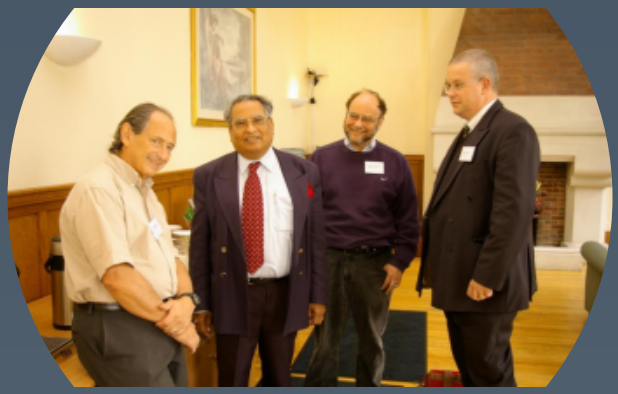


Leeds Annual Statistical Research Workshop (LASR) 2026: Statistics, Deep Learning and Graphical Models

Book of Abstracts

June 29–30, 2026

Shcool of Mathematics

Leeds, UK

Organized by:

Dr. Sofya Titarenko, Dr. Luisa Cutillo, Prof. Kanti Mardia

Letters of Congratulation

Dear Professor Mardia and the LASR Community,

Warmest congratulations to you and LASR as you celebrate 50 years of interdisciplinary collaboration. This milestone reflects a sustained commitment to bringing together diverse perspectives and advancing research at the intersection of complex fields.
It is particularly inspiring to see a focus on protein structure prediction with its 25th Anniversary. As the field continues to evolve rapidly, your work underscores the importance of the foundational discussions and global expertise that LASR has championed for decades.
We also recognise the historic significance of Leeds University, where Astbury's pioneering breakthroughs laid such vital foundations for modern structural biology.
We wish LASR 2026 every success and look forward to your continued achievements in the years ahead.

Best regards,
Pushmeet Kohli, Vice President, Science & Strategic Initiatives, Google DeepMind
30/3/2026

I wish to congratulate the University of Leeds on 50 years of LASR meetings, including 25 years of sustained contributions to protein structure prediction. Protein structure has always been central to my own research, and I have fond memories of participation in LASR 2008 and later returning to Leeds in 2016 to give the Astbury Lecture. Astbury, of course, was one of the pioneers of protein structure.
The 2024 Nobel Prize in Chemistry, awarded for advances in protein structure, highlights how foundational work in this area continues to shape the field. In this context, it is gratifying to see how the LASR community has contributed over the years. In particular, the collaborative PNAS 2008 paper from Leeds (Mardia) and Copenhagen (Thomas Hamelryck) played an important role in the development of AlphaFold1, being explicitly cited as the first probabilistic generative model in this domain.
I warmly congratulate LASR on this memorable 50-year journey and its enduring impact on statistical methodology and structural biology.

Prof. Michael Levitt, Nobel Prize Laureate (Chemistry, 2013),
Professor of Structural Biology, Stanford University
23 April 2026

I want to offer my thanks and warm congratulations to Kanti Mardia and his team at Leeds on the splendid achievement of a half-century of the Leeds Annual Statistical Research workshops. Over this periods, these conferences have been prominent events in the professional calendar, providing an invaluable resource for academic community, and fostering valuable interdisciplinary exchange of ideas across disciplinary boundaries. The enduring success of these events stems from their intimate venues, thoughtfully selected topics, and uniquely collaborative atmosphere balancing regular presenters with newcomers.
Over five decades, LASR has generated not only methodological advances but also institutional legacies. A prominent example is the Mardia Prize (The Royal Statistical Society, 2016), founded to recognise research aligned with the interdisciplinary vision that LASR has championed.

Best regards,
Prof. Peter Green FRS,
Emeritus Professor of Statistics, University of Bristol
30 May 2026

Dear Kanti,

I am writing personally, and as a founder member and past Director of the Astbury Centre for Structural Molecular Biology, to congratulate you for your double celebration of 50 years of LASR and 25 years of Protein Structure Prediction.
As you know solving the structure of proteins and their complexes has been a long road, culminating today in an amazing > 250,000 structures in the pdb, including many large and intricate complexes now solved using the powers of cryoEM. Leeds has played a significant part in this feat, from the diffraction laws of Bragg that enabled protein structures to be solved from crystals, and the pioneering work of Bill Astbury who used X-ray diffraction to understand the architecture of fibrils formed from proteins and DNA. Astbury is also renowned for popularising the term 'molecular biology' (see doi.org/0.1126/science.168.3932.664), not as a method of cloning as it is used today, but through the eyes of understanding biological mechanisms and biological molecules in molecular terms. Astbury knew the importance of protein structure for understanding function, and he coined the terms alpha and beta for the two different structural forms of proteins, some of which he could flip between by pulling on his protein fibrils.
Little did Astbury realise that his so-called cross-beta fibrillar form of a protein would turn out to be the underlying structure of the devastating aberrant protein fibrils that are linked to some of the most challenging of human diseases today, spanning Alzheimer's to Parkinson's, type II diabetes to prions (e.g. doi:10.1016/j.sbi.2026.103245). And it is this concept that forms the logo of the Astbury Centre for Structural Molecular Biology here in Leeds.
Now the quest is on to the solve the structures of these disease-related protein fibrils and to see their interactions in situ in the human brain (e.g. doi: 10.1038/s41586-024-07680-x), so that we can better diagnose, prognose and treat these devastating diseases. And here again structural methods are at the fore, with exciting data again coming from our Astbury Centre, as the papers cited above highlight. So Structural Molecular Biology is still the key, with Alphafold and allied approaches empowering structural work so that it is moving at a pace unparalleled in its history. And working together across the disciplines to make new discoveries is as important today as it was in Astbury's pioneering days in Leeds.
LASR and your community has developed the field of Structure Prediction, that is now so powerful in both discovery science and in its translation for the betterment of humankind. I wish you, and your community, all the very best for a spectacular meeting and look forward to watching the many scientific successes that are bound to come in the years ahead.

