UNIVERSITY OF CAMBRIDGE

Cambridge Mathematics Placements (CMP) Presentation Day

Wednesday 21 August 2024, Centre for Mathematical Sciences

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

	Meeting Room 3	Meeting Room 5
11:00	**Segmentation of partially depolymerised plant microtubules using synthetic data** *Karan Elangovan, Sainsbury Laboratory, University of Cambridge*	**Improving phoneme identification via computational modelling of auditory nerve stimulation and an auto-encoder architecture** *Tomasz Kinowski, MRC Cognition and Brain Sciences Unit*
11:20	**How do plant cells develop, grow and communicate?** *Dora Chen, Sainsbury Laboratory, University of Cambridge*	**Guiding the Generation of Synthetic Speech Corpora** *Erdem Baha Topbas, MRC Cognition and Brain Sciences Unit*
11:40	**Modelling tree growth responses to climate change** *Patrick Thomas, Department of Geography*	**Nuclei segmentation in 3D microscopy images by retraining Cellpose** *Alex Yan, MRC LMB, Cambridge Advanced Imaging Centre, Department of Physiology, Development and Neuroscience*
12:00	**Evaluating key factors affecting the quality of sea-freighted cut roses** *Adaline Wu, Apex Horticulture*	**Nuclei segmentation in 3D microscopy images using fine-tuned SAM** *Xiaoyu Deng, Cambridge Advanced Imaging Centre, Department of Physiology, Development and Neuroscience*

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

	Meeting Room 3	Meeting Room 5
13:30	**Discovery of novel biomarkers using unsupervised statistical learning** *Scott Hislop, MRC Biostatistics Unit*	**Cell Segmentation for Duodenal Biopsies in Celiac Disease** *Rebekah Bryant, Lyzeum*
13:50	**Growth curve time series models for forecasting epidemics and product sales** *Edwin Tang, Judge Business School*	**Developing a database for frameshift peptides** *Arthur Frisk, Department of Pathology*
14:10	**Comparative Analysis of AI Accelerators with Standardised Machine Learning Benchmarks** *Maxwell Ye, Cambridge Open Zettascale Lab*	**Diffusion Models for Splice Site Prediction and Generation** *Somsubhro Bagchi, Novo Nordisk, Digital Science and Innovation, Human Genetics CoE*
14:30	**A data-driven approach to carbon footprint reporting for High Performance Computing (HPC) at the University of Cambridge** *Kell Johnston, Cambridge Open Zettascale Lab*	**Tissue-specific T2D drug target identification with Genetic and Protein Interaction Analysis** *Josefa Stoisser, Novo Nordisk, Digital Science and Innovation, Machine Intelligence*

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

	Meeting Room 3	Meeting Room 5
15:30	**Solving problems in graph theory with photonic quantum computers** *Alexander Makarovskiy, ORCA Computing*	**Protein Identification using Machine Learning** *Jenny Dunstan, EMBL-EBI*
15:50	**Communicating Mathematics: What are groups and why are they useful?** *Justin Chen, Plus (plus.maths.org), Millennium Mathematics Project, DAMTP*	**Detecting Interspecific Positive Selection using Transformers** *Brian Peng, EMBL-EBI*
16:10	**The role of Science Communications in a changing climate** *Lara Grant, Institute of Computing for Climate Science*	**Phylodynamics, but bigger** *Joel Winterton, EMBL-EBI*
		Volatility Surface Fitting: The SABR Model *Amier Hussen, Aspect Capital*

CMP programme review and feedback seminar (Meeting Room 4), 17:15

Philippa Fawcett & Cambridge Mathematics Open Internships review and feedback seminar (Meeting Room 12), 17:15

CMP Presentation Day Dinner (Westminster College)

