



PhD opportunities in Statistics at St Andrews, 2024-2025

(Updated 2nd February 2024.)

Applications are welcomed for students wishing to undertake a PhD in Statistics at St Andrews. Fully funded scholarship places (fees, plus stipend of approx. £19,162) are typically available for well-qualified students. UK, EU and other overseas students are all encouraged to apply. New PhD students would typically start in September 2024, but this is flexible.

Some general information about the Division of Statistics is given below, followed by a list of specific topics that are on offer this year. Finally, more information is given about how to apply.

Statistics at St Andrews

Statistics is a lively area of research at St Andrews. The [Division of Statistics](#) is one of three divisions within the [School of Mathematics and Statistics](#), and currently consists of 15 members of academic staff, 14 research staff and 20 PhD students (plus PhD students co-supervised in other Schools). Our research is consistently rated highly in research assessments – for example 96% of our research outputs submitted to the 2021 [UK Research Excellence Framework](#) (REF) were assessed as being either world-leading or internationally-excellent. Our research environment was rated similarly well in REF2021, while our impact (i.e., the real-world effect of our research on wider society) rated particularly highly with 75% being judged to be world leading.

One major research strength is in the area of statistical ecology: contained within the School is the world-leading [Centre for Research into Ecological and Environmental Modelling](#) (CREEM), which is housed in tailor-made facilities at the St Andrews Observatory on the edge of the town. We are a founding member of the [National Centre for Statistical Ecology](#), a multi-institution consortium that ensures regular intellectual exchange between researchers worldwide with similar interests. Several members of CREEM are also part of the university's multi-school [Centre for Biological Diversity](#).

A second more recent and rapidly developing research focus is the [Statistical Medicine and Molecular Biology](#) group, led by a joint professorial appointment (Prof. Andy Lynch) between the Schools of Mathematics and Statistics and the School of Medicine, four recent appointments (Drs. Giorgos Minas, Nicolò Margaritella, Elham Mirfarah and Benjamin Baer) in this broad area, and research interests from several other staff members.

Many staff members are also active more generally in the field of [Machine Learning and Statistics Methodology](#). Research areas include deep learning (with applications in image, audio and genetic processing), Bayesian statistical inference, bioinformatics, design of experiments, estimation of population size, computer-intensive model fitting techniques, smoothing methods, causal inference, statistical genetics and analysis of clustered and censored data.

A brief summary of the research interest of each member of staff is given at the bottom of this section; more details can be found by following links to [staff members' web pages on the Statistics Division web site](#).

New PhD students join a high calibre but friendly research environment. Training is provided in the first year in the as part of St Andrews' participation in the [Scottish Mathematical Sciences Training Centre](#) and [Academy for Postgraduate Training in Statistics](#), the latter consisting of four one-week residential courses. Students may get the opportunity to become involved in externally-funded research as part of [CREEM's consultancy group](#); they may also be able to assist on statistics training workshops delivered to professional scientists both in the UK and abroad. Some PhDs are supervised jointly with scientists from other institutions, and there may be opportunities for study at those places. PhD studies are expected to last approximately 3.5 years.

St Andrews is a small, vibrant university town. It is situated on the east coast of Scotland and framed by countryside, beaches and cliffs. The town has a rich cultural heritage, having once been at the centre of Scotland's political and religious life. Today it is known around the world as the Home of Golf and a bustling student town with a distinctively cosmopolitan feel, where students and university staff account for more than 30% of the local population. The University is the oldest in Scotland and third oldest in the English-speaking world. It is the top-rated university in Scotland for teaching quality and student satisfaction, and among the top rated in the UK for overall research; it regularly comes in the top few places in UK league tables compiled, for example, by broadsheet newspapers (e.g., [1st place 2024 The Guardian and The Times](#)) and specialist bodies ([1st place 2023 for student positivity about their education](#); [4th place 2024 Complete University Guide – 1st for student satisfaction](#)). Its international reputation for delivering high quality teaching and research and student satisfaction make it one of the most sought-after destinations for prospective students from the UK, Europe and overseas.

