



## PhD opportunities in Statistics at St Andrews, 2018-2019

(Updated 10<sup>th</sup> November 2017.)

Applications are welcomed for students wishing to undertake a PhD in Statistics at St Andrews. Full funding (fees, plus stipend of approx. £14,500) 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 2018, 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 (<http://www.mcs.st-and.ac.uk>), and consists of 9 members of academic staff, 10 research staff and 8 PhD students. It was ranked first in Scotland in the 2008 Research Assessment Exercise (the last for which statistics was ranked separately from mathematics); the School of Mathematics and Statistics was ranked first in Scotland in the 2014 Research Assessment Framework, and is ranked 3<sup>th</sup> in the UK for 2018 by the Complete University Guide (source <http://www.thecompleteuniversityguide.co.uk>). 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; [www.creem.st-and.ac.uk](http://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 ([www.ncse.org.uk](http://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; <http://synergy.st-andrews.ac.uk/research/cbd/>). A second research focus is in the area of Bayesian statistical inference and, relatedly, computer-intensive inference. Members of staff are also active in the fields of datamining, data smoothing, latent state models and statistical genetics. We are also building links with the Mathematical Biology research group within applied mathematics. 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 ([www.smstc.ac.uk](http://www.smstc.ac.uk)) and Academy for Postgraduate Training in Statistics ([www.aps.ac.uk](http://www.aps.ac.uk)), 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 ([creem2.st-andrews.ac.uk/consultancy/](http://creem2.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., 1<sup>st</sup> place 2018 Times and Sunday Times; 3<sup>rd</sup> place 2018 The Guardian). 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/pg/prospectus/research/>. School-specific information about applying is in this pdf: <https://www.st-andrews.ac.uk/media/school-of-mathematics-and-statistics/documents/prospective-students/st-andrews-mathsstats-pgr-info.pdf>

***Brief summary of academic staff interests in the Division of Statistics***

- Rosemary Bailey – design of experiments
- David Borchers – mark-recapture, wildlife estimation, sampling methods
- Steve Buckland – biodiversity, sampling methods, computer-intensive methods
- Carl Donovan – datamining, commercial statistics, multivariate statistics
- Ian Goudie – mark-recapture, plant-capture, sequential inference
- Janine Illian – spatial statistics, biodiversity
- Andy Lynch - statistical genomics and cancer research
- Monique Mackenzie – Random effects models, smoothing methods
- Michail Papathomas – Bayesian methods with application to genetics and biostatistics
- Len Thomas – wildlife (particularly acoustic) surveys, population dynamics modelling

## **Specific projects offered for 2017-18**

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.

### *Spatial modelling of Scottish marine fauna.*

Supervisors: David Borchers, Finn Lindgren (University of Edinburgh), K. Brookes (Marine Science Scotland)

(Note that this PhD is funded under a separate funding stream from the other ones offered.)

This fully-funded PhD will develop and apply innovative statistical methods to model the spatial distribution and abundance of selected Marine fauna species, from survey data. Marine fauna is surveyed in a wide variety of ways around Scotland for management and conservation purposes. This includes visual surveys from ships, visual and camera-based surveys from aircraft, visual surveys from shore-based platforms, acoustic surveys with stationary hydrophones, acoustic surveys with towed hydrophones, and trawl surveys. The PhD will focus on a selection of surveys and species, depending on the candidate's interest, Marine Scotland Science priorities, and suitable data availability, in order to adapt and develop existing statistical methods and software to obtain the best possible estimates of the spatial distribution of species occupancy and/or abundance of the selected species from these data.

It is anticipated that the PhD will build on the methods developed a recent EPSRC-funded project that developed the inlabru software and associated methods (see <https://sites.google.com/inlabru.org/inlabru>).

