



University of
St Andrews

600
YEARS

PhD opportunities in Statistics at St Andrews, 2026-2027

(Updated 17th October 2025.)

Applications are welcomed for students wishing to undertake a PhD in Statistics at St Andrews. Fully funded scholarship places (fees, plus stipend of approx. £19,775) 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 2026, 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 16s members of academic staff, 14 research staff and 27 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 area is the [Statistical Medicine and Molecular Biology](#) group. With more than ten affiliated academic staff and multiple links to the School of Medicine, they work on methods and applications in areas such as Genetics, Cancer Genomics, Infectious Diseases and Neuroscience as well as basic Biology (e.g. decision-making in human cells). Methods for clinical trial design, survival analysis, causal inference, and population size estimation are also researched.

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, computational biology, 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 yet friendly research environment. They are encouraged to participate in the division's weekly seminars and research group meetings. The PhD students in our division frequently arrange social and academic events to create a supportive environment for all students. Training is provided in the first year as part of St Andrews' participation in the [Scottish Mathematical Sciences Training Centre](#) and the [Academy for Postgraduate Training in Statistics](#). Students may get the opportunity to become involved in externally-funded research, e.g. as part of CREEM's consultancy group; they may also be able to assist in statistics training workshops delivered to professional scientists both in the UK and abroad and have the opportunity to be trained for and provide teaching assistance to our taught modules. Some PhD projects 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 in The Times and Sunday Times University Guide 2025](#)) and specialist bodies (e.g. [1st place for student positivity about their education](#)). 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 value diversity and nurture 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 (alphabetical order) interests in the Division of Statistics

- Ben Baer – causal inference, survival analysis, exponential family matrix modelling
- Regina Bispo - environmental statistics, multivariate statistics, spatial statistics
- Fergus Chadwick – modelling complex observation processes in ecology.
- 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 – Bayesian inference for dynamical systems, information theory, supervised learning using time-series and/or high-dimensional biological data
- Michail Papathomas – Bayesian methods with application to genetics and biostatistics
- Ben Swallow - Bayesian statistical inference, stochastic systems of biological processes, spatio-temporal models, applications in ecology and epidemiology
- 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:

- Chris Sutherland – statistical ecology: spatial capture-recapture, spatial occupancy models, multi-species occupancy modelling, optimal survey design
- David Borchers – spatial capture-recapture, camera trap surveys, terrestrial acoustic surveys
- Rui Borges - phylogenetics, population genetics, Bayesian inference, bioinformatics
- Monique Mackenzie – random effects models, smoothing methods

Specific projects offered for 2026-27 (in alphabetical order)

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.

Improved survival analyses in molecular cancer studies

Supervisors: Ben Baer and Andy Lynch

Improved survival (whether this is literal survival, time to a change in symptoms, or time to disease progression) is a key outcome in cancer research. Most applications of survival analysis in the area use standard approaches, but the manner in which events are defined may bias these. Estimating and addressing such biases may be one strand of this project.

The context of these analyses is often one of using large (incorporating thousands to millions of molecular measurements, or different types, per patient) and potentially messy data sets. Other strands of this project might include the development of (asymptotically) efficient methods for survival analysis that make use of machine learning methods to exploit the amount of data present.

Alternatively, there may be investigations into survival-specific dimension reduction methods that can make such data sets more easily used and understood, perhaps incorporating prior information about the relationships between different levels of molecular data (e.g. DNA, RNA, protein) or prior knowledge of gene interaction networks.

Inferring Events from Phylogenetic Analysis of Eyewitness Testimony in the Historical and Forensic Record

Supervisors: Ben Baer and Charles Paxton

This project will consider if we can use phylogenetic analysis to correct for imperfections in eyewitness testimony. Eyewitness testimony from events with multiple witnesses is often inconsistent and sometime contradictory. If we think of these inconsistencies as mutations from a common ancestor with perfect recall (but not necessarily perspective) on the event, can we obtain an objective estimate of what the collected witnesses saw? Here we will consider historical and forensic records of events with multiple witnesses, in order to construct a “family tree” of what was reported and estimate the ancestral condition. Can

statistics tell us what really happened when JFK was assassinated? Could statistics inform analysis of crimes where there are multiple witnesses? Both manual and machine learning approaches will be considered.

Topics arising from the UK Prostate ICGC (International Cancer Genome Consortium) project.