More general information about postgraduate student life at St Andrews is given at the [University postgraduate study web page](#) and in the [postgraduate prospectus](#). School-specific information about applying is given at the [School's postgraduate research page](#).

The Division, School and University values diversity and nurtures an inclusive community where everyone is treated with dignity and respect regardless of individual characteristics such as age, gender, disability, religion or ethnicity. We are committed to equality for all. More details are given on the [School's postgraduate research page](#).

Brief summary of academic staff interests in the Division of Statistics

- Ben Baer – causal inference, survival analysis, exponential family matrix modelling
- Chrissy Fell – deep learning for image interpretation
- Alison Johnston – monitoring biodiversity, citizen science data, species distribution models, data integration
- Andy Lynch – design or analysis of molecular biology experiments, especially applications of DNA/RNA sequencing to cancer research
- Nicolò Margaritella – Bayesian inference, functional data analysis and large-scale inference with application to neuroscience and other applied fields

- Giorgos Minas – stochastic processes and multivariate statistics in molecular biology and medicine
- Elham Mirfarah – multivariate analysis, survival analysis, finite mixture modelling, censored data
- Michail Papathomas – Bayesian methods with application to genetics and biostatistics
- Chris Sutherland – statistical ecology: spatial capture-recapture, spatial occupancy models, multi-species occupancy modelling, optimal survey design
- Ben Swallow - Bayesian statistical inference; stochastic systems of biological processes; continuous zero-inflated data; applications in ecology
- Len Thomas – wildlife (particularly acoustic) surveys, population dynamics modelling
- Hannah Worthington – hidden Markov models for statistical ecology, spatial capture-recapture, movement and behaviour modelling, machine learning and human-in-the-loop identification

Academic staff not taking PhD students in the coming academic year:

- Rosemary Bailey – design of experiments in agriculture, horticulture, ecology and medicine
- David Borchers – spatial capture-recapture, camera trap surveys, terrestrial acoustic surveys
- Monique Mackenzie – random effects models, smoothing methods

Specific projects offered for 2023-24

We are currently looking for candidates for the following projects. In addition, prospective candidates with general interests related to those of staff members (see above) are welcome to contact them to discuss other possible projects.

Efficient estimation in causal inference and survival studies

Supervisor: Benjamin Baer

Causal inference aims to address the question of “what if?”, while survival analysis aims to study event times. In both areas, the observed data is incomplete in formally similar ways. This project will develop asymptotically efficient estimators in statistical problems at the intersection of the two areas, borrowing ideas from each. Causal survival analysis is an exciting and rapidly growing area, so applicants can integrate into ongoing projects or readily develop their own direction.

Distance sampling with milder assumptions

Supervisors: Benjamin Baer and David Borchers

Estimating animal density or abundance is a fundamental challenge in statistical ecology. Distance sampling is a widely employed method for addressing this issue. Recent work has shown that the assumptions underlying distance sampling may be relaxed by leveraging the relative time at which

an animal is detected during a survey. There are several avenues for further methodological development, depending on the interest of the applicant. For instance, parametric assumptions could be relaxed; the methods could also be generalized to allow a surveyor's path to be nonlinear or allow animal movement. Each of these possibilities is motivated by real-world data in need of comprehensive analysis.

Statistical 'omics on graphs

Supervisors: Collin Bleak and Andy Lynch

The field of statistical genomics (and related omics) is rapidly accepting that graphical models have a role to play in addressing the complexity of the molecular systems being investigated.

In St Andrews we have expertise both in graph theory (Bleak) and analysis of cancer genomics (Lynch). In particular we have novel molecular cancer data sets and novel graphical objects that both have the potential to offer new insights into cancer biology. This project will look to take advantage of both.

Quantifying Trade-Offs Between Simple and Complex Models for Decision-Making

Supervisors: Fergus Chadwick and Alison Johnston

Policy-makers in conservation and health frequently rely on relatively simplistic statistical models to inform their decisions. There are good reasons for this: simple models can be easier to explain, they can be easier to automate and they often have lower computational overheads. However, there is relatively little research weighing these benefits against the potential disadvantages of oversimplifying complex systems (e.g. generating biased results, not propagating uncertainty appropriately, over-estimating uncertainty). The aim of this project will be to quantify these trade-offs in a rigorous way for one or two chosen application areas.