### *Spatially explicit capture-recapture method development.*

Supervisor: David Borchers

Spatial capture-recapture (SCR) methods are fast replacing non-spatial capture-recapture for most wildlife surveys. The methods are developing fast and there is enormous scope for innovative method development. A PhD in this area could focus on theoretical problems such as efficient methods of dealing with random effects, modelling territoriality or spatial correlation, modelling movement, births and deaths; or it could focus on issues of important practical relevance such as survey design; or it could focus on methods for particular kinds of surveys: by acoustic detectors, for example, or by DNA sampling, or surveys using drones. Video and acoustic SCR surveys present challenges for individual identification, image and acoustic processing, and dealing with big data. There are many other potential SCR research avenues and the PhD could also focus on some other SCR-related issue of your choosing, subject to agreement with your supervisor.

*Inferring the impact of varying boundary conditions based on spatial statistics and inverse modelling.*

Supervisor: Janine Illian (co-supervisor: Justin Travis, University of Aberdeen)

Species' range boundaries are a consequence of ecological and evolutionary processes and, in turn, can themselves influence ecological and evolutionary processes which together yield observed patterns of individuals and genotypes. Describing and understanding ecotones, biogeographic range limits and hybrid zones are all major current topics in ecology: all involve boundary conditions and all would benefit greatly from the development of spatial statistics that can more effectively describe the unique spatial patterns associated with these boundaries. In this project we will exploit the flexibility of recently developed spatial statistical methods that allow incorporation of different boundary conditions and thus produce spatial patterns that can differ close to boundaries. The first objective will be to develop these models to be useful in an ecological context and use them to describe the spatial patterns present at a case study ecotone, a shifting range limit and a hybrid zone. The second objective is to use the spatial statistics within an inverse-modelling framework to infer the ecological processes (e.g. spatial scale of dispersal, kernel of competitive interactions) that underlie the spatial patterns found close to boundaries.

*Developing flexible spatial models with complex boundary structures.*

Supervisors: Janine Illian (joint with Sophie Smout, School of Biology and Beth Scott, University of Aberdeen)

Policy makers aim to reconcile human socio-economic objectives in the marine environment, and the conservation of protected/sensitive species such as marine mammals. It is increasingly clear that management must take account of the spatial complexity of ocean habitats, with boundaries including coastlines and oceanographic features and that marine species' use of this space is complex and responsive. In order to manage conservation effectively, it is crucial not only that we describe existing distributions of these species, but we must also aim to understand the processes that drive them.

This project will develop Bayesian spatial modelling methods which are flexible, can make use of independent data/knowledge to set informative priors, and facilitate understanding of uncertainty/risk. To make model fitting feasible we will develop methods within the framework of INLA (integrated nested Laplace approximation), Rue et al. 2009) extending existing spatial modelling methods (Lindgren et al., 2011) ) where boundaries have different reflecting/absorbing properties or by considering the distribution of multiple species taking account of dependencies that might exist such as competitive interactions.

References:

Rue H., Martino S. and Chopin N.: Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, 71, 319–392

Lindgren, F., Rue, H. and Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 73 (4), pp. 423-498.

### *Spatially adaptive tiling.*

Supervisor: Monique MacKenzie

The Joint Cetacean Protocol (JCP; <http://jncc.defra.gov.uk/page-5657>) is an ambitious modelling project that involves spatially adaptive smoothing of marine mammal geo-referenced data on a large spatial and temporal scale (approximately 1 million km<sup>2</sup> over about 30 years). Due to computational considerations, current smoothing methods (with targeted flexibility) require the modelling effort to be spread out over the entire spatial area which is likely to be sub-optimal for many locations – more flexibility is likely required in some areas. This PhD project will involve the development of a new ‘tiling’ method to handle large smoothing problems such as these, which permits spatially adaptive pieces of the surface to be individually considered and then sensibly aggregated to return a fitted surface. This project will also involve the treatment of auto-correlated data since data sets of this kind consist of data which are more similar for points close together in time and space (compared with distant points).

### *Bayesian identifiability for log-linear models.*

Supervisor: Michail Papathomas

The description is: Log-linear modelling is the standard approach for investigating the full joint dependence structure between categorical variables 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. 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. Non-identifiability is a major impediment to evaluating how risk 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

### *Object classification from mobile and static sensor feeds*

Supervisors: 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.

## **Stats-Medicine projects**

The following projects, supervised by Prof. Andy Lynch, represent a collaboration between the Division of Statistics and the School of Medicine.

