STOR-i Annual Conference: 9th – 10th January 2025

Titles and Abstracts

Day 1

Sports Scheduling: from consulting to science

Dries Goossens - Ghent University

Any sports competition needs a schedule, specifying when and where teams meet each other. Apart from a number of pioneering theoretical results, most sports timetabling contributions in the literature read as case studies, describing a single problem instance for which a tailored algorithm is developed and compared to a manual solution. While the reported problems are challenging, and the algorithms made an impact in practice, it is hard to assess algorithmic performance. Indeed, real-life problem instances are rarely shared, and few realistic benchmark instances are available. In this talk, we discuss our efforts to obtain insights in the strengths and weaknesses of several state-of-the-art sport scheduling algorithms, and to predict which algorithm to select for which type of problem. The story covers the development of a problem classification and unifying data format, the generation of a set of diverse and realistic benchmark instance, the organization of a timetabling competition, and an instance space analysis for sports scheduling.

Retail Forecasting at The Very Group

Jamie Leigh Chapman – STOR-I Alumni, The Very Group

During the Golden Quarter, The Very Group can have up to 500K products, across 2,000 brands, available to purchase from three different websites. Some of the items we sell are dispatched directly from the supplier, however many of them are stocked in our warehouses. In our warehouses, we want to hold sufficient stock to fulfil the demand from our customers, however we do not want to hold too much stock such that there is excess capital tied up or we succumb to waste.

At The Very Group, the Retail Forecasting Data Science Team generates demand forecasts for all the items we sell on our website. These are used to inform stock buying decisions such that we can improve our inventory management and ultimately have good availability for our customers whilst managing stock levels effectively.

In this talk I will give an overview of some of the practical complexities encountered when generating demand forecasts across such a large and varied assortment of products.

Operational Research for a Sustainable Tomorrow: Tackling poverty, health and climate challenges.

Maria Paola Scaparra – University of Kent

The Sustainable Development Goals (SDGs), established by the United Nations in 2015, aim to end poverty, protect the planet, and ensure peace and prosperity for all by 2030. Achieving these ambitious goals requires innovative approaches and interdisciplinary collaboration. Operational Research (OR) provides powerful tools to optimize decision-making and resource allocation, making it a critical enabler of SDG progress.

This talk highlights the transformative potential of OR through real-world case studies. The first example, the OSIRIS project, demonstrates how data-driven models, interdisciplinary methods, and stakeholder engagement are used to develop optimal flood mitigation strategies for cities in Vietnam, contributing to SDG 11 (Sustainable Cities) and SDG 13 (Climate Action). The second part of the talk explores ongoing healthcare projects in collaboration with CUAMM Doctors with Africa, which use OR to improve accessibility to health services in African countries (SDG 3 - Good Health and Well-being). These include health services in refugee camps in Ethiopia and ambulance services in South Sudan. Alongside these examples, the talk will share key lessons learned and ingredients for delivering successful OR projects in developing countries.

Through these insights, the presentation emphasizes the critical role of the young OR community in developing innovative solutions that can drive meaningful progress toward a more equitable and sustainable future.

Exact MCMC for Intractable Proposals
Dootika Vats – IIT Kanpur
Accept-reject based Markov chain Monte Carlo (MCMC) methods are the workhorse algorithm for Bayesian inference. These algorithms, like Metropolis-Hastings, require the choice of a proposal distribution which is typically informed by the desired target distribution. Surprisingly, proposal distributions with unknown normalizing constants are not uncommon, even though for such a choice of a proposal, the Metropolis-Hastings acceptance ratio cannot be evaluated exactly. Across the literature, authors resort to approximation methods that yield inexact MCMC or develop specialized algorithms to combat this problem. We show how Bernoulli factory MCMC algorithms, originally proposed for doubly intractable target distributions, can quite naturally be adapted to this situation. We present diverse and relevant examples demonstrating the usefulness of the Bernoulli factory approach to this problem.

Anomaly Detection for Nuclear Forensics and Nonproliferation
Kes Ward – STOR-i Impact Fellow
Finding anomalies in radiation count data is important for both nuclear forensics: identifying if a crime has been committed using nuclear material and by who, and nonproliferation: making sure such crimes don't happen in the first place by ensuring that access to such material is restricted. In both these applications, anomalies represent specific radioactive isotopes and their decay chains, present over intervals of time. Our data sources are fine-binned spectroscopic gamma radiation data counts, measured frequently, from large numbers of small sensors, over long periods of time. This all leads to a very large dataset to scan. I will discuss the use of the Functional Online Cumulative Summation method for quickly finding anomalies that occur over intervals of time in a computationally efficient way, and its adaptation to this data type.

Sequential Monte Carlo for Online Clustering
Connie Trojan – STOR-i PhD Student
Sequential Monte Carlo methods give a natural way of updating beliefs over time as data is observed. This talk will discuss how they can be used to perform probabilistic clustering in the online setting, and how existing methods can be adapted to complex, large scale problems like automated knowledge base construction, where incomplete fragments of information must be clustered to existing entries in a database.

STOR-i
excellence with impact

Lancaster University

Explaining Machine Learning Survival Models
Belen Martin-Barragan – University of Edinburgh
In this work, we introduce JointLIME, a novel interpretation method for explaining black-box survival models with endogenous time-varying covariates (TVCs). JointLIME minimises the distances between survival functions predicted by the black-box survival model and those derived from the joint model. The outputs of this minimisation problem serve as explanations to quantify their impact on survival predictions. JointLIME uniquely incorporates endogenous time varying covariates using a spline-based model. We illustrate the explanation results of JointLIME using a US mortgage dataset and compare them with those of SurvLIME.

