New Bridges between Mathematics and Data Science

VALLADOLID

8-11 November 2021

Early registration fee: 50€

Plenary Speakers:

Joan Bruna - U. of New York
Marco Cuturi - Institut Polytechnique de Paris
Alfio Quarteroni - P. di Milano
Coralia Cartis - U. of Oxford
Jeff Goldsmith - Columbia U.

Round Tables:

The government research strategies and funding
The mathematics research community
The Spanish Industry

Minisymposia:

High dimensional statistics
Optimization
Approximation
Computational Mathematics
Statistical analysis of complex data
Industrial applications

IMPORTANT DATES

Grant application
September 20th 2021

Notification of grants
October 1st 2021

Early registration until
October 1st 2021

Scientific Committee:

L. Vega
Ll. Alsedà
E. del Barrio
E. Carrizosa
T.Chacón
J. L. Ferrín
R. Lillo
G. Lugosi
J. J. Nieto Roig

Local Committee:

E. del Barrio
M. P. Calvo
A. Gonzalez Sanz
P. Gordaliza
H. Inouzhe

CONTACT
nbmds@uva.es

GRANTS CONTACT
nbmds.grants@bcamath.org

WEB
http://nbmds.uva.es/

imUVa

RED ESTRATÉGICA EN MATEMÁTICAS

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN

Sponsors:

Red Estratégica en Matemáticas (REM)
Instituto de Investigación en Matemáticas de la Universidad de Valladolid (IMUVa)
Universidad de Valladolid
Ministerio de Economía y Competitividad. Gobierno de España

Scientific Commitee:

Luis Vega (BCAM, Chair)
Lluís Alsedà (CRM/UAB)
Eustasio del Barrio (IMUVa, U de Valladolid)
Emilio Carrizosa (IMUS, U. de Sevilla)
Tomás Chacón (IMUS, U. de Sevilla)
Antonio Falcó (U. CEU Cardenal Herrera)
José Luis Ferrín (U. de Santiago de Compostela)
Mario Fioravanti (U. de Cantabria)
Rosa Lillo (IBiDat, UC3M)
Gabor Lugosi (ICREA-UPF/Barcelona GSE)
David Martín de Diego (ICMAT)
Juan J. Nieto (IMAT, U. de Santiago de Compostela)

Organizing Commitee:

Eustasio del Barrio, *Universidad de Valladolid, Chair*
Miguel A. Benítez, *BCAM*
Mari Paz Calvo, *Universidad de Valladolid*
Alberto González Sanz, *Université Paul Sabatier, Toulouse*
Paula Gordaliza, *BCAM*
Hristo Inouzhe, *BCAM*
Yolanda Larriba, *Universidad de Valladolid*
Paula López Pérez, *Universidad de Valladolid*

Overview

Schedule

	Monday, November 8th	Tuesday, November 9th	Wednesday, November 10th	Thursday, November 11th
9:30-10:00		P3 Alfio Quarteroni	P4 Marco Cuturi	P2 Coralia Cartis
10:00-10:30	Opening Session			
10:30-11:00	Coffee Break	Coffee Break	Coffee Break	Coffee Break
11:00-12:00	Minisymposia MS10, MS13, MS17, MS21, MS23	RT1 Funding & policy making	RT2 Scientific debate	RT3 Data Science and Industry
12:00-13:00				
13:00-15:00	Lunch	Lunch	Lunch	Lunch
15:00-17:00	P1 Joan Bruna	Minisymposia MS7, MS8, MS11, MS18, MS22	Minisymposia MS1, MS16, MS20, MS28, MS30	Minisymposia MS3, MS9, MS25, MS12, MS27
	P5 Jeff Goldsmith			
17:00-19:00		Minisymposia MS6, MS15, MS24, MS26, MS29	Minisymposia MS2, MS4, MS5, MS14, MS19	Closing Session

MS1 High-dimensional Bayesian networks - Felipe II
MS2 Functional data analysis (I) - Cardenal Mendoza
MS3 Spatio-temporal Data Science - Claudio Moyano
MS4 Interpretability and explainability of algorithms - Claudio Moyano
MS5 High-dimensional variable selection - Felipe II
MS6 Fair learning - Real Chancillería
MS7 Optimal transport for data science - Cardenal Mendoza
MS8 Adversarial Machine Learning - Claudio Moyano
MS9 Probabilistic Learning - Paraninfo
MS10 New Approaches in Combinatorial Optimization - Real Chancillería
MS11 Mathematical Optimization Methods for Decision Making - Felipe II
MS12 Decision aid and data science models for disaster management - Real Chancillería
MS13 Mathematical support to the resource and process management in health - Claudio Moyano
MS14 Mathematical Optimization for Data-Driven Decision-Making - Real Chancillería
MS15 Mathematical Optimization, Classification and Regression - Felipe II
MS16 Data Science Applications - Paraninfo
MS17 Non-linear approximation, vision and images - Felipe II
MS18 Neural networks for Mathematicians - Real Chancillería
MS19 Machine learning techniques in control theory and inverse problems - Paraninfo
MS20 Solving inverse problems using data-driven models - Real Chancillería
MS21 New perspectives in Computational Mathematics (I) - Cardenal Mendoza
MS22 New perspectives in Computational Mathematics (II) - Paraninfo
MS23 Statistical analysis of complex data (I) - Paraninfo
MS24 Statistical analysis of complex data (II) - Paraninfo
MS25 Digital Twins - Felipe II
MS26 New Perspectives in Data Science - Claudio Moyano
MS27 Mathematical Optimization in Industry - Cardenal Mendoza
MS28 ML and NLP models: from notebook to production deployment - Cardenal Mendoza
MS29 Functional Data Analysis (II) - Cardenal Mendoza
MS30 Data Science in Action - Claudio Moyano

General Information

Understanding the world through data and computation has always been at the heart of scientific discovery. In the past decade primarily data-driven approaches, such as neural networks, have been very successful. Nevertheless, the reason for this success is to some extent mysterious and raises multiple questions regarding the robustness, explainability, interpretability and fairness of the algorithms used. The answer to these questions is crucial when it comes to decision making.

In recent years, there has been increasing activity in building bridges between these new ideas and other well-established approaches based on models typically derived from first principles. However, the computational cost of the latter makes them unaffordable except in low dimensions, which is a limitation from which neural networks are exempt. Establishing solid connections between these two different points of view have already proved to be extremely fruitful.

The Spanish 'Mathematical Strategic Network' ('Red Estratégica de Matemáticas', REM) organizes a four-day workshop in November 8th-11th 2021 in Valladolid, Spain, with the aim of bringing together researchers in mathematics, machine learning, and data science, to exchange ideas and progress in the construction of new bridges among these fields and make visible the work already done.

The workshop will be structured around five plenary sessions by leading scientists on the international scene, plus three round-tables and 30 thematic minisymposia.

The REM is a network of mathematical research institutes which aims at fostering the international presence of the Spanish research in mathematics, the creation of synergies among the mathematical scientific community and the socioeconomic impact of Spanish mathematical research. It promotes the dissemination and transfer of mathematical technology, orienting research towards the needs of companies, industries and public administrations.

Conference venue

The meeting will be held at Conference centre "Palacio de Congresos Conde Ansúrez"

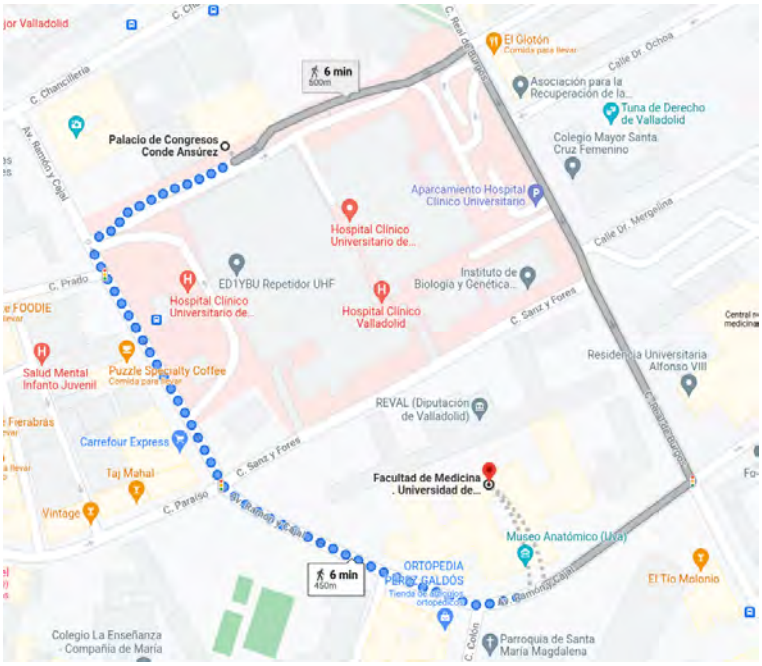
Palacio de Congresos Conde Ansúrez
Calle Real de Burgos, s/n
47011 Valladolid
Spain

The opening session, plenary conferences and round tables will take place at the 'Paraninfo', located on the 2nd floor. Some minisymposia will also be held at the Paraninfo, with the remaining sessions distributed in rooms 'Claudio Moyano', 'Felipe II', 'Cardenal Mendoza' and 'Real Chancillería'. Rooms 'Claudio Moyano' and 'Felipe II' are located on the first floor and rooms 'Cardenal Mendoza' and 'Real Chancillería' on the ground floor. You can check the distribution of minisymposia in time and space on the general schedule on page 3.

Here is a map of the conference building area.

Meals and Coffee Breaks

Lunches for conference participants will be served at the meeting, at the Cafeteria of the 'Facultad de Medicina', located at Avenida Ramón y Cajal, 7, within a short, walkable distance (a 6 minute walk according to Google Maps):



Coffee and a snack will be served during the breaks.

Welcome cocktail

A welcome cocktail will be served on Monday, November 8th, 20:30pm, at the cloister of the 'Patio Herreriano' museum (https://museoph.org/).

Conference dinner

A social dinner will take place on Wednesday 10th (21:00) at the restaurant "Río Luz", located at Plaza del Milenio, 1, Cúpula del Milenio, 47014, Valladolid (https://rioluz.es/). This dinner is not included in the registration fee. The cost of this dinner is 45 euros per person. If you want to attend to this social dinner, please, follow this link and give your personal details (name, surname and institution) as soon as possible.

https://docs.google.com/forms/d/1pPmYHusOtEoFwdsVgUIKxm3CdoWbCly0G8lyQVb1k2A/

Payment for the dinner is required in advance (only cash) and can be done at the registration desk on Monday, Tuesday or Wednesday from 9:30 to 10:00 (ask for Pedro Álvarez).

Program

Monday, November 8th

9:00-10:00 Registration **Registration desk, 2nd floor**

10:00-10:30 Opening Session **Paraninfo**

10:30-11:00 Coffee Break

11:00-13:00 MS10: New Approaches in Combinatorial Optimization **Room: Real Chancillería**

Organizers: Lluís Alsedà/Emilio Carrizosa Chair: José A. Lozano

Taxonomization of Combinatorial Optimization Problems in Fourier Space
Anne Elorza Abstract 40

Optimize your path
David Romero Abstract 109

An offline-online strategy to improve MILP performance via Machine Learning tools
Asunción Jiménez-Cordero Abstract 68

Construct, merge, solve & adapt: a new general algorithm for combinatorial optimization
José A. Lozano Abstract 73

11:00-13:00 MS13: Mathematical support to resource and process management in health services **Room: Claudio Moyano**

Organizer: Fermín Mallor Chair: Fermín Mallor

Problems and challenges of the health management
Isabel Rodrigo-Rincón Abstract 106

Using Electronic Health Record for the management of the patient flow in the Hospital Emergency Department
Marta Cildoz Abstract 31

Deployment and control of rural emergencies resources
Martín Gastón Abstract 50

Analysis of decision-making data for understanding and helping the ICU management
Daniel García de Vicuña Abstract 47

11:00-13:00 MS17: Non-linear approximation, vision and images **Room: Felipe II**

Organizers: Davide Barbieri/Eugenio Hernández Chair: Eugenio Hernández

Theoretical aspects of non-linear approximation
Eugenio Hernández Abstract 63

Analysis of the image inpainting problem using sparse multiscale representations and CNNs
Demetrio Labate Abstract 70

Neuronal models for visual perception in ambiguous visual scenes
Gemma Huguet Abstract 64

Abstract harmonic analysis and image reconstruction in primary visual cortex
Davide Barbieri Abstract 12

11:00-13:00 MS21: New perspectives in Computational Mathematics (I) Room: Cardenal Mendoza

Organizers: Tomás Chacón/Antonio Falcó Chair: Tomás Chacón

Pressure stabilization in Reduced Order Methods for fluid flow problems
Enrique Delgado Ávila Abstract 38

POD stabilized methods for incompressible flows: error analysis and computational results
Samuele Rubino Abstract 111

Tensor Networks from a Quantum Information perspective
David Pérez García Abstract 91

Mechanistic models and machine learning: friends or foes?
Elías Cueto Abstract 34

11:00-13:00 MS23: Statistical analysis of complex data (I) Room: Paraninfo

Organizer: Wenceslao González Manteiga Chair: Wenceslao González Manteiga

Tests of hyperspherical uniformity based on chordal distances
Eduardo García Portugués Abstract 48

Complex regression for complex data
Rosa M. Crujeiras Abstract 33

Kernel methods to cope with the analysis of point processes on road networks
María Isabel Borrajo Abstract 18

Topological data analysis of high-dimensional correlation structures with applications in epigenetics
Sara Prada Abstract 96

13:00-15:00 Lunch

15:00-16:00 P1: Plenary session Paraninfo

Chair: Sebastià Xambó

Prospects and Challenges of Machine Learning in the Physical World
Joan Bruna Abstract 19

16:00-17:00 P5: Plenary session **Paraninfo**

Chair: Rosa Lillo

Functional data methods for wearable device data
Jeff Goldsmith Abstract 52

20:30-22:00 **Welcome cocktail** **Museo Patio Herreriano**

Tuesday, November 9th

9:30-10:30 P2: Plenary session **Paraninfo**

Chair: Tomás Chacón

Physics-Based and Data-Driven-Based Algorithms for the Simulation of the Heart Function
Alfio Quarteroni Abstract 98

10:30-11:00 Coffee Break

11:00-13:00 RT1: Challenges and opportunities for Mathematics in Data Science: the point of view of funding agencies and policy-makers **Paraninfo**

Pilar Dellunde (Universitat Autónoma de Barcelona, IIIA)
David Gómez-Ullate (UCA Datalab Coordinator, vocal de la RSME)
Jesús Marco (Vicepresidente de Investigación y Ciencia y Tecnología del CSIC)
Enrique Playán (Director de la Agencia Estatal de Investigación)
Luis Vega (Coordinador de la REM, Chair)

13:00-15:00 Lunch

15:00-17:00 MS7: Optimal transport for data science **Room: Cardenal Mendoza**

Organizers: Juan A. Cuesta-Albertos/Eustasio del Barrio Chair: Juan A. Cuesta-Albertos

From Multivariate Quantiles to Copulas and Statistical Depth, and Back
Marc Hallin Abstract 62

Consensus-Based Interacting Particle Systems and Mean-field PDEs for Optimization and Sampling
José Antonio Carrillo de la Plata Abstract 23

Central Limit Theorems for General Transportation Costs
Alberto González Sanz Abstract 57

Optimal transport for kernel Gaussian processes
Jean-Michel Loubes Abstract 72

15:00-17:00 MS8: Adversarial Machine Learning **Room: Claudio Moyano**

Organizer: David Ríos Insua Chair: David Ríos Insua

Adversarial Machine Learning. An overview
D. Ríos Insua Abstract 104

Bayesian approaches to protecting classifiers from attacks
F. Ruggeri Abstract 113

Augmented probability simulation for optimization in adversarial machine learning
R. Naveiro Abstract 83

Adversarial machine learning for financial applications
D. García-Rasines Abstract 49

15:00-17:00 MS11: Mathematical Optimization Methods for Decision Making Room: Felipe II

Organizers: Lluís Alsedà/Emilio Carrizosa Chair: José Niño Mora

Optimization for Social Good
Helena Ramalhinho Abstract 99

A new interior-point optimization approach for support vector machines for binary classification and outlier detection
Jordi Castro Abstract 26

Identificación de redes de suministro de energía eléctrica empleando algoritmos de optimización combinatoria
Aritz Pérez Abstract 89