Drinks from 19:00 for dinner at 19:30

UNIVERSITY OF CAMBRIDGE

Cambridge Mathematics Placements (CMP) Presentation Day

Wednesday 21 August 2024, Centre for Mathematical Sciences

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

	Meeting Room 4	Meeting Room 12
11:00	**Combining Physics Informed Neural Networks with Learned Regularisation** *Nina Allmeier, Mathematics of Information (DAMTP)*	**Modelling the Oxidation of Si, SiC: verification and calibration** *Zac Owen, Silvaco Europe*
11:20	**Spectral methods for time-dependent PDEs** *Xane Miles, Numerical Analysis (DAMTP)*	**Finite difference approximation of the Stokes flow with free interfaces on staggered Cartesian grids** *Ewan Heaney, Silvaco Europe*
11:40	**Cosmic strings and boson stars** *Katrina Ng, Relativity and Gravitation (DAMTP)*	**Model comparison and robust estimators for equilibria in supramolecular chemistry** *Daisy Jia, Department of Chemistry*
12:00	**Asymptotic Decomposition of a Charged Scalar Field in de Sitter Space** *Eliza Somerville, Partial Differential Equations / General Relativity (DPMMS)*	**Quantum error correction using belief propagation algorithms** *Weronika Wiesiolek, Riverlane*

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

	Meeting Room 4
13:30	**Canonical Heights on Elliptic Curves** *Helia Kalanaki, Algebraic Geometry (DPMMS)*
13:50	**Monads and distributive laws** *Maia Woolf, Number Theory and Algebra (DPMMS)*
14:10	**Surfaces, graphs and commutators** *Yasna Aminaei Chatroodi, Differential Geometry and Topology (DPMMS)*
14:30	**Gaussian processes in climate science** *Kitty Knight, Institute of Computing for Climate Science*

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

	Meeting Room 4
15:30	**Fluid: data-linked visualisations** *Haofei Chen, Institute of Computing for Climate Science*
15:50	**Data driven modelling of ocean carbon removal strategies** *Hannah Woods, Ocean Dynamics (DAMTP)*
16:10	**Parameter optimisation of ocean biogeochemical models** *Chloe Y. Huang, Ocean Dynamics (DAMTP)*
16:30	**The flash dynamics of bioluminescent biofilms under flow** *Safian Ali, Environmental and Industrial Fluid Dynamics (DAMTP)*

CMP programme review and feedback seminar (Meeting Room 4), 17:15

Philippa Fawcett & Cambridge Mathematics Open Internships review and feedback seminar (Meeting Room 12), 17:15

CMP Presentation Day Dinner (Westminster College)

Drinks from 19:00 for dinner at 19:30

Meeting Room 3, 11:00am - 12:30pm

11:00am

Karan Elangovan

Segmentation of partially depolymerised plant microtubules using synthetic data

Microtubules are polymers that exist throughout the interior of a plant cell and aid in growth. In this project we train a U-Net to segment fluorescent micrographs of microtubule at varying stages of depolymerisation in root hair cells using synthetic data, as part of a wider effort to investigate how the nucleus effects cell growth through the microtubules.

11:20am

Dora Chen

How do plant cells develop, grow and communicate?

In this talk, I shall explain how the research group aims to understand why plant cells divide and form specific geometries more than others. I studied this problem by looking at how different cell geometries affect the diffusion rate of chemicals. I will talk about some simulations we have run, the development of an analytical model and some next steps to undertake.

11:40am

Patrick Thomas

Modelling tree growth responses to climate change

Wood is one of the largest stores of carbon in the world, but the process by which it grows (and hence sequesters carbon) is poorly understood. This project aims to provide a model for how wood grows, and how it will be affected by climate change.

12:10pm

Adaline Wu

Evaluating key factors affecting the quality of sea-freighted cut roses

In this presentation, I will explore the motivations behind sea-freighting roses, the challenges it presents, and the key questions we aim to address. I will also discuss how we identified the factors that influence the visual quality of cut roses and evaluated the impact of different farms and treatments on their overall quality.