The precise structure of the project will be agreed with the successful candidate. However, we expect all projects will involve simulation studies, determining policy-relevant performance metrics, implementing structured cross-validation techniques on simulated and real data, and extensive communication with applied partners. Experience with computational and multivariate statistics, Bayesian hierarchical modelling, or knowledge of an appropriate area of application is desirable. Willingness and capacity to learn and develop these skills is essential.

Identifying complex spatio-temporal biomarkers of brain diseases

Supervisor: Nicolò Margaritella and Michail Papathomas

Bayesian models today are providing the tools to explore the complexity of brain architecture. Therefore, there is a crucial need for leading researchers with an in-depth comprehension of the current challenges in neuroscience and the quantitative skills to develop cutting-edge solutions.

The aim of this project is the development of a modelling framework for the identification of new, complex spatio-temporal brain patterns which can improve our understanding of the functional activity of the brain, our ability to identify early signs of brain diseases and the prediction of their prognosis. In addition, further timely neuroscientific challenges such as the identification of inter-individual variations in brain responses and the inclusion of multiple covariates (e.g. laboratory and clinical) in the identification of complex biomarkers of brain diseases will be researched during the project.

The methodologies developed in the project will provide neuroscientists with innovative analytical tools that will contribute to neuroscientific research on a wide range of brain conditions, from developmental to neurodegenerative diseases, which affect millions of people in the UK and worldwide.

The student will acquire advanced modelling skills in the research areas of Bayesian nonparametrics and functional data analysis which will be essential to develop the innovative modelling framework. New methods will be tested on well-known publicly available neuroscientific datasets and results presented at international conferences in both statistics and neuroscience. The student will be also involved in the development of R packages that will allow immediate access to all methods developed in this project to the wider scientific community.

Stochastic modelling and inference for live-cell gene expression time-series data to unravel the mechanisms of stem cell differentiation

Supervisors: Giorgos Minas and Jochen Kursawe, in collaboration with Cerys Manning (University of Manchester)

This project will develop statistical methodology for noisy time-series data and stochastic computational models to analyse live-cell imaging data provided by the lab of our collaborator Dr Cerys Manning at the University of Manchester. Live-cell imaging is a powerful technique for real-time observation of the activity of genes in single cells. These observations are important in understanding many cellular processes which strongly depend on dynamic gene activity. One of these is the process by which stem cells generate mature cell types (stem cell differentiation). This is a critical biological process not only for embryonic development, but also regeneration, and modern stem cell-based regenerative therapy approaches. Dr Cerys Manning has previously shown that oscillations in gene activity are observed in stem cells of the central nervous system, and these are important for regulating the differentiation process. We now wish to unravel the mechanisms driving these oscillations. We also wish to examine the role of stochasticity in stem cell differentiation and its interplay with oscillations. For this purpose, we will use clustering methods to identify groups of cells that exhibit similar patterns of gene expression. We will also fit stochastic models described by Stochastic Differential Equations to the time-series data and use Bayesian statistics to estimate model parameters, quantify model uncertainty, perform model comparisons, and derive predictions.

The ideal candidate for this project will be interested in Bayesian statistics, stochastic processes, and stem cell differentiation. Background in at least one of the above subjects will be beneficial, but candidates with other backgrounds will be considered.

