UNIVERSITY OF CAMBRIDGE

Cambridge Mathematics Placements (CMP) Presentation Day

Wednesday 27 August 2025, Centre for Mathematical Sciences

Tea & Coffee (Central Core), 10:30 - 11:00

	Meeting Room 4	Meeting Room 3	Meeting Room 5
11:00	**Nested Sampling for ARIMA Model Selection: A Novel Approach to Astronomical Time Series Analysis** *Ajinkya J. Naik, Institute of Astronomy*	**Minimax optimal local polynomial regression with Transformers** *Ioana Popescu, Michelle Ching, and Nico Smith, Statistical Laboratory (DPMMS)*	**Nonlinear eigenvalue problems** *Shota Uka, Cambridge Image Analysis (DAMTP)*
11:20	**Using autoencoders to improve speech clarity for Cochlear Implant users** *Marcus Ng, Deep Hearing Lab, MRC Cognition and Brain Sciences Unit*	**the above talk continues until 11:40**	**Symplectic Neural Flows combining observed trajectories with physics** *Claudia Franco, Cambridge Image Analysis (DAMTP)*
11:40	**Virtual Labelling Using Deep Learning** *Sumer Sener, MRC-LMB & Cambridge Advanced Imaging Centre (PDN)*	**Inferring Diffusion from Killed Molecules** *Paula Horvat, Statistical Laboratory (DPMMS)*	**Functional inequalities for quantum Renyi divergences** *Andreea Silvia Goia, Centre for Quantum Information (DAMTP)*
12:00	**Forecasting Air pollution for pro-active policy intervention** *Samyak Jain, Judge Business School*	**Quantifying simplicity for finitely generated groups** *Niki Hasani, Algebra-Groups (DPMMS)*	**Primordial Gravitational Waves from Cosmic Strings and Inflation** *Charlotte Louw, Relativity and Gravitation (DAMTP)*

Lunch (Central Core), 12:30 - 13:30

	Meeting Room 4	Meeting Room 3	Meeting Room 5
13:30	**Unsupervised learning for data integration and hypotheses generation in flower development** *Shreyas Pandit, Sainsbury Laboratory Cambridge University*	**p-adic Dynamics and Failure of Newton's Method (Part 1)** Jingyi Le, *Number Theory-Algebraic Geometry (DPMMS)*	**Communicating mathematics - uncovering the story behind new research** *Ben Watkins, Plus Magazine, Millenium Maths Project (DAMTP)*
13:50	**Modelling Cellular Kinematics in Self-Similar Plant Growth** *Daniel Baig, Sainsbury Laboratory Cambridge University*	**p-adic Dynamics and Failure of Newton's Method (Part 2)** Annalaura Pegoraro, *Number Theory-Algebraic Geometry (DPMMS)*	**Communicating Climate Science** *Alison Spadaro, Institute of Computing for Climate Science*
14:10	**Cell cycle phase duration inference from time-lapse images** *Yimin Li, Sainsbury Laboratory Cambridge University*	**Combinatorics of KP solitons** *Lilian Contius, Combinatorics-Differential Geometry & Topology (DPMMS)*	**Machine Learning Models for Prediction of Forest Fires** *Michael Lin, Institute of Computing for Climate Science*
14:30	**Modelling the impact of cell division on plant patterning and development** *Ben Handley, Sainsbury Laboratory Cambridge University*		**Identifying the main development regions for sand and dust storms in the Middle East** *Rishul Karia, Institute of Computing for Climate Science*

Tea & Coffee (Central Core), 15:00 - 15:30

	Meeting Room 4	Meeting Room 3	Meeting Room 5
15:30	**Supervised machine learning in digital pathology applied to the detection of disease in oesophageal lesions** *Yibo Wang, VisionLab, Department of Physics*	**Hydrothermal circulation near mid-ocean ridges** *Rebecca Maver, Adam Michaels, and Reuben Price, Fluid Dynamics (DAMTP)*	**Denoising the Results of Monte-Carlo Ion Implantation** *Daniel Chen, Silvaco TCAD*
15:50	**Phylogenetic Networks: Inferring 3-cycle networks** *Anthony Zhao, EMBL-EBI*	**the above talk continues until 16:10**	**Investigation of dopant activation and diffusion in SiC** *Katrina Ng, Silvaco, Process Engineering Team*
16:10	**Hallmarks of cancer regression** *Alex Yang, AstraZeneca PLC*	**Wavefront Propagation in Multiply Scattering Media** *Dyuti Milir, Waves-Fluid Dynamics (DAMTP)*	**Data-Driven Flower Varieties** *Gavriel Harris, APEX Horticulture*