### *Identifying non-linear combinations of gene expression signals*

Supervisor: Andy Lynch

In the history of high-throughput gene expression analysis from microarray to RNA-Seq data, there have been many methods and tools published that make use of the assumption that the combined gene signal of a mixture of two cell-types will be the mixture of their individual gene signals. There are though a couple of scenarios that might cause us to leave behind that scenario. The first is that there may be a genuine difference in signal when cell-types are mixed (e.g. due to cells responding to signals coming from other cell types). The second is that it appears that there is a difference due to selection or censoring (e.g. If we looked across the entire population we might see that two gene expression profiles were additive, but because we are selecting for a particular gene-driven phenotype they may appear not to be).

This project would consider models to detect such scenarios in public data sets, and would seek to quantify the impact of any such phenomena on large genomic health studies.

### *Modelling the interface of metabolism, methylation and mitochondria in prostate cancer*

Supervisor: Andy Lynch

Funding for this project is being sought – meantime, please contact Prof. Lynch ([andy.lynch@st-andrews.ac.uk](mailto:andy.lynch@st-andrews.ac.uk)) for further information about the project.

### *Analysis methods for the proteogenomics of oesophageal cancer*

Supervisor: Andy Lynch (joint with Prof. Russell Petty, Univ Dundee).

*Background.* Oesophageal cancer is a priority area of research. While only the 14th most common cancer in the UK, it is the 6th most common cause of cancer death. This disparity is an indicator of the generally poor survival rates for the disease in spite of a substantial research effort. Much work from projects such as TCGA and ICGC has gone into identifying molecular subtypes of oesophageal (and more general gastric) cancers in order that targeted treatment options might be identified. While considerable numbers of oesophageal cancers have had their genomes sequenced, far fewer have also undergone quantitative proteomic profiling, despite this being the key molecular phenotype by which we should assess cases.

Cell-lines, organoids, and other disease models have a vital role to play in understanding the biology of oesophageal cancer and in developing potential treatments. Inferences from experiments using such models are only as valid as the model is truly representative of the diseases. A number of cell-lines have been profiled for their genomic sequence, but not their proteome.

*The Project.* Supervised by Professor Andy Lynch and in collaboration with Professor Russell Petty (University of Dundee), the project will analyse proteomic data from oesophageal cancer models, placing the results in the context of current knowledge. The successful candidate will apply existing approaches, and develop new methods, for the integration of proteomic and genomic data. The degree to which the disease models represent clinical cases, and the proportion of cases represented will be assessed with a view to developing a framework to predict the extent to which the molecular effects of interventions on the disease model can be expected to translate to clinical samples. In this manner the disease models as tools for understanding clinically-defined disease subtypes will have maximum utility.

*The person.* Applications are encouraged from graduates with backgrounds in a numerate discipline (e.g. bioinformatics, statistics, mathematics, or computer science). The ideal candidate will have an interest in genomics and an enthusiasm for learning about biology. A Masters degree in statistics or similar would be desirable. Some experience of working with genetic data would be desirable but not essential. Experience of coding and scripting in R (and ideally Bioconductor) is desirable.

### ***Stats-MathBio projects***

The following projects represent a collaboration between the CREEM research group within statistics and the Mathematical Biology research group within applied mathematics.

*Continuous time models of animal movement, applied to animal-borne tags and spatial capture recapture*

Supervisors: Len Thomas and Mark Chaplain (and possibly David Borchers)

Most animals must move to forage, avoid predators and locate mates. Hence understanding animal movement is crucial to a full understanding of individual animal success and, ultimately, whole-population dynamics. Ongoing technological developments mean that increasing numbers of animal-borne tags are being applied; depending on the model these may yield noisy observations of position (via ARGOS satellite), or precise positions (via GPS), and other data such as height or depth (for diving animals), direction, acceleration, etc. Although movement occurs in continuous time, analytically tractable continuous time models tend to be overly simplistic (e.g., assuming random movement (Brownian motion)); more complex modelling tends to involve discrete time approximations. This project will investigate the potential for fitting more complex continuous time models using analysis where possible, but simulation-based inference where not. An additional application of such models is in fitting spatial capture-recapture models – these are used to estimate population size from time series of “captures” of marked individuals in a population. The capture does not have to be physical, for example identifying a uniquely-marked tiger on a series of infra-red-triggered camera traps throughout a forest. This PhD will suit students with an interest in (and prior training in) statistics and applied mathematics.