Data-driven dynamic priority allocation: recent advances
José Niño Mora Abstract 84

15:00-17:00 MS18: Neural networks for Mathematicians Room: Real Chancillería

Organizers: Davide Barbieri/Mar González Chair: Davide Barbieri

The continuous formulation of shallow neural networks as Wasserstein-type gradient flows
Xavier Fernández-Real Abstract 42

Generative Adversarial Networks for mathematicians
Ángel González-Prieto Abstract 55

Emergence of Lie symmetries in functional architectures learned by CNNs
Noemi Montobbio Abstract 78

Including Deep Learning into the physical modelling and simulation loop
Jaime López Abstract 71

15:00-17:00 MS22: New perspectives in Computational Mathematics (II) Room: Paraninfo

Organizers: Tomás Chacón/Antonio Falcó Chair: Antonio Falcó

Certified Reduced order Large Eddy Simulation turbulence models
Tomás Chacón Abstract 29

Hybrid Twins: Filling the gap between physics and data
Francisco Chinesta Abstract 30

Open Data Science Task Force against COVID-19: Winning the 500k XPRIZE Pandemic Response Challenge
J. Alberto Conejero Abstract 32

Lebesgue-type estimates for the Thresholding Greedy Algorithm
Pablo Berná Abstract 16

17:00-19:00 MS6: Fair learning **Room: Real Chancillería**

Organizer: Jean-Michel Loubes Chair: Jean-Michel Loubes

From learning with fair regularizers to physics aware models
Adrián Pérez-Suay Abstract 93

Mathematical frameworks for fair learning: review of methods and study of the price for fairness
Paula Gordaliza Abstract 58

Certification Aspects in Future AI-Based High-Integrity Systems
Jaume Abella and Francisco J. Cazorla Abstract 1

Attraction-Repulsion clustering: an approach to fair clustering through diversity enhancement
Hristo Inouzhe Abstract 66

17:00-19:00 MS15: Mathematical Optimization, Classification and Regression **Room: Felipe II**

Organizers: Lluís Alsedà/Emilio Carrizosa Chair: Vanesa Guerrero

On a semidefinite optimization approach to estimate smooth hypersurfaces using P-splines and shape constraints
Manuel Navarro-García Abstract 82

Linear regression analysis on probabilistic-linked data
Sandra Benítez-Peña Abstract 14

Optimal Decision Trees for Complex Data
M Cristina Molero-Río Abstract 77

On some mathematical optimization models to gain insight into complex data
Vanesa Guerrero Abstract 61

17:00-19:00 MS24: Statistical analysis of complex data (II) **Room: Paraninfo**

Organizer: Pedro Delicado Chair: Pedro Delicado

Smart visualization of mixed data
Aurea Grané Abstract 60

Statistical learning of heterogeneous data: a case study and general ideas
Lluís Belanche Abstract 15

Sparse Matrix Classification on Imbalanced Datasets Using Convolutional Neural Networks
Beatriz Pateiro Abstract 88

Finding the optimal soccer player: spatial clustering applied to scouting
Virgilio Gómez-Rubio Abstract 53

17:00-19:00 MS26: New Perspectives in Data Science **Room: Claudio Moyano**

Organizers: Eustasio del Barrio/Rosa Lillo Chair: Luis Ángel García Escudero

Mathematical and Statistical modeling using the FMM approach. The case of the Electrocardiogram.
Cristina Rueda Abstract 112

Modal clustering asymptotics
José Enrique Chacón Abstract 28

Wasserstein gradient flows for machine learning
Anna Korba Abstract 69

Robust modeling of large dimensional time series with cluster structure
Pedro Galeano Abstract 45

17:00-19:00 MS29: Functional Data Analysis (II) **Room: Cardenal Mendoza**

Organizers: Ana Aguilera/Eduardo García Portugués Chair: Ana Aguilera

Solving the multivariate functional ANOVA problem with application to environmental data from COVID-19 pandemic
Christian J. Acal Abstract 2

A goodness-of-fit test for functional time series with applications to diffusion processes
Javier Álvarez Liébana Abstract 7

An overview of functional data analysis contributions to health analytics
Francesca Ieva Abstract 65

Machine learning and statistical methods for clustering in FDA
Belén Pulido Abstract 97

Wednesday, November 10th

9:30-10:30 P4: Plenary session **Paraninfo**

Chair: Carlos Matrán

Learning through ambiguity: differentiable matchings and mappings
Marco Cuturi Abstract 37

10:30-11:00 Coffee Break

11:00-13:00 RT2: New Bridges between Mathematics and Data Science: a Scientifc Debate **Paraninfo**

Joan Bruna (Courant Institute, New York University)
From Mathematics to Artificial intelligence: prospective strategical lines.

Ilaria Chillotti (Director of Research, Zama)
Fully Homomorphic Encryption – end-to-end encryption everywhere.

Jesús María Sanz-Serna (Universidad Carlos III de Madrid)
An evolutionary survey on the mathematical contributions to Artificial Intelligence: An editor perspective.

Sebastià Xambó-Descamps (Chair, Universitat Politècnica de Catalunya)
Two-way bridges between mathematics and machine learning.

13:00-15:00 Lunch

15:00-17:00 MS1: High-dimensional Bayesian networks **Room: Felipe II**

Organizer: Concha Bielza Chair: Concha Bielza

High dimensional hybrid Bayesian networks: Is there life beyond the exponential family?
Antonio Salmerón Abstract 115

Learning and visualizing massive Bayesian networks with FGES-Merge and BayeSuites
José Luis Moreno Abstract 80

Approximation in Value-Based Potentials
Ofelia Paula Retamero Pascual Abstract 103

Convergent and fast natural gradient based optimization method DSNGD and adaptation to large dimensional Bayesian networks
Borja Sánchez-López Abstract 117

15:00-17:00 MS16: Data Science Applications **Room: Paraninfo**

Organizers: Lluís Alsedà/Emilio Carrizosa Chair: Lluís Alsedà

Client scoring for a tourism agency based on Machine Learning and Utility Theory
Víctor Aceña Gil Abstract 3

Data Science success stories
Isaac Martín de Diego Abstract 74

Graph-based approaches for document information extraction
Oriol Ramos Abstract 102

Using a Digital Twin to forecast the SARS-CoV-2 spread in Catalonia
Pau Fonseca Abstract 43

15:00-17:00 MS20: Solving inverse problems using data-driven models **Room: Real Chancillería**

Organizers: Fabricio Maciá/Pedro Caro Chair: Fabricio Maciá

Discussing the paper 'Convolutional neural networks in phase space and inverse problems' by G. Uhlmann and Y. Wang
Pedro Caro Abstract 22

Applying Neural ODE to inverse problems
Pablo Angulo Abstract 9

Wavelet Analysis of the generalized Riemann non-differentiable Function
Luz Roncal Abstract 110

Recommender systems in action
Angel González Prieto Abstract 56

15:00-17:00 MS28: ML and NLP models: from notebook to production deployment **Room: Cardenal Mendoza**

Organizer: David Gómez-Ullate Chair: David Gómez-Ullate

Speech recognition in legal and medical contexts
María Jose Cano Vicente Abstract 21

How Natural Language Processing is helping in Defence and aerospace
Alexandra Aguilar Torres and Jesús Alberto Villa Diez Abstract 5

Zero-shot learning in extremely large Transformer models (GPT and CLIP). Mathematical and computational aspects
Víctor Gallego Alcalá Abstract 46

Data science and Machine Learning for the fishing industry
Alberto Torres Abstract 121

15:00-17:00 MS30: Data Science in Action **Room: Claudio Moyano**

Organizer: Emilio Carrizosa Chair: Adriana Rojas

El reto de aplicar IA en el marco de la Industria 4.0
Juan Jesús Pardo Expósito Abstract 86

Empower and inspire with the most trusted analytics
Adriana Rojas Abstract 108

Tributos inteligentes
César Pérez López Abstract 92

Data Scientificus
Carlos Rivero Antonio Abstract 105

17:00-19:00 MS2: Functional Data Analysis (I) **Room: Cardenal Mendoza**

Organizers: Ana Aguilera/Eduardo García Portugués Chair: Eduardo García Portugués

Penalized methods for functional data with variable domain: application to chronic obstructive pulmonary disease
M. Carmen Aguilera Morillo Abstract 6

On an alternative formulation of the functional logistic model
Antonio Cuevas Abstract 36

Functional depth: Recent progress and perspectives
Stanislav Nagy Abstract 81

Object Oriented Spatial Statistics (O2S2) for densities: an application to the analysis of mortality from all causes in Italy during the COVID-19 pandemic
Piercesare Secchi Abstract 118

17:00-19:00 MS4: Interpretability and explainability of algorithms **Room: Claudio Moyano**

Organizers: Alexandra Cifuentes/Iñaki Úcar Chair: Iñaki Úcar

Explaining Bayesian networks using MAP-independence: Some new properties
Enrique Valero-Leal Abstract 122

Opening the black-box of deep learning architecture with Ranked-LRP
José Luis Salmerón Abstract 116

Can neural networks be explained using polynomial regressions and Taylor series?
Pablo Morala Abstract 79

Counterfactual Explanations via Mathematical Optimization
Jasone Ramírez-Ayerbe Abstract 100

17:00-19:00 MS5: High-dimensional variable selection **Room: Felipe II**

Organizer: Pepa Ramírez-Cobo Chair: Pepa Ramírez-Cobo

Ensemble distance-based regression and classification for large sets of mixed-type data
Amparo Baíllo Abstract 11

Bayesian methods for variable selection. Challenges of the XXI Century.
Anabel Forte Deltell Abstract 44

fPQR: A quantile based dimension reduction technique for regression
Álvaro Méndez Civieta Abstract 76

Variable selection for Naïve Bayes classification
Pepa Ramírez-Cobo Abstract 101

17:00-19:00 MS14: Mathematical Optimization for Data-Driven Decision-Making **Room: Real Chancillería**

Organizers: Rafael Blanquero/Emilio Carrizosa Chair: Víctor Blanco

The Workforce Management Challenge. A mathematical perspective
Iker Barriales Abstract 13

An unsupervised machine learning algorithm to transform waste to biogas
Rocío Vega Martínez Abstract 123

Machine Learning defines innovation
Nuria Gómez-Vargas Abstract 54

Mathematical Optimization approaches to supervised learning with noisy labels
Víctor Blanco Abstract 17

17:00-19:00 MS19: Machine learning techniques in control theory and inverse problems **Room: Paraninfo**

Organizers: Carlos Castro/Francisco Periago Chair: Carlos Castro

Simultaneous control of Neural differential equations
Domenec Ruız-Balet Abstract 114

Optimal control of deep neural networks
Enrique Zuazua Abstract 125

Machine learning algorithms in inverse problems
Carlos Castro Abstract 25

A first step towards numerical approximation of controllability problems via Deep-Learning-based methods
Francisco Periago Abstract 94

21:00-22:30 Social dinner

Thursday, November 11th

9:30-10:30 P2: Plenary session **Paraninfo**

Chair: Emilio Carrizosa

Challenges and improvements in optimization algorithms for machine learning
Coralia Cartis Abstract 24

10:30-11:00 Coffee Break

11:00-13:00 RT3: Mathematics, Data Science and transfer: the point of view of industry **Paraninfo**

Manuel Castro (Presidente de SeMA)
Álvaro Fernández Velando (Head Models & Data, Santander España)
Jesús López-Fidalgo (Presidente de SEIO)
Peregrina Quintela (Chair, Vicepresidenta de EU-MATHS-IN)
Rocío Vega (Investigadora en Reganosa)

13:00-15:00 Lunch

15:00-17:00 MS3: Spatio-temporal Data Science **Room: Claudio Moyano**

Organizer: Lola Ugarte Chair: Lola Ugarte

Calibration of Spatial Forecasts from Citizen Science Urban Air Pollution Data with Sparse Recurrent Neural Networks
Stefano Castruccio Abstract 27

Scalable Bayesian models for spatio-temporal count data
Aritz Adin Abstract 4

DeepKriging: Spatially Dependent Deep Neural Networks for Spatial Prediction
Ying Sun Abstract 120

Large-Scale Spatial Data Science with ExaGeoStat
Marc Genton Abstract 51

15:00-17:00 MS9: Probabilistic Learning **Room: Paraninfo**

Organizer: Santiago Mazuelas Chair: Santiago Mazuelas

Minimax Classification with 0-1 Loss and Performance Guarantees
Santiago Mazuelas Abstract 75

What if causal models were imprecise?
Rafael Cabañas Abstract 20

Concentric Mixtures of Mallows Models for Top-k Rankings
Ekhine Irurozki Abstract 67

Learning decomposable models by coarsening
Aritz Perez Abstract 90

15:00-17:00 MS12: Decision aid and data science models for disaster management **Room: Real Chancillería**

Organizer: Begoña Vitoriano Chair: Begoña Vitoriano

Evacuation and supply distribution facing a natural disaster
M. Teresa Ortuño Abstract 85

Optimisation models for wildfire suppression
Bibiana Granda Abstract 59

Wildfire risk measurement for fuel management decision-making using stochastic scenarios and Bayesian networks
Adán Rodríguez Martínez Abstract 107

Strategic and tactical preparedness in humanitarian logistics based on scenario generation from historical data
Begoña Vitoriano Abstract 124

15:00-17:00 MS25: Digital Twins **Room: Felipe II**

Organizers: Iván Area/Adrián Fernández/Francisco Fernández Chair: Francisco J. Fernández

Concept and Solution of Digital Twin based on a Stieltjes Ordinary Differential Equation
Iván Area Abstract 10

A perspective on digital twins from the point of view of Stieltjes Parabolic Partial Differential Equations
Francisco J. Fernández Abstract 41

On the use of scientific inductive biases for the construction of digital (hybrid) twins
Elías Cueto Abstract 35

Digital Twin for the human cornea: curvature estimation
Lukasz Plociniczak Abstract 95

Stieltjes Bochner spaces and strong damping wave equation: application to digital twin of discrete dynamic systems
Jean-Daniel Djida Abstract 39

15:00-17:00 MS27: Mathematical Optimization in Industry **Room: Cardenal Mendoza**

Organizers: Lluís Alsedà/Emilio Carrizosa Chair: Francisco Parreño

Grid operation-based outage maintenance planning
Ramón Álvarez-Valdés Abstract 8

A mathematical model for the slitting problem in a Spanish steel industry
María Sierra-Paradinas Abstract 119

Solving a large cutting problem in the glass manufacturing industry
Francisco Parreño Abstract 87

List of Abstracts

Abstract Index

Abstracts

1 Certification Aspects in Future AI-Based High-Integrity Systems

Jaume Abella and Francisco J. Cazorla

Barcelona Supercomputing Center

The trend towards increased autonomy functions in high-integrity systems, like those in planes and cars, causes disruptive changes to the certification process. At software level, the challenge relates to the increasing use of Artificial Intelligence (AI) based software to provide the levels of accuracy required. At the hardware level, it relates to the use of high-performance heterogeneous multi-core processors to provide the required level of computing performance and the impact multi-cores have on functional safety including software timing aspects. In this talk we will cover some of the main challenges brought by both, AI software and multi-cores, to the certification process of high-integrity systems. We will also discuss potential research paths to address those challenges.

Session MS6

2 Solving the multivariate functional ANOVA problem with application to environmental data from COVID-19 pandemic

Christian J. Acal (joint work with Ana M. Aguilera, Annalina Sarra, Adelia Evangelista, Tonio Di Battista and Sergio Palermi)

Universidad de Granada

The analysis of variance problem for functional data (FANOVA) concerns to test the equality of several mean functions. This statistical technique is highly used in many field of sciences such as medicine, environment or engineering, in which the experimental data are usually functions (curves or images) instead of vectors. Even though there are many works from a univariate perspective, there is a lot to be done in the multivariate context (more than one functional response variable in the analysis). At present, the most common procedures to solve the multivariate FANOVA problem are focused on permutation random projections tests. In this talk, a new approach based on multivariate functional principal component analysis is introduced. Specifically, the aim is to solve the multivariate FANOVA problem through testing the equality of the mean vectors of the most explicative principal components scores. Parametric and non-parametric procedures are considered depending on whether the multivariate normality is suitable or not. Besides, the statistics available in the literature for FANOVA problem with repeated measures (the information of each subject is measured in different periods of time or under different conditions) are extended by assuming a basis representation of the sample curves. Both methodologies has been used to better understand the behaviour of air pollution in the Region of Abruzzo (Italy) during the COVID-19 pandemic. In particular, the temporal evolution of concentrations of four pollutants is measured in two different periods of time (before and during lockdown period established by the Italian Government) for several monitoring stations classified by their location (background and traffic stations). The objective is to detect possible differences, on the one hand, between both timespan and, on the other hand, between the geographical situations of stations.