Stochastic simulation, analysis, and inference of non-linear dynamical systems

Supervisor: Giorgos Minas

This project will develop a novel framework for studying the dynamics of systems presenting oscillations and multi-stabilities. This type of dynamics is abundant in many fields and especially in molecular biology, epidemiology, ecology, sociology. For instance, the developed methods will apply to gene expression oscillations for biological time-keeping and cell-to-cell communication, multi-stabilities in cell development, epidemic oscillations driven by public awareness, ecological oscillations driven by species competitions, and many other settings. To build this framework, we will use powerful results from the theory of dynamical systems to decompose large, non-linear dynamical systems. This decomposition will allow us to break down systems into simpler components. For the non-linear components presenting oscillations or multi-stabilities, we will now be able to study their dynamics in detail and develop methods for controlling their non-linear variation, while for the linear components use standard models described by stochastic differential equations. The combination of the two components will provide fast, and long-time accurate models for a much wider range of problems than ever before. We will then be able to use these models for fast, long-time accurate simulation and parameter estimation algorithms.

The ideal candidate for this project will be interested in stochastic processes and dynamical systems. Background in stochastic processes (e.g. Markov processes, stochastic differential equations) or non-linear dynamical systems will be beneficial, but candidates with strong background in other mathematical subjects will be considered.

Supervised learning methods to measure information transfer in biology

Supervisor: Dr Giorgos Minas

Information theory is widely used as the basis of communication channels to transfer information through the Internet and other platforms. The study of information transfer is also hugely important in many other fields (e.g. marketing, epidemics control, molecular and cell signalling). For instance, molecular biology is all about how biological cells respond to information coming from their environment to translate genetic code to functional macromolecules that in turn transfer information to other molecules through their interactions. This project has two main objectives: (a) to fill a gap in how this amazing theory of information flow originally derived for communication channels applies to other fields and especially but not exclusively molecular biology, (b) to study the use of supervised learning methods in estimating information theoretic quantities, and particularly mutual information. Supervised learning (e.g. classification) is naturally embedded into information theory, and the emergence of machine learning provides new tools that seem to be powerful but are also poorly understood, at least in this setting. Therefore, the project will explore the performance

of supervised learning methods in estimating information theoretic quantities, but also in terms of prediction accuracy for a wide range of simulated data.

The ideal candidate for this project will be interested in information theory (see textbook below), and supervised learning. Background in one of those fields will be beneficial, but candidates with strong background in other mathematical or computational subjects will be considered.

Reference: Elements of Information Theory, Thomas M. Cover, and Joy A. Thomas, Wiley Series in Telecommunications

Incorporating Mixture of Expert Models for Longitudinal Data with Missing and Censoring

Supervisor: Elham Mirfarah

Longitudinal data analysis is a powerful tool for studying changes in subjects over time. However, the presence of missing data and censoring poses significant challenges. Traditional methods may not fully utilize the information available in such data. This research project aims to address these challenges by incorporating a mixture of expert models, which can account for complex patterns such as heterogeneity in longitudinal data, missingness, and censoring using Bayesian methodology.

We will develop and adapt a mixture of expert models to accommodate longitudinal data when data are suffering from a kind of limitation. This may involve the creation of Bayesian hierarchical models that can capture both within-expert and between-expert variations. The result of the developed models can be implemented in real-world datasets from diverse fields to demonstrate the practical utility of the methodologies such a cancer data.

Propagation of uncertainty for signatures of mutational processes

Supervisors: Michail Papathomas and Andy Lynch

There is a trend, especially in cancer research, to i) take a set of DNA mutations ii) cross-categorize them by patient and mutational characteristic and iii) decompose the resulting counts matrix into two sets of vectors – one set representing the mutational impact of specific mutagens and one set representing the exposure of individuals to those mutagens.

We have previously worked [1] on the question of uncertainty in that decomposition, and in the resulting estimates of exposure, but the uncertainty that goes into building the count matrix in the first instance remains neglected.

In this project the student will examine the uncertainties and biases that feed into the construction of the original data set, and develop a basis for predicting their effects on both the final decomposition, but also the believed uncertainty about that decomposition.

The project is supervised by Michail Papathomas who has extensive experience in the evaluation of uncertainty and the analysis of genetic data, and Andy Lynch who has extensive experience in methods for and analysis of cancer sequencing data.

