



PhD opportunities in Statistics at St Andrews, 2022-2023

(Updated 12th January 2022.)

Applications are welcomed for students wishing to undertake a PhD in Statistics at St Andrews. Full funding (fees, plus stipend of approx. £15,840) is available for well-qualified students; we encourage applications as soon as possible to maximize your chances of being funded. UK, EU and other overseas students are all encouraged to apply. New PhD students would typically start in September 2022, 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 within the School of Mathematics and Statistics (<https://www.mcs.st-and.ac.uk>), and consists of 17 members of academic staff, 16 research staff and 13 PhD students (plus PhD students co-supervised in other Schools). Our research is consistently ranked towards the top of UK universities in research assessments – for example we were ranked 11th in the UK and 1st in Scotland in the most recent (2014) UK Research Excellence Framework.

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; <https://www.creem.st-and.ac.uk>), 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 (<https://www.ncse.org.uk>), 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 (CBD; <https://synergy.st-andrews.ac.uk/cbd/>).

A second more recent and rapidly developing research focus is statistical medicine and molecular biology group (<https://sites.google.com/view/smmb/home>), led by a joint professorial appointment (Prof. Andy Lynch) between the Schools of Mathematics and Statistics and the School of Medicine, three recent appointments (Drs. Giorgos Minas, Nicolò Margaritella and Rachel Sippy) in this broad area, and research interests from several other staff members.

Many staff members are also active more generally in the field of statistical methodology. Research areas include Bayesian statistical inference and, relatedly, computer-intensive inference, data mining, data smoothing, latent state models and experimental design.

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 on the school and CREEM web sites.

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 (<https://smstc.ac.uk/>) and Academy for Postgraduate Training in Statistics (www.apt.ac.uk), the latter consisting of four one-week residential courses (although these may be online in the 2022/23 academic year due to COVID). Students may get the opportunity to become involved in externally-funded research as part of CREEM's consultancy group (<https://www.creem.st-andrews.ac.uk/consultancy/>); 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 2022 Times and Sunday Times; 3rd place 2022 The Guardian) and specialist bodies (2nd place 2021 in National Student Survey; 4th place 2022 Complete University Guide). 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 web site <https://www.st-andrews.ac.uk/subjects/study-options/pg/> and in the PhD prospectus <https://www.st-andrews.ac.uk/study/prospectus/pg-prospectus/>. School-specific information about applying is given at <https://www.st-andrews.ac.uk/mathematics-statistics/prospective/pgpr/>.

Brief summary of academic staff interests in the Division of Statistics

- David Borchers – spatial capture-recapture, wildlife surveys, spatial modelling
- Carl Donovan – data mining, commercial statistics, multivariate statistics
- 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
- Théo Michelot – animal movement, time series, latent state models
- Giorgos Minas – stochastic processes and multivariate statistics in molecular biology and medicine
- Michail Papatomas – Bayesian methods with application to genetics and biostatistics
- Valentin Popov – time series and hidden process models
- Rachel Sippy – statistical modelling of infectious diseases, particularly dengue virus.

- Chris Sutherland – statistical ecology: spatial capture-recapture, spatial occupancy models, multi-species occupancy modelling, optimal survey design
- Len Thomas – wildlife (particularly acoustic) surveys, population dynamics modelling
- Hannah Worthington – hidden Markov models for mark-recapture analyses, links between ecological and epidemiological/medical methods

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

- Rosemary Bailey – design of experiments in agriculture, horticulture, ecology and medicine
- Steve Buckland – biodiversity, sampling methods, computer-intensive methods
- Richard Glennie – latent or partially observed stochastic processes , with applications in statistical ecology
- Monique Mackenzie – random effects models, smoothing methods

Specific projects offered for 2021-22

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.

Statistical models for digital ecological surveys

Supervisor: David Borchers

Ecological surveys are the foundation on which evidence-based conservation of the planet's biodiversity and wildlife resources is built. Wildlife surveys that provide the evidence base have traditionally been conducted by humans and the resulting data treated as "snapshots" in time. However, digital survey devices like camera traps, acoustic arrays or aerial video platforms generate streams of data, not snapshots. These data are more appropriately viewed as time-to-event data, with the events being detections of the species of interest. Using the event times has the potential to yield much richer inferences about the populations under study than snapshot data can, but very little work has been done in this area. If you are interested in developing and applying new statistical methods to exploit the power of digital survey devices, this may be the PhD for you. Specific applications include camera trap surveys of large cats, acoustic surveys of gibbons, chimpanzees, and various bird species, drone-borne aerial surveys of snow leopard prey, and many others.