Supervisor: Andy Lynch

The UK Prostate ICGC project has generated many data on a cohort of men with the prostate cancer that have the following notable features

- Data from a range of technologies (DNA sequencing, methylation sequencing, RNA sequencing, proteomics mass-spec data)
- Exploration of the heterogeneity of prostate
- Data from matched benign and disease-free prostate samples
- High-quality linked clinical data

While these data are already being analysed, existing methods do not exploit all of these characteristics. This project will develop methods or analyses of the data to address fundamental questions on the biology and treatment of the disease.

Special attention could be given to the question of integrating the proteomic data into other analyses, or to meta-analysing these data with other large clinical or experimental data sets. Indeed, linking in to other data sets (including those of the Pan-prostate cancer group, or other cancer types) will be key to exploiting the resource.

Efficient design of high-throughput molecular studies.

Supervisor: Andy Lynch

A typical experiment might consist of taking one sample from each of 100 patients, measuring many molecular properties, and then performing an analysis. At the analysis stage, some samples may be dropped because of the sample quality (e.g. a 'tumour sample' may have contained very little tumour). Since the experiments remain relatively expensive, this is not ideal.

Given a cheap technology (e.g. Abujudeh et al) that could inform on these characteristics, other options become available to us, and this project will look at different ways of exploiting such a technology in the design of experiments. The same concept applies to the selection of representative samples from a heterogeneous patient

Alternatively, for some questions it may be appropriate to consider whether one could look at trading off the quality of data generated for an individual, in order to increase the chances of saying something about a group of patients, and this can be investigated also.

Some analysis of experimental data could form part of the project. As could extension of the methods of Abujudeh to non-human cases.

Abujudeh et al. (2022) Low-cost and clinically applicable copy number profiling using repeat DNA. BMC Genomics 23. 599

Obesity, Prostate Cancer and DNA methylation

Supervisors: Andy Lynch, Ben Baer

Prostate Cancer is the most common cancer in men in the US and UK. Obesity is a known risk factor for many cancers and levels of obesity are increasing. Despite the disease and potential risk factor being so common, the evidence for obesity as a risk factor for prostate cancer is inconsistent – as evidenced by a recent meta-analysis (despite the meta-analysis itself not being entirely satisfactory).

DNA methylation involves the addition of methyl groups to certain positions in the genome that contribute to the regulation of genetic activity. DNA methylation patterns can be copied from one cell to another, but differ between cell types as part of the ‘epigenetic’ code that allows e.g. a prostate cell to appear and behave differently from a skin cell behaviour despite both cells sharing the same genetic sequence. Methylation changes are known to be early and recurrent in prostate cancers (Massie et al 2017), and obesity is known also to alter methylation patterns.

We might therefore wonder whether consideration of methylation patterns is required to understand the relationship between obesity and prostate cancer. However, including high-dimensional molecular data as a mediating factor in causal analyses is difficult. Moreover, we have to acknowledge that other factors may be involved such as diabetes and genetics.

This project will use large public data sets (e.g. Genomics England, UK Biobank, Generation Scotland, ICGC, PPCG) to investigate the relationship between obesity and prostate cancer and the potential role of methylation within that.

The project may involve (but will not be restricted to)

- Improved meta-analysis of obesity risk for prostate
- Developing methods for incorporating high-dimensional data in a causal mediation analysis.
- Developing novel methods for dimension reduction in methylation data sets.
- Developing methods for meta-analysis of high-dimensional methylation data.

Tzenios, Tazanios, Chahine (2022) The impact of body mass index on prostate cancer: An updated systematic review and meta-analysis. *Medicine (Baltimore)* 101(45):e30191

Massie Mills Lynch (2017) The importance of DNA methylation in prostate cancer development. *The Journal of steroid biochemistry and molecular biology.* 166 1-15

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

Supervisor: Giorgos Minas

This project aims to create a new framework for studying the dynamics of systems that exhibit periodic behaviour or multiple equilibria. These types of dynamic behaviours are common in various fields such as molecular biology and epidemiology. Examples include the circadian clocks, oscillatory responses to stress signals, and the specialisation of stem cells, as well as epidemic oscillations driven by public awareness. To build this framework, the project will utilize the theory of dynamical systems, which allows for the decomposition of large, non-linear dynamical systems into simpler components of smaller dimensions. The project will also develop stochastic models that accurately describe stochastic dynamics, while being computationally fast for simulation, sensitivity analysis and Bayesian inference of model parameters using time-series data. The ideal candidate for this project will have strong interest in stochastic dynamical systems, molecular biology and/or epidemiology, and

will possess strong programming abilities. While a background in stochastic processes (e.g. Markov processes, stochastic differential equations) or non-linear dynamical systems will be beneficial, candidates with a strong background in other mathematical subjects will also be considered.

