. *Stochastic Environmental Research and Risk Assessment*, **27**: 1021–1038, 2013. <https://doi.org/10.1007/s00477-012-0640-7>.
- [2] A. C. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, New York, 2009.
- [3] F. C. Pereira, A. M. Gonçalves and M. Costa. Short-term forecast improvement of maximum temperature by statespace model approach: the study case of the TO CHAIR project. *Stochastic Environmental Research and Risk Assessment*. DOI: 10.1007/s00477-022-02290-3.
- [4] G. Petris, S. Petrone and P. Campagnoli. *Dynamic linear models with R*. Springer, New York, 2009.

## Simulation study of time series with outliers in the context of state space models

F. Catarina Pereira<sup>1</sup>, A. Manuela Gonçalves<sup>2</sup> and Marco Costa<sup>3</sup>

<sup>1</sup>University of Minho and Centre of Mathematics, Campus de Azuré, 4800-058 Guimarães, Portugal

<sup>2</sup>University of Minho, Department of Mathematics and Centre of Mathematics, Campus de Azurém, 4800-058 Guimarães, Portugal

<sup>3</sup>University of Aveiro, ESTGA - Águeda School of Technology and Management and CIDMA - Centre for Research and Development in Mathematics and Applications, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

**E-mail addresses:** *id9976@alunos.uminho.pt; mneves@math.uminho.pt; marco@ua.pt*

---

In time series analysis, the presence of outliers is common. When not treated, these outliers can affect parameter estimation and prediction quality. This study aims to investigate the impact of outliers on the performance of state space modeling. In this paper, several simulation scenarios are tested, considering Gaussian and exponential errors with non-contaminated and contaminated data, as well as contaminated data free from outliers.

### Keywords

Outliers, State space model, Simulation study.

---

State space models are a very flexible class of dynamic models. These models are able to integrate several data features, allowing the analysis of dynamic phenomena that vary significantly over time. State space models have wide applicability, and are typically used to model economic [4] and environmental [2] time series.

In time series analysis, it is common to deal with data with structural changes that occur due to change points and outliers. Most methodologies cannot deal with outliers and failure to deal with them may produce problems related to misspecification of parameter estimates, which consequently affect the efficiency of the forecast model. Even simple linear Gaussian state space models can have estimation problems, especially when the measurement error is greater than the stochasticity of the process [1,3].

Furthermore, the assumption that Kalman filter solutions are optimal, in the sense that they have the minimum mean square error, is only true when the assumption of normality is verified and the model is fully specified. The presence of outliers can affect the assumption of normality, and in most cases the true parameters of the model are not known, and are therefore replaced by estimates, which will also be affected by the existence of outliers.

This paper presents several simulation scenarios, namely the linear state space model with Gaussian/exponential errors with non-contaminated and contaminated data, and the linear state space model with Gaussian/exponential errors with contaminated data free from outliers. For each of the four studies, several scenarios were tested, in which 2,000 samples with valid estimates of size  $n$  ( $n = 50, 200, 500$ ) were simulated. The results obtained are evaluated in terms of the difference between the maximum likelihood estimate and the true value of the parameter and the convergence rate (%), which is given by the number of valid simulated estimates (in this case, 2,000) divided by the number of total simulations.

**Acknowledgements:** F. Catarina Pereira was financed by national funds through FCT (Fundação para a Ciência e a Tecnologia) through the individual PhD research grant UI/BD/150967/2021 of CMAT-UM. A. Manuela Gonçalves was partially financed by Portuguese Funds through FCT (Fundação para a Ciência e a Tecnologia) within the Projects UIDB/00013/2020 and UIDP/00013/2020 of CMAT-UM. Marco Costa was partially supported by The Center for Research and Development in Mathematics and Applications (CIDMA) through the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), references UIDB/04106/2020 and UIDP/04106/2020.

## References

- [1] M. Auger-Méthé, C. Field, C. M. Albertsen, A. E. Derocher, M. A. Lewis, I. D. Jonsen and J. M. Flemming. State-space models' dirty little secrets: even simple linear Gaussian models can have estimation problems. *Scientific Reports* **6**, 26677, 2016. <https://doi.org/10.1038/srep26677>
- [2] F. C. Pereira, A. M. Gonçalves and M. Costa. Short-term forecast improvement of maximum temperature by state-space model approach: the study case of the TO CHAIR project. *Stoch Environ Res Risk Assess*, 2022. <https://doi.org/10.1007/s00477-022-02290-3>
- [3] F. C. Pereira, A. M. Gonçalves and M. Costa. Outliers Impact on Parameter Estimation of Gaussian and Non-Gaussian State Space Models: A Simulation Study. *Eng. Proc* **18(1)**: 31, 2022. <https://doi.org/10.3390/engproc2022018031>
- [4] F. Z. C. Rose, M. T. Ismail and M. H. Tumin. Outliers detection in state-space model using indicator saturation approach. *Indonesian Journal of Electrical Engineering and Computer Science* **22(3)**, pp. 1688–1696, 2021. <http://doi.org/10.11591/ijeecs.v22.i3.pp1688-1696>

## Projections of amputations in diabetics in Portugal with count regression model

Elisabete Carolino<sup>1,2,3</sup>, M. Rosário Ramos<sup>4,5</sup> and José Pedro Matos<sup>3</sup>

<sup>1</sup>H&TRC - Health & Technology Research Center, ESTeSL-Escola Superior de Tecnologia da Saúde, Instituto Politécnico de Lisboa, Portugal

<sup>2</sup>ISAMB - Instituto de Saúde Ambiental, Faculdade de Medicina da Universidade de Lisboa, Portugal

<sup>3</sup>ESTeSL - Escola Superior de Tecnologia da Saúde, Instituto Politécnico de Lisboa, Portugal

<sup>4</sup>Universidade Aberta and CEG (Centro de Estudos Globais-UAb)

<sup>5</sup>CEAUL, Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Portugal

**E-mail addresses:** *etcarolino@estestl.ipl.pt; mariar.amos@uab.pt; fulgenciomatos@estestl.ipl.pt*

---

Establishing the incidence and estimate the future trends in limb loss is important for health care planning and for care of diabetic patients in particular. This study explores two models of Poisson regression family to predict the number of amputations in the diabetic population of Portugal, mainland, stratified by sex and age group. Both models point to an increase in the amputations incidence. Some limitations and further research are discussed.