Reference:

[1] Velasco-Pardo, V., Papathomas, M., Lynch, A.G. (2022). Statistical Challenges in Mutational Signature Analyses of Cancer Sequencing Data. In: Bispo, R., Henriques-Rodrigues, L., Alpizar-Jara, R., de Carvalho, M. (eds) Recent Developments in Statistics and Data Science. SPE 2021. Springer Proceedings in Mathematics & Statistics, vol 398. Springer, Cham.

Understanding the uncertainty in the decomposition of cancer gene or protein expression data

Supervisors: Michail Papathomas and Andy Lynch

Many tools exist that will take the expression profile of a tumour sample and decompose that signal into components arising from different tissue types (e.g. tumour cells, benign cells, immune cells, fibroblasts, stromal cells etc.).

Few of these tools consider formally the uncertainty inherent in the problem, and those that provide a measure of uncertainty about the final reported decomposition tend to do so in a manner that takes into account only some sources of variability.

In this project the student will consider a general approach to the question and look to develop methods to quantify the uncertainty inherent in the solutions provided. Methods will be motivated by problems identified in the literature, as well as the analysis of primary data arising from consortia in which the supervisors of the project are involved.

The project is supervised by Michail Papathomas who has extensive experience in the evaluation of uncertainty and the analysis of genetic data, and Andy Lynch who has extensive experience in methods for and analysis of cancer sequencing data.

Fully funded NERC DTP Studentship: Developing efficient and accessible “data-to-decision” pipelines for remotely sensed biodiversity data.

Supervisors: Chris Sutherland; external supervisors Xavier Lambin, Carl Chalmers, Kenny Kortland

Note: Closing deadline 9th February 2024 (open to students worldwide)

Ecological monitoring involves multiple tasks across several disciplines. Practically every step of what we term the ‘data-to-decision’ pipeline has the potential to benefit from recent technological advances. Remote sensors collect vast quantities of data; novel and sophisticated AI can convert ‘big data’ into usable ecological data; statistical ecology involves developing methods that keep pace with novel data streams; and evidence is critical to support decision-making to prevent global biodiversity loss. Despite this nascent potential, existing workflows suffer from inefficiencies occurring at disciplinary borders that have limited uptake by the intended beneficiaries: the global community of scientists and practitioners working across taxa and systems to quantify ecological responses to environmental and anthropogenic change. There is an urgent need therefore to capitalise on recent developments and develop more efficient and accessible ‘data-to-decision’

pipelines that target multidisciplinary innovations at disciplinary transitions, that encourage dialogue between domain-specific experts, and that are accessible and lead to accelerated uptake by the user community.

We identify three barriers of existing pipelines/workflows that limit the translation of technological benefits into improved biodiversity science, and these are the focus of this project.

Goal 1: Removing data processing bottlenecks – Working at the interface of ecology and computer science, we will develop improved interfacing with the Conservation AI platform to make the process of data storage and image transfer that will lead directly to increased accessibility to, and uptake of, transformative machine learning technologies. In doing so, we will develop much needed data collection, storage, and reporting, standards for sensor data that will contribute to improved and transparent use of technology in biodiversity science.

Goal 2: Performance of AI for ecological inferences – Using a combination of simulation and camera and audio monitoring data collected in the Cairngorms Connect landscape, we will first quantify how variation in classification performance relates to how, where, and when data are collected, and second, quantify how sensitive ecological inferences about single and multi-species occupancy rates and biodiversity metrics are to varying levels of error. Developing a better understanding of how aspects of study design influences classification performance, and if and when biases exist when making inferences from sensor data should alleviate the reluctance to embrace technological advances.

Goal 3: Develop Statistical methods for classification data – Acknowledging that existing occupancy models were not developed with machine learning classification data in mind, we will continue the development of ‘continuous score’ occupancy models that are designed specifically for non-binary and uncertain classification output and that explicitly account for both false negative (not detecting the species on a sensor) and false positive (uncertainty in species identity). Specifically, we will extend these early ML-occupancy models to include ‘coupled classification’ where the processes of classification and statistical inference are integrated leading to appropriate treatment of important sources of uncertainty and increased precision in both processes.