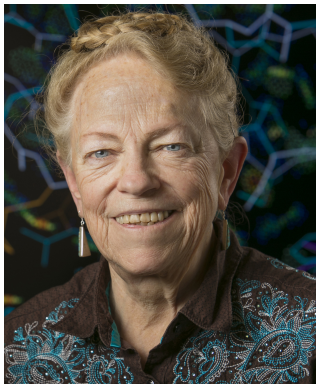
Warm regards,
Prof. Sheena E. Radford OBE FRS FMedSci,
Astbury Professor of Biophysics; Royal Society Professorial Research Fellow,
University of Leeds
8 June 2026

Contents

1

Keynote Speakers

Keynote Presentation



Prof. Jane S. Richardson

Professor in biochemistry

Duke University, USA

Biography

Jane Richardson and her husband David have worked together for over 50 years on research to understand the 3D structure of protein and RNA molecules. They were early pioneers in protein crystallography, protein de novo design, and molecular graphics. They developed a method that calculates hydrogen-atom contacts to quantify the details of modeled packing interactions, widely used on their MolProbity website and elsewhere to improve the accuracy of macromolecular structures by crystallography, cryoEM, and now AI predictions. Jane developed the ribbon representation of protein structures, described many common features of overall protein folds and local motifs (Greek key beta barrels, helix caps, etc.), and works to spread molecular 3D literacy at Duke and around the world. From a Swarthmore B.A. in philosophy she has become a biophysicist, a MacArthur Fellow, a member of the National Academies of Sciences & of Medicine, Hollaender Award in Biophysics, and has three honorary degrees. For details, see Ann. Rev. Biophysics 4:1-28.

More recently she and Dave have been busy responding to the CryoEM Resolution Revolution,

3D structures of SARS CoV-2 macromolecules, and AlphaFold structure predictions including what happens when they fail.

Title of Talk

The Good and the Bad of AlphaFold2 and AlphaFold 3

Abstract

Experimental structures usually start from high-accuracy, high-pLDDT (the estimate of AlphaFold prediction confidence) AlphaFold models and are then refined further to better match the data, which greatly speeds up the process since there is no need to solve the phase problem or to build a new initial model into a map of the structure. However, at pLDDT < 70, AlphaFold2 does not fail gracefully, with a majority of models looping out into "barbed-wire" structures containing such a large number of serious outliers that they are not physically plausible.

AlphaFold3, which can handle structures with multiple components such as proteins, nucleic acids, ligands, and carbohydrates, produces fewer instances of barbed-wire. However, it instead appears to produce more "hallucinations", or regions that resemble normal secondary structure but are actually incorrect.

I have examined a limited number of AlphaFold3 predictions in detail so far, reflecting its current availability primarily through a webserver for individual structures rather than a large-scale database such as the AlphaFold Protein Structure Database, which provides predictions for millions of proteins. These observations are therefore based on a relatively small sample but nevertheless offer initial insights into the successes and limitations of the method.

Keynote Presentation

Dr Emma Robinson

Head of Research Department of Biomedical Computing

King's College London, UK

Biography

Dr Robinson's research focuses on the development of computational methods for brain imaging and neuroscience, spanning classical as well as AI methods. Most notably, her software for cortical surface registration (Multimodal Surface Matching, MSM) was central to the development of the Human Connectome Project's (HCP)"Multi-modal parcellation of the Human Cortex " (https://www.nature.com/articles/nature18933), and has featured as a central tenet in the https://www.nature.com/articles/nn.4361 for neuroimage analysis. Her current research interests are focused on the development of geometric deep learning tools for modelling functions on the cortical and cardiac surfaces, optimising image processing pipelines open data collections such as the HCP, developing HCP and UK Biobank, and building learning-based models of neurological processes (neurological Digital Twins).

Title of Talk

Geometric deep learning for precision cortical modelling and simulation

Abstract

The shape of any object is inextricably linked to its function. Whether it be electrical signalling, chemical diffusion, fluid flow, or biomechanical deformations, these physical processes are inherently constrained by the topology and geometry of that object. This explains why shape is so often a vital biomarker of disease. It also explains why physics simulations must be solved from meshes or graphs that have been carefully tailored to capture object structure (e.g. surface curvature or network topology); they cannot be effectively run from standard Euclidean grids.

This presents a problem when seeking to address these questions using artificial intelligence (AI); since most modern AI frameworks are designed for Euclidean domains (e.g. images/text/audio) and do not generalise easily to curved surfaces or irregular graphs. Yet - this is a missed opportunity - since, as has already proven for imaging, AI has the potential to resolve many of the challenges faced by the shape modelling and simulation community.

In this talk I will present the background to geometric deep learning over meshes and graphs. I will show how we have used it to improve the precision with which we

model neurological processes. I will touch on physics-informed AI and how together these tools can be used to stimulate biological processes.

Keynote Presentation

Dr Daniela De Canditiis

Istituto per le Applicazioni del calcolo CNR

Rome, Italy

Biography

Dr Daniela De Canditiis graduated in Mathematics in 1995 from the University of Naples "Federico II," where she also obtained a Ph.D. in Applied Mathematics and Computer Science in 2002. Since then, she has been with the Institute for Applied Calculus (IAC-CNR), where she currently holds the position of Senior Researcher. Her research focuses on mathematical modelling and statistical inference for both low-dimensional and high-dimensional data. In particular, she develops novel statistical and machine learning methods for the analysis of complex data, with expertise spanning nonparametric regression, graphical model inference, survival analysis, and clustering techniques. She has contributed to numerous interdisciplinary projects across diverse application domains, including atmospheric physics, genetics, medicine, neuroscience, cultural heritage, and environmental sciences.