### Keywords

Portuguese diabetic population, Amputations, Poisson Regression

---

### Introduction

Diabetes is known as one of the leading causes of morbidity and mortality worldwide. Portugal is the fourth country in the European Union with the highest incidence rate of diabetes, according to the IDF Atlas (International Diabetes Federation, 2021). The person with diabetes has an increased risk for related health problems leading eventually to the need of lower limb amputation, such as finger, foot or leg amputation. This causes functional limitation for the individual, and have a general negative socio-economic impact for families and governments.

Projections of the future limb loss are important for predicting future needs for health care services and help the public health programs to reduce the future burden of the disease. It is estimated that about 50 % of amputations and ulcerations can be prevented by evaluating the foot, degree of risk of ulceration, thus allowing the implementation of preventive measures. Studies on the projection of amputations are scarce, as so the aim of this study is to contribute to describe the evolution of the amputations in diabetics and calculate projections for 2025 and 2030.

### Methods and Analysis

Data of the study are hospitalizations records related to amputations in public hospitals of the National Health Service (SNS) of Portugal mainland. Is under a retrospective observational cross-sectional study using the Hospital Morbidity Database (BDGDH), for episodes with amputations, provided by the Central Administration of the Health System, IP. (ACSS), supervised by the Ministry of Health. The registers includes the calendar

year, age of patient, district (place of birth), gender, etiology and level of amputation. The Poisson regression model was fitted to the time series of annual counts of amputations available from 2000. The effects of covariates sex and age group were added to the equation. Assumption violations for the standard Poisson model such as overdispersion were addressed taking alternatives such as the negative binomial regression. Measures of performance of the models include Information Criteria types and Pearson Deviance. Empirical prediction intervals with 95% confidence were derived for 2025 and 2030 projections. Projections were adjusted to the projected diabetic population and resident population in the Portugal mainland, considering the statistical projections of population (National Institute of Statistics - INE).

**Acknowledgements:** Elisabete Carolino was supported by FCT/MCTES (UIDB/05608/2020 and UIDP/05608/2020). M. Rosário Ramos was partially financed by national funds through FCT - Fundação para a Ciência e a Tecnologia under the project UIDB/00006/2020.

## References

- [1] L. Coffey, P. Gallagher and D. Desmond. Goal Pursuit and Goal Adjustment as Predictors of Disability and Quality of Life Among Individuals With a Lower Limb Amputation: A Prospective Study. *Arch Phys Med Rehabil.* **95(2)**: 244–252, 2017
- [2] N. Holman, R. J. Young and W. J. Jeffcoate. Variation in the recorded incidence of amputation of the lower limb in England. *Diabetologia* **55(7)**: 1919–1925, 2012
- [3] M. Sousa-Uva, L. Antunes, B. Nunes, A. P. Rodrigues, J. A. Simões, R. T. Ribeiro, J. M. Boavida and C. Matias-Dia. Trends in diabetes incidence from 1992 to 2015 and projections for 2024: A Portuguese General Practitioner’s Network study. *Primary Care Diabetes.* **10(5)**: 329–333, 2016.
- [4] K. Ziegler-Graham, E. J. MacKenzie, P. L. Ephraim, T. G. Travison and R. Brookmeyer. Estimating the Prevalence of Limb Loss in the United States: 2005 to 2050. *Arch Phys Med Rehabil.* **89(3)**, 422–429, 2008.

## Organized Session 9

### Statistical and computational methods in applications

Organizer: Manuela Neves



## Optimism and stress as predictors of burnout in college students: PLSc-SEM approach

Luís M. Grilo<sup>1,2,3,4,5</sup> and M. Cristina Costa<sup>1,3</sup>

<sup>1</sup> Instituto Politécnico de Tomar (IPT), Portugal

<sup>2</sup> Universidade Aberta, Portugal

<sup>3</sup> Ci2 (Centro de Investigação em Cidades Inteligentes), IPT, Portugal

<sup>4</sup> NOVAMath (Centro de Matemática e Aplicações), FCT NOVA, Universidade NOVA de Lisboa, Portugal

<sup>5</sup> CIICESI, ESTG, Politécnico do Porto, Portugal

**E-mail addresses:** *lgrilo@ipt.pt; ccosta@ipt.pt*

---

The promotion of well-being either at the academic or professional level is a growing trend these days in order to provide better performance for both students and workers, thus reducing the failure rate of students, as well as the number of worker absences. In this study, a questionnaire was applied to students from a Portuguese Polytechnic that included questions (observed/manifest variables on a Likert-type scale) about the dimensions optimism, stress and burnout (latent variables, unobserved). To investigate the relationships between manifest and latent variables Structural equation modeling (SEM) was considered. A theoretical reflective SEM was proposed and using the consistent Partial Least Squares (PLSc) estimator a model was estimated and analyzed. The results obtained allow a better understanding of the situation of the institution's students in relation to the variables under study, and may lead to the development of mechanisms that contribute to greater student success.

### Keywords

Emotional exhaustion, Mental health, Survey, Well-being, Reflective model.

---

An increasing number of studies focus on well-being and performance efficiency both at an academic and professional level. In this sense, several dimensions can be taken into account such as optimism, self-efficacy, resilience, stress and burnout, among others. The main aim of this paper is to estimate a path model that includes the latent variables optimism, stress and burnout, regarding college students. On the one hand, optimism can be defined as inclination to look forward to positive life events related to psychological, social and physical well-being [1]. On the other hand, stress is a state of emotional or physical tension that results from adverse circumstances and when excessive and prolonged it can lead to burnout syndrome [2,3]. Burnout syndrome can result from a state of chronic stress, which overwhelms an individual's capacity to cope with professional or academic demands [4]. This is a state of emotional, mental, and physical exhaustion caused by negative stress (usually called distress). Regarding students, it may be related to the organisation of higher education where they are inserted, and can cause harmful effects on their academic performance, that in a final stage can lead to dropping their studies [5]. It occurs in three dimensions: emotional exhaustion (which is considered the core component of burnout); cynicism and a detached attitude towards school (manifested by indifference); and a feeling of inadequacy and inefficacy as a student (this is related to limited sense of competence and academic achievement). Based on the specialized literature, namely [1], we consider in this study a theoretical structural equation model (SEM), where optimism and