Specific application instructions for this project: The application should include:

- Statement of Purpose (max 800 words)
- CV
- Transcripts (undergraduate and/or master’s degrees)
- Names and contact details of two referees
- Evidence of English language capability: English language test scores as per Profile 6-D, unless exempt, see English language requirements.

In the statement of purpose (max 800 words), applicants must: 1) demonstrate their interests in specific aspects of the project, including interdisciplinary interests; and 2) outline their skills and experience that make them suited to undertake the research outlined. Note that it is not our expectation that candidates already have an ecology, statistics and AI background, instead we are looking for candidates who are interested in working at, and developing skills in, the interface of

these disciplines. Shortlisted applicants will be interviewed in late February or early March. For informal enquiries, please contact Dr Chris Sutherland (css6@st-andrews.ac.uk).

Causal inference and trial emulation for ecological observational data

Supervisor: Ben Swallow and Hannah Worthington

Conducting formal causal inference for ecological data is challenging due to the complex observational processes that are usually involved. The propensity score method allows the estimation of causal effects in non-experimental studies, however is dependent on constructing emulated experiments to ensure independence between observation process and treatment allocation. This PhD will develop and apply novel methodology in causal inference for ecological observational studies, for example changes in spatio-temporal distribution and movement patterns driven by environmental disturbance and impact assessment studies.

Uncovering the rules of dispersal and the consequences for the distribution, dynamics, and persistence of pathogens and their hosts.

Supervisors: Chris Sutherland and Oscar Gaggiotti (Biology)

*EASTBIO Doctoral Training Centre-funded PhD – **closing date Monday November 27th 2023** – see <http://www.eastscotbiotdp.ac.uk/how-apply-0> and <https://www.findaphd.com/phds/project/eastbio-uncovering-the-rules-of-dispersal-and-the-consequences-for-the-distribution-dynamics-and-persistence-of-pathogens-and-their-hosts/?p162395>*

Note that complete applications comprise several parts, including 2 references, so please allow sufficient time.

Managing disease risk requires mechanistic understanding of the dynamics and persistence of pathogens and their hosts. Host population dynamics can influence pathogen eco-evolutionary dynamics, but spatial (connectivity, barriers to dispersal) and demographic (colonisation and extinction) processes likely dictate host-pathogen interactions, and hence their emergent dynamics and persistence with far-reaching consequences for livestock or human health.

This project seeks to advance understanding of the links between the distribution, dynamics, and diversity of pathogens, and host dispersal using data from a long-term landscape-scale study of a multi-host model metacommunity. Bartonella infections in rodents are ideal for examining these issues.

Bartonella species, including several associated with human disease, have high genetic diversity and prevalence, and exhibit a wide range of host specificity. In our metacommunity system (Assynt, Northwest Scotland), we study bartonella and two bartonella hosts: the water vole, the primary host and a habitat specialist restricted to <1% of suitable riparian habitat on the landscape and hence represent a true metapopulation; and the field vole, a secondary host and a predominantly

grassland species (~33% of the landscape). In our system, water voles occur only in the riparian flushes while field voles are found in the riparian and non-riparian grassland, meaning that both the distribution of hosts and their interactions are spatially structured.

Bartonella persists in this spatially structured multi-host population representing an ideal setting for examining how cross-scale multi-host dynamics shapes pathogen distribution, dynamics, and persistence. We have assembled a unique 25-year dataset consisting of individual capture-recapture data of water voles ($n > 5000$) and field voles ($n > 1500$); host-specific *Bartonella* infection information; and sequencing-based host-specificity of different *Bartonella* genotypes. Integrating data and knowledge from genomics, epidemiology, metacommunity theory, and statistical ecology, the student will develop and apply state-of-the-art molecular and statistical methods to advance ecological and epidemiological theory. The specific aims are to:

1. Infer disease-transmission corridors by reconstructing dispersal routes using host-specific pedigree analysis.
2. Quantify spatiotemporal patterns of *Bartonella* diversity and persistence using whole-genome-sequencing.
3. Integrate objectives 1&2 to develop spatially-explicit multihost-pathogen metacommunity models to predict emergent transmission and infection landscapes.