Meeting Room 3, 1:30pm - 3:00pm

1:30pm

MRC Biostatistics Unit

Scott Hislop

Discovery of novel biomarkers using unsupervised statistical learning

The goal of my project was to develop some new signals (biomarkers) for whether a particular patient had cancer (and if so, how severe it was). In this talk I'll discuss some extensions / modifications to common unsupervised statistical techniques (primarily Mixture Models) that were used to create the signals. We'll also cover a more supervised approach that's loosely based off of a Bayes classifier.

1:50pm

UNIVERSITY OF CAMBRIDGE Judge Business School

Edwin Tang

Growth curve time series models for forecasting epidemics and product sales

Forecasting peaks in cases and hospitalizations accurately is crucial during the COVID-19 pandemic. Harvey and Kattuman (2021) developed time series models for cases and hospitalizations inspired by the Gompertz curve, showing incredible accuracy. Actual R implementation of these models in the tsgc package will be discussed, allowing you to make your own predictions following their theory with applications beyond epidemiology.

2:10pm

CAMBRIDGE OPEN ZETTASCALE LAB

Maxwell Ye

Comparative Analysis of AI Accelerators with Standardised Machine Learning Benchmarks

2:30pm

CAMBRIDGE OPEN ZETTASCALE LAB

Kell Johnston

A data-driven approach to carbon footprint reporting for High Performance Computing (HPC) at the University of Cambridge

My presentation explains how I approached the problem of calculating the carbon footprint for users of the HPC facilities at the UoC. I explain carbon footprint reporting and the structure of HPC. Before describing how I tackled the problems that occurred in calculating, handling and reporting the data from a user centred perspective.

Meeting Room 3, 3:30pm - 5:00pm

3:30pm

Alexander Makarovskiy

Solving problems in graph theory with photonic quantum computers

Many interesting real-world problems can be posed in terms of graph theory. However, these problems are often too complicated to be able to be solved by a classical computer. This talk will present an exciting idea about how we could use photonic quantum computers to solve these problems, and discuss the practicality of implementing these methods.

3:50pm

Justin Chen

Communicating Mathematics: What are groups and why are they useful?

How can mathematicians communicate to a non-expert audience what maths looks like to them? Using the example of groups, I talk about how abstract mathematics can be made more accessible, as well as how existing communications content can be organised on a webpage to tell a cohesive story.

4:10pm

Lara Grant

The role of Science Communications in a changing climate

Science communications for the Institute of Computing for Climate Science: I translated complex climate science and software engineering topics into accessible content, managed social media, and supported major events like the Cross-VESRI conference and the ICCS Summer School. This experience deepened my understanding of science communication's crucial role in making research more accessible to the public.

Meeting Room 5, 11:00am - 12:30pm

11:00am

MRC Cognition and Brain Sciences Unit

Tomasz Kinowski

Improving phoneme identification via computational modelling of auditory nerve stimulation and an auto-encoder architecture

Cochlear Implants (CI) are a vital tool in aiding individuals with impaired hearing. Notwithstanding their major successes, certain problems still arise for CI recipients, such as issues with identifying particular phonemes. In this project, we developed a novel approach for improving phoneme discrimination by enhancing phonemic contrasts using an autoencoder neural network framework coupled with a computational model of CI hearing.

11:20pm

MRC Cognition and Brain Sciences Unit

Erdem Baha Topbas

Guiding the Generation of Synthetic Speech Corpora

Hearing tests are one of the key tools in testing the capabilities of an individual's auditory senses. The challenges associated with developing new ones lead to errors regarding outdated content and underserved populations. This project focuses on evaluating the feasibility of using machine-learning based synthetic speech as a tool for accelerating hearing test corpora generation.

11:40am

UNIVERSITY OF CAMBRIDGE
Department of Physiology, Development and Neuroscience

MRC Laboratory of Molecular Biology

Alex Yan

Nuclei segmentation in 3D microscopy images by retraining Cellpose

In recent years, many deep learning–based approaches have been introduced for segmentation tasks in bioimage analysis. In this talk, I focus on one popular method for cell and nuclei segmentation, Cellpose (Stringer et al, 2021), and I describe our attempts to use Cellpose on our highly anisotropic 3D images of mouse embryos. In addition, I demonstrate that the CNN architecture used by Cellpose can also be trained to detect dividing cells, a functionality that has the potential to improve the accuracy of cell-tracking.