stress appear as possible predictors of academic burnout. The reflective estimated model was obtained using the consistent Partial Least Square (PLSc) estimator, which corrects for bias to consistently estimate SEM with common factors [6], and considering the sample collected through an online questionnaire survey. As main results we have that the latent exogenous construct optimism has a direct negative effect on perceived stress (as expected) and also indirect effects through this mediator construct on the constructs exhaustion, cynicism and efficacy (considered as the three dimensions of burnout). On the other hand, perceived stress has statistically significant direct effects on exhaustion, cynicism, and effectiveness. However, these last two path coefficients (cynicism and efficacy) have an effect size that is not statistically significant, which has implications for goodness-of-fit measures (namely, the value of SRMR - standardized residual root mean square). Thus, the eventual removal of these two links will be discussed, which will lead to a simpler model with a better quality of fit.

**Acknowledgements:** Research partially supported by National Funds through FCT - Fundação para a Ciência e a Tecnologia, project UIDB/05567/2020 (Ci2/IPT).

This work is funded by national funds through the FCT - Fundação para a Ciência e Tecnologia (Foundation for Science and Technology), I.P., under the scope of the projects UIDB/00297/2020 and UIDP/00297/2020 (Center for Mathematics and Applications).

## References

- [1] E. C. Chang, K. L. Rand and D. R. Strunk. Optimism and risk for job burnout among working college students: Stress as a mediator. *Personality and Individual Differences* **29(2)**: 255–263, 2000.
- [2] L. M. Grilo and A. Mubayi. Data-driven Approach with Structural Equation Modeling Monitors Psychosocial Community Risks. SIAM news, 2022. <https://sinews.siam.org/>.
- [3] L. M. Grilo, A. Mubayi, K. Dinkel, B. Amdouni, J. Ren and M. Bhakta. Evaluation of Academic Burnout in College Students. Application of SEM with PLS approach. AIP Conf. Proc. 2186, 090008, 2019 <https://doi.org/10.1063/1.5138004>.
- [4] J. Martínez-Líbano, M. M. Yeomans, N. González Campusano and E. Campos Flores. Emotional exhaustion and mental health in a sample of Chilean social science students during the COVID-19 pandemic. *Psocial* **7(2)**: 69–81, 2021.
- [5] D. J. Madigan and T. Curran. Does burnout affect academic achievement? A meta-analysis of over 100,000 students. *Educational Psychology Review* **33**: 387–405, 2021.
- [6] J. F. Hair, G. T. M. Hult, C. M. Ringle and M. Sarstedt. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). 2nd Ed., Thousand Oakes, CA: Sage, 2017.

## Revisiting computational procedures in extreme value parameter estimation

Manuela Neves<sup>1,2</sup> and Dora Prata Gomes<sup>3,4</sup>

<sup>1</sup>ISA, Universidade de Lisboa, Portugal

<sup>2</sup>Centro de Estatística e Aplicações (CEAUL), Portugal

<sup>3</sup>FCT, Universidade Nova de Lisboa, Portugal

<sup>4</sup>Centro de Matemática e Aplicações (CMA), Portugal

**E-mail addresses:** *manela@isa.ulisboa.pt; dsrp@fct.unl.pt*

The main objective of statistics of extremes is the estimation of parameters of rare events. The extreme value index,  $\xi$ , and the extremal index,  $\theta$ , play a central role by themselves and because their influence in the estimation of other parameters, such as high quantiles. Several estimators of  $\xi$  and  $\theta$  have appeared in the literature. However those estimators depend on tuning parameters that need to be adequately chosen. Computational procedures have been considered and present nice results. They will be revisited and new results will be shown.

### Keywords

Extremal index, Extreme value index, Extreme value theory, Resampling methodology, Semiparametric estimation.

Extreme Value Theory (EVT) aims to study and to predict the occurrence of extreme or even rare events, outside of the range of available data. These events are part of the real world but environmental extreme or rare events may have a massive impact on everyday life and may have catastrophic consequences for human activities.

There is then a need for an adequate estimation of parameters that appear in EVT. The classical assumption in EVT, is that we have a set of independent and identically distributed (i.i.d.) random variables (r.v.'s),  $X_1, \dots, X_n$ , from an unknown cumulative distribution function (c.d.f.)  $F$  and we are concerned with the limiting behaviour of  $M_n \equiv X_{n:n} = \max(X_1, \dots, X_n)$  as  $n \rightarrow \infty$ . Whenever it is possible to linearly normalize  $M_n$  so that we get a non-degenerate limit, as  $n \rightarrow \infty$ , such a limit is of the type of the extreme value (EV) d.f.,

$$EV_{\xi}(x) := \begin{cases} \exp[-(1 + \xi x)^{-1/\xi}], & 1 + \xi x > 0 \text{ if } \xi \neq 0 \\ \exp[-\exp(-x)], & x \in \mathbf{R} \quad \text{if } \xi = 0. \end{cases} \quad (6)$$

The parameter  $\xi$  is the *extreme value index* (EVI) and it measures essentially the weight of the right tail function,  $\bar{F} = 1 - F$ .

In most fields of applications, the independence assumption is not valid. Stationary sequences are realistic for many real problems and dependence in stationary sequences can assume several forms. Provided that a stationary sequence  $\{X_n\}_{n \geq 1}$  verifies some conditions, the stationary sequence  $\{X_n\}_{n \geq 1}$  is said to have an *extremal index* (EI),  $\theta$ ,  $0 < \theta \leq 1$ , if for each  $\tau > 0$ , we can find a sequence of levels  $u_n = u_n(\tau)$  such that, with  $\{Y_n\}_{n \geq 1}$  the associated i.i.d. sequence (i.e. from the same  $F$ ),

$$\mathbf{P}(Y_{n:n} \leq u_n) = F^n(u_n) \xrightarrow[n \rightarrow \infty]{} e^{-\tau} \quad \text{and} \quad \mathbf{P}(X_{n:n} \leq u_n) \xrightarrow[n \rightarrow \infty]{} e^{-\theta\tau}. \quad (7)$$

Under the validity of those conditions the *extremal index* can also be defined as

$$\begin{aligned}\theta &= \frac{1}{\text{limiting mean size of clusters}} = \lim_{n \rightarrow \infty} \mathbf{P}(X_2 \leq u_n | X_1 > u_n) = \\ &= \lim_{n \rightarrow \infty} \mathbf{P}(X_2 > u_n | X_1 \leq u_n),\end{aligned}\tag{8}$$

where  $u_n : F(u_n) = 1 - \tau/n + o(1/n)$ , as  $n \rightarrow \infty$ , with  $\tau > 0$ , fixed.