Object classification from mobile and static sensor feeds

Supervisor: Carl Donovan

The demand for video processing is rapidly increasing, driven by greater numbers of sensors with greater resolution, new types of sensors, new collection methods and an ever wider range of applications. For example, video surveillance, vehicle automation or wildlife monitoring, with data

gathered in visual/infra-red spectra or SONAR, from multiple sensors being fixed or vehicle/drone-mounted.

This project will focus on a specific application – object (animal) extraction and classification from extremely high-resolution aerial video from moving platforms. Issues of data size, dynamic backgrounds, rapid platform and target movement and classification errors will all need to be resolved and propagated into the final goal – inferring the densities of target species.

The project will require solving substantive computational bottlenecks and creative programming e.g. GPU and distributed file systems. Elements can be found in Erichson & Donovan (2016), but is only a tiny fraction of what is required.

References:

Erichson, N. B. & Donovan, C. R. (2016) Randomized low-rank Dynamic Mode Decomposition for motion detection. *Computer Vision and Image Understanding*. Vol. 146 pp 40-50.

Trading in Peer-to-peer (P2P) markets

Supervisor: Carl Donovan

Peer-to-peer market trading is becoming more prevalent in many areas that were traditionally the domain of large companies – stocks, foreign exchange, gambling. This project will look at a variety of statistical fundamentals for these areas including the practical application of statistical arbitrage, traditional arbitrage, algorithmic and high-frequency trading.

The project deals with large data issues, data-mining/machine-learning methods, cloud-computing and the dynamics of automated trading via APIs. The project will be heavily computational in either R or python, with potentially compiled languages for computational bottlenecks. The ideal candidate would have a good grasp of practical computing and statistics.

Automated evaluation of geo-political risk

Supervisor: Carl Donovan

There are massive amounts of data presented to the internet in real or near-to-real-time that allow monitoring of economic and societal conditions, amongst other things. This has been used to some effect in the automated monitoring of stock fluctuations, used to inform algorithmic trading. The evaluation of data-sources and curation of these on the basis of predictive power is an area requiring exploration. The project here will focus on a multitude of data sources that would allow the real-time evaluation of geo-political conditions around the globe, with the intention to predict various market shifts and impending political flash-points. Data will be captured by a wide range of sources, including multiple languages, print, audio and video. Text-mining methods will be developed to generate topic models and monitoring these over time, including those emergent. Methods will be explored that evaluate predictive performance for the purposes of curating data-sources and selection of modelling techniques.

The project deals with large data issues, data-mining/machine-learning methods, cloud-computing and interaction with APIs. The project will be heavily computational in either R or python, with potentially compiled languages for computational bottlenecks. The ideal candidate would have a good grasp of practical computing and statistics.

The molecular biology of (prostate) cancer

Supervisor: Andy Lynch

Cancer is caused by mutations in the DNA of cells that drive changes in molecular processes. To investigate cancer, for studies such as The Cancer Genome Atlas and The International Cancer Genome Consortium, it is not only the DNA of samples that is profiled but additional orthogonal data are generated to profile the epigenetics, transcriptomics and proteomics of the same samples.

We research cancer to answer questions about the fundamental biology (why cancers arise, why some people are more at risk), but also to impact clinical pathways (can we better predict outcomes, can we identify new treatment options). This work in turn often requires the development of new methods and new software (e.g. to maximize the data from new technologies, to combine different types of data, to meta-analyse previously published data with our own) or new experimental designs. Prof Lynch welcomes applications from students wanting to work on the applied analysis of molecular data from prostate cancer, or the development of novel methods for cancer research.

Investigating modes of action of genetic risk variants through integrated analysis of multiple high-dimensional "omics" data.

Supervisors: Andy Lynch and Michail Papatomas

Genome-wide association studies have identified thousands of genomic loci that are associated with higher risk of a trait (often a disease such as breast cancer). While these associations may have been identified, the mechanisms through which they act have tended not to be elucidated. This despite the growing number of diverse data sets potentially available for the purpose (see for example The American Journal of Human Genetics 103, 637–653 for discussion).