Title of Talk

Structure Learning of Complex Systems: Advanced Methods for Inferring Networks from Multiple Count Datasets

with Francesco Migliaccio[1], *Claudia Angelini*[2], *Annamaria Carissimo*[2]

[1]University of Naples Federico II, Italy, [2]University of Seoul, South Korea

Abstract

The objective of network inference from observational data is to uncover the underlying statistical dependencies among variables, thereby elucidating the structure of complex systems. In modern genomics, high-throughput technologies such as single-cell next-generation sequencing (scRNA-seq) generate data consisting of discrete counts of sequencing reads. These data, while rich in biological information, pose substantial statistical challenges: they are typically overdispersed (their variance often exceeds the mean), zero-inflated (a large proportion of zeros arises from technical or measurement effects rather than absolute biological absence of the transcripts), and high-dimensional (involving tens of thousands of genes or features measured on the same conditions for each cell). Furthermore, integrating information from multiple datasets to infer a shared underlying structure is a pervasive and increasingly important challenge in modern data analysis—particularly, though not exclusively, in network infer-

ence. In this talk, after an overview of existing approaches proposed in the literature, we will present **a new joint inference method for count data** that extends the penalized Zero-Inflated Negative Binomial (ZINB) model of Park et al. (2021) from the single-dataset to the multi-dataset setting. The proposed framework simultaneously addresses overdispersion, zero inflation, and high dimensionality, while enabling coherent joint inference across datasets. It thereby provides a novel alternative to the few existing methods that can address all these challenges within a unified statistical model. Finally, results obtained from both synthetic and real datasets will be shown to illustrate the performance and potential of the proposed approach.

Key References:

[1] B. Park, H. Choi, and C. Park (2021).
Negative binomial graphical model with excess zeros. Statistical Analysis and Data Mining: The ASA Data Science Journal, 14, 449–465. link.

2

Featured Speakers

Featured Talk

Prof. Kanti Mardia OBE

University of Leeds

Biography

Kanti V. Mardia is currently a Senior Research Professor, a position which has been specially created by the University of Leeds after ending in 2000 his position as Chair of Applied Statistics which he had held since 1973. He was also a Visiting Professor at Oxford University from March 2013 to September 2023. From October 2017-2024, he was the Leverhulme Emeritus Fellow.

Professor Kanti Mardia has received many prestigious honours, including the Guy Medal in Silver of the Royal Statistical Society in 2003, and the Wilks Memorial Medal by the American Statistical Society in 2013.

In 2019, Mardia received the **Life Achievement Award** of the International Indian Statistical Association for recognition of his achievements. He was awarded the Mahatma Gandhi Medal of Honour 2019-20 by the NRI Institute on the occasion of the 150th Birth Anniversary "for his outstanding contributions". He received OneJAIN Life Achievement Award 2021 by the Jain All Party Parliament Group for his work on Science and Jainism.

Professor Mardia was awarded an OBE (the Officer of the Most Excellent Order of the British Empire) for “Services to Statistical Sciences” in the New Year Honours List of 2023. For some details see this article. Her Royal Highness Princess Anne performed the Investiture Ceremony for the OBE at Buckingham Palace on 3rd March 2023.

One of his very high profile research works is of saving lives by analysing the shape of the brain: methods developed for shape analysis by him in Leeds, with other collaborators, have been used on brain images to assess the extent of brain damage in people suffering from fetal

alcohol spectrum disorders (FASD). One of the new high impact projects in progress is related to craniofacial surgery for facial deformities such as cleft lip. One measure of success is the ability of the patient to make "normal-looking" facial expression such as a smile.

Professor Kanti Mardia also contributed towards the establishment of the Centre of Medical Imaging Research (CoMIR) in 1993, the Centre of Statistical Bioinformatics (CoSB) Leeds in 2006 and he was also the founding Vice-President of the International Indian Statistical Association (IISA). He founded and has organised the Leeds Annual Statistics Research workshops (LASR) which have grown since 1973 into international conferences. With the aim of making his vision of LASR broader by encouraging cutting-edge interdisciplinary work between statisticians and other science communities, the **Mardia Prize** has been created. This annual prize was inaugurated by the Royal Statistical Society on 7 September 2016 to fund a grant programme supporting events or workshops in emerging interdisciplinary areas.

His research monographs Statistical Shape Analysis, Statistics of Directional Data, Multivariate Analysis and Spatial Analysis are classics.

Title of Talk

LASR at 50: Statistics, 25 Years of Structure, and the Nobel Prize

Abstract

This talk reviews key themes that have shaped LASR over the past 50 years, with particular emphasis on the long-standing focus on Protein Structure Prediction and how this area ultimately contributed to the 2024 Nobel Prize in Chemistry. The prize recognised the breakthrough achievements of Demis Hassabis and John Jumper of Google DeepMind, and David Baker, who solved the long-standing protein folding problem.

A central part of the talk will describe how progress in protein structure prediction has been assessed through CASP, the biennial international competition in which research groups submit predictions for new protein sequences whose experimental structures are known only to the organisers. The dramatic performance of DeepMind's AlphaFold1 at CASP13 (2018) marked a turning point in the field and demonstrated the power of deep learning for structural biology.

Our own contributions — including early LASR work on graphical models and directional statistics, the TorusDBN generative model (PNAS 2008), and subsequent developments of the Reference Ratio Method (LASR 2011) — helped establish statistical and generative approaches to protein geometry. These ideas form part of the intellectual lineage that impliitly fed into modern deep-learning methods (AlphaFold1).

The talk will also outline future directions for LASR, reflecting on how its interdisciplinary mission continues to evolve. One spin off is the establishment of the Annual Mardia Prize by the Royal Statistical Society in 2016, created to promote research aligned with the LASR vision.

Key References:

[1] W. Boomsma et al. (2006). *Graphical models and directional statistics capture protein structure.* In *LASR 2006: Statistical Modelling and Bioinformatics in the New Era*, Leeds University Press, pp. 91–94.

[2] T. Hamelryck, J. T. Kent, and A. Krogh (2006). *Sampling realistic protein conformations using local structural bias.* PLoS Computational Biology, 2(9), e131.