Semi-parametric estimators of  $\xi$  and  $\theta$  depend on  $k$ , the number of upper order statistics and on a threshold,  $u_n$ , respectively.

Bootstrap and Jackknife methodologies and also an heuristic computational procedure have been studied in several works and will be here revisited with new challenges.

**Acknowledgements:** Research partially supported by National Funds through FCT| Fundação para a Ciência e a Tecnologia, projects UIDB/MAT/0297/2020 (CMA/UNL) and UIDB/MAT/0297/2020 (CEA/UL).

## Approximation methods for the likelihood function in parameter estimation for SDE mixed models

Nelson T. Jamba<sup>1,4</sup>, Patrícia A. Filipe<sup>1,3</sup>, Gonçalo Jacinto<sup>1,2</sup> and Carlos A. Braumann<sup>1,2</sup>

<sup>1</sup>Centro de Investigação em Matemática e Aplicações, Instituto de Investigação e Formação Avançada, Universidade de Évora

<sup>2</sup>Departamento de Matemática, Escola de Ciências e Tecnologia, Universidade de Évora

<sup>3</sup>Iscte - Instituto Universitário de Lisboa, ISCTE Business School, Departamento de Métodos Quantitativos para Gestão e Economia

<sup>4</sup>Liceu n. 918 do Município dos Gambos, Chiange, Gambos, Angola

**E-mail addresses:** *d39830@alunos.uevora.pt; patricia.filipe@iscte-iul.pt; gjcj@uevora.pt; braumann@uevora.pt*

We study stochastic differential equations mixed models to describe the evolution of cattle weight. In order to incorporate individual characteristics of the animals, we will consider that the model parameters  $\alpha$  (average transformed weight at maturity) or  $\beta$  (growth parameter) vary randomly from animal to animal. The application of the maximum likelihood estimation method involves some complexity in obtaining a closed-form expression for the likelihood function. Here we present and compare the existing Laplace approximation method with our proposed delta method to overcome this problem.

### Keywords

Stochastic differential equations, Mixed models, Maximum likelihood, Delta method, Laplace approximation method.

To describe the growth of an individual in a randomly fluctuating environment, stochastic differential equation (SDE) models have been studied, for instance in [2,4,3], and applied to model the evolution of cattle weight in [2,4]. To incorporate individual characteristics of the animals, we have considered that model parameters may vary randomly from animal to animal, which results in SDE mixed models.

We use a class of SDE to model the evolution of cattle weight which, by an appropriate transformation of the weight, results in a variant of the Ornstein-Uhlenbeck model. The parameters are the average transformed size at maturity  $\alpha$ , a growth parameter  $\beta$  and the intensity of the effect of environmental fluctuations  $\sigma$ . Obtaining a closed-form expression for the likelihood function in order to apply the maximum likelihood estimation method is a difficult, sometimes impossible, task. We propose the use of the Laplace method (LA) or our proposed adaptation of the delta method (DA) to approximate the integrals involved in the likelihood function. Our goal is to compare both approximation methods for the case where just one parameter  $\alpha$  or  $\beta$  is considered random, which is useful in so many applications. This approach allows the estimation of the parameters even when the requirement of most known packages (for instance, [1]) fails. Namely, they require having the same time vector of observations for all trajectories. This is not the case in our application, since in our real data the animals are not weighted at the same ages, and for this reason the existing packages could not be used.

Here we present the results for the case where  $\alpha$  or  $\beta$  are Gaussian random variables,  $\alpha_i \sim N(\mu, \theta^2)$  or  $\beta_i \sim N(\lambda, \omega^2)$ , ( $i = 1, \dots, M$ ). We have results for real weight data from

a large and heterogeneous sample of  $M = 16029$  Mertolengo cattle males, where each animal has several observations with a minimum of 3 and a maximum of 33 weights at ages varying between birth and a maximum age that ranges from 0.2 until 16 years, totalling 96204 observations. The estimates of all parameters through the LA method and the DA method for random  $\alpha$  are very similar, except for a slight difference in  $\theta$ . The approximate confidence intervals based on Fisher's information matrix are also quite similar. Regarding the estimation using the DA method for random  $\beta$ , the estimates for the asymptotic weight of the animal, and for  $\lambda$  and  $\sigma$  are also very close to those obtained through the LA method.

To illustrate the performance of our proposed DA method we also compare, using simulated datasets (to guarantee that animals have the same number of weights obtained at the same time instants), the estimated parameters obtained with the DA method and the LA method with the estimates obtained from existing R packages (MsdeParEst and mixedside that use numerical approximation methods for the likelihood). The results show a very good performance of the proposed approximate methods, having similar estimates for some parameters and often outperforming the existing methods for the remaining parameters.

**Acknowledgements:** The Centro de Investigação em Matemática e Aplicações is supported by the FCT, project UID/MAT/04674/2020. This work was developed within the Operational Group PDR2020-1.0.1-FEADER-031130 - Go BovMais, funded by PDR 2020.

## References

- [1] M. Delattre and C. Dion. MsdeParEst. Parametric Estimation in Mixed - Effects Stochastic Differential Equations. *R package version 1.7 1*, <https://cran.r-project.org/web/packages/MsdeParEst/index.html> 2017 (accessed on 25 January 2022).
- [2] P. A. Filipe, C. A. Braumann and C. Carlos. Profit optimization for cattle growing in a randomly fluctuating environment. *Optimization* **64(6)**: 1393–1407, 2015.
- [3] N. T. Jamba, G. Jacinto, P. A. Filipe and C. A. Braumann. Likelihood Function through the Delta Approximation in Mixed SDE Models. *Mathematics* **10(3)**: 385, 2022.
- [4] U. Picchini, A. De Gaetano and S. Ditlevsen. Stochastic differential mixed-effects models. *Scandinavian Journal of Statistics* **37(1)**: 67–90, 2010.