CMP programme review and feedback seminar (Meeting Room 4), 16:45

Philippa Fawcett & Cambridge Mathematics Open Internships review and feedback seminar (Meeting Room 12), 16:45

CMP Presentation Day Dinner (Christ's College),

Drinks from 19:00 for dinner at 19:30

Meeting Room 4, 11:00 - 12:30

11:00

Ajinkya J. Naik

UNIVERSITY OF CAMBRIDGE
Institute of Astronomy

Nested Sampling for ARIMA Model Selection: A Novel Approach to Astronomical Time Series Analysis

The next generation astronomical surveys are expected to produce unprecedented volumes of time series data, demanding an efficient statistical analysis. ARIMA is a class of powerful models for analysing such time series data, but their flexibility requires a method to select the optimal model. This talk discusses the use of Nested Sampling as a possible solution to this problem.

11:20

Marcus Ng

UKRI MRC Cognition and Brain Sciences Unit

Using autoencoders to improve speech clarity for Cochlear Implant users

Auto-encoders can represent complex data as meaningful vectors in an embedding space.
Matrix subspace projection identifies a subspace of this embedding with an orthogonal basis aligned to specified features.
With this approach, sound representations for Cochlear Implant listeners can be biased to enhance speech clarity.
This talk presents results of applying the method to Neurograms generated from the TIMIT speech corpus.

11:40

Sumer Sener

UNIVERSITY OF CAMBRIDGE
Department of Physiology, Development and Neuroscience
UKRI MRC Laboratory of Molecular Biology

Virtual Labelling Using Deep Learning

Exploring the applications of U-Nets and flow matching for virtual labelling

12:00

Samyak Jain

UNIVERSITY OF CAMBRIDGE
Judge Business School

Forecasting Air pollution for pro-active policy intervention

Annually, Delhi faces a rapid rise in air pollution, PM2.5 being a key pollutant with serious health risks. In this project I predict PM2.5 concentrations up to 14 days ahead, allowing emergency policies like GRAP to be implemented proactively. By shifting from a reactive to a predictive model, interventions can be timed to control severe pollution spikes and limit exposure, enhancing GRAP's effectiveness and limiting the health risks.

Meeting Room 3, 11:00 - 12:30

11:00

Ioana Popescu, Michelle Ching, and Nico Smith

UNIVERSITY OF CAMBRIDGE
Department of Pure Mathematics and Mathematical Statistics

Statistical Laboratory

Minimax optimal local polynomial regression with Transformers

Transformers have recently shown strong in-context learning capabilities: given examples in the prompt, they can generalize to new unseen tasks. By reframing this as a regression problem, we study the expressiveness of transformers. We will show that well-designed transformers achieve minimax optimal rates for nonparametric regression, using only a moderate number of training samples and parameters.

11:20

The above talk continues until 11:40

11:40

Paula Horvat

UNIVERSITY OF CAMBRIDGE
Department of Pure Mathematics and Mathematical Statistics

Statistical Laboratory

Inferring Diffusion from Killed Molecules

Molecular motion in cells can be modeled by certain PDEs called diffusion equations, which can have varying parameters. Under certain conditions, molecules can be “killed,” meaning their movement stops. Given only the information about the final positions of killed molecules, the goal is to investigate which assumptions enable one to infer the true diffusion parameters.

12:00

Niki Hasani

UNIVERSITY OF CAMBRIDGE
Department of Pure Mathematics and Mathematical Statistics

Algebra
Groups

Quantifying simplicity for finitely generated groups

Sturmian sub-shifts are an interesting family of dynamical systems to study. These shifts arise from simple settings but are in fact very rich in terms of dynamical behavior. We start by defining these shifts and the space they live in and then discuss some of their unique features. By the end, we introduce a group that can be constructed from these systems and explore their relations.

Meeting Room 5, 11:00 - 12:30

11:00

Shota Uka

Nonlinear eigenvalue problems

UNIVERSITY OF CAMBRIDGE
Department of Applied Mathematics and Theoretical Physics

Cambridge Image Analysis

In recent years, nonlinear analogues of eigenvalue problems have emerged in machine learning and image analysis. We aim to generalize eigenvalues and eigenvectors to the nonlinear case while preserving key properties and ensuring efficient computation. Generalisations of the power and the inverse power method are proposed. Using the KL property, we show convergence to stationary points of the generalized Rayleigh quotient.