References

Frederick Truman-Williams and Giorgos Minas (2025), Simulating stochastic population dynamics: The Linear Noise Approximation can capture non-linear phenomena, <https://arxiv.org/abs/2504.15166>

Swallow, B., Rand, D. A., & Minas, G. (2024). Bayesian Inference for Stochastic Oscillatory Systems Using the Phase-Corrected Linear Noise Approximation. *Bayesian Analysis.*, 1(1). <https://doi.org/10.1214/24-BA1471>

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.

Supervised learning methods to measure information transfer in biology

Supervisor: 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 powerful 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

computational methods (e.g. supervised learning) in estimating information theoretic quantities, and particularly mutual information.

The ideal candidate for this project will be interested in information theory and machine learning methods, and will possess strong programming abilities. Background in one of those fields will be beneficial, but candidates with strong background in other mathematical or computational subjects will be considered.

Bayesian identifiability for log-linear models.

Supervisor: Michail Papathomas

Log-linear modelling is the standard approach for investigating the full joint dependence structure between categorical variables. Many applications exist. For instance, such as phenotypes and SNPs. Complex dependence structures can be easily discerned using graphical log-linear models (Papathomas and Richardson, 2016). This can potentially lead to identifying functionally important pathways. Another application is discerning the size of hidden populations, such as victims of modern slavery (Cruyff, M., Overstall, Papathomas, McCrea (2020)). The number of cells in the associated contingency table increases rapidly with the number of variables, creating sparse contingency tables with a number of zero cell counts, even for a large number of subjects. The presence of zero cell counts can potentially make some model parameters non-estimable, also referred to as non-identifiable (Sharifi Far, Papathomas, King, 2022). Non-identifiability is a major impediment to evaluating how risk factors interact, and understanding important biological or other mechanisms. Problems associated with identifiability are currently not sufficiently understood, and have not been addressed in a systematic manner. The aim of this project is to develop methods that will utilize information pertaining to the Bayesian identifiability of interaction parameters, towards choosing the best log-linear model given the data.

References:

Papathomas, M. and Richardson, S. (2016): Exploring dependence between categorical variables: benefits and limitations of using variable selection within Bayesian clustering in relation to log-linear modelling with interaction terms. *Journal of Statistical Planning and Inference*. 173, 47-63

Sharifi Far, S., Papathomas, M. & King, R. (2022). Parameter redundancy and the existence of maximum likelihood estimates in log-linear models. *Statistica Sinica*.

Cruyff, M., Overstall, A., Papathomas, M. & McCrea, R. (2020) Multiple system estimation of victims of human trafficking: model assessment and selection. *Crime and Delinquency*. Online First

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.

Inferring the fitness effects of new mutations across evolutionary scales

Supervisor: Rui Pinto Borges

The fate of new mutations has long intrigued evolutionary biologists, as understanding it is key to a fundamental question in evolutionary biology: how do species adapt in the face of ever-changing environments? It is generally accepted that mutations can be classified as deleterious, advantageous, or neutral, depending on whether they decrease, increase, or do not affect the fitness of the individual carrying them, respectively. This understanding is of fundamental importance and provides a basis for explaining why genome sequences exhibit varying levels of diversity, with some regions highly conserved and others less so.

However, there is ongoing debate about the proportion of emerging deleterious, advantageous, and neutral mutations. More recently, the debate resurfaced, with some researchers arguing that positive selection is more pervasive than traditionally assumed, challenging the prevailing idea that most mutations are neutral or nearly neutral. Others, however, contend that positive selection is being overestimated due to methodological limitations, and that genetic drift and weak selection remain the predominant forces shaping genetic variation.

Progress in this debate depends on our ability to characterise the distribution of fitness effects (DFE) of new mutations - that is, the probability distribution associated with each fitness effect. Advances in sequencing techniques have enabled the collection of vast amounts of genomic data, raising hopes that we may achieve a better characterisation of this distribution. However, significant challenges remain that are mainly related to the difficulty of modelling

and estimating the DFE: (i) confounding factors that mimic selection can bias its estimation; (ii) strongly advantageous and deleterious mutations are particularly difficult to observe and may be underappreciated; (iii) over longer timescales, the accumulation of recurrent mutations further complicates the modelling of fitness effects. These challenges make it difficult to determine the shape of the DFE, let alone how it evolves over time.