Session MS29

3 Client scoring for a tourism agency based on Machine Learning and Utility Theory

Víctor Aceña Gil

Universidad Rey Juan Carlos

A travel agency's main resource is its agents. They provide quotes to all clients who request them, whether the purchase is consolidated or not. However, depending on the purpose of the trip, this can be time-consuming for the agent and waste time and money for the agency. Typically, a manager is responsible for allocating each budget to one agent or another, based on his or her expert judgement and knowledge of each agent's capabilities. This talk will present a scoring method that uses utility functions based on the expected net profit of each potential customer. The net profit per travel package is easily estimated, taking into account the cost per agent over all the days it takes to prepare each quote and the price of the travel package. In addition, a machine learning model estimates the probability of purchase for each customer needed to construct the utility function.

Session MS16

4 Scalable Bayesian models for spatio-temporal count data

Aritz Adin (joint work with Erick-Orozco Acosta and María Dolores Ugarte)

Universidad Publica de Navarra

Spatio-temporal disease mapping studies the geographical distribution of a disease in space and its evolution in time. Many statistical techniques have been proposed during the last years for analyzing disease risks, most of them including spatial and temporal random effects to smooth risks borrowing information from neighbouring regions and time periods. Despite the enormous expansion of modern computers and the development of new software and estimation techniques to make fully Bayesian inference, dealing with massive data is still computationally challenging. In this work, we propose a scalable Bayesian modeling approach to smooth mortality or incidence risks in high-dimensional spatio-temporal disease mapping context. The method is based on the well-known "divide and conquer" approach, so that local models can be simultaneously fitted reducing the computational time substantially. Model fitting and inference is carried out using the well-known integrated nested Laplace approximation (INLA) technique. The methods and algorithms proposed in this work are being implemented in the R package "bigDM" available at https://github.com/spatialstatisticsupna/bigDM. We illustrate the models's behaviour by estimating lung cancer mortality risks in almost 8000 municipalities of Spain during the period 1991-2015. A simulation study is also conducted to evaluate the performance of this new scalable modeling approach in comparison with usual spatio-temporal models in disease mapping.

Session MS3

5 How Natural Language Processing is helping in Defence and aerospace

Alexandra Aguilar Torres and Jesús Alberto Villa Diez

Airbus DS

Airbus is a global aerospace-and-defense corporation known for developing military and commercial aircraft. Like many other traditional industries, it is delving into a digital. In this process, AI and machine learning techniques are playing a key role. We'll provide an overview of the different areas where NLP techniques are being used at Airbus Defense & Space, their benefits, challenges and lessons learnt, as well as the technical approach for one of our projects.

Session MS28

6 Penalized methods for functional data with variable domain: application to chronic obstructive pulmonary disease

M. Carmen Aguilera Morillo (joint work with Pavel Hernández Amaro and María Durban)

Universitat Politècnica de València

Most statistical techniques for functional data analysis have been developed for situations where all functions have the same domain. However, many real datasets do not stand for this assumption and then, new approaches to estimate functional regression models for functional data with variable domain are required. In this work we focus on variable-domain functional regression models, which estimation is based on basis representations with B-splines and a discrete penalty. This research is motivated by a real study, carried out in collaboration with the Hospital de Galdakao (Vizcaya) and the Universidad del País Vasco, whose objective is to study the impact of physical activity in patients with Chronic Obstructive Pulmonary Disease on the progression of the disease in terms of the number of hospitalisations.

Session MS2

7 A goodness-of-fit test for functional time series with applications to diffusion processes

Javier Álvarez Liébana (joint work with Alejandra López Pérez, Wenceslao González and Manolo Febrero Bande)

Universidad Complutense de Madrid

Within the burgeoning Functional Data Analysis framework, the analysis of intra-day high-frequency

data is currently one of the topics of greatest interest in financial research. In this context, the Functional Linear Model with Functional Response is one of the most relevant models to assess the relation between two functional random variables. A particular case arises when functional responses are given by their own past values, in which functional errors and responses are (linearly) correlated. In this talk, a novel goodness-of-fit test for autoregressive Hilbertian (ARH) models is presented. Furthermore, we also provide a new specification test for stochastic diffusion models, such as Ornstein-Uhlenbeck processes, illustrated with an application to intra-day currency exchange rates. In particular, a two-stage methodology is proffered: firstly, we check if functional samples and their past values are related via ARH(1) model; secondly, under linearity, we perform a functional F-test.

Session MS29

8 Grid operation-based outage maintenance planning

Ramón Álvarez-Valdés (joint work with Francisco Parreño)

University of Valencia

RTE (Réseau de Transport d'Électricité) is the operator of the French electricity transmission system, with a network of 100,000 km. When planning maintenance operations, some interventions require the power supply to be cut off. When this happens, the electricity supply must be guaranteed, so maintenance operations must be carefully planned. To tackle this issue, RTE decided to apply a three-step approach. First, risk values are calculated for different future scenarios. Second, these computed values are used to find a good schedule. Eventually, a third step validates the obtained planning. Our optimization problem arises in the second step of this approach: given the risk values, the goal is to find an optimal planning regarding a risk-based objective. Moreover, this planning must be consistent with all job-related restrictions such as resource constraints. The objective function includes the average risk, over time and scenarios, and a measure of the cost variability, expressed by a quantile of the risk distribution. Our approach generates first a set of good solutions by solving integer linear models whose objective functions are approximations of the actual objective of the problem. These solutions then go through an improvement phase, which includes a Variable Neighborhood Search and a Path Relinking algorithm. The computational study, on a set of instances provided by the company, shows that the complete procedure obtains high quality solutions.

Session MS27

9 Applying Neural ODE to inverse problems

Pablo Angulo

Universidad Politécnica de Madrid

Neural ODE are the natural evolution of ResNets, allowing for very deep learning neural networks. Evaluation of the Neural ODE amounts to the numerical integration of a ODE system and the gradient of the loss function can be obtained through the adjoint method instead of backpropagation. We survey the applications of this technique to inverse problems, and the caveats that must be taken into account in order to get meaningful answers, with a special focus on continuous normalizing flows.

Session MS20

10 Concept and Solution of Digital Twin based on a Stieltjes Ordinary Differential Equation

Iván Area (joint work with Francisco J. Fernandez, Juan J. Nieto and F. Adrian F. Tojo)

Universidade de Vigo

In this work we introduce the concept of a Digital Twin by using Stieltjes Ordinary Differential Equations (SODE). A precise mathematical definition of solution to the problem is presented. We also analyze the existence and uniqueness of solutions and introduce the concept of Main Digital Twin. As a particular case, the classical compartmental SIR (Susceptible, Infected, Recovered) epidemic model is considered and we study the interrelation between the digital twin and the system. In doing so, we use Stieltjes derivatives to feed the data from the real system to the virtual model which, in return, improves it in real time. Numerical simulations with real data of the COVID-19 epidemic, show the accuracy of the proposed ideas

Session MS25

11 Ensemble distance-based regression and classification for large sets of mixed-type data

Amparo Baíllo

Universidad Autónoma de Madrid

The distance-based linear model (DB-LM) extends the classical linear regression to the framework of mixed-type predictors or when the only available information is a distance matrix between regressors (as it sometimes happens with big data). The main drawback of these DB methods is their computational cost, particularly due to the eigendecomposition of the Gram matrix. In this context, ensemble regression techniques provide a useful alternative to fitting the model to the whole sample. This work analyzes the performance of three subsampling and aggregation techniques in DB regression on two specific large, real datasets. We also analyze, via simulations, the performance of bagging and DB logistic regression in the classification problem with mixed-type features and large sample sizes.

Session MS5

12 Abstract harmonic analysis and image reconstruction in primary visual cortex

Davide Barbieri

Universidad Autónoma de Madrid

Human vision has inspired several advances in harmonic analysis, especially wavelet analysis, and it has been the main source of heuristics for the development of neural network architectures devoted to image processing. One of the most studied neural structures in brain's visual cortex is area V1, where neurons perform a wavelet-like analysis that is generally considered to be associated with the group structure of rotations and translations. It is indeed possible to model part of the perceptual behavior of the network of neural cells in V1 as a projection of the image onto one, or more, orbits of that group, and consequently associate to each neuron in V1 a parameter of the group. However, due to the physical constraint of having a neural displacement onto a two dimensional layer, the group is not fully, nor uniformly, represented in V1. The represented subset of the group has however a characteristic geometric structure, that has been modeled over the physiological measurements of what are called orientation preference maps. A natural question posed by this empirical observation is whether the missing part of the group, and of the corresponding wavelet coefficients, has perceptual consequences, and if, on the other hand, it is possible to recover or estimate in some stable way the missing information. The ability to perform such a task would allow one to effectively learn a full group representation from a partial set of well positioned detectors. We will propose such a mechanism, and briefly discuss its possible physical implementation.

Session MS17

13 The Workforce Management Challenge. A mathematical perspective.

Iker Barriales (joint work with Paula Terán)

MAPAL-OS

One of the most complex and limiting jobs a company operations manager must deal with is Workforce Management: forecasting the future activity, translating this activity into the staff and work needs, making sure the company workforce is well designed and dimensioned and scheduling the staff correctly is key to provide a good service while keeping costs under control. This job is particularly challenging in Hospitality and other similar sectors. In Mapal OS the Optimisation team within the OPT&DS department has developed a mathematical model that provides the best weekly schedule for a certain location. With an expected workload and a given available staff, the model provides back the best schedule in terms of cost and demand fitness subject to all the operational and legal constraints applicable.

Session MS14

14 Linear regression analysis on probabilistic-linked data

Sandra Benítez-Peña (joint work with Rafael Blanquero, Emilio Carrizosa and Pepa Ramírez-Cobo)

Universidad Carlos III de Madrid

Data linkage is a task used for merging data sets that contain information of the same entities, but lack of unique identification codes. Real datasets come from an exact matching, however, the procedure of data merging does not need to be exact: a single entity can be linked to two or more instances if they are similar enough. In this talk, we present a novel Non-Linear Programming model that integrates, in a single formulation, the task of obtaining a probabilistic matching and that also performs linear regression using such obtained linked data. Numerical results are presented for both simulated and real data sets, demonstrating the power of our methodology. Also, heuristics for providing good initial solutions are presented here.

Session MS15

15 Statistical learning of heterogeneous data: a case study and general ideas

Lluís Belanche

Universitat Politècnica de Catalunya

In data analysis, it is known that the chosen data representation is a crucial factor for a successful learning process, and yet current practice advocates for a change in representation, to adapt the data to the chosen modeling technique, instead of otherwise. Kernel methods offer a principled way for statistical learning when confronted with mixed data, even when faced with added difficult situations, like missing values. In this contribution we illustrate these assertions with the study of a difficult problem, the Horse Colic data set. Moreover, we give some advice and general ideas on the matter.

Session MS24

16 Lebesgue-type estimates for the Thresholding Greedy Algorithm

Pablo Berná

Universidad CEU San Pablo

The approximation theory with respect to bases in Banach spaces consists in the study of different ways to approximate a function by a finite linear combination of elements of that basis. The idea behind the nonlinear approximation theory is that the elements used in the approximation do not come from a prefixed vector space. The Thresholding Greedy algorithm builds approximations of each function by selecting the largest coefficients (in absolute value) in the series expansion with respect to the basis. In that talk we present new results about the efficiency of the greedy algorithm.

Session MS22

17 Mathematical Optimization approaches to supervised learning with noisy labels

Víctor Blanco (joint work with Alberto Japón and Justo Puerto)

Universidad de Granada

The primary goal of supervised classification is to find patterns from a training sample of labeled data in order to predict the labels of out-of-sample data, in case the possible number of labels is finite. Among the most relevant applications of classification methods are those related with security, as in spam filtering or intrusion detection. The main difference of these applications with respect to other uses of classification approaches is that malicious adversaries can adaptively manipulate their data to mislead the outcome of an automatic analysis. In this work we propose novel methodologies to optimally construct classifiers that take into account that label noises occur in the training sample. We propose different alternatives based on solving Mixed Integer Linear and Non Linear models by incorporating decisions on relabeling some of the observations in the training dataset. This feature is adequatelly embedded into different types of optimization-based classifiers, as SVM or Decision Trees. Extensive computational experiments are reported based on a battery of standard datasets taken from UCI Machine Learning repository, showing the effectiveness of the proposed approaches.

Session MS14

18 Kernel methods to cope with the analysis of point processes on road networks

María Isabel Borrajo

Universidad de Santiago de Compostela

In this talk we explain a statistically principled method for kernel smoothing of point pattern data on a linear network when the first-order intensity depends on covariates. In particular, we present a consistent kernel estimator for the first-order intensity function that uses a convenient relationship between the intensity and the density of events location over the network, which also exploits the theoretical relationship between the original point process on the network and its transformed process through the covariate. The performance of the estimator is analysed through a simulation study under inhomogeneous scenarios. We also present a real data analysis on wildlife-vehicle collisions in a region of North-East of Spain.

Session MS23

19 Prospects and Challenges of Machine Learning in the Physical World

Joan Bruna

Courant Institute, New York University

The last decade has witnessed an experimental revolution in data science, led by the huge empirical success of deep learning methods across many areas of science and engineering. In order to capitalise on these successes, it has become increasingly important to provide a mathematical foundation that gives guiding design principles, and mitigates the current data 'hunger' of these DL architectures, to enable further applications within computational science. In this talk, we will describe the crucial role that data structure plays in constructing such foundations. Existing mathematical models are mostly agnostic to data structure, and as a result rely on strong assumptions in order to break the curse of dimensionality. Alternatively, we will present a geometrical perspective that unifies all successful DL architectures (CNNs, RNNs, Transformers, GNNs) from the principles of symmetry and scale separation, providing a viable mathematical picture where the curse of dimensionality is avoided under more realistic assumptions. We will cover theoretical analyses that highlight the role of data structure, as well as applications of these models to computational science, emphasizing the incipient area of 'Scientific Machine Learning'.

Session P1

20 What if causal models were imprecise?

Rafael Cabañas

Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA), Lugano, Switzerland

Causality is currently an emerging direction for data science with a wealth of potential applications in diverse domains such as Artificial Intelligence, Economics, Social Science or Medicine. Pearl's structural causal models are a natural formalism for causal inference, in particular for their appealing graphical representation. However, the peculiar features of causal models may render them not always easy to access to a traditional audience, which is instead familiar with pre-existing graphical tools and related procedures. Structural causal models can be then transformed into equivalent credal networks. This means that every query on the causal model can be reformulated as a query on the imprecise model, which can then be solved by standard algorithms for the latter. Moreover, this also allows producing bounds for unidentifiable queries.

Session MS9

21 Speech recognition in legal and medical contexts

María Jose Cano Vicente

Vócali

Transcribing audio speech segments into text has many different applications these days. From simple transcription, subtitles, keyword searching in media files. . . to dictation, human machine interaction and all kinds of Natural Language Processing applications. All of them have in common necessities and difficulties. Speech recognition is typically based on different pieces to model different parts of the natural understanding process of a speech. It is necessary to process sound into a numerical representation that allows associate frames of audio to elemental parts. If these elemental parts are phonemes, they must be bundled into words among a lexicon, which can be general or very specific to the domain. To enhance probability

of choosing word combinations into coherent phrases also language is modelled. Several approximations are making good results over last years. Classical approximations are improving their performances even in low resources cases, as well as new varieties of sequence-to-sequence models are integrating the latest advances in neural networks architectures. Vócali INVOXMedical and INVOXLegal make it possible to adapt these speech recognition technologies into domain-specific lexicons and speech contexts.

Session MS28

22 Discussing the paper "Convolutional neural networks in phase space and inverse problems" by G. Uhlmann and Y. Wang

Pedro Caro

BCAM

In this talk I will discuss the results on the paper "Convolutional neural networks in phase space and inverse problems" by G. Uhlmann and Y. Wang. The goal will be to analyse if some of these ideas could be transferred to the resolution of some other inverse problems.