This project will look to develop a flexible modelling framework to incorporate many potential sources of evidence in suggesting and evaluating mechanisms of action. This could potentially build upon work of Papatomas et al. (2012) and related efforts in producing a flexible Bayesian approach for the analysis of GWAS data. The evaluation of potential mechanisms may then feed back into the detection and prioritization of association loci through, e.g., specification of prior probabilities.

Reference:

Papatomas, M., Molitor, J., Hoggart, C., Hastie, D. and Richardson, S. 2012. Exploring data from genetic association studies using Bayesian variable selection and the Dirichlet process: application to searching for gene-gene patterns. Genetic Epidemiology 36, 663-674

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.

Statistical inference for stochastic dynamical systems in biology I

Supervisors: Giorgos Minas, Jochen Kursawe

Stochastic dynamical systems can describe the interactions governing biological processes. In many applications, such as the circadian clock or embryonic development, researchers are collecting time-course data to gain insights to dynamic behaviours and regulation. Statistical inference can be applied to these data to identify parameters and properties that would otherwise not be experimentally measurable. Key challenges for statistical inference in large dynamical systems are parameter identifiability and computational speed. Which parameters can be inferred given a specific type of data? Can we optimise the experimental design to make it most informative? Can we generate faster algorithms for a specific inference problem?

This project will use theoretical approaches to answer these questions. The candidate will develop new methodology that can help practitioners decide on their data collection and analysis routines. While this project focusses on dynamical systems in general, the results will be applicable to many real-world applications, including embryonic development, circadian rhythms, and dynamic regulation of physiology.

Statistical inference for stochastic dynamical systems in biology II

Supervisors: Giorgos Minas, Jochen Kursawe

This project will develop stochastic models describing the oscillatory dynamics of gene expression during embryonic stem cell differentiation. It will also develop statistical methodology and computational algorithms to estimate model parameters using 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 expression of targeted genes in single cells. These observations are important in understanding cellular processes, such as stem cell differentiation, which strongly depend on dynamic gene expression. Stem cell differentiation is a critical biological process for embryonic development, regeneration, and regenerative therapy approaches. Dr Cerys Manning has previously shown that gene expression oscillations 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 develop stochastic models that will be fitted to the highly variable live-cell imaging data. Bayesian statistical methodology will be employed to estimate model parameters, quantify model uncertainty, perform model comparisons, and derive predictions. The ideal candidate will be interested in Bayesian statistics, dynamical systems, and stem cell differentiation. Background in at least one of the above subjects will be beneficial, but candidates with other backgrounds will be considered.

Strategies for detecting high probability dependence structures

Supervisor: Michail Papathomas

The aim of the project is to investigate and infer the complex dependence structure between a large number of categorical and continuous variables. This can also be in the context of an outcome that is explained by a number of covariates. It has been shown that non-parametric Bayesian modelling such as the Dirichlet Process Mixture Model can inform such an investigation, due to the fact that variable selection within mixture modelling selects covariates that combine to form homogeneous groups of subjects, rather than covariates with a strong marginal signal (Papathomas et al. 2012, Papathomas and Richardson, 2016). A model search algorithm informed by results from a procedure that combines Bayesian mixture modelling and variable selection could potentially detect relatively low probability areas in the space of possible models, thus enabling a quicker determination of high probability areas. This project will extend the results in Papathomas and Richardson (2016), by utilizing results in Jing, Papathomas and Liverani (2021), considering more complex mixture models as well as the effect of including both continuous and categorical variables within the analysis. This research has a wide range of potential applications, but the focus will be on applications in Biostatistics, in particular the detection of gene-gene and gene-environment interactions.

References:

Jing, W. and Papathomas M. (2021) Challenges and proposals for Dirichlet process mixture models

with Gaussian kernels. In preparation.

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

Liverani, S., Hastie, D. I., Azizi, L., Papathomas, M. and Richardson, S. (2015) PReMiuM: An R package for Profile Regression Mixture Models using Dirichlet Processes. *Journal of Statistical Software*. 64, Issue 7, pp 1-30.