12:10am

UNIVERSITY OF CAMBRIDGE
Department of Physiology, Development and Neuroscience

MRC Laboratory of Molecular Biology

Xiaoyu Deng

Nuclei segmentation in 3D microscopy images using fine-tuned SAM

This talk aims to use the Segment Anything Model (SAM) for nuclei segmentation. We explore how well the pre-trained SAM performs on our custom 3D microscopy images, and how fine-tuning it can further improve its performance in this specific context. Additionally, we expand the application of SAM to membrane segmentation, demonstrating its versatility in handling different types of biological imaging challenges.

Meeting Room 5, 1:30pm - 3:00pm

1:30pm

UNIVERSITY OF CAMBRIDGE
Department of Pathology

Lyzeum Ltd

Rebekah Bryant

Cell Segmentation for Duodenal Biopsies in Celiac Disease

This talk explores the development of an interpretable deep learning model for the segmentation of histopathology images in celiac disease diagnosis. Building on previous work, this project focuses on automating the identification of Intraepithelial Lymphocytes (IELs) and Enterocytes, aiming to improve diagnostic accuracy and efficiency while reducing the workload on healthcare professionals.

1:50pm

UNIVERSITY OF CAMBRIDGE
Department of Pathology

Arthur Frisk

Developing a database for frameshift peptides

I will discuss the development of a database and website cataloguing what are known as frameshift mutations. Such mutations play a role in the development of MSI-H (microsatellite instability-high) cancers, and the database is part of research into a new method of screening for such cancers. After some biological background, I will go over the design of the database, and challenges which came up in automatically validating and normalising the data.

2:10pm

novo nordisk®

Somsubhro Bagchi

Diffusion Models for Splice Site Prediction and Generation

Splicing is a crucial biological process in the cell. Alternative splicing can be the cause of diseases and reversing such splicing could have therapeutic benefits. However, we still do not have a good understanding of when/where splicing occurs. In this talk, I'll present an ML approach to solving this problem using diffusion models.

2:30pm

novo nordisk®

Josefa Stoisser

Tissue-specific T2D drug target identification with Genetic and Protein Interaction Analysis

Identifying potential drug target candidates for Type 2 Diabetes (T2D) is challenging due to the complex nature of the disease. In this project we use genetic and tissue-specific data to address this challenge. More precisely, we evaluate the impact of integrating genetic information into protein-protein interaction (PPI) networks, both in a tissue specific context, by analyzing diffusion approaches, centrality metrics, node embeddings and implementing machine learning models.

Meeting Room 5, 3:30pm - 5:00pm

3:30pm

Jenny Dunstan

Protein Identification using Machine Learning

EMBL-EBI

Given an unknown protein, how do we tell what it is? Being able to identify proteins quickly and accurately is crucial for a wide range of applications, including in medicine. This project looks at a new way of doing this, by feeding the protein through a nanopore and then employing the latest machine learning techniques.

3:50pm

Brian Peng

Detecting Interspecific Positive Selection using Transformers

EMBL-EBI

Traditional statistical methods utilising maximum likelihood and Bayesian inference can detect positive selection from multiple sequence alignments but are subject to false positives due to errors in alignment. Exploiting the innate ability of self-attention mechanism for information exchange, we propose a solution resistant to misalignments. We have trained the model and leveraged attention maps to understand what the model learns.

4:10pm

Joel Winterton

Phylodynamics, but bigger

EMBL-EBI

Determining how to use pandemic-scale genomic datasets in disease models is key to determining important information about a disease, like it's origin and spatial patterns of transmission, but there are several challenges to using data at this scale. This talk introduces how genomic data can be used in modelling infectious diseases, and describes how this project gets the ball rolling on addressing some challenges in this field.