[3] W. Boomsma et al. (2008). *A generative, probabilistic model of local protein structure*. PNAS, 105(26), 8932–8937.

[4] K. V. Mardia et al. (2011). *A statistical view on the Reference Ratio Method*. In *LASR 2011 Proceedings*, Leeds University Press, pp. 55–61.

[5] A. W. Senior et al. (2018). *Protein structure prediction using deep learning: AlphaFold*. Proteins, 87(12), 1141–1149.

[6] A. Kryshtafovych et al. (2019). *CASP—Round XIII*. Proteins, 87(12), 1011–1020.

[7] LASR Proceedings (2006, 2015, 2021).

Featured Talk

Prof. Thomas Hamelryck

University of Copenhagen

Biography

Thomas Hamelryck is Full Professor of Machine Learning at the University of Copenhagen, holding joint appointments in the Department of Biology and the Department of Computer Science. His research expertise lies in Bayesian modeling, probabilistic machine learning, and deep probabilistic programming, with particular focus on applications to protein structure prediction (joint work with Kanti V. Mardia and John T. Kent, Leeds) and protein evolution (a groundbreaking deep model of phylogenetics was published in ICLR, 2022; joint work with Jotun Hein, Oxford). His work on generative models of protein structure based on directional statistics and Bayesian networks from the 2010s was acknowledged by DeepMind as a direct precursor of AlphaFold. In addition, he is the author of the widely used Bio.PDB structural bioinformatics module of the open source software package Biopython, and teaches courses in data science, structural bioinformatics and probabilistic machine learning.

Title of Talk

AlphaFold's Bayesian Roots in Directional Statistics and Probability Kinematics
with Kanti V. Mardia[1]
[1]University of Leeds, UK

Abstract

In 2018, CASP13 (the 13th edition of the "Olympic Games" of protein structure prediction) reported "[...] unprecedented progress in the ability of computational methods to predict protein 3D structure. The reasons are not yet fully clear [...]". That reason was, of course, the first version of DeepMind's AlphaFold, which fundamentally changed protein structure prediction and arguably science itself. AlphaFold1 was published as "Improved protein structure prediction using potentials from deep learning" by Nature in 2020, which revealed that the AI-based method used deep networks to parameterize potentials over dihedral angles and pairwise distances, combining the two using a third, so-called reference potential. This approach of combining potentials was heuristically justified by referring to the classic work by the physicist John

Kirkwood from 1935 on potentials of mean force for liquids. But there's more to the story, and it builds on some fascinating work done well after 1935. I will relate how AlphaFold's paradigm-shifting first version is rooted in the classic work on the directional statistics of proteins by Kanti V. Mardia and collaborators from around 2010 (highlighting the role of the LASR meetings) and the generalized Bayesian inference calculus called probability kinematics pioneered by Richard Jeffrey in the 1950s, which was first brought to protein structure prediction by the seminal work of Manfred Sippl in the early 1990s as a heuristic. In contrast to AlphaFold2 and 3, AlphaFold1 is an SE(3)-invariant probabilistic model based on generalized Bayesian updating of a directional prior. This overall approach is still of considerable relevance for future applications in probabilistic machine learning.

Key References:

[1] W. Boomsma, K. V. Mardia, C. C. Taylor, J. Ferkinghoff-Borg, A. Krogh, and T. Hamelryck (2008). *A generative, probabilistic model of local protein structure*. PNAS, 105, 8932–8937.

[2] T. Hamelryck, K. V. Mardia, and J. Ferkinghoff-Borg (Eds.) (2012). *Bayesian methods in structural bioinformatics*. Springer, Statistics for Biology and Health.

[3] M. Golden, E. Garcia-Portugues, M. Sørensen, K. V. Mardia, T. Hamelryck, and J. Hein (2017). *A generative angular model of protein structure evolution*. Molecular Biology and Evolution, 34, 2085–2100.

[4] L. S. Moreta, O. Rønning, A. S. Al-Sibahi, J. Hein, D. Theobald, and T. Hamelryck (2022). *Ancestral protein sequence reconstruction using a tree-structured Ornstein-Uhlenbeck variational autoencoder*. ICLR.

[5] T. Hamelryck and K. V. Mardia (2026). *AlphaFold's Bayesian Roots in Probability Kinematics*. AISTATS. link.

[6] K. V. Mardia, B. Eltzner, and S. F. Huckemann (2025). *Constrained Shape Analysis with Applications to RNA Structure*. arXiv:2502.16270.

[7] H. Wiechers, C. J. Williams, B. Eltzner, F. Hoppe, M. G. Prisant, V. B. Chen, E. Miller, K. V. Mardia, J. S. Richardson, and S. F. Huckemann (2026). *RNAprecis: Prediction of full-detail RNA conformation from the experimentally best-observed sparse parameters*. PLOS Computational Biology. link.

Featured Talk

Prof. Stephan Huckemann

Georg August University of Göttingen, Germany

Biography

Stephan Huckemann received his degree in mathematics from the University of Giessen (Germany) in 1987. He was a visting lecturer and scholar at the University of Michigan, Ann Arbor (1987 - 1989) and a postdoctoral research fellow at the ETH Zürich, Switzerland (1989 - 1990). He then worked as a commercial software developer (1990 - 2001) and returned to academia as a contributor to the computer algebra system MuPAD at Sciface Software and the University of Paderborn, Germany (2001 - 2003). Working with H. Ziezold (2004 - 2006, Univ. of Kassel), P. Mihailescu (2007, Univ. of Göttingen), P.T. Kim (2009, Univ. of Guelph, Canada) and A. Munk (2007 - 2010, Univ. of Göttingen) he completed his Habilitation at the Univ. of Göttingen (Germany) in 2010 and was awarded a DFG Heisenberg research fellowship. As such he continued

at the Institute for Mathematical Stochastics, Univ. of Göttingen while being a research group leader at the Statistical and Applied Mathematical Sciences Institute (2010/11 SAMSI, Durham, NC, USA). After substituting (2012 - 2013) for the Chair of Stochastics and Applications at the Univ. of Göttingen he holds the new Chair for Non-Euclidean Statistics.