Session MS20

23 Consensus-Based Interacting Particle Systems and Mean-field PDEs for Optimization and Sampling

José Antonio Carrillo de la Plata

Mathematical Institute, University of Oxford

We will start by doing a quick review on consensus models for swarming. Stability of patterns in these models will be briefly discussed. Then we provide an analytical framework for investigating the efficiency of a consensus-based model for tackling global optimization problems. We justify the optimization algorithm in the mean-field sense showing the convergence to the global minimizer for a large class of functions. An efficient algorithm for large dimensional problems is introduced. Theoretical results on consensus estimates will be illustrated by numerical simulations.

We then develop these ideas to propose a novel method for sampling and also optimization tasks based on a stochastic interacting particle system. We explain how this method can be used for the following two goals: (i) generating approximate samples from a given target distribution, and (ii) optimizing a given objective function. This approach is derivative-free and affine invariant, and is therefore well-suited for solving complex inverse problems, allowing (i) to sample from the Bayesian posterior and (ii) to find the maximum a posteriori estimator. We investigate the properties of this family of methods in terms of various parameter choices, both analytically and by means of numerical simulations.

This talk is a summary of works in collaboration with Y.-P. Choi, O. Tse, C. Totzeck, F. Hoffmann, A. Stuart and U. Vaes.

Session MS7

24 Challenges and improvements in optimization algorithms for machine learning

Coralia Cartis

University of Oxford

We will discuss some key challenges to optimization algorithm development arising from machine earning. In particular, we investigate dimensionality reduction techniques in the variable/parameter domain for local and global optimization; these rely crucially on random projections. We describe and use sketching results that allow efficient projections to low dimensions while preserving using properties, as well as other useful tools from random matrix theory and conic integral geometry. We focus on functions with low effective dimensionality – that are conjectured to provide an insightful proxy for neural networks landscapes. Time permitting we also discuss first- versus second-order optimization methods for training, and/or stochastic variants of classical optimization methods that allow biased noise, adaptive parameters and have almost sure convergence.

Session P2

25 Machine learning algorithms in inverse problems

Carlos Castro

Universidad Politécnica de Madrid

Neural networks software have serious difficulties to solve specific simple inverse problems concerning partial differential equations. We illustrate such difficulties and show how mathematical analysis for such problems can improve their efficiency

Session MS19

26 A new interior-point optimization approach for support vector machines for binary classification and outlier detection

Jordi Castro

Universitat Politècnica de Catalunya

In this work we present a new interior-point optimization method for the solution of 2-class and 1-class linear support vector machines (SVMs), which are, respectively, used for binary classification and outlier detection. Unlike previous interior-point approaches for SVMs, which were only practical when the dimension of the points was small, the new proposal can also deal with high-dimensional data. The new approach is compared with state-of-the-art solvers for SVMs, either based on interior-point algorithms (such as SVM-OOPS), or specific algorithms developed by the machine learning community (such as LIBSVM and LIBLINEAR).

Session MS11

27 Calibration of Spatial Forecasts from Citizen Science Urban Air Pollution Data with Sparse Recurrent Neural Networks

Stefano Castruccio

University of Notre Dame, USA

With their continued increase in coverage and quality, data collected from personal air quality monitors has become an increasingly valuable tool to complement existing public health monitoring systems over urban areas. However, the potential of using such 'citizen science data' for automatic early warning systems is hampered by the lack of models able to capture the high-resolution, nonlinear spatio-temporal features stemming from local emission sources such as traffic, residential heating and commercial activities. In this work, we propose a machine learning approach to forecast high-frequency spatial fields which has two distinctive advantages from standard neural network methods in time: 1) sparsity of the neural network via a spike-and-slab prior, and 2) a small parametric space. The introduction of stochastic neural networks generates additional uncertainty, and in this work we propose a fast approach for forecast calibration, both marginal and spatial. We focus on assessing exposure to urban air pollution in San Francisco, and our results suggest an improvement of 35.7% in the mean squared error over standard time series approach with a calibrated forecast for up to 5 day.

Session MS3

28 Modal clustering asymptotics

José Enrique Chacón

Universidad de Extremadura

In nonparametric density-based clustering, clusters are understood as regions of high concentration of probability mass, separated from each other by regions of lower density. Therefore, clusters are naturally associated to density modes and this approach is called modal clustering. The population goal of modal clustering can thus be defined in terms of the domains of attraction of the true density modes, and that allows framing the clustering problem in a standard inferential setting. In this talk we show some recent results concerning the asymptotic properties of data-based modal clusterings, constructed via the usual plug-in methodology, employing a density estimator. Limit theorems are shown for the unidimensional case, but their multivariate extensions stand out as a challenging open problem.

Session MS26

29 Certified Reduced order Large Eddy Simulation turbulence models

Tomás Chacón (joint work with Cristina Caravaca, Enrique Delgado Ávila and Macarena Gómez)

Universidad de Sevilla

This talk deals with the construction of reduced-order turbulence models with targeted error levels. We consider Large Eddy Simulation (LES) models of Smagorinsky kind, for which a complete mathematical and numerical analysis is known. This analysis allows the rigorous derivation of a-posteriori error estimators are built. On this basis, reduced basis are built by greedy algorithms, yielding error below targeted levels (certified method). We present the mathematical derivation of the a posteriori error estimators, as well as some application to benchmark flows as well as applications to thermal analysis of transition spaces in buildings.

Session MS22

30 Hybrid Twins: Filling the gap between physics and data

Francisco Chinesta

ENSAM ParisTech

World is changing very rapidly. Today we do not sell aircraft engines, but hours of flight, we do not sell an electric drill but good quality holes, . . . We are nowadays more concerned by performances than by the products themselves. Thus, the new needs imply focusing on the real system subjected to the real loading that it experienced until the present time in order to predict the future responses and in this manner, anticipate any fortuity event or improve the performances. Here, usual modeling and simulation techniques are limited because of the fact that a model is sometimes no more than a crude representation of the reality. Artificial Intelligence irrupted and became a major protagonist in many areas of technology and society at the beginning of the third millennium, however many times it requires impressive training efforts (incredible amount of data, most of them inexistent, difficult to collect and manipulate, extremely expensive in time and resources). A highway to circumvent these difficulties and successfully accomplishing the most efficient (fast, accurate and frugal) generation of information and knowledge facilitating a real-time decision-making in engineering in general, and in forming processes in particular, consists of a hybrid paradigm combining real-time physics ad real-time physics-aware data-driven modelling.

Session MS22

31 Using Electronic Health Record for the management of the patient flow in the Hospital Emergency Department

Marta Cildoz (joint work with Fermín Mallor)

Research group q-UPHS

Emergency Departments (EDs) work in a stochastic environment, with unplanned patient arrivals and unknown healthcare resources necessary for patient treatment. These characteristics, in addition to the increasing demand, lead to overcrowding problems. The flow of patients in the ED is impacted by two phases of sequencing decisions: assigning each patient to a physician after the triage, and determining the order in which patients are seen once they are under the responsibility of a physician. In this talk, we analyze the use of electronic health records to improve patient flow at both stages. First, we propose new rules to equitably assign patients to physicians taking into account not only patients waiting time by priority but also physicians' stress and workload. The improvement achieved by the new proposals is investigated by using simulation models. We also report the success of an intervention conducted at the Hospital Compound of Navarre. Second, investigating new queue disciplines to assist the physicians to select the next patient to be seen among those waiting for their first consultation or their second one (after some necessary diagnostic medical tests were carried out). The management of the physician's portfolio of patients has to accomplish several objectives: not to exceed the first consultation waiting time limit, minimize the length of stay of each patient and minimize the number of patients in the ED (giving different importance to the patients according to their severity score).

Session MS13

32 Open Data Science Task Force against COVID-19: Winning the 500k XPRIZE Pandemic Response Challenge

J. Alberto Conejero

Universitat Politècnica de València

When the COVID-19 arrive to Spain, the Valencian Government created a Data Science Task Force to fight the pandemics, where the scientific community (through the Group of Experts) collaborate with the public administration (through the Commissioner at the level of the Presidency. After some time in which data was scarce and hard to obtain, we achieve to develop accurate computational epidemiological models that were complemented with human mobility studies, and information from a citizen survey called COVID19 impact survey.

Our work has received national and international recognition, including being the global winners of the 500k XPRIZE Pandemic Response Challenge, a four-month global competition organized by the XPRIZE Foundation. The challenge had two main goals: The first one was to foster the development of advanced AI models to forecast the evolution of the pandemics by combining different data sources. The second one was to prescribe Non-Pharmaceutical Intervention Plans that governments, business leaders and organizations could implement to minimize harm when reopening their economies. We will briefly describe these models and how information systems can feed these models to help against the pandemics.

Session MS22

33 Complex regression for complex data

Rosa M. Crujeiras

Universidad de Santiago de Compostela

There is a diverse range of practical situations where one may encounter random variables which are not defined on Euclidean spaces, as it is the case for circular data. Circular measurements may be accompanied by other observations, either defined on the unit circumference or on the real line, and in such cases it may be if interest to model the relationship between the variables from a regression perspective. It is not infrequent that parametric models fail to capture the underlying model given their lack of flexibility, but it may also happen that the usual paradigm of (classical) mean regression. We will present in this talk some recent advances in nonparametric multimodal regression, showing an adaptation of the mean-shift algorithm for regression scenarios involving circular response and/or covariate. Real data illustrations will be also presented. This is a joint work with María Alonso-Pena.

Session MS23

34 Mechanistic models and machine learning: friends or foes?

Elías Cueto (joint work with Quercus Hernández, Beatriz Moya, Alberto Badías, Iciar Alfaro, David Gonzalez and Francisco Chinesta)

Universidad de Zaragoza

In this talk we will explore the interplay between well-known mechanistic physical laws and data science in the framework of the fourth paradigm of science. While the former have proved their success for centuries, they are also well-known to be difficult to distill, maintain, validate and apply, due to their inherent computational cost in many cases. In the last years we face an increasing interest in the leverage of the capabilities of data science to obtain predictive surrogates to these mechanistic models. However, the validity of these black-box surrogates is always under scrutiny: sensitivity to noise in the data, extrapolation capability, compliance with existing models, . . . It can be shown, however, that first principles can be easily incorporated into the learning machinery, giving rise to a promising family of techniques that satisfy by construction these physical laws. For instance, it is straightforward to impose a symplectic structure for systems that are conservative, thus leading to a learning procedure that guarantees energy conservation. But it is not so easy to develop learning methods for dissipative systems: what is the appropriate framework for them? We will show that imposing a metriplectic structure to the learning system guarantees the satisfaction of the laws of thermodynamics, thus opening the possibility of developing systems able to learn autonomously and still preserve the physics of the system under scrutiny. Examples will be show that prove the interest of such an approach.

Session MS21

35 On the use of scientific inductive biases for the construction of digital (hybrid) twins

Elías Cueto

Universidad de Zaragoza

Hybrid twins are a particular type of digital twins able to detect systematic biases between their predictions and experimental data, and therefore to correct themselves. For this to be possible, they must be able to learn from data this discrepancy. To this end, we employ scientific machine learning, and particularly, physics-informed neural networks. Trying to avoid as much as possible their black-box character, we employ inductive biases constructed by well-known laws of physics, and particularly, the"physics of physics": thermodynamics.

We show how a neural network constructed so as to take into account the laws of thermodynamics outperforms classical neural networks in this learning procedure, while minimizing the risk of wrong predictions, that are subjected to fulfill conservation of energy and non-negative entropy production by construction.

Session MS25

36 On an alternative formulation of the functional logistic model

Antonio Cuevas (joint work with José R. Berrendero, Beatriz Bueno-Larraz, Antonio Coín)

Universidad Autónoma de Madrid

The problem of predicting a binary response Y from a functional explanatory variable $X = X(t)$ arises very often in practice. A common approach, considered by several authors in the recent literature, is the L^2-based functional logistic model. We explore here an alternative approach based on the theory of Reproducing Kernel Hilbert Spaces. We will show how this alternative model offers some theoretical and practical advantages in terms of generality (as it encompasses, as particular cases, many easy-to-interpret models, including the L^2-one) and ease of estimation of the involved parameters.

Session MS2

37 Learning through ambiguity: differentiable matchings and mappings

Marco Cuturi

Google Brain/ENSAE, Institut Polytechnique de Paris

Data points (or, more generally, entire datasets) studied under the lens of ML undergo shifts of all types. In many machine learning tasks, such as fair classification, domain adaptation or robustness against attacks, crafting classifiers that can handle such shifts is crucial. In several fields of science (such as single cell genomics or neuroscience) understanding and modeling the dynamics of such shifts is of the essence. For all of these examples, a canonical strategy is to lift the ambiguity inherent to such shifts through some form of matching, assignment, registration step that helps build correspondences across datasets (i.e., match a cell observed at time t1 with another at time t2; an individual from group A to another in minority B, an image in setup 1 to another in setup 2, a word or a sentence in a language, paragraph to another in a different language, paragraph). These correspondences can be costly to compute, and even more so difficult to differentiate; it can be also challenging to extend them to out-of-sample settings. I will show in this work how recent ideas in the fields of optimal transport and differentiable programming have joined forces to provide scalable software solutions, providing new ways to learn through the ambiguity of shifts with the flexibility of end-to-end learning pipelines. Entropic regularization, quantile transforms, input-convex neural networks and the implicit function theorem will each play a role in these approaches.

Session P4

38 Pressure stabilization in Reduced Order Methods for fluid flow problems

Enrique Delgado Ávila

Universidad de Sevilla

In this work we present a Reduced Basis Model for a pressure stabilized Finite Element fluid flow. We perform the construction of an a posteriori error estimator for the selection of the basis functions via the Greedy algorithm, and we discuss the consideration of the inner pressure supremizer for the pressure recovery. In our model, we deal with some non-linearities that we solve in the Reduced Order framework with the Empirical Interpolation Method. Finally, we present some numerical results in which we show the speed-up in the computation of the reduced basis solution.

Session MS21

39 Stieltjes Bochner spaces and strong damping wave equation: application to digital twin of discrete dynamic systems

Jean-Daniel Djida

African Institute for Mathematical Sciences (AIMS), Limbe, Cameroon

Dynamic and nanosystems are well understood across engineering and data science and represent a convenient platform for exploring the various aspects of a digital twin design. The aim is to create a mathematical framework accessible to engineering sciences related to mechatronics, quantum machine learning, mechanical and computational systems with impulses. The virtual model of those prototypes of a physical system is expressed as a second-order differential equation in two-time scales. The concept of a slow time is used to separate the evolution of the system properties from the instantaneous time. The first part of this discussion is devoted to the mathematical analysis of a strong damping wave equation with Stieltjes time derivative in Stieltjes Bochner spaces. This novel formulation allows us to study harmonic oscillators type equations that involve impulses when the system does not evolve. We present several theoretical results related to the existence of a solution. In the second part, we employ a discrete damped dynamic system to investigate the emerging concept of a digital twin and show some concrete applications.

Session MS25

40 Taxonomization of Combinatorial Optimization Problems in Fourier Space

Anne Elorza

Universidad del País Vasco UPV/EHU

In the field of permutation-based Combinatorial Optimization Problems, those classified as NP-hard represent a major challenge, since the cost of exact algorithms becomes prohibitive. As an alternative, metaheuristic algorithms have been proposed. However, there still exists a major difficulty in their application: given a specific problem instance, and considering the great variety of possible algorithms, how could we select the most appropriate algorithm for solving it? A first step to try to solve this problem would be to create a taxonomy that groups together problem instances that can be solved efficiently by the same algorithms. In this talk, we explain the theoretical framework that we have adopted in order to construct such a taxonomy, by making use of the Fourier characteristics of each problem instance. As with the classical Fourier transform over the real line, which decomposes a function into a sum of sines and cosines, the Fourier transform over the symmetric group decomposes a permutation-based function into a linear combination of basis functions. Therefore, an objective function can be described through its Fourier coefficients, and we plan to use this information to create the taxonomy.

Session MS10

41 A perspective on digital twins from the point of view of Stieltjes Parabolic Partial Differential Equations

Francisco J. Fernández (joint work with Iván Area, Juan J. Nieto and F. Adrián F. Tojo

Instituto de Matemáticas, USC

In this work we introduce the concept of a Digital Twin by using Stieltjes Parabolic Partial Differential Equations (SPPDE). We present a mathematical definition of the solution and we also analyze the existence and uniqueness of solution. The advantage of considering the spatial variable in the mathematical model allows us to study situations in which ordinary differential equations are not adequate, such as heat transfer problems in N-dimensional domains, fluid mechanics problems, population dynamics problems where the spatial distribution is relevant, etc.