4:30pm

Amier Hussen

Volatility Surface Fitting: The SABR Model

aspect capital

Traditional options pricing models like Black-Scholes assumes that the volatility is constant, which results in a flat volatility surface and fails to account for current market dynamics. The SABR model addresses this limitation by introducing stochastic volatility and incorporating parameters which are easy to interpret. We discuss the model, ways in which it is calibrated in practice. We also discuss how this model can help us manage risk in a more accurate and robust way.

Meeting Room 4, 11:00am - 12:30pm

11:00am

Nina Allmeier

UNIVERSITY OF CAMBRIDGE
Department of Applied Mathematics and Theoretical Physics

- Mathematics of Information

Combining Physics Informed Neural Networks with Learned Regularisation

As the number of CT scans and their potential applications increase, reducing the dosage is important. Consequently, reconstructions are noisy and provide little insight into medical issues. Advances in machine learning have made it a promising tool for noise reduction. This talk introduces the concepts of learned regularisation and shows how ML approaches can be used to improve denoising methods.

11:20am

Xane Miles

UNIVERSITY OF CAMBRIDGE
Department of Applied Mathematics and Theoretical Physics

- Numerical Analysis
- Partial Differential Equations

Spectral methods for time-dependent PDEs

A variety of important problems in physics can be written as time-dependent PDEs. We explore a method of numerically approximating the solution to these PDEs as a series of orthonormal functions, with particular regard to systems derived from particular weight function and Dirichlet, Neumann and higher-order boundary conditions, at first on a line and then on a triangle.

11:40am

Katrina Ng

UNIVERSITY OF CAMBRIDGE
Department of Applied Mathematics and Theoretical Physics

- Relativity and Gravitation

Cosmic strings and boson stars

Cosmic strings are a class of topological defect, found from solutions to the wave equation for a complex scalar field with a 'wine bottle' potential, and a potential source of axion dark matter. Boson stars are gravitationally-bound compact objects, and are solitonic solutions of the Einstein-Klein-Gordon equations with a similar potential. This project will investigate the mathematical links between these two phenomena.

12:10pm

Eliza Somerville

UNIVERSITY OF CAMBRIDGE
Department of Pure Mathematics and Mathematical Statistics

- Partial Differential Equations
- General Relativity

Asymptotic Decomposition of a Charged Scalar Field in de Sitter Space

In general relativity, certain spacetimes may be conformally compactified by bringing infinity to a finite region. The behaviour of a field on the boundary of the compactified spacetime may then be translated into the asymptotic behaviour of the field in the physical spacetime. In this project, we use the conformal method to investigate the existence of an asymptotic expansion of a scalar field in de Sitter space.

Meeting Room 4, 1:30pm - 3:00pm

1:30pm

Helia Kalanaki

Canonical Heights on Elliptic Curves

UNIVERSITY OF CAMBRIDGE
Department of Pure Mathematics and Mathematical Statistics

Algebraic Geometry

In this talk, we will explore an unexpected encounter within a family of elliptic curves. By tweaking certain numbers, we noticed surprising patterns in the ranks and regulators of these curves. We then integrated the patterns into deeper concepts in arithmetic geometry. This talk will be offering fresh insights, even for those new to the world of elliptic curves!

1:50pm

Maia Woolf

Constructing algebraic theories via distributive laws

UNIVERSITY OF CAMBRIDGE
Department of Pure Mathematics and Mathematical Statistics

Number Theory & Algebra

Monads are abstract, category theoretic objects that can be used to study entire theories of algebraic structures at once. For example, there is a monad describing abelian groups. Distributive laws allow composing monads to form new monads, and are a vast generalisation of the familiar $a(b + c) = ab + ac$. I will introduce these concepts, assuming few prerequisites.

2:10pm

Yasna Aminaei Chatroodi

Surfaces, Graphs, Commutators

UNIVERSITY OF CAMBRIDGE
Department of Pure Mathematics and Mathematical Statistics

Differential Geometry & Topology

A *commutator* is an element $aba^{-1}b^{-1}$ in a group G. The notation *num(g)* denotes the number of ways that an element g can be written as a commutator up to some equivalence relation. We investigated the maximum *M(F) = max(num(g))* as g ranges over all non-trivial elements of a free group F, and gave an upper bound for it.