The project would suit students interested in multidisciplinary science at the interface of spatial population ecology, disease spread and dynamics, and statistical ecology. Purposefully broadly defined, the student will have control over the direction taken to achieve the objectives. Collaboration between domain experts will provide interdisciplinary training including next generation sequencing and advanced statistical models.

Fully funded PhD studentship: Estimating spatial and temporal patterns in density of harbour porpoises in the North Sea using datasets from passive acoustic ocean observation networks with data flow to EU Digital Twin Ocean database

Closing date: 4 March 2024

Supervisors: Len Thomas; Peter Tyack and Jamie Macaulay (School of Biology)

We welcome applications for a fully-funded inter-disciplinary PhD in statistics and marine mammal bioacoustics at the University of St Andrews. This PhD project will use existing underwater acoustic datasets to monitor the distribution and abundance of harbour porpoises throughout the North Sea. The PhD project will run alongside a recently started multi-year, international project, called DTO Bioflow (<https://dto-bioflow.eu>). The project will be tailored to suit the student's specific skills and interests, building on work developed by DTO partners to: 1) process available acoustic and visual sighting datasets to detect porpoise; 2) implement animal density estimation methods on processed datasets; 3) compare results using visual vs acoustic detections with an effort to integrate data from both sources; and (4) use spatial and temporal modelling methods to investigate how patterns in porpoise densities vary with biological and physical oceanographic variables.

Project supervisors are Professor Len Thomas (CREEM), Professor Peter Tyack (Sea Mammal Research Unit, SMRU) and Dr. Jamie Macaulay (SMRU). The student will matriculate in the School of Mathematics and Statistics.

The ideal applicant will have excellent academic credentials, strong university-level training in statistics and signal processing, a demonstrated interest in statistical ecology and marine mammal bioacoustics and evidence of contributions to the wider research community. Applications are particularly welcome from candidates belonging to a group that has been historically under-represented in statistics and biology programmes at St Andrews. For more details of the position, our selection criteria and how to apply, please see

https://stats.wp.st-andrews.ac.uk/files/2024/02/DTO_bioflow_studentship.pdf

Movement through space and time, realistic movement for species abundance methods

Supervisors: Hannah Worthington and David Borchers

Evidence-based conservation and ecology are reliant on wildlife surveys. As a result, there exists a range of methods that have been developed specifically with the aim of estimating animal abundance and animal distribution. Traditionally data have been collected by humans and are often modelled as a 'snapshot' of the system at a particular moment in time. However, we often have more information available to us such as a precise time of observation. This is particularly true when the data are instead collected using digital devices such as camera traps or acoustic arrays that are generally recording and collecting data continuously. The combination of spatial and temporal information lets us consider the potential for incorporating realistic animal movement by considering the spatio-temporal clustering of observations. This PhD would explore some of these ideas such as: movement models for spatial capture-recapture data; temporal clustering to assist identification in spatial count models; self-exciting processes to model observation hotspots; or integrating movement and the hazard function in distance sampling and other observation processes. This PhD is likely to appeal to candidates with a keen interest in statistical simulation and computation, an interest in stochastic differential equations would likely also be beneficial.

Exploring synergies between statistical ecology and statistical genomics.

Supervisors: Hannah Worthington and Andy Lynch

While superficially different, these two areas of research share several questions in common (How many species? How many of each species? How are the species distributed spatially? How should we sample?) that differ fundamentally only in whether the species in question are flora and fauna or nucleic acids and proteins.

In addition, direct application of genomics within an ecology setting is now becoming more common, with environmental DNA a hot topic for research.

In this project the student will examine the scope for application of mature methods from one of the two disciplines within the other (e.g. formalizing the use of capture-recapture methods in genomics, or taking models from cancer analyses and applying them to eDNA questions). There will be

particular interest in questions, such as population size, where traditional methods in statistical ecology and methods using molecular data can be used to answer a question, and the project may consider how the approaches should be combined. The student may choose to focus on the theoretical aspects of the questions or be driven by a particular application.