Session MS25

42 The continuous formulation of shallow neural networks as Wasserstein-type gradient flows

Xavier Fernández-Real

EPFL

It has been recently observed that the training of a single hidden layer artificial neural network can be reinterpreted as a Wasserstein gradient flow for the weights for the error functional. In the limit, as the number of parameters tends to infinity, this gives rise to a family of parabolic equations. This talk aims to discuss this relation, focusing on the associated theoretical aspects appealing to the mathematical community and providing a list of interesting open problems.

Session MS18

43 Using a Digital Twin to forecast the SARS-CoV-2 spread in Catalonia

Pau Fonseca

Universitat Politècnica de Catalunya

We explore a Digital Twin approach to model the spread of SARS-CoV-2 in Catalonia. Our Digital Twin is composed of three different dynamics models. These three models are used to perform validation using the Model Comparison approach. In this talk, we will discuss the Digital Twin structure, and how we use the Validation process to obtain knowledge from the system. This allows us to understand the effects of the nonpharmaceutical interventions. To simplify the maintenance of the dynamic compartmental model for the SARS-CoV-2 spread forecast we use Specification and Description Language (SDL) to represent it. This simplifies the model assumptions understanding by the different specialists involved in the Digital Twin maintenance and use; assumptions that must be validated continuously following a Solution Validation approach. We will discuss the Digital Twin adoption in the decision-making process and the implications of the discussion based on models.

Session MS16

44 Bayesian methods for variable selection. Challenges of the XXI Century

Anabel Forte Deltell

Universitat de Valencia

Model selection and, in particular Variable selection is, without doubt, one of the most difficult procedures in science. Along history it has been approached from different points of view as well as from different paradigms such as Frequentist or Bayesian statistics. Specifically in this talk we will review how Bayesian Statistics can deal with variable selection, trying to understand the advantages of this paradigm. Also we will try to point to the new challenges that the Era of high dimensional data adds to this already difficult task and how Bayes may deal with it.

Session MS5

45 Robust modeling of large dimensional time series with cluster structure

Pedro Galeano (joint work with Andrés M. Alonso and Daniel Peña)

Universidad Carlos III de Madrid

Large dimensional time series can be appropriately modeled with dynamic factor models. However, these data often have heterogeneity and cluster structure and the formulation and estimation of dynamic factor models should be adapted to these features. This article presents a procedure to fit Dynamic Factor Models with Cluster Structure (DFMCS), where some of the factors are global and others group-specific, to heterogeneous data that may include multivariate additive outliers and level shifts. The procedure starts with an initial cleaning of the times series from outlying effects. Then a first estimation of the possible factors is applied to the cleaned data and these factors are used to build the common component of each series. The groups are found by studying the joint dependency of these common components. Then additional factors are estimated by using the series in each cluster and, finally, all the factors found are classified as global or group-specific. We show in a Monte Carlo study that the procedure works well and seems to be better than other alternatives in terms of estimation of factors and loadings as well as in terms of misclassification rates for the series. An example of an electricity market is presented to illustrate the advantages of cleaning for outliers and taking into account the cluster structure for understanding and forecasting.

Session MS26

46 Zero-shot learning in extremely large Transformer models (GPT and CLIP). Mathematical and computational aspects

Víctor Gallego Alcalá

ICMAT & Komorebi AI

The rise of neural models such as BERT, GPT-3 or CLIP, trained on huge amounts of data at scale, has led an undergoing paradigm shift in Artificial Intelligence. These deep learning models, leveraged with transfer learning, have been proved to be adaptable to a wide range of downstream tasks in both Natural Language Processing and Computer Vision. Traditionally, models were pretrained on a large corpus of data, and then fine-tuned on a specific dataset and task. However, scaling up language models vastly improves few- (and zero-)shot performance in different tasks, sometimes reaching comparable results to the state of the art. In this talk, after reviewing the underlying the mathematical aspects of these models, we will showcase several approaches towards zero/few-shot learning, such as prompt engineering or prompt tuning. Then, we will show several industrial applications, like text generation for content creation and SEO optimization, and semantic search for navigating large datasets of raw, non-annotated images. A demo of the language model can be found at http://api.vicgalle.net:8000/

Session MS28

47 Analysis of decision-making data for understanding and helping the ICU management

Daniel García de Vicuña (joint work with Laida Esparza and Fermín Mallor)

Research group q-UPHS

Management Flight Simulators allow researchers to study decision-making in real-time by requesting input from participants. Using a web-based model which recreates a real ICU, we simulate the arrival and clinical evolution (by using 275 variables) of two types of patients (emergency and scheduled patients). The user manages the simulated ICU by deciding about their admission or diversion and which inpatients are discharged. The data collected consists of a sequence of decisions that physicians made about the admission and discharge of patients over the period the ICU is simulated. Each decision made by a physician affects the situation of the ICU in which the following decisions are made. Therefore, every physician makes decisions under unique ICU scenarios, which difficult the comparison. In this talk, we present the simulator and several ways of performing the data analysis of the recorded data to characterize how physicians' decisions are made.

Session MS13

48 Tests of hyperspherical uniformity based on chordal distances

Eduardo García Portugués

Universidad Carlos III de Madrid

We provide a general and tractable family of tests of uniformity on the hypersphere of arbitrary dimension. The family is constructed from powers of the chordal distances between pairs of observations. The asymptotic null distributions of the new family of tests are obtained, as well as their explicit powers against sequences of generic local alternatives. The family of tests connects and extends three especially interesting particular cases. Numerical experiments corroborate the theoretical results. Two real data applications on the two-dimensional sphere are given.

Session MS23

49 Adversarial machine learning for financial applications

D. García-Rasines (joint work with C. Guevara and S. Rodríguez-Santana)

ICMAT

Numerous business applications entail dynamic competitive decision environments under uncertainty. We shall sketch how adversarial machine learning methods

may be used in such domains, illustrating the ideas with problems in relation to pension funds, loans and the stock market.

Session MS8

50 Deployment and control of rural emergencies resources

Martín Gastón (joint work with Daniel García de Vicuña, Marta Cildoz, Laura Frías, Cristina Azcárate and Fermín Mallor)

Research group q-UPHS

Healthcare managers are challenged with providing and planning rural services upon the current context of changes in demographic, in communication networks and of new ways of providing healthcare (medicalized ambulances, helicopters, etc.). In this work we address the problem of reorganizing the continuous and urgent healthcare resources in rural areas. We consider an existing service network organization and we propose an optimization model to determine the geographic location and opening hours of the fixed care centers with the support of mobile resources (ambulances), with the double aim of providing a care of quality while keeping the cost at minimum. The model rationalizes the scarce resources as well as guaranteeing quality criteria measured by patient travel times and balancing the workload of the fixed centers. We propose an integer lineal optimization model that extends classic location problems by integrating time-dependent demand and both fixed and mobile resources. This model has been applied to the current rural healthcare network in a region of Spain. Besides, we have implemented a graphical visualization tool of the solutions to help stakeholders in the analysis and understand of them. Therefore, it serves to health managers decision process to choose the preferred solution to be implemented in practice.

Session MS13

51 Large-Scale Spatial Data Science with ExaGeoStat

Marc Genton

KAUST University

Spatial data science aims at analyzing the spatial distributions, patterns, and relationships of data over a predefined geographical region. For decades, the size of most spatial datasets was modest enough to be handled by exact inference. Nowadays, with the explosive increase of data volumes, High-Performance Computing (HPC) can serve as a tool to handle massive datasets for many spatial applications. Big data processing becomes feasible with the availability of parallel processing hardware systems such as shared and distributed memory, multiprocessors and GPU accelerators. In spatial statistics, parallel and distributed computing can alleviate the computational and memory restrictions in large-scale Gaussian process inference and prediction. In this talk, we will describe cutting-edge HPC techniques and their applications in solving large-scale spatial problems with the new software ExaGeoStat.

Session MS3

52 Functional data methods for wearable device data

Jeff Goldsmith

Columbia University

In the last ten years, technological advances have made many activity- and physiology-monitoring wearable devices available for use in large-scale epidemiological studies. This trend is likely to continue and even expand as devices become cheaper and more reliable. These developments open up a tremendous opportunity for clinical and public health researchers to collect critical data at an unprecedented level of detail, while posing new challenges for statistical analysis of rich, complex data. This talk will present a collection of approaches in functional data analysis for identifying and interpreting variability in activity trajectories within and across participants, for building regression models in which activity trajectories are the response, and for understanding shifts in the circadian rhythms that underly the timing of activity. We'll draw on several applications, including the Baltimore Longitudinal Study of Aging and data collected through the Columbia Center for Children's Environmental Health.

Session P5

53 Finding the optimal soccer player: spatial clustering applied to scouting

Virgilio Gómez-Rubio (joint work with Jesús Lagos)

Universidad de Castilla-La Mancha

Soccer teams face the problem of replacing players throughout the season. This is often due to injuries or some players leaving the team. Looking for new players is known as 'scouting' and it is a challenging problem as many times specific characteristics in the players are required, which means that a large number of characteristics need to be compared. From a statistical point of view, this problem can be tackled in a number of ways. If the desired player's characteristics can be expressed as a (numerical) vector, then a distance can be defined so that the player with the smallest distance to the desired characteristics is the desired match. However, there are other restrictions that may apply such as players already under a contract, etc. One of the characteristics that defines a players role is the location in the field, as this is indicative of the main position within the team. Modern devices allow recording this position throughout the game, so that this can be exploited to develop 'spatial profiles' for the players. However, clustering these spatial profiles may be difficult due to a number of problems: (spatially) correlated data, different levels of spatial and temporal aggregation, etc. We have developed a novel way of exploiting location information about the players' location in the field by means of spatial statistical methods. In particular, we have used an estimate of the time spent at every position in the field (obtained with a specific personal device) so that we can compare any two players by means of Lee's test of spatial autocorrelation. The p-values obtained with these tests are then used as similarity functions in a hierarchical cluster so that different groups of players can be identified. We have applied this method to more than 4000 soccer player's profiles from different soccer leagues worldwide.

Session MS24

54 Machine Learning defines innovation

Nuria Gómez-Vargas (joint work with Rafael Blanquero Bravo, Elisa Isabel Caballero Ruiz, Emilio Carrizosa Priego, Marina Enguidanos Weyler, Ana Gema Galera Pozo and Jasone Ramírez-Ayerbe)

Instituto de Matemáticas de la Universidad de Sevilla (IMUS)

The presence of companies on the internet has been fundamental for theirgrowth in recent years. For this reason, the exploitation of their webpages isproposed as a way of characterizing them. However, the vast magnitude ofthe variables that can be extracted from these sites makes their treatment aproblem.In this respect, we have developed a machine learning tool in order tocharacterize the innovation of a company. First, we have defined a preprocessingstep applying text mining techniques to the respective webpages, followed bydifferent dynamics of grouping and selecting words and html tags that bringout their relevance. Finally, we classify companies according to their innovationusing random forests. With this methodology, we obtain not only a distinctionbetween companies that are innovative or not, but also a definition of innovationaccording to the importance of the variables.

Session MS14

55 Generative Adversarial Networks for mathematicians

Ángel González-Prieto (joint work with A. Mozo, E. Talavera and S. Gómez-Canaval)

Universidad Complutense de Madrid

Since their inception, Generative Adversarial Networks (GANs) have revolutionized the field of generative models due to their flexibility and ability to generate fully synthetic samples of very complex phenomena with high resolution. However, as they lie in the half-way between mathematics and engineering, sometimes is hard to unravel the mathematical properties of GANs and to translate them to implementations. In this talk, we shall review the mathematical fundamentals of GANs, with special attention on how GANs are formulated as a competitive game and their optima as Nash equilibria. We will comment some of the known results about the convergence of GANs and their relation to the minimization of the Jensen-Shannon divergence and optimum transport problems. Time permitting, we will discuss some of the recent developments in the study of the GAN convergence. In particular, we will focus on the interplay between the topology of the parameter space and the induced dynamical system, as well as the use of probabilistically inspired activation functions to improve the accuracy and convergence of GANs.

Session MS18

56 Recommender systems in action

Angel González Prieto (joint work with Jesús Bobadilla, Raúl Lara-Cabrera and Fernando Ortega)

UCM/ICMAT

In the present-day information society, people are exposed to a massive amount of data from different sources. When we want to watch a series, the streaming platforms offer thousands of possibilities; when we want to travel abroad, search engines return hundreds of suitable flights with multiple companies; when we want to go out for dinner, innumerable restaurants are proposed through the booking platforms. This continuous bombing of information is certainly overwhelming. To sort out this mess, recommender systems arose as machine learning models able to find the right item to be recommended to any user. Since their very inception, recommender systems have been a very active research area whose results have been quickly incorporated to almost all costumer-focused platforms such as Netflix, Spotify, Facebook, Amazon, Tinder... In this talk, we will review the fundamental concepts and models of collaborative filtering-based recommender systems. This are state-of-art methods which, in a way or another, encode an inverse problem, namely, to extract the fundamental latent features of both users and items and to analyze how these hidden characteristics affect the recommendation. In particular, we shall focus on the main two approaches: matrix factorization-based systems and deep learning-based models, reaching some of our most recent proposals in both trends.

Session MS20

57 Central Limit Theorems for General Transportation Costs

Alberto González Sanz

Institut de Mathématiques de Toulouse and ANITI

One of the main ways to quantify the distance between distributions is the well known Wasserstein metric. In Statistics and Machine Learning applications it is increasingly common to deal with measures supported on a high dimensional space. Some recents results show that the Wasserstein metric suffers from the curse of dimensionality, which means that its empirical approximation becomes worse as dimension grows. We will explain a new method based on the Efron-Stein inequality and on the sequential compactness of the closed unit ball in $L^2(P)$ for the weak topology that improves a result of del Barrio and Loubes (2019) and states that, even if the empirical Wasserstein metric converges with slow rate, its oscillations around its mean are asymptotically Gaussian with rate $\sqrt{n}$, n being the sample size, which means that the curse of dimensionality is avoided in such a case. Finally, we will present some applications of these results to statistical and data science problems.

Session MS7

58 Mathematical frameworks for fair learning: review of methods and study of the price for fairness

Paula Gordaliza

BCAM

A review of the main fairness definitions and fair learning methodologies proposed in the literature over the last years is presented from a mathematical point of view. Following an independence-based approach, we consider how to build fair algorithms and the consequences on the degradation of their performance compared to the possibly unfair case. This corresponds to the price for fairness given by the criteria statistical parity or equality of odds. Novel results giving the expressions of the optimal fair classifier and the optimal fair predictor (under a linear regression gaussian model) in the sense of equality of odds are presented.

Session MS6

59 Optimisation models for wildfire suppression

Bibiana Granda (joint work with Javier León, Begoña Vitoriano and John W. Hearne)

Universidad Complutense de Madrid

Wildfires are recurrent natural disasters that are increasing in frequency and severity over the last decades, threatening human lives and damaging ecosystems and infrastructure, leading to high recovery costs. To deal with wildfires, several activities must be

managed and coordinated in order to develop a suitable response that is both effective and affordable, considering the resources available and the safety of the personnel involved. This includes actions taken before (mitigation, prevention, and preparedness), during (response) and after the event (recovery). In the response phase of a wildfire management scheme two main problems can be distinguished: deployment and dispatch of resources. A review of models and methodologies that, applying operations research and optimization techniques, deal with the management of the two problems mentioned above will be presented.

Session MS12

60 Smart visualization of mixed data

Aurea Grané

Universidad Carlos III de Madrid

In this work, we propose a new protocol that integrates robust classification and visualization techniques to analyze mixed data, which is based on the combination of the Forward Search Distance-Based (FS-DB) algorithm and robust clustering. The methodology is illustrated on a real dataset related to European COVID-19 numerical health data, as well as policy and restriction measurements of the 2020-2021 COVID-19 pandemic across the EU Member States.

Session MS24

61 On some mathematical optimization models to gain insight into complex data

Vanesa Guerrero

Universidad Carlos III de Madrid

Mathematical Optimization plays a crucial role to extract knowledge from data and cope with nowadays requirements in decision making processes. The increase in data complexity has made, in some cases, the classical statistical tools obsolete and more sophisticated frameworks are thus needed. In particular, dimensionality reduction techniques demand an update to face the new challenges posed by different data structures and to make the new features interpretable. In this talk, we review some mathematical optimization approaches which have helped to enhance the interpretability of the low-dimensional embeddings produced by different dimensionality reduction techniques and in different contexts.