11:20

Claudia Franco

Symplectic Neural Flows combining observed trajectories with physics

UNIVERSITY OF CAMBRIDGE
Department of Applied Mathematics and Theoretical Physics

Cambridge Image Analysis

Hamilton's equations are fundamental for modelling complex physical systems, where preserving key properties like energy and momentum is crucial for reliable long-term simulations. In this presentation, we start by analysing SympFlow, a time-dependent symplectic neural network designed using a parametrised Hamiltonian flow map. We present how to train this model in a supervised and unsupervised setup. We then explore two new hybrid training setups leveraging both data and the knowledge of the physics ruling the system. Does this hybrid training methodology actually help?

11:40

Andreea Silvia Goia

Functional inequalities for quantum Renyi divergences -1 -1

UNIVERSITY OF CAMBRIDGE
Department of Applied Mathematics and Theoretical Physics

Centre for Quantum Information

Quantum divergences are a central aspect in characterizing the time evolution of open quantum systems. In this work, we briefly introduce the theory behind open quantum systems. We will then particularly focus on new results for the geometric Rényi divergence and the Belavkin-Staszewski relative entropy, which are two lesser-used, but equally interesting, quantum measures.

12:00

UNIVERSITY OF CAMBRIDGE
Department of Applied Mathematics and Theoretical Physics

Relativity and Gravitation

Charlotte Louw

Primordial Gravitational Waves from Cosmic Strings and Inflation

Meeting Room 4, 13:30 - 15:00

13:30

Shreyas Pandit

Unsupervised learning for data integration and hypotheses generation in flower development

The process of morphogenesis is driven by a combination of chemical, mechanical and morphological signals. To understand how these different factors work together to shape the developing flower, we must analyse data that is multimodal, e.g. gene expression profiles along with image data. This project discusses two unsupervised learning approaches to analyse and integrate such datasets for studying flower development.

13:50

Daniel Baig

Modelling Cellular Kinematics in Self-Similar Plant Growth

Some plant structures exhibit self-similar growth, that is, they maintain stable geometries as their cells grow, divide, and flow within them. This offers an analytically tractable system for exploring their growth kinematics and the coupling between genetic regulation and morphoelasticity in plant development. We create models to emulate these forms, capturing cellular growth and proliferation while preserving overall geometry.

14:10

Yimin Li

Cell cycle phase duration inference from time-lapse images

This talk presents a semi-Markov model for inferring cell cycle phase durations in plants from time-lapse imaging. By modelling G1 and non-G1 sojourn times with Gamma distributions and defining a distance function between simulated and observed lineages, we fit parameters using a Metropolis–Hastings algorithm, enabling inference of division and transition dynamics under varying environmental conditions.

14:30

Ben Handley

Modelling the impact of cell division on plant patterning and development

Plant cells communicate by diffusion through cell walls - but during development, they have options for where and how often divisions form new walls. Could these divisions help to shape cell communication and coordination? Through numerical modelling and analysis of experimental data, I investigate the impact these divisions have on plant growth and patterning mechanisms.

Meeting Room 3, 13:30 - 15:00

13:30

Jingyi Le

p-adic Dynamics and Failure of Newton's Method (Part 1)

Based on Newton's method, I'll first illustrate how p-adic dynamics and bad primes shape convergence. Then, extend to another superconvergent algorithm, McMullen's method, to introduce two main theorems about convergence and non-convergence.

UNIVERSITY OF CAMBRIDGE
Department of Pure Mathematics and Mathematical Statistics

- Number Theory
- Algebraic Geometry

13:50

Annalaura Pegoraro

p-adic Dynamics and Failure of Newton's Method (Part 2)

Building on Part 1, I will discuss the density of primes that yield p-adic convergence in Newton's and McMullen's maps, supported by some computational data. I will also present further results on the dynamics of rational maps with superattractors, highlighting how these phenomena arise in a broader context

UNIVERSITY OF CAMBRIDGE
Department of Pure Mathematics and Mathematical Statistics

- Number Theory
- Algebraic Geometry

14:10

Lilian Contius

Combinatorics of KP solitons

The Kadomstev-Petviashvili equation u_{xxxx}^{-1} models non-linear motion of shallow water waves in two dimensions. Waves described by solutions to this PDE are approximated by the locus where a certain function is not linear. We study the combinatorial structure of these solutions via the dual relationship between hyperplane arrangements and projections of cubes, and generalize to higher dimensions.

UNIVERSITY OF CAMBRIDGE
Department of Pure Mathematics and Mathematical Statistics

- Combinatorics
- Differential Geometry & Topology

Meeting Room 5, 13:30 - 15:00

13:30

Ben Watkins

Communicating mathematics - uncovering the story behind new research

This talk will be split into two parts, the first discussing maths communication: what it is? and why is it important? The second section will then discuss the growing field of proof assistants: what are they and why are they important? This will be a non-technical talk and open to all regardless of specialities.