Title of Talk

Torus PCA and Constrained Shape for Rare Data

with Benjamin Eltzner[1] and Kanti V. Mardia[2]
[1]Georgia-Augusta-University of Göttingen, Germany, [2]University of Leeds, UK

Abstract

Highly curated ground truth gold standard data–while difficult to obtain, and hence rare–is essential for realistic statistical learning. Amending for scarceness of data, we include specific structure into model and algorithm building. We illustrate mathematical details for the task of reconstructing high detailed RNA structure from low detail measurements. One major ingredient for obtaining high detail structure classes is Torus PCA. Another major ingredient is Constrained Shape for modeling low detail structure.

Key References:

[1] K. V. Mardia, H. Wiechers, B. Eltzner, and S. F. Huckemann (2022). *Principal component analysis and clustering on manifolds*. Journal of Multivariate Analysis. link.

[2] K. V. Mardia, B. Eltzner, and S. F. Huckemann (2025). *Constrained Shape Analysis with Applications to RNA Structure*. arXiv:2502.16270.

[3] H. Wiechers, C. J. Williams, B. Eltzner, F. Hoppe, M. G. Prisant, V. B. Chen, E. Miller, K. V. Mardia, J. S. Richardson, and S. F. Huckemann (2026). *RNAprecis: Prediction of full-detail RNA conformation from the experimentally best-observed sparse parameters*. PLOS Computational Biology. link.

Featured Talk

Dr. Chuoxin Ma

Beijing Normal-Hong Kong Baptist University, Cina

Biography

Dr. Ma is an Associate Professor in the Department of Statistics and Data Science at Beijing Normal–Hong Kong Baptist University (BNBU). Prior to joining BNBU, she was a Postdoctoral Research Associate at the University of Cambridge, supervised by Dr. William Astle and Professor John Aston. She completed her PhD at the University of Manchester under the supervision of Professor Jianxin Pan. Dr Ma's research focuses on statistical modeling of imaging data, survival analysis, and longitudinal data analysis. Her research has been published in journals such as the Annals of Applied Statistics, the Scandinavian Journal of Statistics, Statistics in Medicine, the Biometrical Journal, BMJ Open, and others.

Title of Talk

Low-rank structure reshape for tensor regression with application in neuroimaging-based reading comprehension analysis

with Tianyi Lu[1]*, Bohai Zhang*[1]*, Haiyan Liu*[2] *and Jianxin Pan*[1]
[1]Beijing Normal-Hong Kong Baptist University, China [2]University of Leeds, UK

Abstract

A common challenge in magnetic resonance imaging (MRI) analysis is the estimation of high-dimensional parameters from limited data, as sample sizes are often small to moderate. To solve this problem, we propose a novel method based on reshaping the structure of the Tucker decomposition in tensor regression. By recasting the tensor into a higher-dimensional space with significantly shorter mode lengths, our approach effectively relaxes the requirement on sample size. Unlike approaches that simply reshape the tensor covariate as a preprocessing procedure, our method directly reshapes the low-rank structure of the Tucker decomposition of the regression coefficient. We establish the algorithmic convergence and statistical convergence properties of the proposed estimator. Simulation results demonstrate that the proposed method achieves lower estimation error, lower prediction error, and a higher signal recognition rate than the existing methods. We also apply our method to a reading comprehension analysis.

Featured Talk

Dr. Bailey Andrew

University of Manchester, UK

Biography

Bailey Andrew completed his BSc in Artificial Intelligence and Mathematics at the University of Edinburgh in 2021. Until March of this year, he was a PhD candidate in the Centre for Doctoral Training in Artificial Intelligence for Medical Diagnosis and Care at the University of Leeds - although he has now started a post-doc at the University of Manchester. His research interests include the development of graphical models and the application of statistics to medical research.

Title of Talk

Developing practical multi-axis graphical models

with Bailey Andrew[1]*, David Westhead*[2]*, Luisa Cutillo*[2]
[1] University of Manchester, UK, [2]University of Leeds, UK

Abstract

Multi-axis graphical models allow one to learn networks describing all axes of a dataset. A typical use case would be "single-cell RNA sequencing" dataset (scRNA-seq), which often takes the

form of a matrix whose entries represent the expression of a given gene (column) in a given cell (row) - in such a case, multi-axis models would simultaneously learn a gene network and a cell network. Doing simultaneous inference allows one to avoid independence assumptions (i.e. assuming cell independence to learn a gene network, like most methods) and take advantage of inter-task transfer. Unfortunately, multi-axis models have historically been slow and restricted to simple (Gaussian) scenarios. In this talk, we present our work making such models scalable to millions of samples/features as well as work relaxing and generalizing the assumptions made by such models. We ground the talk in real examples, and conclude with a discussion of interesting directions for future research.

Key References:

[1] Alfredo Kalaitzis, John Lafferty, Neil D. Lawrence, Shuheng Zhou (2013) *The Bigraphical Lasso*. Proceedings of the 30th International Conference on Machine Learning.

[2] Andrew, B., Westhead, D.R. and Cutillo, L., 2024, January. *GmGM: a fast multi-axis Gaussian graphical model*. In Proceedings of AISTATS 2024 (Vol. 238).

[3] Andrew, B., Westhead, D.R. and Cutillo, L., 2025, April. *Gaussian graphical modelling without independence assumptions for uncentered data*. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 15, pp. 15391-15398).

[4] Andrew, B., Westhead, D.R. and Cutillo, L., 2026. *The Cartesian Gaussian additive noise model for directed network inference in omics data*. BMC Medical Research Methodology.