Hidden Markov models for spatially structured populations

Supervisors: Hannah Worthington and Chris Sutherland

Hidden Markov models (HMMs) offer a very powerful, flexible, and efficient structure for likelihood computation. They're a popular tool for problems in statistical ecology since the structure often has an associated, and insightful, ecological interpretation for the system being modelled. This PhD will look to develop a general and unified framework for a collection of models that have yet to be expressed as an HMM formulation. One such group of methods are dynamic occupancy models that can be applied to spatially structured populations, surveys with spatially structured sampling, or the combination of both. In particular, this project will focus on approaches that can be applied to metapopulations with well-defined patches that experience colonisation-extinction dynamics, continuously distributed populations where data are collected through spatially structured sampling, and multi-state/hierarchical extensions that allow for inference on the spread and prevalence of disease and pathogen transfer within spatially structured populations. The methods will be explored through a combination of simulation and application to real-world case studies.

Application procedure

Although there is no fixed deadline (unless noted otherwise for a particular topic), you are strongly encouraged to make your application as early as possible! The first round of funding decisions will be made in mid-January 2024.

Many details of the general requirements and admissions procedure are given at the [University postgraduate research application web page](#).

Applicants should have a good first degree in mathematics, statistics or another discipline (e.g., biology, computer science), with substantial statistical component. A masters' level degree (MSc, etc.) is an advantage, as is any other relevant professional experience. A major criterion for selection is academic excellence: most successful applicants (particularly those who are awarded scholarships) have a good to very good 1st class undergraduate degree and/or a distinction at MSc level. Those who do not have English as a first language, and who have not undertaken an undergraduate or graduate degree taught in English, should provide evidence of English proficiency (minimum IELTS 6.5 or equivalent).

A full list of the criteria we look at when assessing candidates is as follows:

- Academic merit (degree type and classification)
- Research potential (e.g. previous research experience or employment, published papers)
- Alignment of research interests with PhD topic applied for
- Personal and professional development (e.g. non-research work experience)
- Outreach (public communication of mathematics and statistics)

Applicants should explain how they meet these criteria in their application materials and personal statement. The personal statement may also address other issues such as why the applicant wants to study in St Andrews, in a particular research area, or with a specific supervisor. We will consider the accomplishments of prospective students in the context of their background. We also consider factors internal to the School, such as whether the proposed supervisor is a new supervisor (weighted positively) and whether the research group applied to has relatively few PhD students.

Please note that applications for PhD places and for funding are usually considered separately, both using the above criteria. Offers of PhD places are not always accompanied by offers of funding.

Potential applicants are encouraged to contact the Postgraduate Officer responsible for PhDs in Statistics, in advance of making a formal application. He is: Len Thomas, email len.thomas@st-andrews.ac.uk, tel. 01334 461801.

To make a formal application, complete the appropriate online form at <https://www.st-andrews.ac.uk/study/pg/apply/research/> (click on "Apply Now" on that page). You also need to provide the following supporting documentation: CV, evidence of qualifications and evidence of English language (if applicable); you should also provide a personal statement. You don't need to provide a research proposal unless you are proposing your own project, or sample of academic written work. You will need to ask two referees to provide academic references for you – once you fill in their name on the form, they will be sent emails asking them to upload their references. Please note that we give serious consideration to both the stature of your referees and the remarks

that they make about you. More details about the application procedure are given at <https://www.st-andrews.ac.uk/study/pg/apply/research/>

Further School-specific information is on this page <https://www.st-andrews.ac.uk/mathematics-statistics/prospective/pgr/> and links from that page.

In addition to the scholarships mentioned on those pages:

- The Centre of Research into Ecological and Environmental Modelling has a small scholarship fund; all students applying for School funding with an intended PhD topic in the field of statistical ecology are automatically considered.
- An up-to-date list of external scholarships is given at <https://www.st-andrews.ac.uk/study/fees-and-funding/postgraduate/scholarships/research-scholarships/>.

We look forward to hearing from you!