Optimising fresh food loadings for in-flight retail
Paul Sharkey – STOR-i Alumni, Datasparq
In-flight retail, where passengers can purchase duty-free items, luxury goods, and food and beverages, has become a major revenue stream for airlines. Most products are reusable stock; however, food menus typically have a significant proportion of fresh items that have a short shelf-life. Airlines must choose stock quantities carefully; enough items need to be loaded so that passenger demand can be satisfied, but if too many are loaded then they go to waste and the airline incurs the cost of a missed sale.

At Datasparq, we have recently collaborated with a major European airline to design a loading solution for fresh food quantities that maximises profit but controls for acceptable levels of waste and stockouts. The approach has two components: a forecaster to predict flight-level demand for a fresh food item, and an optimiser to select a loading that will optimise for profit subject to waste and stockout constraints. In this talk, we will discuss the methodology behind this model, as well as our approach to demonstrating the value of our model to our airline partners.

Unlocking Insights in Optimisation with the Empirical Attainment Function
Manuel López-Ibáñez – University of Manchester
The Empirical Attainment Function (EAF) estimates the probability that a random finite set of d-dimensional points attains an arbitrary point in the same space. We say that a point attains another if the former is not worse in any dimension than the latter, where "worse" is defined according to some desirable order on each dimension. The most popular application of the EAF is perhaps in the context of multi-objective optimisation, where it is used to describe the probabilistic distribution of the Pareto front approximations returned by optimisation algorithms. Visualisations of the differences between EAFs can provide insights into the relative ability of two algorithms in attaining particular regions of the multi-objective space of an optimisation problem. In addition, human preferences over EAF differences can be transformed into an optimisation objective that, for example, guides the fine-tuning of algorithmic parameters. The EAF is also useful when analysing the anytime behaviour of single-objective algorithms, where it generalises the target-based ECDF often employed in benchmarking continuous optimisers. This talk will introduce the EAF and related statistical concepts, and describe several applications of these concepts within the context of optimisation as well as open research questions.

Day 2

Parametric, nonparametric and repulsive mixture models for ecological data
Eleni Matechou – University of Kent

Ecological surveys often track individuals or species to monitor time-varying processes such as migration patterns and changes in behavioural or life states. Mixture models are a suitable and flexible approach for analyzing such data, and they have been used extensively in the field. I will discuss parametric, nonparametric, and repulsive mixture models for different types of ecological data and present results for case studies on species monitored using different observation processes and data types.

Sequential optimisation of stochastic systems using streaming simulation
Robert Lambert – STOR-i PhD Student

When optimising complex stochastic systems, analytical intractability often necessitates the need to implement simulation models. In settings where observations of a physical system are obtained sequentially, we have the opportunity to re-tune simulation input parameters, allowing for a simulation model that improves over time. These improvements can also enable improved decision-making capabilities.

As a motivating example, we consider the problem of determining the optimal number of servers in a multi-server queueing system with an unknown arrival rate. The objective is to achieve the best possible balance between system running costs and user congestion levels. We consider the data streaming setting, where observations of the target system arrive in sequential batches, allowing for the incremental improvement of arrival rate estimates. Decisions on the number of servers to employ are made after each successive observation period.

We focus on a greedy decision-making policy that always selects the number of servers in order to optimise the objective function, under the assumption that the latest arrival rate estimate is correct. Necessary conditions are established in order to provide asymptotic optimality of our policy.

Instance space analysis for the vehicle routing problem
Danielle Notice – STOR-i PhD Student

In the operations research literature, new algorithms are often introduced and declared to be state-of-the-art by showing that they outperform previous approaches on a set of well studied test instances. However, if a wider set of problem instances is assessed, the new algorithm may not be as dominant. This is the case for the capacitated vehicle routing problem (CVRP). In this talk, we perform algorithm selection for the CVRP via instance space analysis. We will discuss the steps involved in instance space analysis and how the set of problem instances used affects the selection model.

Patrolling Dispersed Locations
Edward Mellor – STOR-i Impact Fellow

Organisations are sometimes required to protect assets from sabotage. If these assets are geographically dispersed, it can be prohibitively expensive to protect all of them simultaneously. In this talk, we consider the following patrol model.

Attackers arrive over time according to a Poisson process. Upon arriving, each can attack one of n locations. Once an attack has started, it continues until the attacker is apprehended. A single patroller protects these locations. While at a location, she can spend any amount of time searching it. Searching detects each attacker present independently with a fixed detection rate. Moving between locations takes time. The patroller's objective is to minimise the expected time an attacker stays undetected regardless of where the attack occurs.

We define two cycle types based on common patrol practice for perimeter patrol and border patrol, respectively, and derive formulae for the expected time to detect an attack in each case. We also provide an algorithm for finding the best parameters for each cycle type subject to some unimodality conditions.

Forecasting food insecurity using Gaussian processes

Francesca Panero – Sapienza University

"End hunger, achieve food security and improved nutrition and promote sustainable agriculture" is one of the 17 goals of the 2030 Agenda for Sustainable Development, adopted by all United Nations Member States in 2015. Governments and international organisations are interested in estimating the percentage of households that are considered food insecure, and to do so agencies like the UN World Food Programme (WFP) carry out daily surveys in person or by telephone. Collecting these data is expensive, sometimes dangerous and might lead to biased estimated. To overcome such difficulties, WFP has developed ways to do nowcasting of food insecurity using machine learning methods on secondary variables coming from economic, climate and conflict assessments. In this talk, I will explain why and how we propose to use Bayesian methods, in particular Gaussian processes (GP), to perform such forecasting. GPs are interpretable methods that allow to model space, time and covariates effects, include expert information and provide clear uncertainty estimates. After explaining our model, I will present and discuss some forecasting results on Nigeria and compare with current proposals in the literature.