Featured Talk

Dr Valeria Policastro

University of Naples Federico II, Italy

Biography

Valeria Policastro obtained her Master's degree in Statistics from the University of Naples Federico II in 2018. She subsequently completed a second-level Master's degree in Medical and Genomic Statistics at the University of Pavia. During her training, she carried out an internship at the "Mauro Picone" Institute for Applied Mathematics of the National Research Council of Italy (CNR), working on statistical methods for genomic data, network robustness, and validation. In 2022, she received her PhD from the University of Campania "Luigi Vanvitelli" with a thesis entitled Network Methods for Biomedical Discoveries. During her PhD, she won a CNR Short-Term Mobility grant for a project on network integration, which allowed her to spend a research period at Uppsala University, Sweden, working with Prof. Matteo Magnani. She received the Best Presentation Award, supported by the VADISTAT Association, for her presentation "INet for Network Integration" delivered at the ARS'23 conference. She was also awarded funding under the Erasmus+ KA1 Higher Education Staff Mobility Programme for a research period in 2025 at Uppsala University, Sweden, collaborating with Prof. Davide Vega D'Aurelio. This project investigated the influence of gender diversity in academic groups through the analysis of multilayer networks. Since 2023, she has been lecturer in Social Statistics at the University of Naples Federico II, where she teaches Network Science and Computational Biostatistics in the Master's degree programme in Statistics.

Title of Talk

Assessing community robustness in weighted networks

with Domenico Sgambato[1,2] , Dario Righelli [3], Vincenzo Della Gatta [1], Annamaria Carissimo [1] and Luisa Cutillo[4]
[1]Institute for Applied Mathematics "Mauro Picone", National Research Council, Italy
[2]Department of Electrical Engineering and Information Technology, University of Naples "Federico II", Italy
[3] Department of Biology, University of Padua, Italy
[4]School of Mathematics, University of Leeds,United Kingdom.

Abstract

Community detection is widely used to identify meaningful groups in network-based representations of high-dimensional data. However, the reliability of the detected communities should be assessed through appropriate validation procedures. The robin framework provides a strategy to evaluate the stability of community partitions under graph perturbations, but its original formulation is limited to unweighted networks. In this work, we extend `robin` to weighted networks, motivated by the role that edge weights play in many real-world applications. To address this limitation, we introduce three novel perturbation strategies, Rewire, Sum, and Shuffle, specifically designed to evaluate community stability in weighted networks. The proposed extension enables the assessment of community detection methods under progressive perturbations of weighted graph structures. We demonstrate the applicability of the framework on weighted social and biological networks.

Key References:

[1] V. Policastro, D. Righelli, L. Cutillo, A. Carissimo. *robin2: accelerating single-cell data clustering evaluation*, Bioinformatics Advances, Volume 5, Issue 1, 2025, vbaf184, link.

[2] V. Policastro, D.Righelli, L. Cutillo, A. Carissimo, I. De Feis. *ROBustness In Network (robin): an R package for Comparison and Validation of communities.* The R Journal, 7 June 2021.

[3] A. Carissimo, L. Cutillo, I. De Feis (2018). *Validation of community robustness.* Computational Statistics and Data Analysis, 120, 1–24. doi:10.1016/j.csda.2017.10.006.

3

Contributed Talks

Introducing a Physical Cylindrical Galton Board for Directional Data: Concepts, Challenges, and Applications

Colin Goodall[*1], Kanti V Mardia[1], John Rubbo[2],

[1] University of Leeds, UK
[2] Independent

* Presenting author

Abstract

The Galton Board is a device invented about 150 years ago by Francis Galton for use in his Royal Society lectures. The board demonstrates how the binomial distribution tends toward the normal distribution. It consists of a triangular array of pegs through which balls fall, producing the normal distribution. The device has been extremely successful and widely used for many decades, and today it is commonly used in schools to illustrate the central limit theorem. There will be a display of few physical versions. However, the classical Galton Board is designed for linear data, and until now there has been no comparable device to illustrate the behaviour of angular or directional distributions. We introduce an angular Galton Board, a cylindrical version suitable for circular data. Angular or directional data have become increasingly important, especially in molecular biology and protein structure prediction. One of the key methods in structure prediction depends on modelling the core dihedral angles of the protein backbone. A physical device that illustrates directional distributions is therefore extremely valuable for teaching and intuition. We will describe how our cylindrical model is constructed. One of the challenges, compared with the classical Galton Board, is that the peg layout on the cylinder is highly constrained by the number of columns, and

for large numbers of peg rows the resulting wrapped normal distribution can even become uniform. Our talk will highlight these points, demonstrate our physical model, and show how the cylindrical board can be simulated, including interactive computer simulations. We will also discuss how the model can be extended to toroidal data, which arise naturally in protein structure prediction and more generally in molecular biology, where multiple angles must be modelled jointly.

Key References:

[1] K. V. Mardia, C. Goodall, and J. Rubbo (2026). *A Cylindrical Galton Board at the Galton Board's 150th Anniversary*. Available at: arXiv:2603.07742.

Differential Dynamic Causal Nets: Model Construction, Identification and Group Comparisons

Kang You[*1], Gary Green[1], Jian Zhang[1]

[1] University of Kent, UK

* Presenting author

Abstract

Pathophysiolpgical modelling of brain systems from microscale to macroscale remains difficult in group comparisons partly because of the infeasibility of modelling the interactions of thousands of neurons at the scales involved. Here, to address the challenge, we present a novel approach to construct differential causal networks directly from electroencephalogram (EEG) data. The proposed network is based on conditionally coupled neuronal circuits which describe the average behaviour of interacting neuron populations that contribute to observed EEG data. In the network, each node represents a parameterised local neural system while directed edges stand for node-wise connections with transmission parameters. The network is hierarchically structured in the sense that node and edge parameters are varying in subjects but follow a mixed-effects model. A novel evolutionary optimisation algorithm for parameter inference in the proposed method is developed using a loss function derived from Chen-Fliess expansions of stochastic differential equations. The method is demonstrated by application to the fitting of coupled Jansen-Rit local models. The performance of the proposed method is evaluated on both synthetic and real EEG data. In the real EEG data analysis, we track changes in the parameters that characterise dynamic causality within brains that demonstrate epileptic activity. We show evidence of network functional disruptions, due to imbalance of excitatory-inhibitory interneurons and altered epileptic brain connectivity, before and during seizure periods.