Session MS15

62 From Multivariate Quantiles to Copulas and Statistical Depth, and Back

Marc Hallin

ECARES and Department of Mathematics, Université libre de Bruxelles

The univariate concept of quantile function – the inverse of a distribution function – plays a fundamental role in Probability and Statistics. In dimension two and higher, however, inverting traditional distribution functions does not lead to any satisfactory notion. In their quest for the Grail of an adequate definition, statisticians dug out two extremely fruitful theoretical pathways: copula transforms, where marginal quantiles are privileged over global ones, and depth functions, where a center-outward ordering is substituting the more traditional South-West/North-East one. We show how a recent center-outward redefinition, based on measure transportation ideas, of the concept of distribution function reconciles and fine-tunes these two approaches, and eventually yields a notion of multivariate quantile matching, in arbitrary dimension d, all the properties that make univariate quantiles a successful and vital tool of statistical inference.

Session MS7

63 Theoretical aspects of non-linear approximation

Eugenio Hernández

Universidad Autónoma de Madrid

A popular tool in non-linear approximation is the Greedy Algorithm used to approximate a signal efficiently by a finite number of coefficients. I will review the main results concerning this Algorithm, focusing in showing conditions to ensure a fixed rate of convergence.

Session MS17

64 Neuronal models for visual perception in ambiguous visual scenes

Gemma Huguet

Universidad Politécnica de Cataluña

When observers view for an extended time an ambiguous visual scene (admitting two or more different interpretations), they report spontaneous switching between different perceptions. The most studied case is perceptual bistability (two interpretations), which includes binocular rivalry (alternation of two different images, one presented to each eye), but there are other cases in which ambiguous images may show phenomena of tristability and much more complex dynamics. Models of multistable perception include models with multiple attractors and with heteroclinic cycles. In both models, noise is added to account for the irregular oscillations. In this talk, we will discuss the main features of these models and we will show how they can account for the dynamical properties (transition probabilities, distributions of percept durations, etc) observed in the experiments. Finally, we discuss the role of noise and we show that in the heteroclinic network models it can be replaced by quasi-periodic perturbations, assuming that the system is receiving events, either internal or from other brain areas, that include only a finite number of (incongruent) frequencies.

Session MS17

65 An overview of functional data analysis contributions to health analytics

Francesca Ieva

Politecnico di Milano

The healthcare setting often presents situations where dynamic monitoring of biological or vital signals is required, or models for longitudinal observations and covariates are needed. In these cases, Functional Data Analysis (FDA) may be used as a proficient support to precision medicine, since it allows for developing powerful models which account not only for baseline or cross sectional information, but also for the dynamic of the process. In this talk, an overview of clinical applications where models exploiting FDA techniques are used will be presented, with the aim of highlighting FDA potential in supporting clinical practice.

Session MS29

66 Attraction-Repulsion clustering: an approach to fair clustering through diversity enhancement

Hristo Inouzhe

BCAM

We consider the problem of diversity enhancing clustering, i.e, developing clustering methods which produce clusters that favour diversity with respect to a set of protected attributes such as race, sex, age, etc. In the context of fair clustering, diversity plays a major role when fairness is understood as demographic parity. To promote diversity, we introduce perturbations to the distance in the unprotected attributes that account for protected attributes in a way that resembles attraction-repulsion of charged particles in Physics. These perturbations are defined through dissimilarities with a tractable interpretation. Cluster analysis based on attraction-repulsion dissimilarities penalizes homogeneity of the clusters with respect to the protected attributes and leads to an improvement in diversity. An advantage of our approach, which falls into a preprocessing set-up, is its compatibility with a wide variety of clustering methods and whit non-Euclidean data. We illustrate the use of our procedures with both synthetic and real data and provide discussion about the relation between diversity, fairness, and cluster structure.

Session MS6

67 Concentric Mixtures of Mallows Models for Top-k Rankings

Ekhine Irurozki

LTCI, Telecom Paris, Institut Polytechnique de Paris

Mixtures of two Mallows models for top-k rankings with equal location parameters but with different scale parameters arise when we have a heterogeneous population of voters formed by two populations, one of which is a subpopulation of expert voters. They are denoted as concentric mixtures of Mallows models. We show the identifiability of both components and the learnability of their respective parameters. These results are based upon, first, bounding the sample complexity for the Borda algorithm with top-k rankings. Second, we characterize the distances between rankings, showing that an off-the-shelf clustering algorithm separates the rankings by components with high probability -provided the scales are well-separated. As a by-product, we include an efficient sampling algorithm for Mallows top-k rankings. Finally, since the rank aggregation will suffer from a large amount of noise introduced by the non-expert voters, we adapt the Borda algorithm to be able to recover the ground truth consensus ranking which is especially consistent with the expert rankings.

Session MS9

68 An offline-online strategy to improve MILP performance via Machine Learning tools

Asunción Jiménez-Cordero (joint work with Juan Miguel Morales and Salvador Pineda)

Universidad de Málaga

Solving large-scale Mixed Integer Linear Problems (MILP) is well known to be a challenging task. To alleviate their computational burden, several works in the literature have proposed Machine Learning techniques to identify and remove constraints. However, all these techniques report that a non-negligible percentage of the obtained solutions are infeasible since they violate some of the removed constraints. This talk presents an offline-online strategy that improves the quality of the available data to significantly reduce the number of infeasible solutions. By linking Mathematical Optimization and Machine Learning, our approach leads to substantial performance improvements in terms of feasibility and computational time, which we demonstrate through synthetic and real-life MILP problems.

Session MS10

69 Wasserstein gradient flows for machine learning

Anna Korba

ENSAE, Paris

An important problem in machine learning and computational statistics is to sample from an intractable target distribution, e.g. to sample or compute functionals (expectations, normalizing constants) of the target distribution. This sampling problem can be cast as the optimization of a dissimilarity functional, seen as a loss, over the space of probability measures. In particular, one can leverage the geometry of Optimal transport and consider Wasserstein gradient flows for the loss functional, that find continuous path of probability distributions decreasing this loss. Different algorithms to approximate the target distribution result from the choice of the loss, a time and space discretization; and results in practice to the simulation of interacting particle systems. Motivated in particular by two machine learning applications, namely bayesian inference and optimization of big neural networks, we will present recent convergence results obtained for algorithms derived from Wasserstein gradient flows.

Session MS26

70 Analysis of the image inpainting problem using sparse multiscale representations and CNNs

Demetrio Labate

University of Houston

Image inpainting is an image processing task aimed at recovering missing blocks of data in an image or a video. In this talk, I will show that sparse multiscale representations offer both an efficient algorithmic framework and a well-justified theoretical setting to address the image inpainting problem. I will start by formulating inpainting in the continuous domain as a function interpolation problem in a Hilbert space, by adopting a formulation previously introduced by King et al. [2014]. As images found in many applications are

dominated by edges, I will assume a simplified image model consisting of distributions supported on curvilinear singularities. I will prove that the theoretical performance of image inpainting depends on the microlocal properties of the representation system, namely exact image recovery is achieved if the size of the missing singularity is smaller than the size of the structure elements of the representation system. A consequence of this observation is that a shearlet-based image inpainting algorithm – exploiting their microlocal properties – significantly outperforms a similar approach based on more traditional multiscale methods. In the second part of the talk, I will apply this theoretical observation to improve a state-of-the-art algorithm for blind image inpainting based on Convolutional Neural Networks.

Session MS17

71 Including Deep Learning into the physical modelling and simulation loop

Jaime López

Repsol

This talk will cover deep learning algorithms for PDE surogate modelling, with examples and emphasis on practical applications, PDE constrained optimization, physical-constrained Deep Learning based PDE solvers, state and challenges.

Session MS18

72 Optimal transport for kernel Gaussian processes

Jean-Michel Loubes

Institut de Mathématiques de Toulouse and ANITI

We propose to define Gaussian Processes indexed by multidimensional distributions. In the framework where the distributions can be modeled as i.i.d realizations of a measure on the set of distributions, we prove that the kernel defined as the quadratic distance between the transportation maps, that transport each distribution to the barycenter of the distributions, provides a valid covariance function. In this framework, we study the asymptotic properties of this process, proving micro ergodicity of the parameters.

Session MS7

73 Construct, merge, solve & adapt: a new general algorithm for combinatorial optimization

Jose A. Lozano

BCAM

In this talk we present Construct, Merge, Solve & Adapt (CSMA): A Recent Hybrid Approach for Combinatorial Optimization. This algorithm provides a means for taking profit from exact techniques (such as, for example, general-purpose integer linear programming (ILP) solvers) in the context of problem instances that are much too large for solving them with the exact technique directly. In this presentation, we introduce the algorithm and show its successful application in the solution of several combinatorial optimization problems.

Session MS10

74 Data Science success stories

Isaac Martín de Diego

ICMAT, CSIC

Data science is defined at the intersection of three broad areas: mathematics, computer science and an application domain. Typically, the academic environment provides a high degree of expertise in the first two, and a low degree of interaction with industry. In this talk we address some of the success stories that the data science laboratory of the King Juan Carlos University has achieved in domains as diverse as cattle breeding, health, the chemical energy sector, telecommunications, and tourism.

Session MS16

75 Minimax Classification with 0-1 Loss and Performance Guarantees

Santiago Mazuelas

Basque Center for Applied Mathematics (BCAM)

Supervised classification techniques use training samples to find classification rules with small expected 0-1 loss. Conventional methods achieve efficient learning and out-of-sample generalization by minimizing surrogate losses over specific families of rules. This talk presents minimax risk classifiers (MRCs) that do not rely on a choice of surrogate loss and family of rules. MRCs achieve efficient learning and out-of-sample generalization by minimizing worst-case expected 0-1 loss w.r.t. uncertainty sets that are defined by linear constraints and include the true underlying distribution. In addition, MRCs' learning stage provides performance guarantees as lower and upper tight bounds for expected 0-1 loss. We also present MRCs' finite-sample generalization bounds in terms of training size and smallest minimax risk, and show their competitive classification performance w.r.t. state-of-the-art techniques using benchmark datasets.

Session MS9

76 fPQR: A quantile based dimension reduction technique for regression

Álvaro Méndez Civieta (joint work with M. Carmen Aguilera-Morillo and Rosa E. Lillo)

Universidad Carlos III de Madrid

Partial least squares (PLS) is a well known dimensionality reduction technique used as an alternative to ordinary least squares (OLS) in collinear or high dimensional scenarios. Being based on OLS estimators, PLS is sensitive to the presence of outliers or heavy tailed distributions. Opposed to this, quantile regression (QR) is a technique that provides estimates of the conditional quantiles of a response variable as a function of the covariates. The usage of the quantiles makes the estimates more robust against the presence of heteroscedasticity or outliers than OLS estimators. In this work, we introduce the fast partial quantile regression algorithm (fPQR), a quantile based technique that shares the main advantages of PLS: it is a dimension reduction technique that obtains uncorrelated scores maximizing the quantile covariance between predictors and responses. But additionally, it is also a robust, quantile linked methodology suitable for dealing with outliers, heteroscedastic or heavy tailed datasets. The median estimator of the PQR algorithm is a robust alternative to PLS, while other quantile levels can provide additional information on the tails of the responses.

Session MS5

77 Optimal Decision Trees for Complex Data

M Cristina Molero-Río (joint work with R. Blanquero, E. Carrizosa and D. Romero Morales)

IMUS - Instituto de Matemáticas de la Universidad de Sevilla

In this talk, we tailor optimal decision trees to deal with complex data including functional data. A compromise between prediction accuracy and interpretability is sought. Whilst fitting the tree model with first- and higher-order information of the functional data provided by their derivatives, the detection of a reduced number of time intervals that are critical for prediction, as well as the control of their width, is performed through the inclusion of LASSO-type regularization terms. The resulting optimization problem is formulated as a nonlinear continuous model with linear constraints. We illustrate the performance of our approach on real-world datasets.

Session MS15

78 Emergence of Lie symmetries in functional architectures learned by CNNs

Noemi Montobbio

Italian Institute of Technology

Convolutional Neural Networks (CNNs) are a powerful tool providing outstanding performances on image classification tasks, based on an architecture designed in analogy with information processing in biological visual systems. The functional architectures of the early visual pathways have often been described in terms of geometric invariances, and several studies have leveraged this framework to investigate the analogies between CNN models and biological mechanisms. Remarkably, upon learning on natural images, the translation-invariant filters of the first layer of a

CNN have been shown to develop as approximate Gabor functions, resembling the orientation-selective receptive profiles found in the primary visual cortex (V1). With a similar approach, we modified a standard CNN architecture to insert computational blocks compatible with specific biological processing stages, and studied the spontaneous development of approximate geometric invariances after training on natural images. In particular, inserting a pre-filtering step mimicking the Lateral Geniculate Nucleus (LGN) led to the emergence of a radially symmetric profile well approximated by a Laplacian of Gaussian, which is a well-known model of receptive profiles of LGN cells. Moreover, we introduced a lateral connectivity kernel acting on the feature space of the first network layer. We then studied the learned connectivity as a function of relative tuning of first-layer filters, thus re-mapping it into the roto-translation space. This analysis revealed orientation-specific patterns, which we compared qualitatively and quantitatively with established group-based models of V1 horizontal connectivity.

Session MS18

79 Can neural networks be explained using polynomial regressions and Taylor series?

Pablo Morala

Universidad Carlos III de Madrid

While neural networks are one of the main actual trends in machine learning and artificial intelligence, they are still considered not easily interpretable and therefore they are usually referred as black boxes. Here we present a new approach to this problem by finding a relationship between the weights of a trained feed forward neural network and the coefficients of a polynomial regression that performs almost equivalently as the original neural network. This is achieved through Taylor expansion at the activation functions of each neuron, and then the resulting expressions are joint in order to obtain a combination of the original network weights that are associated with each term of a polynomial regression. The order of this polynomial regression is determined by the order used in the Taylor expansion and the number of layers in the neural network. This proposal has been empirically tested covering a wide range of different situations, showing its effectiveness and opening the door to extending this methodology to a more broad range of types of neural networks. This kind of relationship between modern machine learning techniques and more traditional statistical approaches can help solve interpretability concerns and provide new tools to develop their theoretical foundations. In this case, polynomial regression coefficients have a much easier interpretation than neural network weights and it significantly reduces the number of parameters.

Session MS4

80 Learning and visualizing massive Bayesian networks with FGES-Merge and BayeSuites

Jose L. Moreno (joint work with Nikolas Bernaola, Pedro Larrañaga and Concha Bielza)

Technical University of Madrid, Spain

In this work we present a new algorithm, FGES-Merge, for learning massive Bayesian networks of the order of tens of thousands of nodes by using properties of the topology of the network and improving the parallelization of the arc search procedure. We use the algorithm to learn a network for the full human genome using expression data from the brain and to aid with the interpretation of the results, we present the BayeSuites web tool, which allows for the visualization of the network and gives a GUI for inference and search over the network avoiding the typical scalability problems of networks of this size.

Session MS1

81 Functional depth: Recent progress and perspectives

Stanislav Nagy

Charles University, Prague

The depth is a tool of nonparametric statistics. Its objective is to generalise quantiles, rankings, and orderings to multivariate and non-Euclidean data. While a rich body of literature on various depths and depth-like procedures exists, many open problems still stimulate research in the area. We consider the depth of random functions. We revisit the very definition of the standard depths for functional data and introduce procedures allowing adaptive selection of a depth in functional data analysis. Secondly, we draw connections of the func-

tional depth research with topics firmly established in the statistical machine learning literature.

Session MS2

82 On a semidefinite optimization approach to estimate smooth hypersurfaces using P-splines and shape constraints

Manuel Navarro-García

Universidad Carlos III de Madrid

In this talk, we address the problem of estimating smooth hypersurfaces in a regression problem for data lying on large grids, and where the fit of the data has to satisfy shape constraints such as non-negativity or monotonicity in a certain direction. We assume that the smooth hypersurface to be estimated is defined through a tensor product of reduced-rank basis (B-splines) and fitted by means of P-splines. In order to incorporate these requirements, a semidefinite programming approach is developed which, for the first time, successfully conveys out-of-range constrained forecasting. The usefulness of our methodology is illustrated in simulated and real data related to demography as well as data arising in the context of the COVID-19 pandemic.