Posters

---

# Big Data Analytics in Water Consumption

João M. Alves<sup>1</sup> and Patrícia A. Filipe<sup>1,2</sup>

<sup>1</sup>Iscte Business School, Iscte - University Institute of Lisbon, Portugal

<sup>2</sup>Research Centre for Mathematics and Applications, Institute for Advanced Studies and Research, University of Évora, Portugal

**E-mail addresses:** *jmgas1@iscte-iul.pt; patricia.filipe@iscte-iul.pt*

---

Water is an essential and limited resource nowadays, so it is necessary to distribute and conserve it efficiently. This study aims to develop a methodology to predict the water consumption of domestic consumers, through the application of big data and machine learning techniques for this purpose. Clustering techniques are applied to the time series to group consumers according to consumption patterns, and forecasting techniques, such as Random Forest and autoregressive models, are applied to predict future consumption.

## Keywords

Big Data Analytics, Machine Learning, Water Consumption.

---

Freshwater is an essential and scarce resource nowadays, as it represents only 0.014% of all the water on the planet [2]. With the help of the new digital meters, it is possible to measure consumption on an hourly basis and, through machine learning techniques, produce information on consumption patterns, consumer behavior, detection of events in water distribution networks and feedback to customers. Currently, companies that manage the distribution of drinking water to the population have, with the advance of technology, new ways to predict water consumption in a given location, and therefore manage in a more efficient way the amount of water injected into the network.

We intend to apply machine learning techniques based on time series, in order to predict the water consumption of domestic consumers.

The data used for this study were provided by Infraquinta, a company based in Quinta do Lago, Algarve, which manages the infrastructure and water distribution at this location. This company was concerned with modernizing this sector, and for that installed digital water meters on all properties in this location. In this way, it was possible to provide hourly customer consumption data, among other variables, such as customer information.

This study was carried out on the Knime Analytics Platform, a software with an intuitive interface, without the need for programming, with enormous potential for projects in Big Data Analytics. First of all, the variables are selected and then are made the necessary transformations and cleaning of the data, such as treatment of missing values, time series alignment and data organization. Due to the large number of consumers (about 201 consumers) and the impossibility of analyzing each one individually, clustering techniques are applied to group consumers according to their consumption patterns, thus creating clusters. The series of each cluster are created through the average of the customers' consumption, corresponding and organized in hourly and daily cadence. After choosing the cluster for analysis, some models are applied, such as Random Forest, Gradient Boosted,



ARIMA, and others, in order to make predictions about the time series of water consumption, taking into account the seasonality, trend and cycles of the sample[1]. Finally, all the models considered are presented and the one that had the best performance is chosen.

**Acknowledgements:** The second author is a member of Centro de Investigação em Matemática e Aplicações (CIMA) a research centre supported by the Fundação para a Ciência e a Tecnologia, project UID/MAT/04674/2020. The authors would like to thank Infraquinta for providing the data.

## References

- [1] P. I. Karamaziotis, A. Raptis, K. Nikolopoulos, K. Litsiou and V. Assimakopoulos. An empirical investigation of water consumption forecasting methods. *International Journal of Forecasting* **36**: 588–606, 2020.
- [2] M. S. Rahim, K. A. Nguyen, R. A. Stewart, D. Giurco and M. Blumenstein. Machine Learning and Data Analytic Techniques in Digital Water Metering: A Review. *Water* **12**: 294, 2020.

## Fitting mixtures of linear mixed models: a case study

Susana Faria<sup>1</sup> and Luísa Novais<sup>2</sup>

<sup>1</sup>Department of Mathematics and Centre of Mathematics (CMAT), University of Minho, Guimarães, Portugal

<sup>2</sup>Department of Mathematics, University of Minho, Guimarães, Portugal

**E-mail addresses:** *sfaria@math.uminho.pt; luisa\_novais92@hotmail.com*

---

In this study, we discuss the procedure for fitting mixtures of linear mixed models to fishery data, by means of maximum likelihood estimation using the EM algorithm.

We pretend to study the relationship between the price per kilo of mackerel caught in the coast of Portugal and other covariates, like as, region of fishing, body size of mackerel and year of fishing. The results highlight that landings differences lead to two distinct components. These components are used to characterize the distribution of the price per kilo of mackerel caught in Portugal.

### Keywords

Mixtures of linear mixed models, EM algorithm and Fishery data.

---

Finite mixture models are a widely known method for modelling data that arise from a clustered and heterogeneous population. Within the family of mixtures of regression models, mixtures of linear mixed models have also been applied in different areas since, besides taking into consideration the heterogeneity in the population, they also allow to take into account the correlation between observations from the same individual [3].

In this work, we consider fitting a finite mixture of linear mixed models to fishery data, to study the relationship between the price per kilo of mackerel caught in the coast of Portugal and other covariates.[3].

In order to obtain parameter estimates for the proposed model, we perform maximum likelihood estimation via the EM algorithm. The EM algorithm was first proposed as a method for estimating parameters in the setting of missing data, however, it is commonly used for maximum likelihood estimation in finite mixture modeling.

The number of components in the mixture model was deemed to be unknown and estimated from the data. [1]

A mixture of linear mixed models with two components was fitted to the data. The first component contains the highest number of vessels and concerns vessels with a higher quantity caught, whose price per kilo of mackerel is low and which landed almost entirely in North and Center regions of Portugal. A second component has a slightly lower number of vessels and contains vessels that have caught more mackerel whose price per kilo of mackerel is high. These vessels have landed mostly in the South region of Portugal.

**Acknowledgements:** Supported by Portuguese funds through the Portuguese Foundation for Science and Technology, within project PTDC/MAT-STA/28243/2017.

## References

- [1] G. J. McLachlan and D. Peel. *Finite Mixture Models*. Wiley, New York, 2000.
- [2] L. Novais and S. Faria. Selection of the number of components for finite mixtures of linear mixed models. *Journal of Interdisciplinary Mathematics*, **24(8)**: 2237–2268, 2021.
- [3] D. H. Ogle. *Introductory Fisheries Analyses with R*. CRC Press, UK, 2016.