*Parameterizing mathematical models of angiogenesis and diffusion, and selecting among competing models.*

Supervisors: Mark Chaplain and Len Thomas

Mathematical models of angiogenesis (blood vessel development) and the diffusion of chemicals from blood to tissues are both crucial components of our ability to predict whether new drugs will work – for example drugs that target cancer. These models are typically formulated as stochastic differential equations, with model parameters that govern the how the quantities modelled evolve over time. These model parameters are traditionally set in an ad hoc manner, using “reasonable” values and tuning them to achieve an approximate match to data. This project aims to improve upon this process, by leveraging fitting and optimization techniques from statistics. In particular, Approximate Bayesian Computation (ABC) appears to offer strong potential for a rigorous yet flexible fitting mechanism for such complex models. In addition to model fitting, the project will also address the question of how to choose among competing models in a rigorous manner, leveraging model selection techniques from statistics. This PhD will work at the interface between mathematical biology and applied statistics.

***Centre for Biological Diversity-specific projects***

The following projects are based within the Centre for Biological Diversity; the primary supervising school may not be Mathematics and Statistics. Please contact us for more details if you are interested in one of these projects.



*Visualizing movement interaction for biodiversity.*

Supervisors: Urška Demšar and Jed Long (Geography and Geosciences), Janine Illian (Statistics)

Recent developments in positioning sensor technologies have enabled location-tracking at an unprecedented level of detail. These advances have allowed for great strides in the field of movement ecology, which studies the movement of individuals and groups across different animal species, different scales of space and time and different levels of complexity. Observing animal movement leads to improved understanding not only of properties of movement of an individual animal but also in how observed individuals interact with each other and with other co-located species. These are crucial factors in understanding the spatio-temporal dynamics of inter- and intra-species interactions in the competition for resources. Methods for visualisation of interaction are lagging behind sensor development due to the complexity of the task, in particular since movement data are often complemented with additional animal-borne sensor data (e.g. oceanographic sensors and accelerometers). This project will use advances in information visualisation to develop new methods for visualising interaction in movement. The aim is to develop new visualisations of movement and other sensor data which will allow identification of spatio-temporal patterns in interaction between animal species and thus allow new insights into the dynamics of biodiversity. The methods will be tested on a case study on interaction of two species of seals: grey seals and harbour seals.

*The influence of body condition on functional behavioural decisions of animals.*

Supervisors: Nathan Bailey and Patrick Miller (Biology), Len Thomas (Statistics)

The goal of this project is to combine theoretical development with laboratory experiments with an appropriate model organism to predict and evaluate the role of body condition on behaviour. In animals, there is expected to be a fundamental trade-off between foraging and anti-predator vigilance and behaviour. Energy-store body condition of individuals is predicted based upon existing theory to influence such functional behavioural decisions (Houston et al., 1993; Proc Roy Soc B). By more explicitly developing and testing fundamental theoretical predictions, this project will add value by providing more advanced tools to interpret the biological significance of measures of body condition made with free-ranging animals across a wide range of taxa. Thus, the project will improve our ability to monitor the health status of living animals, aiding in conservation applications.

**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 (UK class 2:1 or better, or international equivalent) in mathematics, statistics or another scientific discipline with a substantial numerical component. Applicants with degrees in other subjects, such as biology, are invited to discuss their qualifications with the Postgraduate Officer (contact details below). 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 have a good to very good 1<sup>st</sup> 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](mailto: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/math/prospective/pg/phdprogrammes/>

an in this pdf

<https://www.st-andrews.ac.uk/media/school-of-mathematics-and-statistics/documents/prospective-students/st-andrews-mathsstats-pgr-info.pdf>

We look forward to hearing from you.