This research aims to develop new theory and methods to estimate the DFE and address these challenges. We will develop a model for fitness effects that integrates population and phylogenetic timescales of evolution. Then, we will model the DFE using extreme value theory and further employ Bayesian inference to infer it. We will apply our methods to microbial genomic data, including bacterial and viral case studies of epidemiological and medical importance. Comparing the DFE estimates across these systems will ultimately allow us to explore key questions: How conserved is the DFE among microbes? Do all species exhibit similar patterns, or is substantial variation observed between clades and over time? To what extent do closely related species share similar DFE shapes, and how much of this similarity is driven by shared evolutionary history?

Improving estimates of uncertainty in wildlife population assessments

Supervisors: Len Thomas and Laura Marshall

Reliable estimates of wildlife population size are fundamental to effective management and conservation. One important component of reliability is the precision (or, conversely, variance) of the estimate. Systematic survey designs (where survey lines or points are laid out in a regular pattern over the study area) tend to produce estimates with lower variance than alternatives – but it can be hard to accurately estimate this variance with currently-implemented estimators tending to over-estimate it. The goal of this PhD will be to investigate and improve variance estimation for systematic survey designs. A particular focus will be on distance sampling surveys, which are widely used to monitor populations from tiny geckos to very large whales. We will test new methods using both simulation studies and real-world datasets. Within this broad topic there is considerable scope for the student to develop their own interests.

The impact of such research would lead to better variance estimation techniques being available to researchers via our distance sampling analysis software. Current distance sampling analysis methods rely on either assuming random designs or on the variance estimators of Fewster et. al. (2009), both of which can over-estimate variability to different degrees depending on the survey characteristics. Fewster (2011) has demonstrated how improved variance estimation can be achieved for some designs using a ‘stripset’ approach. However, this has yet to be incorporated into our analysis software. Doing so would allow investigation via a simulation tool within the software. There are also many designs for which the approach needs extended or other approaches developed – for example point designs, systematic line segment designs and camera trap surveys. The ability to better optimise survey design and more accurately estimate the variability of population estimates will allow for more cost efficient surveys and more effective wildlife management and conservation.

References

Fewster, R. M. (2011) Variance Estimation for Systematic Designs in Spatial Surveys, *Biometrics* 67(4):1518-31. <https://doi.org/10.1111/j.1541-0420.2011.01604.x>

Fewster, R. M., Buckland, S.T., Burnham, K.P., Borchers, D.L., Jupp, P.E., Laake, J.L. and

Thomas, L. (2009) Estimating the Encounter Rate Variance in Distance Sampling, *Biometrics* 65(1):225-236. <https://doi.org/10.1111/j.1541-0420.2008.01018.x>

Exploring methodological 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. We are interested in talking to candidates about projects in either direction, for example:

Analysing proteomic data in prostate cancer – As well as answering ‘standard’ questions such as which proteins differ between cancer types, can we borrow information from statistical ecology to estimate what we aren’t observing, or to better design future experiments?

Accounting for DNA degradation in eDNA studies. DNA will degrade at different rates depending on environment (temperature etc). Can we use laboratory investigations of DNA to inform models that use environmental DNA to make inference about species populations?

We are open to discussing alternative projects and ideas in this area.

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

Supervisors: Hannah Worthington and Michail Papatomas

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.

Causal inference and trial emulation for ecological and/or environmental observational data

Supervisor: Hannah Worthington and Ben Swallow

Conducting formal causal inference for ecological and/or environmental 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 this is dependent on constructing emulated experiments to ensure independence between the observation process and treatment allocation. This PhD will develop and apply novel methodology in causal inference and causal discovery for ecological and/or environmental

observational studies, for example changes in spatio-temporal distribution and movement patterns driven by environmental disturbance and impact assessment studies.

Application procedure

Although there is no fixed deadline (unless noted otherwise for a particular topic), you are strongly encouraged to make your application and contact the potential supervisors as early as possible! The first round of funding decisions will be made in mid-January 2026.

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: Giorgos Minas, email gm256@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/postgraduate-research/apply/> (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/postgraduate-research/apply/>

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/postgraduate/fees/scholarships/>
- We look forward to hearing from you!