---

## Mixtures of linear mixed models: variable selection using CEM algorithm

Luísa Novais<sup>1</sup> and Susana Faria<sup>2</sup>

<sup>1</sup>Department of Mathematics, University of Minho, Guimarães, Portugal

<sup>2</sup>Department of Mathematics and Centre of Mathematics (CMAT), University of Minho, Guimarães, Portugal

**E-mail addresses:** *sfaria@math.uminho.pt; luisa\_novais92@hotmail.com*

---

Variable selection plays an important role in any modelling study, requiring the search for the simplest possible model that adequately describes the observed data.

In this work we analyze the problem of variable selection in mixtures of linear mixed models using the Classification Expectation-Maximization (CEM) algorithm through a simulation study and a real-world application.

### Keywords

Mixtures of linear mixed models, Variable Selection and Classification Expectation-Maximization (CEM) algorithm.

---

Finite mixture of linear mixed models provide a flexible tool for modeling data that arise from a clustered and heterogeneous population, where the relationship between the response and the covariates varies across the sub-populations [3].

In the applications of these models, a large number of covariates are often used and their contributions toward the response variable vary from one component to another of the mixture model. For this reason, variable selection assumes a great relevance for mixture models, something particularly notorious in the last few years.

Variable selection via penalized likelihood has attracted great attention in recent literature. In particular, [2] have investigated the variable selection problem for finite mixture regression models with versions of the penalty functions.

In this work we analyze the problem of variable selection in mixtures of linear mixed models in the presence of a large number of explanatory variables. We study the performance of a penalized likelihood approach in identifying the most relevant subset of covariates using the Classification Expectation-Maximization algorithm (CEM) [1].

**Acknowledgements:** Supported by Portuguese funds through the Portuguese Foundation for Science and Technology, within project PTDC/MAT-STA/28243/2017.

### References

- [1] G. Celeux and G. Govaert. A classification EM algorithm for clustering and two stochastic versions. *Computational Statistics and Data Analysis* **14(3)**: 315–332, 1992.
- [2] A. Khalili, J. Chen. Variable selection in finite mixture of regression models. *Journal American Statistical Association* **102**: 1025–1038, 2007.
- [3] G. J. McLachlan, D. Peel. *Finite Mixture Models*. Wiley, New York, 2000.

# Extremal Index: estimation and resampling

Marta Ferreira<sup>1</sup>,

<sup>1</sup>Center of Mathematics, University of Minho, Portugal

**E-mail address:** *msferreira@math.uminho.pt*

## Keywords

Extreme value theory, Stationary sequences, Statistics of extremes, Bootstrap and jack-knife.

---

The duration of extremes in time leads to a phenomenon known as clustering of high values, with a strong impact on risk assessment. The extremal index is a measure developed within Extreme Value Theory that quantifies the degree of clustering of high values. In this work we will consider the cycles estimator introduced in [1]. A reduced bias estimator based on the Jackknife methodology will be presented. The bootstrap technique will also be considered in the inference and will allow to obtain confidence intervals. The performance will be analyzed based on simulation. An application of the methods to real data will also be presented.

**Acknowledgements:** The research at CMAT was partially financed by Portuguese Funds through FCT (Fundação para a Ciência e a Tecnologia) within the Projects UIDB/00013/2020 and UIDP/00013/2020 and PTDC/MAT-STA/28243/2017.

## References

- [1] H. Ferreira and M. Ferreira. Estimating the extremal index through local dependence. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques* **54(2)**: 587–605, 2018.

## Survival analysis of women diagnosed with breast cancer in Angola

Jaime Jerónimo<sup>1,2</sup>, Patrícia A. Filipe<sup>2,4</sup> and Dulce Gomes<sup>2,3</sup>

<sup>1</sup>Instituto Superior Privado Alvorecer da Juventude, Angola

<sup>2</sup> Research Centre for Mathematics and Applications, Institute for Advanced Studies and Research, University of Évora, Portugal

<sup>3</sup>Department of Mathematics, School of Science and Technology, University of Évora, Portugal

<sup>4</sup> Iscte Business School, Iscte - University Institute of Lisbon, Portugal

**E-mail addresses:** *d35936@alunos.uevora.pt; patricia.filipe@iscte-iul.pt; dmog@uevora.pt*

Breast cancer is the most common cancer among women worldwide and in the African continent it has higher mortality rates than malaria and HIV combined. Based on records of 1227 women diagnosed with breast cancer in Angola in the period 2013-2018, several descriptive and inferential statistical techniques were applied to the data, as well as specific survival analysis techniques to study the time elapsed from the diagnosis of the disease to the outcome "death" taking into account some associated factors, such as age, stage of the disease, alcohol consumption, among others.

### Keywords

Angola, Breast cancer, Survival analysis.

Breast cancer is the most common cancer among women and accounts for 24.5% of all cancer cases worldwide [1]. It is estimated that in 2020 there will be 19.3 million new cases of cancer worldwide and 9.9 million deaths from cancer, and that in Africa there will be a 16.8% increase in new cancer cases and 12.1% of African women will die from this disease [4]. On the African continent breast cancer has higher mortality rates than malaria and HIV combined [3].

According to the annual report from Angola's National Statistics Institute, breast cancer was one of the main causes of death in Angolan women in 2020, accounting for 11.5% of deaths from all oncological diseases in the country [2].

We intend to study modelling techniques, in particular Survival Analysis, in order to describe the evolution of breast cancer diagnosed in Angolan women in the period 2013-2018, taking into account some associated factors, such as age, stage of the disease, family history of breast cancer, alcohol consumption, among others.

The study population is composed of a retrospective population-based cohort obtained from the Angolan Institute for Cancer Control (IACC) and refers to the records of 1227 women from the 18 provinces of the country. Several descriptive and inferential statistical techniques were applied to characterize the data. Specific survival analysis techniques were applied to the time elapsed from the diagnosis of the disease to the outcome "death". The analyses were performed using the R software.