Millennium Mathematics Project

13:50

Alison Spadaro

Communicating Climate Science

Why should we care about communication in climate science? One of the biggest problems facing society is the growing chasm between the scientific community and the outside world. As scientific issues become more and more relevant to the public, there is a worrying lack of accessible resources on such issues. To combat this, we need to emphasize communication not just with our academic peers, but also with the general public.

Institute of Computing for Climate Science

14:10

Michael Lin

Machine Learning Models for Prediction of Forest Fires

Forest fires are becoming increasingly prevalent, and as such the importance of models for their prediction is also rising. My project looked at antecedent proxy variables (read: rolling averages) to replace local variables (soil moisture) in such models, and whether an effective model may be trained using these

Institute of Computing for Climate Science

14:30

Rishul Karia

Identifying the main development regions for sand and dust storms in the Middle East

Sand and dust storms play an important role in geopolitics in the Middle East, with this project seeking to model the genesis regions of such events using satellite data. This is paired with a neural network model which trains on various climatic variables to investigate how these vary across each source region.

Institute of Computing for Climate Science

CAMBRIDGE ZERO

Meeting Room 4, 15:30 - 16:30

15:30

Yibo Wang

Supervised machine learning in digital pathology applied to the detection of disease in oesophageal lesions

15:50

Anthony Zhao

Phylogenetic Networks: Inferring 3-cycle networks

Phylogenetic Networks are graphs describing the evolution of species, like phylogenetic trees, but are able to describe “horizontal” evolution events such as hybridisation. We are interested in learning about Markov models of genome evolution placed on phylogenetic networks. One approach we are using is that from algebraic statistics, where models are viewed as varieties from algebraic geometry. We are interested in using this approach to develop fast methods of inferring phylogenetic networks from DNA sequence data. Our main focus in the project is to infer 3-cycle network structures, which do not correspond uniquely to any variety.

16:10

Alex Yang

Hallmarks of cancer regression

This project integrates TCGA colorectal cancer data with the Hallmarks of Cancer framework to build interpretable survival models. We compared Cox regression with Bayesian approaches using the horseshoe prior to handle sparsity, extended to gene-level hierarchies and treatment–interaction models, and explored how such approaches may detect true hallmark effects while controlling false positives.

Meeting Room 3, 15:30 - 16:30

15:30

Rebecca Maver, Adam Michaels, and Reuben Price

UNIVERSITY OF CAMBRIDGE
Department of Applied Mathematics and Theoretical Physics

Hydrothermal circulation near mid-ocean ridges

A volume of water equivalent to the entire global ocean circulates through the ocean floor in around ten million years. This hydrothermal circulation is qualitatively understood, but the precise mechanics and dependence on key parameters are still largely unexplored. Through building and simulating a mathematical model informed by geophysical data, we aim to understand these crucial and complex dynamics.

- Fluid Dynamics

15:50

The above talk continues until 16:10

16:10

Dyuti Milir

UNIVERSITY OF CAMBRIDGE
Department of Applied Mathematics and Theoretical Physics

Wavefront Propagation in Multiply Scattering Media

We investigate formation and persistence of high-intensity “ribbons” in random media. A fast 2D split-step solver provides a reference simulation over ensembles of Gaussian phase-screen media. We link local curvature (lens-like structure) to ribbon paths and introduce a discrete geometrical-optics/Markov transport surrogate. Correlation analyses—including transverse fourth-moment statistics—quantify persistence. We sketch applications to transcranial ultrasound planning.

- Fluid Dynamics
- Waves

Meeting Room 5, 15:30 - 16:30

15:30

Daniel Chen

Denoising the Results of Monte-Carlo Ion Implantation

SILVACO

One of the important processes during semiconductor manufacturing is implantation, in which a wafer is bombarded with ions to modify its electrical properties. Since physical testing is expensive, engineers rely on computer simulations - but these tend to produce very noisy results. In this project, a novel smoothing algorithm is developed that halves the mean-squared error compared to previous methods.

15:50

Katrina Ng

Investigation of dopant activation and diffusion in SiC

SILVACO

In this talk, I investigate dopant activation and diffusion in SiC-4H. Silicon carbide is a semiconductor, with many attractive properties that makes it suitable for high power applications. The process of annealing at a high temperature allows the dopant to activate and diffuse, hence it is an important step in manufacturing SiC devices. Literature review was performed to calibrate the simulations of this step.

16:10

Gavriel Harris

Data-Driven Flower Varieties

APEX HORTICULTURE