Shape analysis of AF segments for rapid assessment of Mohs layers for BCC presence by AF-Raman microscopy

Alexey A. Koloydenko[*1], Ioan Notingher[1], Radu Boitor[1], Jüri Lember[1]

[1] Royal Holloway, University of London, UK

* Presenting author

Abstract

An automated method to detect Basal Cell Carcinoma (BCC) relies on Autofluorescence (AF) imaging guiding Raman microscopy to obtain biochemical information for tissue classification. The guidance is provided via an image segmentation technique aiming to reduce the risk of missing cancer. Recently we presented evidence that shape of an AF segment may be useful for 'trimming' a fraction of non-BCC segments in an essentially coordinate free manner, without compromising BCC detection. By allowing the AF-Raman method to focus the more time consuming Raman analysis on remaining regions, the proposed trimming could ultimately improve the overall accuracy. The presented shape analysis used Smooth Weighted Euler Curve Transform (SWECT). SWECT embeds segments in a space of real matrices of fixed dimensions, where a shape is an equivalence class of segments with SWECTs matching under a cyclic permutation of matrix columns. The induced rotation invariant distance is non-Hilbertian, which requires special care when using kernel-based methods (e.g. Kernel PCA, Kernel LDA, SVMs). Our previously reported best results were achieved by a semisupervised L_1 SVM based on the Laplace 'kernel' and additionally regularized with respect to the symmetrically normalized Graph Laplacian eigenmaps. With more data, however, computational issues, particularly the ones related to the lack of positive-definiteness, make the approach less attractive. Instead, we now use an 1-norm SVM that can be solved via linear programming even when the underlying similarity measure is indefinite, such as our 'kernels' given by the rotation invariant quasi dot product of SWECTs or the Laplace and Gaussian RBFs with any value of the scale parameter. We address the issue of comparing segments of significantly different sizes, as the fixed resolution of the SWECT-based approach may result, e.g., in spurious similarity of shapes. Specifically, we experiment with redefining the shape distance to be infinite for segments of incomparable sizes and assess robustness of our findings to such alterations. Tissue specimens used in this work were obtained from patients undergoing Mohs surgery at the Nottingham NUH NHS Treatment Centre.

Key References:

[1] K. Kong, C. J. Rowlands, et al. (2013). *Diagnosis of tumors during tissue-conserving surgery with integrated autofluorescence and Raman scattering microscopy*. PNAS, 110(38), 15189–15194. link.

[2] R. Boitor, S. Varma, et al. (2024).
Ex vivo assessment of basal cell carcinoma surgical margins in Mohs surgery by autofluorescence-Raman spectroscopy: A

pilot study. JEADV Clinical Practice, 3(2), 498–507. link.

[3] R. A. Boitor, S. Varma, et al. (2024). *Diagnostic accuracy of autofluorescence-Raman microspectroscopy for surgical margin assessment during Mohs micrographic surgery of basal cell carcinoma*. British Journal of Dermatology, 191(3), 428–436. link.

[4] R. Boitor, K. Kong, et al. (2017). *Automated multimodal spectral histopathology for quantitative diagnosis of residual tumour during basal cell carcinoma surgery*. Biomedical Optics Express, 8(12), 5749–5766. link.

[5] R. Boitor, C. de Wolf, et al. (2021). *Clinical integration of fast Raman spectroscopy for Mohs micrographic surgery of basal cell carcinoma*. Biomedical Optics Express, 12(4), 2015–2026. link.

[6] A. A. Koloydenko, I. Notingher, et al. (2024). *Shape analysis of AF segments for rapid assessment of Mohs layers for BCC presence by AF-Raman microscopy*. In *Developments in Statistical Modelling*, pp. 211–217. link.

[7] L. Crawford, A. Monod, et al. (2020). *Predicting clinical outcomes in glioblastoma: An application of topological and functional data analysis*. Journal of the American Statistical Association, 115(531), 1139–1150. link.

[8] Q. Jiang, S. Kurtek, et al. (2020). *The weighted Euler curve transform for shape and image analysis*. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 3685–3694. link.

[9] I. Alabdulmohsin, M. Cisse, et al. (2016). *Large margin classification with indefinite similarities*. Machine Learning, 103(2), 215–237. link.

[10] J. Zhu, S. Rosset, et al. (2003). *1-norm support vector machines*. In *Advances in Neural Information Processing Systems*, Vol. 16, MIT Press, pp. 49–56. link

Spectral Power Analysis with M/EEG Data

Jian Zhang[*1]

[1] University of Kent, UK

* Presenting author

Abstract

Magnetoencephalography (MEG) neuroimaging technique has provided a non-invasive tool for facilitating early detection, diagonosis and treatment of brain conditions such as mild traumatic brain injury, dimentia and epilepsy. One of important limitations of commonly used two-samples diagnostic tests in clinical neuroscience is that control subjects are homogeneous and normally distributed subjects. This assumption is invalid in the single-case neuroimaging study where spetrum data of cases/controls are heterogeneous and not normally distributed. To tackle this issue, we develop a novel two-sample testing procedure for diagnosis of individuals against heterogeneous controls. The proposed procedure involves a region-wise likelihood ratio test for abnormality in the whole brain. To adjust bias, due to hetegeneity in the control data, we calibrate the resulting p-values by use of parametric bootstrap resampling. Applications to both the real and synthetic data show that the new procedure is promising when the between-case-control discrepancy overtakes the within-control discrepancy.