The covariates that proved significant in the model were treatment, age, and stage of disease. In terms of treatment, only radiotherapy was not significant compared to surgery (reference class). All the other therapies (chemotherapy, palliative and hormone therapy) proved to be significant and contribute to the increased risk of death compared to surgery. Women diagnosed at older ages have the higher risk of death. And, as expected, the risk

of a patient suffering from breast cancer dying is higher when at the time of diagnosis she was in stage IV.

**Acknowledgements:** Research Centre for Mathematics and Applications, Research and Advanced Training, University of Évora, Project UID/MAT/04674/2022 funded by FCT (Fundação para a Ciência e a Tecnologia).

## References

- [1] J. A. Gomes. Estimativa 2020: incidência do Câncer no Brasil. Rio de Janeiro: INCA. <https://www.inca.gov.br/controlado-cancer-de-mama/conceito-e-magnitude> (accessed in 22.Sep.2022)
- [2] Instituto Nacional de Estatística de Angola (INE). Registo Nacional de Câncer, Relatório 2021.
- [3] L. V. Lopes (2018). Cancro da Mama em Angola: Caracterização clínica patológica e recursos necessários para a definição e concretização de um programa de acção, 2016. <https://repositorio-aberto.up.pt/bitstream/10216/90869/2/157381.pdf>
- [4] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71(3)**: 209–249, 2021.

## Retail store object detection using Yolo

Miguel C. Lobo<sup>1</sup> and Patrícia A. Filipe<sup>1,2</sup>

<sup>1</sup>Iscte Business School, Iscte - University Institute of Lisbon, Portugal

<sup>2</sup>Research Centre for Mathematics and Applications, Institute for Advanced Studies and Research, University of Évora, Portugal

**E-mail addresses:** *mclo@iscte-iul.pt; patricia.filipe@iscte-iul.pt*

---

The analysis of images and videos for understanding customer and location behaviours have the ability to link more data together and may be able to, not only adopt specific strategies, but also to measure their effects more accurately.

It was developed a deep learning algorithm using Yolo that provides object detection in real time, in order to extract insights of consumer behaviours in retail environments. In this follow-up, it is necessary to understand how one can take advantage of the analysis of images regarding people behaviours inside the shops.

### Keywords

Deep learning, Big data, Object detection, Retail.

---

At the beginning of the year 2018, recent advances in the field of artificial intelligence and its fields in computer vision, machine learning, and especially deep learning have led to a new paradigm in the retail area [3]. Probably the most prominent example is the Amazon Go store chain, in which the entire checkout process is automated and does not require the existence of cashiers in its operation.

This provided new ideas and capabilities for stores to explore new types of information and reduce constraints on the consumer. This was due to multiple cameras and sensors deployed inside the store that will provide new data and information. Several studies, including [3,4] state that having the ability to understand consumer behaviours is one of the keys to success for retailers and extol the importance of taking this direction. Additionally, the study [4] assumes the perceptions that consumer behaviours are perpetual concerns that must always be the target of analysis in studies of this scope.

That being said, in this research, a study is developed that simulates insights into customer behaviour within these stores and the information that can be gleaned from the analysis of the cameras without concerns about data privacy. In this follow up, it is necessary to understand how one can take advantage of the analysis of images regarding consumer behaviours inside the stores and for that, have the ability to recognize actors in the shopping process in images by parameterizing the best model using deep learning techniques. For this purpose, data was collected from two public videos taken from the YouTube platform [1,2], relative to security camera images of two distinct retail establishments. From each video, hundreds of frames were generated for analysis, and a figure of the space in one of the video frames was activated to represent the layout of the space.

With the research objectives in mind, extensive research is conducted on the area of Big Data Analytics applied to computer vision and deep learning, its history and evolution, and the real-world applications that can be found today in the retail area. In the same way, it is deepened through the analysis of sensor analysis methods applied to these same areas and to congruent researches to the one developed here.

Yolo, is a deep learning algorithm that can be run in real time and captures the entire image at once, as opposed to analyzing the proposals of a region generated in previous



methods. This algorithm has an even faster detection time than CNNs (Convolutional Neural Network) due to its simplistic architecture [4].

According to the interpretation of the literature for this study, Yolo is the algorithm chosen due to its presence in modern studies, such as [3]. Additionally, it is considered the fact that these types of models (one-stage detectors) have the advantage of predicting the probabilities of classes and bounding boxes from complete images with a single convolutional neural network as opposed to two-stage detectors that analyze parts of an image at a time.

For that purpose, the Python library ImageAI is used, a library that helps students, researchers and developers to build applications with deep learning and computer vision capabilities. From this library, the model is programmed to identify people from a 75% confidence level due to the image quality of the videos.

Additionally, track masks and bounding boxes are generated for each object/person detected in order to highlight consumption patterns and affluences inside the store, and later a heatmap are built over those results in order to present that data over the frame, that would represent the layout of the space. The results are evaluated against the metrics present in the research, but also in their practical scope in real world applications.

**Acknowledgements:** The second author is a member of Centro de Investigação em Matemática e Aplicações (CIMA) a research centre supported by the Fundação para a Ciência e a Tecnologia, project UID/MAT/04674/2020.

## References

- [1] HDCCTV Cameras: HD CCTV Camera video 3MP 4MP iProx CCTV HDCCTV-Cameras.net retail store. <https://www.youtube.com/watch?v=KMJS66jBtVQ> 2017.
- [2] CBS. News: Apple Store robbery caught on video. <https://www.youtube.com/watch?v=H9Qqr-pKrn8> 2018.
- [3] K. Nguyen, M. Le, B. Martin, I. Cil and C. Fookes. When AI meets store layout design: A Review. *Artificial Intelligence Review*. **55**: 5707–5729, 2022.
- [4] M. Paolanti, R. Pietrini, A. Mancini, E. Frontoni and P. Zingaretti. Deep understanding of shopper behaviours and interactions using RGB-D vision. *Machine Vision And Applications*. **31**: 2020.