Papathomas, M., Molitor, J., Hoggart, C., Hastie, D. and Richardson, S. (2012) Exploring data from genetic association studies using Bayesian variable selection and the Dirichlet process: application to searching for gene-gene patterns. *Genetic Epidemiology*. 36, 663-674

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. Applications include discerning the relation between phenotypes and environmental, anthropometric or genetic risk factors. Complex dependence structures can be easily discerned using graphical log-linear models (Papathomas and Richardson, 2016). This can lead to the identification of functionally important pathways. Another application concerns the size of hidden populations, such as victims of modern slavery (Cruyff, M., Overstall, Papathomas, McRea (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, 2019). Non-identifiability is a major impediment to evaluating how factors interact, and understanding important biological 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. (2019). Parameter redundancy and the existence of maximum likelihood estimates in log-linear models. *Statistica Sinica*, 31, 1125-1143

Cruyff, M., Overstall, A., Papathomas, M. & McRea, R. (2020) Multiple system estimation of victims of human trafficking: model assessment and selection. *Crime and Delinquency*. Online First, 17p

Detecting prevalent clusters of Multimorbidity with uncertainty evaluation using Bayesian Mixture Modelling and subsampling

Supervisors: Michail Papathomas and Nicolò Margaritella

Multimorbidity is the presence of at least two diseases in an individual. MM is one of the major challenges faced by health systems across the world. In the UK, the care of multimorbid patients accounts for more than half of the primary and secondary care costs, with an expectation that the cost will rise. This project will utilise the Ninewells Hospital & Medical School dataset that follows the Scottish Tayside and Fife population. We will first identify MM patients, focusing on those with complex MM (more than 4 conditions). We will then utilise Bayesian Mixture Modelling to detect the prevalent clusters of MM (Papathomas et al. 2012, Jing and Papathomas 2021), with patients in the same cluster characterised by conditions of a homogenous range and nature. The use of Bayesian modelling will allow for the evaluation of uncertainty with regard to the number of clusters and also the typical subject profile within each one of the MM clusters. Profile Regression (Liverani et al. 2015), a popular extension of Bayesian Mixture Modelling, will be used to model the effectiveness of different treatments for these patients. As this concerns a large dataset, subsampling techniques will be used, which will lead to further developments in the field of Bayesian Mixture Modelling when handling large datasets.

References:

Jing, W. and Papathomas M. (2021) Challenges and proposals for Dirichlet process mixture models with Gaussian kernels. In preparation.

Liverani, S., Hastie, D. I., Azizi, L., Papathomas, M. and Richardson, S. (2015) PReMiuM: An R package for Profile Regression Mixture Models using Dirichlet Processes. *Journal of Statistical Software*. 64, Issue 7, pp 1-30.

Papathomas, M., Molitor, J., Hoggart, C., Hastie, D. and Richardson, S. (2012) Exploring data from genetic association studies using Bayesian variable selection and the Dirichlet process: application to searching for gene-gene patterns. *Genetic Epidemiology*. 36, 663-67

[How do recreational activities alter spatiotemporal species interactions networks, and can this knowledge assist in promoting pro-environmental behaviour?](#)

Supervisor: Chris Sutherland

Fully funded scholarship (via Scottish Universities Partnership for Environmental Research (SUPER) Doctoral Training Programme).

Ecological communities are characterised by complex interaction networks which define their structure and function, and effective conservation requires understanding the strength and direction of these species associations. Predicting the fate of ecological communities requires a network approach, explicitly acknowledging interspecific interdependencies to understand how network structure varies with environmental change. To date, network approaches that treat species as nodes and the strength of interactions as weighted links have omitted humans, despite species responding differently to non-lethal human impact. In this very applied PhD project, the student will collect and analyse a variety of monitoring data (eg, camera traps, audio devices) to identify how recreational disturbance influences species interaction networks and whether providing this information to park visitors can promote pro-environmental behaviour. This will involve the

development and application of statistical methods for quantifying ecological networks and animal movement patterns, and methods from social psychology to quantify people's beliefs, attitudes, intentions, and behaviours.

This is a fully funded PhD through the SUPER Doctoral Training Program (<https://superdtp.st-andrews.ac.uk/how-to-apply/>).