Session MS15

83 Augmented probability simulation for optimization in adversarial machine learning

R. Naveiro (joint work with T. Ekin and A. Torres)

ICMAT

Adversarial machine learning from an adversarial risk analysis perspective entails a cumbersome computational procedure in which one first simulates from the attacker problem to forecast attacks and then includes such forecasts in the Defender problem to be optimized. We shall present how the procedure may be streamlined with the aid of augmented probability simulation approaches.

Session MS8

84 Data-driven dynamic priority allocation: recent advances

José Niño Mora

Universidad Carlos III de Madrid

This talk will present recent advances on data-driven dynamic priority allocation models based on the restless bandit framework and on dynamic priority indices. The focus is on partially observed Markov decision models with a Bayesian data-incorporation mechanism, motivated by diverse application areas. The results include approaches for establishing existence of the indices and for computing them efficiently. Evidence will be presented of the practical value of the proposed approach.

Session MS11

85 Evacuation and supply distribution facing a natural disaster

M. Teresa Ortuño (joint work with Inmaculada Flores and Gregorio Tirado)

Universidad Complutense de Madrid

Disasters have been striking human-beings from the beginning of history and their management is a global concern of the international community. Minimizing the impact and consequences of these disasters, both natural and human-made, involves many decision and logistic processes that should be optimized. A crucial logistic problem is the evacuation of the affected population and the appropriate sheltering. In this talk we will focus on the planning of supported evacuation of vulnerable people to safe places when necessary, as well as the simultaneous supply distribution to those shelters. A lexicographic goal programming model for supported evacuation is proposed, which introduces dynamism regarding the arrival of potential evacuees to the pickup points, according to their own susceptibility about the disaster.

Session MS12

86 El reto de aplicar IA en el marco de la Industria 4.0

Juan Jesús Pardo Expósito

Tecdesoft

La industria actual, tanto la manufacturera como la de proceso se enfrenta a enormes retos ligados especialmente a la personalización del producto y a un menor tiempo para lanzar soluciones al mercado. En estos procesos la digitalización de los sistemas es clave, pero a veces no llega. Debemos tomar decisiones en tiempo real sobre la fabricación que afectarán directamente en la cuenta de resultados de la compañía. La gran cantidad de datos que aporta la digitalización de procesos implica la necesidad de grandes recursos de personal para su interpretación y análisis de conclusiones. Las técnicas de analítica e inteligencia artificial nos pueden allanar este camino. La charla abordará algunos de los problemas y retos a los que se enfrenta la industria actual a la hora de aplicar inteligencia artificial en sus procesos de fabricación. Para ello caminaremos en el proceso de tratamiento del dato a través de 2 ejemplos reales, uno enfocado a la fabricación con calidad cero defectos y otro orientado al mantenimiento predictivo de grandes activos.

Session MS27

87 Solving a large cutting problem in the glass manufacturing industry

Francisco Parreño (joint work with Ramón Álvarez-Valdés)

Universidad de Castilla-La Mancha

The two-dimensional glass cutting problem to be solved by Saint Gobain, one of the world's largest producers of flat glass, includes some specific constraints that prevent the direct application of procedures developed for the standard cutting problem. On the one hand, the sheets to be cut have defects that made them unique and must be used in a specific order. On the other hand, the pieces are grouped in stacks and the pieces in each stack must be cut in order. There are also some additional characteristics due to the technology used, especially the requirement for a three-staged guillotine cutting process. We have developed heuristic and exact procedures. First, we have developed a beam search algorithm, using a tree structure in which at each level the partial solution is augmented by adding some new elements until a complete solution was built. We developed a randomized constructive algorithm for building these new elements and explored several alternatives for the local and the global evaluation. An improvement procedure, specifically designed for the problem, was also added. The computational study, using the datasets provided by the company, shows the efficiency of the proposed algorithm for short and long running times.

Session MS27

88 Sparse Matrix Classification on Imbalanced Datasets Using Convolutional Neural Networks

Beatriz Pateiro

Universidad de Santiago de Compostela

In this work we deal with a class imbalance problem in the context of the automatic selection of the best storage format for a sparse matrix with the aim of maximizing the performance of the sparse matrix vector multiplication (SpMV) on GPUs. Our classification method uses convolutional neural networks (CNNs) trained using images that represent the sparsity pattern of the matrices, whose pixels are colored according to different matrix features. The experiments conducted show that our classifiers are able to select the best performing format 92.8% of the time, obtaining 98.3% of the maximum attainable SpMV performance. A comparison to other state-of-the-art classification methods is also provided, demonstrating the benefits of our proposal.

Session MS24

89 Identificación de redes de suministro de energía eléctrica empleando algoritmos de optimización combinatoria

Aritz Pérez

Basque Center for Applied Mathematics

La energía eléctrica se transfiere entre proveedores y consumidores empleando una red de distribución de energía. Dicha red es cambiante a lo largo del tiempo debido a que los consumidores pueden cambiar de proveedor. Debido a la naturaleza cambiante de la red de distribución, se desconocen las conexiones entre proveedores y consumidores de energía eléctrica. En el proyecto realizado junto a Ormazabal S.L. hemos reformulado el problema de la identificación de la red de suministro como un problema de optimización combinatoria. El problema de optimización consiste en asociar a cada consumidor un único proveedor de manera que se minimice la diferencia entre la energía consumida y la producida por cada proveedor. El problema de optimización se ha abordado empleando algoritmos genéticos y la búsqueda local, así como diversas variantes de los mismos.

Session MS11

90 Learning decomposable models by coarsening

Aritz Pérez

Basque Center for Applied Mathematics (BCAM)

During the last decade, some exact algorithms have been proposed for learning decomposable models by maximizing additively decomposable score functions, such as Log-likelihood, BDeu, and BIC. However, up to the date, the proposed exact approaches are practical for learning models up to 20 variables. In this work, we present an approximated procedure that can learn decomposable models over hundreds of variables with a remarkable trade-off between the quality of the obtained solution and the quantity of computational resources required. The proposed learning procedure iteratively constructs a sequence of coarser decomposable (chordal) graphs. At each step, given a decomposable graph, the algorithm adds the subset of edges due to the actual minimal separators that maximize the score function while maintaining the chordality. The proposed procedure has shown competitive results for learning decomposable models over hundred of variables using a reasonable amount of computational resources. Finally, we empirically show that it can be used to reduce the search space of exact procedures, which would allow them to address the learning of high-dimensional decomposable models.

Session MS9

91 Tensor Networks from a Quantum Information perspective

David Pérez García

UCM/ICMAT

I will introduce Tensor Networks and their use in quantum information and condensed matter physics. I will then review some of the main results and open problems.

Session MS21

92 Tributos inteligentes

César Pérez López

Instituto de Estudios Fiscales

TBA

Session MS30

93 From learning with fair regularizers to physics aware models

Adrián Pérez-Suay

Universitat de València

In recent years, Machine Learning (ML) models have increased its capability and lead to solutions of real world problems. Some of those problems directly affect people's lives, like for instance autonomous driving cars, learning from social networks or bank loan prediction. When dealing with real data scenarios, Machine Learning models could lead to biased decisions over protected variables, this could incur in moral and/or legal violations. In this talk we cover some independence regularizers to overcome these model limitations. In particular, we revise the Fair Kernel Learning (FKL) method and introduce its probabilistic formulation, the

Fair Gaussian Process. Furthermore, we introduce a new setting for using that FKL method to obtain more physically plausible models.

Session MS6

94 A first step towards numerical approximation of controllability problems via Deep-Learning-based methods

Francisco Periago

Universidad Politécnica de Cartagena

This presentation is concerned with Deep-Learning-based algorithms for numerical approximation of controllability problems for PDEs. As a first step, and with the aim of having some feeling on accuracy of the proposed methods, two toy (low-dimensional) models for the heat and wave equations are considered. Error estimates for generalization error are presented. Implementation details and numerical simulation results are showed. Finally, the extension of the proposed methods to high-dimensional problems is also discussed.

Session MS19

95 Digital Twin for the human cornea: curvature estimation

Lukasz Plociniczak

Wrocław University of Science and Technology

Cornea is one of the most essential constituents of human eye accounting for about 2/3 of refracting power. It is a transparent, shell-like structure that makes up a large portion of the frontal part of the eye. To fully understand the corneal biomechanics it is necessary to formulate mathematical models that help to describe this organ and allow us to investigate various intrinsic properties associated with it. One of the most important features of the cornea is its geometry since certain anomalies in corneal topography are responsible for many seeing disorders like myopia or astigmatism.

This talk is about our work in progress concerning construction of a cornea digital twin. Ultimately, we would like to design a virtual clone of a real cornea that encodes all the important information about it. This includes material and visual properties along with measurement data of the intraocular pressure and several other parameters. We start with our previously devised simplified model of corneal topography in order to understand difficulties that may arise during the construction of the full model. This version is based on a nonlinear prescribed curvature equation that has already been extensively studied in the literature. Here, we focus on an inverse problem: knowing the topography (we can measure it) find material properties of the cornea. This immediately leads to some ill-posedness where the lack of stability is the most serious. If one would like to calculate the curvature in a naive way by differentiating the data they would arrive at severe noise amplification. We investigate some other ways of determining curvature and connect it to corneal digital twin.

Session MS25

96 Topological data analysis of high-dimensional correlation structures with applications in epigenetics

Sara Prada

Clinipace WorldWide Company, Madrid

There is currently a lack of standard and efficient analytical tools to deal with the great quantities and varieties of high-dimensional data as the genetic one. Particularly, the analysis of big-dimensional correlation structures is a pending topic in the epigenetics field. The topological analysis of the large and complex correlation structures contributes greatly to their understanding and interpretation. Generally, the application of algebraic topology in data analysis through topological data analysis (TDA) provides with an efficient perspective, as the study and representation of the "shape" of the data is key to extract underlying data characteristics doing minimal prior assumptions about their distribution and reducing the dimension of the dataset. Using the topological data analysis idea, our main proposal was to study the correlation of high-dimensional epigenetic datasets through the topological properties of the associated correlation networks or graphs, which represents a novel method to describe and model those structures. This analysis was done locally and globally to cover distinct complexity levels, designing different mathematical strategies and topological data analysis methodologies for each aim, as

a computational algorithm called MultiNet. This algorithm is able to quickly represent the correlation structure and extract substantial information from it, as epigenetic patterns associated with a sample condition (as a disease). This work opens the door to the application of these methodologies to other non-biological fields too.

Session MS23

97 Machine learning and statistical methods for clustering in FDA

Belén Pulido (joint work with Alba M. Franco-Pereira and Rosa E. Lillo)

Universidad Carlos III de Madrid

Clustering is considered as one of the most commonly used techniques in Data Science. Clustering functional data is a challenging problem since it involves working with an infinite dimensional space. This problem is addressed by applying the epigraph and the hypograph indexes to a functional dataset and thereby, converting it from a functional data problem into a multivariate problem where typical techniques of multivariate statistics or techniques more linked to machine learning can be applied. Our procedure is applied to different datasets, both simulated and real ones, and it is also compared to some clustering techniques originally designed for functional data. In view of the results, we conclude that the proposed methodology is competitive in terms of computational time and performance.

Session MS29

98 Physics-Based and Data-Driven-Based Algorithms for the Simulation of the Heart Function

Alfio Quarteroni

Politecnico di Milano

In this talk, I will present a mathematical model that is suitable to simulate the cardiac function, thanks to its capability to describe the interaction between electrical, mechanical, and fluid-dynamical processes occurring in the heart. The model comprises a system of nonlinear differential equations (either ordinary and partial) featuring a multi-physics and multi-scale nature. Efficient numerical strategies are devised to allow for the analysis of both heart function and dysfunction. These strategies rely on both classical physics-based numerical discretization methods and machine-learning algorithms, as well as on their interplay.
Acknowledgment: The work presented in this talk is part of the project iHEART that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 740132)

Session P3

99 Optimization for Social Good

Helena Ramalhinho

Universitat Pompeu Fabra

Analytics focuses on transforming data into insights by applying advanced analytical method, based on mathematics, statistics, operations research and artificial intelligent models and algorithms, with the objective to improve the performance of an organization. One of the main tools in Analytics is Optimization. In this talk, we present the optimization tools and methodologies applied to NonProfit Organizations (NPO). We will describe applications of Mathematical Programming Models and Metaheuristics Algorithms to Social Care, Healthcare, Humanitarian Logistics and Environmental organizations. Examples of applications of Optimization in these organizations are: home health care logistics and scheduling; planning disaster response and preparedness to improved decision-making; location of the primary health care centers or schools; planning the humanitarian aid distribution; planning a sustainable transportation; location of electrical charge stations, etc. We will discuss also the main aspects of these models and algorithms, and the main differences to other more frequent applications, as in manufacturing and retailing industries.

Session MS11

100 Counterfactual Explanations via Mathematical Optimization

Jasone Ramírez-Ayerbe (joint work with Emilio Carrizosa and Dolores Romero Morales)

Universidad de Sevilla

Due to the increasing use of complex machine learning models, often seen as "black boxes", it has become more and more important to be able to understand and explain their behaviour, and thus ensure transparency and fairness. An effective class of post-hoc explanations are counterfactual explanations, i.e. minimal perturbations of the predictor variables to change the prediction for a specific instance. We propose a multi-objective mathematical formulation for different state-of-the-art models based on scores, including tree ensemble classifiers and linear models. We formulate the problem at individual and group level. Real-world data has been used to illustrate our method.

Session MS4

101 Variable selection for Naïve Bayes classification

Pepa Ramírez-Cobo (joint work with and Rafael Blanquero, Emilio Carrizosa, M. Remedios Sillero-Denamiel)

Universidad de Cádiz

The Naïve Bayes has proven to be a tractable and efficient method for classification in multivariate analysis. However, features are usually correlated, a fact that violates the Naïve Bayes' assumption of conditional independence, and may deteriorate the method's performance. Moreover, datasets are often characterized by a large number of features, which may complicate the interpretation of the results as well as slow down the method's execution.

In this paper we propose a sparse version of the Naïve Bayes classifier that is characterized by three properties. First, the sparsity is achieved taking into account the correlation structure of the covariates. Second, different performance measures can be used to guide the selection of features. Third, performance constraints on groups of higher interest can be included.

Session MS5

102 Graph-based approaches for document information extraction

Oriol Ramos

Universitat Autònoma de Barcelona

Document information extraction is a classic task in image processing and computer vision. From the first OCR systems, currently distributed within home scanners, to unconstrained handwriting recognition, the main challenge in this field is not the "simple" content transcription but to extract information to feed database systems. To this end, it is needed to understand the document content context. In this talk, we will briefly review the main techniques developed in this field and we will focus on two particular ones such as table detection and table understanding. For these two tasks we will explain some recent graph-based approaches that we have recently developed using latest advances in deep learning. We will also discuss some of the main difficulties when dealing with real data, like for instance, the lack of (annotated) data and some of the most successful strategies currently used to deal with.

Session MS16

103 Approximation in Value-Based Potentials

Ofelia Paula Retamero Pascual (joint work with Manuel Gómez-Olmedo, Andrés Cano Utrera)

University of Granada

When dealing with complex models (i.e., models with many variables, a high degree of dependency between variables, or many states per variable), the efficient representation of quantitative information in probabilistic graphical models (PGMs) is a challenging task. To address this problem, Value-Based Potentials (VBPs) leverage repeated values to reduce memory requirements when managing Bayesian Networks or Influence Diagrams. In this work, we propose how to approximate VBPs to achieve a greater reduction in the memory space required and thus be able to deal with more complex models.

Session MS1

104 Adversarial Machine Learning. An overview

D. Ríos Insua (joint work with R. Naveiro and J. Poulos)

ICMAT

Adversarial machine learning aims at robustifying machine learning algorithms against possible actions from adversaries. Most earlier work in AML has modelled the confrontation between learning systems and adversaries as a 2-agent game from a game theoretic perspective. After briefly overviewing previous work, we shall present an alternative framework based on adversarial risk analysis.

