

Research Summary

My research focuses on developing hierarchical models for ecological and environmental data, and Bayesian methods for fitting these models, with an emphasis on data from mark-recapture studies of animal populations. Hierarchical models are important in the analysis of ecological and environment data because they allow researchers to test hypotheses about the unobservable processes driving complex systems. The use of these models within classical statistics has been limited because the likelihood functions often include intractable sums and integrals; however, advances in computation have allowed many of these models to be fit with Bayesian methods. I develop novel hierarchical models to address new problems in ecology and environmental science, and I construct Markov chain Monte Carlo (MCMC) sampling algorithms to obtain inference from these models. The models I develop often incorporate latent variables that are continuous functions of time and this has led me to secondary research in the area of functional data analysis.

Hierarchical Modelling of Mark-Recapture

Data from mark-recapture studies are complicated by the simple fact that animals in the wild cannot be observed at will. Individual life histories will never be known completely, and the missing information – which is not missing at random – must be modelled directly. In the past, this has limited the questions that could be addressed using mark-recapture data because the necessary models were too difficult to fit with classical methods. Researchers began developing Bayesian methods for the analysis of mark-recapture data starting in the mid-1990's, and my work has been at the forefront of these efforts.

One significant problem I have addressed concerns the use of continuous, time varying, individual covariates in mark-recapture models. Continuous measures of fitness like body mass are important predictors of survival for many species, but these variables can only be observed when individuals are captured. Papers I have published in *Biometrics* in collaboration with Carl Schwarz, Byron Morgan, and Ruth King develop hierarchical models to account for the missing covariate values, describe MCMC methods to fit these models, and compare the results with competing methods. A further publication in the refereed proceedings of the EURING Technical Meeting (the principal forum for the discussion of mark-recapture methods) extends these models to allow for more complex relationships which may be representative of complex evolutionary processes (e.g., stabilizing and disruptive selection). I am continuing to develop these models to provide methods for estimating abundance from large mark-recapture data sets with continuous covariates, as discussed in my future research plans below.

In other work, I have developed new models to improve estimates of abundance from mark-recapture data stratified by time. Populations are often stratified to reduce bias in abundance estimates caused by varying capture rates, but stratification also decreases precision. A paper I published in *Biometrics* describes new models that incorporate Bayesian penalized splines to fit mean abundance as a smooth function of time to maintain precision by sharing information between neighbouring strata. Carl Schwarz and I jointly maintain an R-package implementing these methods which is currently being used by fisheries scientists studying several threatened populations of Pacific salmon.

Most recently, I have become interested in problems of imperfect identification in mark-recapture studies. This research developed from collaborations with researchers at EcOcean USA who study whale sharks using automated photo-identification. Photographs in the catalogue may have been taken from either side but the skin patterns on the left and right side of the same shark cannot always be matched. Events will be duplicated if all photographs are used without correction but information will be lost if data from only the right or left side photographs are considered. A solution to this problem had been published previously, but estimates of uncertainty from this method were severely biased. I published a response to this paper explaining the cause of this bias, and developed a new, Bayesian method to produce valid inferences from all photographs. I was invited to talk about this work at the 2012 International Biometrics Conference and a paper with *Biometrics* is in press. In continuing work with Matthew Schofield and Ruriko Yoshida at the University of Kentucky, I am developing new MCMC algorithms to fit models with more complicated forms of misidentification.

Smoothing and Functional Data Analysis

My secondary research concerns applications of semiparametric regression and functional data analysis in modelling ecological and biological data. This interest was originally motivated by my work with time stratified mark-recapture data, and I have applied these methods in two other areas.

First, I have developed semiparametric methods to assess the links between teleconnections (periodic patterns in atmospheric pressure) and variations in long-term temperature observations. In a paper published in *Environmental and Ecological Statistics* which I co-authored with Nancy Heckman at the University of British Columbia and Nathaniel Newlands at Agriculture/Agri-Foods Canada, I developed new models under the assumption that spatial variations in long-term observations in temperature can be explained by a small number of functional principal components. In a three-stage approach using multiple simulations to account for uncertainty, I modelled the time series with Bayesian splines, computed the functional principal components, and then identified links between these components and the teleconnections. By applying these methods to data collected from weather stations across British Columbia over a 50-year period I showed links with several major teleconnections and also identified a province-wide warming trend matching the effect of climate change.

I have also been developing semiparametric models to help Vincent Cassone in the Department of Biology at the University of Kentucky study the entrainment of circadian rhythms. In his experiments, birds are kept in darkness for long periods, and different stimuli are introduced for several weeks at a time. Entrainment occurs when a bird's daily activity synchronizes with an external stimulus. Existing methods to test for entrainment apply simple ANOVA to compare point estimates of the mean period (time between waking) across treatments, but these methods are inappropriate because they do not account for uncertainty in the estimated period and because the assumptions of ANOVA are not satisfied. Most importantly, the mean period is not constant within each treatment but drifts gradually when stimuli are changed. I have developed hierarchical models that treat the period as a latent variable to allow for uncertainty and that also incorporate splines to model the mean period as a smooth function of time. One of Dr. Cassone's graduate students is currently using these methods in his work, and I am preparing a manuscript for submission to the *Annals of Applied Statistics*.

Future Research Plans

My future research will focus on the following three projects.

1. Analysis of Large Mark-Recapture Data Sets:

Recent developments in Bayesian computation have allowed more and more complex models to be fit to mark-recapture data, but these methods are often too computationally intensive for large data sets. A particular challenge in mark-recapture concerns estimating abundance when capture probabilities depend on individual covariates. Existing methods model the entire population through data augmentation, but the computational complexity of the algorithms for fitting these models grows with the size of the population.

I have proposed two alternative methods whose complexity depends only on the number of individuals observed. The first uses Monte Carlo estimation within the MCMC algorithm to approximate the posterior density numerically. I was invited to present preliminary results from this work at the 2013 EURING Technical Meeting and a paper is currently in press with *Methods in Ecology and Evolution*. My second method uses MCMC output from models of the observed data to estimate abundance using an adjusted Horvitz-Thompson estimator. I am still investigating the theoretical properties of this estimator and have begun preparing a manuscript for submission to *Ecological and Environmental Statistics*.

2. Misidentification in Mark-Recapture and Irreducibility of Markov chains:

Although I was able to address the problem of data duplication from photo-identification in mark-recapture, my work raised concerns about methods for fitting models to mark-recapture data with other forms of misidentification. In particular, I have shown that simple MCMC algorithms may produce irreducible chains. I am currently working with Matthew Schofield and Ruriko Yoshida from the University of Kentucky to develop new MCMC algorithms using Markov bases to construct irreducible chains for fitting these models. I am the principal investigator on a grant application that we are submitting to the National Science Foundation's statistics program. Our proposal addresses three issues. First, we will determine what aspects of misidentification in mark-recapture require more involved MCMC algorithms to construct irreducible Markov chains. Second, we will investigate methods to construct Markov bases and Markov sub-bases for our specific models, as general algorithms are too computationally demanding. Third, we will explore alternative methods using multiple-chain MCMC or modified posterior distributions to construct valid sampling algorithms.

3. Functional Data Analysis of Circadian Rhythms:

My third goal is to incorporate functional data analysis into my methods for analyzing data on circadian rhythms. The methods I have developed so far apply to data from a single individual, and I would like to compare behaviour across individuals (e.g., between males and females or among groups exposed to different stimuli). I will extend the models I have developed using methods from functional data analysis to compare responses among different groups, and I will be applying to the National Science Foundation's mathematical biology program this fall for funding to conduct this work.