Developing Novel Methods for Estimating the Abundance of Breeding Grey Seals

Supervisors: Len Thomas, Deborah Russell (Biology), Eiren Jacobson

Fully funded scholarship (University of St Andrews World-Leading Doctoral Scholarship). **Closing date 7 February 2022.**

Interdisciplinary approaches, such as statistical ecology, are increasingly needed to tackle pressing environmental challenges. Uniting the disciplines of biology and statistics can help us to better understand and ultimately conserve the environment. For example, monitoring the abundance of animal populations over time is important for effective conservation and management, including sustainable resource acquisition from the environment. However, estimating abundance is difficult for many species that are not always observable (e.g., when at sea or migrating). This is an area of active research and development within the field of statistical ecology.

The grey seal population in the UK presents an ideal opportunity for the development of statistical methods for abundance estimation using a comprehensive long-term data set. This PhD project will lead to innovations in statistical ecology and make real-world contributions to the management of grey seals in the UK.

The UK hosts approximately 40% of the global population of grey seals and the population is protected under both national and international legislation. The Sea Mammal Research Unit (SMRU) at the University of St Andrews has monitored the UK grey seal population for over 30 years. Their findings feed into the NERC Special Committee on Seals (SCOS) reports which are used by UK and devolved governments to inform sustainable management of seal populations and marine spatial planning. However, accurate estimates of population size and trends (e.g., Thomas et al. 2019) depend on reliable estimates of grey seal pup production (i.e., the number of pups born each year). Multiple counts of breeding colonies are conducted over a season, and are combined with information on life history parameters to derive a birth curve and estimate pup production (Russell et al. 2019). With modern statistical methods and computational capabilities, the student, with support from their supervisors, will develop a new pup production model that can account for recent changes in survey methods and sources of observational uncertainty, and ultimately provide more robust estimates of pup production.

The project will be tailored to suit the student's specific skills and interests, within the overall research topic. We envision that the student will build on a preliminary pup production model developed by the supervisory team to: 1) explore Bayesian approaches to model fitting that incorporate different sources of information (e.g., data, information from previous studies, expert opinion) to improve inference; 2) create a hierarchical multi-year, multi-colony models so that

available information from data-rich colonies and years is effectively shared with data-poor colonies and years; 3) develop statistical methods for estimating uncertainty at a sub-population (e.g., management area, regional units) level; and 4) conduct sensitivity analyses via simulation to determine the timing and number of surveys that would maximize the robustness of pup production estimates given available resources. Statistical modelling will be conducted in the statistical software R with model development through packages such as TMB, nimble, and RStan.

For more details about this project, and for application instructions, please see <https://www.st-andrews.ac.uk/study/fees-and-funding/postgraduate/scholarships/world-leading-02/>

References:

Jacobson, EK, Boyd, C, McGuire, TL, Shelden, KEW, Himes Boor, GK & Punt, AE. (2020). Assessing cetacean populations using integrated population models: an example with Cook Inlet beluga whales. *Ecol App*, 30(5), e02114.

Russell DJF, Morris CD, Duck CD, Thompson D, Hiby L. (2019) Monitoring long-term changes in UK grey seal pup production. *Aquatic Conserv: Mar Freshw Ecosyst*. 29: 24-39.

Thomas L, Russell DJF, Duck CD, et al. (2019) Modelling the population size and dynamics of the British grey seal. *Aquatic Conserv: Mar Freshw Ecosyst*. 29:6-23.

Modelling local population dynamics for whale sharks in the Maldives

Supervisor: Hannah Worthington

Data for this project will be provided by the Maldives Whale Shark Research Programme. The broad aim is to further the understanding of the local population dynamics of this enigmatic species.

Potential areas for investigation include:

- Robust stopover models for long-term mark-recapture data. The extension of these models to include biologically realistic behaviour (temporary migration, heterogeneous captures, individual covariates, environmental covariates). Robust estimation of recruitment probabilities and stopover duration on different temporal scales.
- Daily estimation of abundance and/or animal density. Modelling seasonal variation caused by, for example, lunar cycles, monsoon season, time-varying environmental conditions.
- Modelling animal movement and identifying habitual behaviour patterns. Investigating differences in behaviour for sharks of different ages or regions.
- The integration of different sampling methods including emerging technologies (drone surveys, sonar, eDNA). The inclusion of citizen science data.
- The potential extension of existing mark-recapture methods to a continuous-time framework.

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!

Many details of the general requirements and admissions procedure are given at the university web site <https://www.st-andrews.ac.uk/study/pg/apply/research/>

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. Please note that our primary 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).

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 are welcome to include a covering letter. 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!