## An application of logistic regression model to the study of constitutional imbalances in human chromosomal fragile sites

Inês J. Sequeira<sup>1</sup>, Giovanni L. Silva<sup>2</sup>, Dora Prata Gomes<sup>1</sup>, José Rueff<sup>3</sup>  
and Aldina Brás<sup>3</sup>

<sup>1</sup>Center for Mathematics and Applications (NovaMath), FCT NOVA and Department of Mathematics, FCT NOVA, Portugal

<sup>2</sup>CEAUL and Department of Mathematics, Instituto Superior Técnico, Universidade de Lisboa, Portugal

<sup>3</sup>TOXOMICS, NOVA Medical School/Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Portugal

**E-mail addresses:** *ijs@fct.unl.pt; giovani.silva@tecnico.ulisboa.pt; dsrp@fct.unl.pt; jose.rueff@nms.unl.pt; aldina.bras@nms.unl.pt*

### Keywords

Human chromosomal fragile sites, Logistic regression model.

Chromosomal fragile sites (CFSs) are loci or regions susceptible to spontaneous or induced occurrence of breaks and rearrangements. They are classified in two main categories, rare and common, depending on their frequency in the population. In order to identify which CFSs are influential or significant in the occurrence of deletions and duplications (chromosomal constitutional imbalances), we propose a logistic regression analysis for the CFS data set, since the underlying response variable is categorical, specifically binary (deletion or duplication). Some results are presented here as an informative preliminary contribution to understand the frailty of these CFS in increasing/decreasing of the deletion odds. This study has implications for our comprehension of human pathogenesis.

**Acknowledgements:** This study makes use of data generated by the DECIPHER community. A full list of centres who contributed to the generation of the data is available from <http://decipher.sanger.ac.uk> and via email from [decipher@sanger.ac.uk](mailto:decipher@sanger.ac.uk). This work is funded by national funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P., under the scope of the projects UIDB/00297/2020 and UIDP/00297/2020 (Center for Mathematics and Applications), UIDB/00006/2020 (CEAUL) and UID/BIM/00009/2016 (Centre for Toxicogenomics and Human Health (TOXOMICS)).

### References

- [1] J. J. Faraway *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. 2nd Edition, Chapman and Hall/CRC, Boca Raton, 2016.
- [2] D. W. Hosmer Jr., S. Lemeshow and R. X. Sturdivant *Applied Logistic Regression*. 3rd Edition, Wiley, New York, 2013.
- [3] K. Mrasek, C. Schoder, A. C. Teichmann, K. Behr, B. Franze, K. Wilhelm, N. Blau-rock, U. Claussen, T. Liehr and A. Weise. Global screening and extended nomenclature for 230 aphidicolin-inducible fragile sites, including 61 yet unreported ones. *International Journal of Oncology* **36**: 929–940, 2010.

- [4] I. Sequeira, J. Mexia, J. Santiago, R. Mamede, E. Silva, J. Santos, D. Faria, J. Rueff and A. Brás. Predominance of constitutional chromosomal rearrangements in human chromosomal fragile sites. *Open Journal of Genetics* **3**: 8–13, 2013.

## Index of authors

- Afonso, Afonso, *45*  
Alves, Carlos, *6*  
Alves, João M., *67*  
Antunes, Patrícia, *38*  
Araújo, Adérito, *46*  
Arulandu, Alvan Caleb, *31*  
Arunachalam, Viswanathan, *35, 36*
- Bagwan, Arif S., *9*  
Banerjee, Sohini, *33*  
Barbeiro, Sílvia, *44, 46*  
Brás, Aldina, *77*  
Braumann, Carlos A., *64*
- Cabral, Ivanilda, *16*  
Caeiro, Frederico, *4, 16*  
Cantarinha, Ana, *24*  
Carolino, Elisabete, *57*  
Coelho, Carlos A., *2*  
Costa, M. Cristina, *60*  
Costa, Marco, *54, 55*
- DebRoy, Swati, *49*  
Dias, Ana, *25*  
Dias, Cristina, *27, 28*
- Esquível, Manuel L., *39*
- Fabián, Zdeněk, *11*  
Faria, Susana, *69, 71*  
Ferreira, Dário, *38*  
Ferreira, José Augusto, *3*  
Ferreira, Marta, *72*  
Ferreira, Sandra S., *38*  
Figueiredo, Adelaide, *17, 19*  
Figueiredo, Fernanda Otilia, *4, 17, 19*  
Filipe, Patrícia A., *64, 67, 73, 75*  
Frederico, Caeiro, *21*  
Fung, Isaac, *50*
- Gomes, Dora Prata, *62, 77*  
Gomes, Dulce, *73*  
Gomes, M. Ivette, *4, 16, 21*  
Gonçalves, A. Manuela, *54, 55*  
Grilo, Luís M., *60*
- Henriques, Rafael, *44*  
Henriques-Rodrigues, Lígia, *4, 21*
- Jacinto, Gonçalo, *64*
- Jamba, Nelson T., *64*  
Jerónimo, Jaime, *73*
- Khanday, Mukhtar Ahmad, *12*  
Kiselák, Jozef, *13*
- Lobo, Miguel C., *75*
- Martins, Nuno F.M., *6*  
Matos, José Pedro, *57*  
Mendes, Luzia, *41*  
Mexia, João T., *24, 27, 28, 39*  
Moreira, Elsa, *24*  
Mubayi, Anuj, *51*
- Neves, Manuela, *62*  
Nguyen, Long, *33*  
Novais, Luísa, *69, 71*  
Nunes, Célia, *28, 39*
- Ogueda, Alonso Gabriel, *35*  
Oliveira, Inês, *41*  
Oliveira, Teresa, *41*  
Opoku-Ameyaw, Kwaku, *39*
- Pereira, F. Catarina, *54, 55*  
Pereira, José, *41*  
Pinto, Jorge, *47*
- Ríos-Gutiérrez, Andres, *36*  
Ramos, M. Rosário, *57*  
Rincón-Prieto, Andres Felipe, *36*  
Rueff, José, *77*
- Salinas, Erika Johanna Martínez, *35*  
Santos, Carla, *25, 27, 28*  
Santos, José, *44*  
Santos, Milene, *46*  
Saucedo, Omar, *52*  
Sequeira, Inês J., *77*  
Serrão, Rodrigo, *6*  
Seshaiyer, Padmanabhan, *31, 33, 35*  
Silva, Giovani L., *77*  
Silvestre, Ana Leonor, *6*  
Singh, Rohan, *33*  
Stehlík, Milan, *14*
- Torres, Delfim F. M., *47*
- Vaz, Sandra, *47*