Investigating causal networks of dementia using causal discovery and natural language processing models

Hui Guo[*1]

[1] University of Manchester, UK

* Presenting author

Abstract

Comprehensively studying modifiable risk factors to understand their contributions to dementia mechanisms is imperative. We used natural language processing (NLP) models to pre-select candidate risk factors for dementia from over 5000 baseline variables in the UK Biobank. We then applied causal discovery approaches to examine the relationships among the selected variables and their links to dementia in later life, presenting these connections in a causal network. We identified eight risk factors that directly or indirectly influence dementia, with mental disorders due to brain dysfunction acting as direct causes and mediators in pathways from other neurological disorders to dementia. Although evidence for the direct link between biological age and dementia was less pronounced, its potential value in dementia management remains non-negligible. This study advances our understanding of dementia mechanisms and highlights the potential of NLP and machine learning for the causal discovery of complex diseases from high-dimensional data.

Key References:

[1] X. Yu, A. Lophatananon, V. Holmes, K. R. M., and H. Guo (2025). *Investigating causal networks of dementia using causal discovery and natural language processing models*. npj Dementia, 1, 4. link.

Detection, coverage and percolation in dynamic Boolean models with random radii based on α-stable processes

Peter Gracar[*1], Benedikt Jahnel[2], Lukas Lüchtrath[3], Anh Duc Vu[4]

[1] School of Mathematics, University of Leeds, UK
[2] Technische Universität Braunschweig, Germany
[3] Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany
[4] Unaffiliated

* Presenting author

Abstract

We consider a dynamic network in continuum time and space in which nodes, with initial locations given by a Poisson point process, move according to i.i.d. isotropic α-stable processes. Each node is additionally equipped with an i.i.d. detection radius. Inspired by corresponding results by Peres et. al. on mobile networks based on Brownian sausages with fixed width, we investigate the tail behaviour of three stopping times: The detection time of the first discovery of a designated node, the first coverage of an entire set, and the first discovery of a node by the infinite connected component of the system. Broadly speaking, we discover that the stability index as well as the random radii manifest themselves only in constants in the otherwise exponential decay rates. The proofs rest on heat-kernel bounds for the underlying Lévy processes and a detailed multiscale analysis allowing us to control the space-time correlations of the system.

Key References:

[1] P. Gracar, B. Jahnel, L. Lüchtrath and A. D. Vu (2026). *Detection, coverage and percolation in dynamic Boolean models with random radii based on α-stable processes*. Available at: arXiv:2602.22109.

Functional Data Analysis for Time-varying Binary Networks

Puchong Paophan[*1], Georgios Aivaliotis[1], Leonid Bogachev[1], Haiyan Liu[1]

[1] School of Mathematics, University of Leeds, UK

* Presenting author

Abstract

Time-varying binary networks arise in fields such as neuroscience, social science, and biology, yet statistical tools for their analysis remain underdeveloped. We propose a novel method for analysing such networks using a binary functional data framework. Each time-varying binary network is represented by a sequence of adjacency matrices. By vectorising, we convert each matrix into a binary vector, yielding multivariate binary functional data. Classical multivariate functional principal component analysis (MFPCA) is not directly applicable due to the discrete nature of the data. To overcome this, we analyse each binary function separately using logistic functional principal component analysis (LFPCA). The resulting univariate LFPCA components are then combined into multivariate principal components, which are mapped back to the network space via inverse vectorisation. This approach yields low-dimensional, interpretable representations of time-varying binary networks and supports downstream tasks such as clustering and visualisation. Our method provides a principled, scalable framework for analysing binary network-valued functional data. **Key References:**

[1] C. Happ and S. Greven (2018). *Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains*. Available at: .

[2] R. Zhong, S. Liu, H. Li, and J. Zhang (2023). *Sparse logistic functional principal component analysis for binary data*. Available at: .

Explainability and Interpretability in Graph Neural Networks for Complex Systems

Brian Daniel Bernhardt[*1], Mario Guarracino[1], Lucia Maddalena[2]

[1] University of Cassino and Southern Lazio,Italy [2] National Research Council, Italy

* Presenting author

Abstract

Graph Neural Networks (GNNs) have emerged as powerful tools for modeling complex relational systems, enabling learning tasks on non-Euclidean data such as social, biological, and economic networks. Despite their strong predictive capabilities, many GNN architectures still suffer from limited interpretability, making their adoption challenging in domains where transparency and explainability are essential. This work explores recent approaches aimed at improving explainability and interpretability in GNNs while maintaining predictive performance. In particular, we present two complementary directions. First, we introduce a modified GraphSAGE aggregation framework based on entropy and feature similarity, designed to mitigate over-smoothing and improve robustness in low-homophily settings. Second, we discuss an extension of Graph Neural Additive Networks, where probabilistic transition structures are used to obtain more interpretable node representations and feature-topology interactions. The proposed approaches are evaluated on synthetic graph classification settings under different structural conditions, including balanced and imbalanced class distributions and varying homophily regimes. Results show improvements in predictive performance and provide interpretable insights into how neighborhood structure and feature contributions influence model behavior.

Key References:

[1] M. Bechler-Speicher, A. Globerson, and R. Gilad-Bachrach (2024). *The Intelligible and Effective Graph Neural Additive Network*. Advances in Neural Information Processing Systems, 37, 90552–90578. Available at: NeurIPS Proceedings.

[2] C. Marciano, B. D. Bernhardt, L. Maddalena, and M. R. Guarracino (2025). *EGNAN: An Enhanced Method for Interpretable Graph Neural Networks*. In E. di Bella, V. Gioia, C. Lagazio, and S. Zaccarin (Eds.), *Statistics for Innovation III*, pp. 457–462. Cham: Springer Nature Switzerland. Available at: doi:10.1007/978-3-031-95995-0_76.

[3] B. D. Bernhardt, C. Marciano, and M. R. Guarracino (2026). *GraphSAGEnES: A Graph Neural Network Designed for Graphs with Low Feature Discriminability*. In G. Nicosia, V. Ojha, S. Giesselbach, M. P. Pardalos, R. Umeton, E. La Malfa, and G. La Malfa (Eds.), *Machine Learning, Optimization, and Data Science*, pp. 279–289. Cham: Springer Nature Switzerland. Available at: doi:10.1007/978-3-032-21480-5_19.

Speaker Index