2:30pm

Kitty Knight

Gaussian processes in climate science

Institute of Computing for Climate Science

Gaussian processes are a Bayesian machine learning method which can be used to fit flexible models to noisy data. In this project Gaussian processes are used to study the effects of chemical composition on rate of leaf litter decay, by encoding prior scientific knowledge and using decay data to assess which variables have a significant impact on decay rate.

Meeting Room 4, 3:30pm - 5:00pm

3:30pm

Haofei Chen

Fluid: data-linked visualisations

Institute of Computing for Climate Science

Understanding data visualisations means understanding what visual elements represent: how specific elements relate to the data they were computed from. Fluid is an experimental programming language that uses bidirectional dynamic analysis to equip visualisations with information about how their parts relate to the data input, allowing users to explore those connections without the programmer having to build those features explicitly.

3:50pm

Hannah Woods

Data driven modelling of ocean carbon removal strategies

UNIVERSITY OF CAMBRIDGE
Department of Applied Mathematics and Theoretical Physics

Ocean Dynamics

Ocean modelling could provide an effective way to explore the impact of ocean carbon dioxide removal (OCDR) strategies. Macroalgae play a role in carbon sequestration, and cultivation of macroalgae could be an important way to increase carbon export to the deep ocean. This talk discusses the development of the ocean modelling tool OceanBioME, and its application to testing OCDR strategies.

4:10pm

Chloe Y. Huang

Parameter optimisation of ocean biogeochemical models

UNIVERSITY OF CAMBRIDGE
Department of Applied Mathematics and Theoretical Physics

Ocean Dynamics

The Ensemble Kalman Filter and its related algorithms are a class of derivative free computational techniques which can be used to perform parameter optimisation on inversion problems. In this case, it is used to optimise biogeochemical models using real data and which can inform estimations of Carbon exchange between the ocean and atmosphere.

4:10pm

Safian Ali

The flash dynamics of bioluminescent biofilms under flow

UNIVERSITY OF CAMBRIDGE
Department of Applied Mathematics and Theoretical Physics

Fluid Dynamics

Meeting Room 12, 11:00am - 12:30pm

11:00am

Zac Owen

Modelling the Oxidation of Si, SiC: verification and calibration

SILVACO

The oxidation of both Si and SiC plays a key role in the manufacturing of semiconductor devices, so an understanding of oxidation is vital in this industry. This talk builds the theory and mathematics behind oxidation from the ground up (no chemistry needed!), before comparing simulated data to experimental results, with the aim being to then alter the model parameters to get a better fit.

11:20am

Ewan Heaney

Finite difference approximation of the Stokes flow with free interfaces on staggered Cartesian grids

SILVACO

This talk begins by outlining what Stokes flow is and why we are modelling it. The talk progresses onto two problems encountered in the numerical approximation of Stokes flow and how they've been overcome. The talk finishes with some numerical results on the evolution of a polymer under Stokes flow.

11:40am

Daisy Jia

Model comparison and robust estimators for equilibria in supramolecular chemistry

UNIVERSITY OF CAMBRIDGE | Yusuf Hamied Department of Chemistry

The presentation focuses on statistical modelling, addressing challenges like model misspecification, errors in X, and data outliers. Different than normal applications of statistics, here independent variables are calculated from some guessed parameters rather than accurately measured. The PE method is designed, which assigns weights to data based on prediction errors, enhancing the accuracy of parameter estimates. The discussion includes limitations and potential improvements of this mathematical approach.

12:00pm

Weronika Wiesiolek

Quantum error correction using belief propagation algorithms

RIVERLANE

Quantum error correction uses software to detect errors on qubits and correct for them. One of the QEC algorithms is Belief Propagation (BP). BP estimates error likelihoods using a graph algorithm based on Bayesian probability. I will cover error correction basics, explain BP, and highlight its challenges and advantages.