Session MS8

105 Data Scientificus

Carlos Rivero Antonio

CESCE Chief Analytics Officer

TBA

Session MS30

106 Problems and challenges of the health management

Isabel Rodrigo-Rincón

Complejo Hospitalario de Navarra

Hospitals are organizations that deal with complex issues in their day-to-day life. In general terms we can speak of two different types of management processes. One is clinical management, when we are dealing with individual patients and their process of care. The other is system management, when we are balancing scarce resources (people, knowledge, hospital beds, budget, just to mention some) in order to maximize the outcomes, that is, the amount and quality of care provided to the entire population. Nowadays managers are using classical tools, such as standards based on good practices from scientific clinical societies, when working on resource planning. Is in this field where mathematics based tools, tailored for the specific needs of a system (hospital, unit, region, etc.) can improve the decision making process with the result of a better balance and assignment of resources, thus leading to a better and faster health care provision. This field of work can have a deep and long lasting impact on the health of entire populations. Some examples of challenges where mathematical tools can support the resolution of these problems are: sizing specific beds needs (for neonatology, stroke,...), intelligent planning of professionals' schedules, smart overbookings for patient appointment management identifying the activity that should be prioritized to minimize the total wait for patients... Perhaps they seem less glamorous topics than other more clinical ones, but without a doubt of great importance for all patients.

Session MS13

107 Wildfire risk measurement for fuel management decision-making using stochastic scenarios and Bayesian networks

Adán Rodríguez Martínez (joint work with Begoña Vitoriano and Gonzalo Barderas)

Universidad Complutense de Madrid

Forest fires are natural disasters whose impact has increased in recent decades due to land use and climate change. Prevention measures play a very important role in the fight against fires, a risk measure will be presented that allows quantifying the impact of the different preventive actions. The risk measure is based on a probability model that has been shown to be a Bayesian network under certain conditions. In addition, it is necessary to consider wind scenarios to improve the efficiency of the algorithm. A methodology for obtaining such scenarios is proposed, taking into account that the scale of the study area does not allow the assumption of a constant wind throughout the region. Finally, a case study of the Filabres mountain range located in southern Spain is presented.

Session MS12

108 Empower and inspire with the most trusted analytics

Adriana Rojas

SAS Academics Alliances

TBA

Session MS30

109 Optimize your path

David Romero

Centre de Recerca Matemàtica

Urban mobility has become through the last decades as one of the key pillars of sustainability and energy efficiency as urban population increases and by consequences the number of private vehicles does it too. However, the problem of offering a public transport that is sustainable, optimizes the maximum resources available and, at the same time, offers the best possible service to users is a complicated problem. One possible solution is to consider a flexible transport service strategy is considered, based on the idea behind ride-sharing services. This model is based on the concept of a tailored ride with a starting and an ending point previously agreed by the user and the bus company with the expected departure, arrival and fare but shared with other users who would have the same experience, making the journey faster, more comfortable and sustainable. The aim of this talk is to explain, by using a concrete situation, how one can deal with that routing problem by using optimization algorithms tailored for our needs.

Session MS10

110 Wavelet Analysis of the generalized Riemann non-differentiable Function

Luz Roncal

BCAM

We will report on recent progress done in showing the multifractality of a family of graphs that include Riemann non-differentiable function using wavelet analysis.

Session MS20

111 POD stabilized methods for incompressible flows: error analysis and computational results

Samuele Rubino (joint work with Julia Novo)

Universidad de Sevilla

Proper orthogonal decomposition (POD) stabilized methods for the Navier-Stokes equations are considered and analyzed. We consider two cases: the case in which the snapshots are based on a non inf-sup stable method and the case in which the snapshots are based on an inf-sup stable method. For both cases we construct approximations to the velocity and the pressure. For the first case, we analyze a method in which the snapshots are based on a stabilized scheme with equal order polynomials for the velocity and the pressure with local projection stabilization (LPS) for the gradient of the velocity and the pressure. For the POD method we add the same kind of LPS stabilization for the gradient of the velocity and the pressure as the direct method, together with grad-div stabilization. In the second case, the snapshots are based on an inf-sup stable Galerkin method with grad-div stabilization and for the POD model we also apply grad-div stabilization. In this case, since the snapshots are discretely divergence-free, the pressure can be removed from the formulation of the POD approximation to the velocity. To approximate the pressure, needed in many engineering applications, we use a supremizer pressure recovery method. Error bounds with constants independent of inverse powers of the viscosity parameter are proved for both methods. Numerical experiments show the accuracy and performance of the schemes, also combined with a data-driven approach.

Session MS21

112 Mathematical and Statistical modeling using the FMM approach. The case of the Electrocardiogram.

Cristina Rueda

Universidad de Valladolid

Oscillatory systems arise in the different biological and medical fields. Mathematical and statistical approaches are fundamental to deal with these processes. The FMM approach, the acronyms refer to Frequency Modulated Mobius, reviewed here, is one of these approaches that competes with the Fourier and Wavelets decompositions. Little known as it has been recently developed, solves a variety of exciting questions with

real data; some of them, such as the decomposition of the signal into components and their multiple uses, are of general application others are specific. Among the exciting specific applications is the FMMecgmodel that solves the forward and reverse problem in electrocardiography providing a sound automatic interpretation method of the ECG signal.

Session MS26

113 Bayesian approaches to protecting classifiers from attacks

F. Ruggeri (joint work with V. Gallego and A. Redondo)

CNR-IMATI

A major area within adversarial machine learning deals with producing classifiers that are robust to adversarial data manipulations. This talk will present formal Bayesian approaches to this problem considering settings in which robustification takes place at training time and at operation time.

Session MS8

114 Simultaneous control of Neural differential equations

Domenec Ruız-Balet

Universidad Autónoma de Madrid

The contents of this lecture have been developed together with Enrique Zuazua. In this talk we will analyze the simultaneous controllability property of Neural differential equations. We will construct strategies for controlling N distinct data points to their corresponding targets for continuous time versions of residual neural networks (Resnets), momentum resnets and some models involving memory.

Session MS19

115 High dimensional hybrid Bayesian networks: Is there life beyond the exponential family?

Antonio Salmerón (joint work with Helge Langseth, Thomas D. Nielsen and Andrés R. Masegosa)

University of Almería

Within the context of hybrid Bayesian networks, the problem of high dimensionality is challenging both from the point of view of parameter estimation and probabilistic inference. While parameter estimation can be efficiently carried out, specially for models within the exponential family of distributions, it sometimes comes along with limitations on the network structure or costly probabilistic inference/Bayesian updating schemes. On the other hand, probabilistic models based on mixtures of truncated basis functions (MoTBFs) have turned out to be compatible with efficient probabilistic inference schemes. However, MoTBFs do not belong to the exponential family, which makes the parameter estimation process more problematic due to, for instance, the non-existence of fixed dimension sufficient statistics (but the sample itself). In this work we explore some reparameterizations of MoTBFs distributions that make possible the use of efficient likelihood-based parameter estimation procedures.

Session MS1

116 Opening the black-box of deep learning architecture with Ranked-LRP

José Luis Salmerón

Universidad Pablo de Olavide de Sevilla

Understanding what Deep Learning models are doing is not always trivial. This is especially true for complex models such as Deep Neural Networks, which are the best-suited algorithms for modeling very complex and nonlinear relationships. But this need to understand has become a must since privacy regulations (GDPR and others) are hardening the use of these models in specific industries. There are several methods to address the explainability issues that Machine Learning models arises. This paper is focused on opening the so-called Deep Neural architectures black-box. This research extends the technique called Layerwise Relevant Propagation (LRP) enhancing its properties to compute the most critical paths in different deep neural architectures using multicriteria analysis. We call this technique Ranked-LRP and it was tested on four different datasets and tasks, including classification and

regression tasks. The results show the worth of our proposal.

Session MS4

117 Convergent and fast natural gradient based optimization method DSNGD and adaptation to large dimensional Bayesian networks

Borja Sánchez-López (joint work with Jesús Cerquides)

IIIA-CSIC

Information geometry has shown that probabilistic models are twisted and distorted manifolds compared to standard Euclidean spaces. In such cases where every point of the manifold describes a probability distribution, Fisher information metric (FIM) becomes handy to correctly observe the space and their local measure as it actually is. For example, the gradient of a function defined on such manifold is not even well defined until we apply metric information to it. Once FIM is considered, the steepest ascent direction is available and well defined, this is the so-called natural gradient.

Dual stochastic natural gradient descent (DSNGD) is our version of a natural gradient based algorithm to optimize the conditional log-likelihood of a class variable Y given features X. It is convergent and its computational complexity is linear, when X is discrete. We define DSNGD and take a glance to its convergence property. Some experiments are discussed paying special attention to the performance enhancement acquired after convergence property, with respect to standard non convergent stochastic natural gradient descent (SNGD). We extend DSNGD to Bayesian networks where the log-odds ratio of P(Y—X) is an affine function of features. Since DSNGD is showing low computational complexity, it scales nicely as dimension of the manifold grows.

Session MS1

118 Object Oriented Spatial Statistics (O2S2) for densities: an application to the analysis of mortality from all causes in Italy during the COVID-19 pandemic.

Piercesare Secchi (joint work with Alessandra Menafoglio, Laura Sangalli and Riccardo Scimone)

Politecnico di Milano

Along the unifying perspective offered by Object Oriented Spatial Statistics (O2S2), we analyze the densities of the time of death during the calendar year for the Italian provinces and municipalities in the year 2020, the first of the COVID-19 pandemic. The official daily data on mortality from all causes are provided by ISTAT, the Italian National Institute of Statistics. Densities are regarded as functional data belonging to the Bayes space B^2. In this space, we use functional-on-functional linear models to predict the expected mortality densities in 2020, based on those observed in the previous years, and we compare predictions with actual observations, to assess the impact of the pandemic. Through spatial downscaling of the provincial data, we identify spatial clusters of municipalities characterized by mortality densities anomalous with respect to the surroundings. The analysis could be extended to indexes different from death counts, measured at a granular spatio-temporal scale, and used as proxies for quantifying the local disruption generated by the pandemic.

Session MS2

119 A mathematical model for the slitting problem in a Spanish steel industry

María Sierra-Paradinas (joint work with Óscar Soto-Sánchez, F. Javier Martín-Campo, Micael Gallego, A. Alonso Ayuso)

IDOM consultoría

From an economic point of view, the steel industry plays an important role and, when it comes to responding to new challenges, innovation is a crucial factor. This paper proposes a mathematical methodology to solve the slitting problem in a steel company located in Spain. This process involves longitudinally slitting steel coils into narrower coils, known as strips, to meet customer orders. One of the main challenges in this problem is the large amount of data to be handled. The company has thousands of coils in the warehouse and for each of them dozens of parameters are considered that make it unique (thickness, width, length, quality are the most important, but there are many others)

and the orders are characterised by tolerances for each of the above parameters, which complicates the allocation problem (there is no clear correspondence between orders and coils). In addition, the demand is given in weight and can be served in one or more strips, not necessarily of the same length or weight. Furthermore, the characteristics of the machines used for cutting (which are not compatible with all coils and, moreover, depending on the speed at which they are used and each specific coil, may include a different number of cutting blades) must be taken into account. This makes the amount of data to be handled excessively high and means that the current operation, carried out manually, takes several hours and achieves very low utilisation of the coil (about 50%). For this reason, we have developed a Mixed Discrete Mathematical Optimisation model that has been validated with real data and it outperforms the results obtained by the company in different ways: by adjusting the orders that are to be served, by reducing the amount of scrap and by using the retails for future orders. Furthermore, planning times are reduced to only a few minutes, while the company needs several hours to prepare the scheduling in the current operating process.

Session MS27

120 DeepKriging: Spatially Dependent Deep Neural Networks for Spatial Prediction

Ying Sun

KAUST University

In spatial statistics, a common objective is to predict the values of a spatial process at unobserved locations by exploiting spatial dependence. In geostatistics, Kriging provides the best linear unbiased predictor using covariance functions and is often associated with Gaussian processes. However, when considering non-linear prediction for non-Gaussian and categorical data, the Kriging prediction is not necessarily optimal, and the associated variance is often overly optimistic. We propose to use deep neural networks (DNNs) for spatial prediction. Although DNNs are widely used for general classification and prediction, they have not been studied thoroughly for data with spatial dependence. In this work, we propose a novel neural network structure for spatial prediction by adding an embedding layer of spatial coordinates ¿ with basis functions. We show in theory that the proposed DeepKriging method has multiple advantages over Kriging and classical DNNs only with spatial coordinates as features. We also provide density prediction for uncertainty quantification without any distributional assumption and apply the method to PM2.5 concentrations across the continental United States.

Session MS3

121 Data science and Machine Learning for the fishing industry

Alberto Torres

Komorebi AI

In this talk we will present several mathematical problems related to the fishing industry. The first one is related to drifting Fishing Aggregating Devices or dFADS, that are floating objects drifting in the sea that attract hundreds of marine species. Modern dFADs are equipped with an echo-sounder and GPS communication devices, so they can transmit both the position and an estimation of the biomass located under the FAD. We will explore the use of external data sources and machine learning models to improve this biomass estimation. The second problem is related to leveraging weather information to improve the fuel consumption of big ships while at sea. In this sense the routes can be optimized by using currents and waves forecasts. Finally, if we combine the two previous problems we can try to optimize the full route of fishing vessels taking into account both the location of the FADs with more promising biomass estimations and the time it takes to reach those FADs, including also the expected currents, waves and other oceanographic information. This becomes a novel and interesting problem in operations research, with huge potential benefits for the spanish fishing fleet.

Session MS28

122 Explaining Bayesian networks using MAP-independence: Some new properties

Enrique Valero-Leal (joint work with Pedro Larrañaga and Concha Bielza)

Universidad Politécnica de Madrid

In discrete Bayesian networks, MAP-independence tries to define a notion of variable relevance in a probabilistic inference and uses it as an explanation. In our work, we deepen further into this idea, exploring some properties of the original proposal, expanding them to the continuous domain and lying the ground for new methodologies for explaining Bayesian networks.

Session MS4

123 An unsupervised machine learning algorithm to transform waste to biogas

Rocío Vega Martínez

REGANOSA

A circular economy aims to maintain the value of products, materials and resources for as long as possible by returning them into the product cycle at the end of their use, while minimizing the generation of waste. Among all the opportunities to promote the circular economy, we highlight the production of biogas from manure to be the topic of this talk. Creating an optimal network to collect and manage the manure from farms and produce biogas in digestion plants is a complex and challenging mathematical problem. One of the most critical points is to locate the plants to dehydrate the manure and the digestion plants. Our approach to suggest optimal locations is to use an unsupervised machine learning algorithm named weighted K-means clustering. In this talk, we introduce the weighted K-means Clustering algorithm and disccuss how it is used to help promote the circular economy.

Session MS14

124 Strategic and tactical preparedness in humanitarian logistics based on scenario generation from historical data

Begoña Vitoriano (joint work with Adán Rodríguez-Martínez and M. Teresa Ortuño)

Universidad Complutense de Madrid

The disaster management cycle is a process involving several phases, some before a disaster occurs (Prevention/Mitigation and Preparedness) and others after (Response, Recovery and Assessment). In the preparedness phase, logistics processes for establishing the logistics network (strategic planning) and the resources pre-positioned to be used in disaster response (tactical planning) are developed under high uncertainty. Mathematical models for decision support can incorporate this uncertainty through quantified and valued scenarios of potential disasters in the targeted area. This presentation introduces a methodology for generating scenarios for a multi-stage stochastic model for the location and sizing of warehouses (strategic decisions) and the budget allocation and pre-positioning of relief aid (tactical decisions), considering response scenarios (operational decisions). The methodology is based on historical data, which are usually scarce and incomplete, especially for disasters in developing countries. This difficulty, together with the need to keep the number of scenarios limited for their subsequent inclusion in optimisation models, leads to the use of different methodologies for classification and aggregation of historical cases. The methodology is illustrated in a case study of Mozambique.

Session MS12

125 Optimal control of deep neural networks

Enrique Zuazua

Chair for Dynamics, Control and Numerics – Alexander von Humboldt-Professorship FAU, Erlangen (Germany); Chair of Computational Mathematics, Fundacion Deusto, Bilbao; Universidad Autónoma de Madrid.

We discuss the training process for Deep Neural Networks (DNN) from an optimal control perspective. In particular we analyze the turnpike phenomena, and how it emerges, as a function of the cost functional to be optimized, and that guarantees that, in the deep layer regime, the DNN experiences the tendency to become steady.

This lecture is inspired on recent joint work with Borjan Geshkovski, Carlos Esteve-Yagüe and Dario Pighin.

Session MS19