

# Dr. Simon Bonner

Department of Statistics  
University of British Columbia  
333-6356 Agricultural Road  
Vancouver, BC V6T 1Z2

Phone: W. 604-822-0570  
H. 604-422-0056  
E-mail: s.bonner@stat.ubc.ca  
Web-page: www.simon.bonniers.ca

## Research Summary

Some of the biggest challenges facing science today involve understanding, evaluating, and managing humans' impact on the natural environment. The most obvious effects in recent years are the changes in climate caused by the release of greenhouse gases, the modification of ecosystems for human uses including forestry, agriculture and aquaculture, and the pollution of land, air and water. My research focuses on developing statistical methods to study these changes in the environment and their impacts on wild populations.

## Capture-Recapture Methodology

My primary area of expertise is the development of models for analysing data from capture-recapture (CR) studies of wild animal populations. Simple CR models enforce the assumptions that sampling probabilities and demographics parameters, like survival probabilities or migration rates, are the same for all individuals in the population. These assumptions are rarely realistic, and my research has focused on models that account for two common sources of variation in the population: temporal changes and the effects of certain types of covariates.

Populations of anadromous fish, like the salmon species native to Pacific North America, are commonly monitored by estimating the number of young fish leaving or adults returning to their freshwater spawning grounds. These migrations last for long periods, weeks or even months, over which time the sampling probabilities may change considerably so that assuming a single probability for all individuals can lead to severely biased estimates of abundance. One strategy is to stratify the data – essentially, estimating the sampling probabilities and abundance independently for each day or week of the study. This addresses the problem of bias, but counts in the small time periods are often very low and estimates of abundance very imprecise. As part of my PhD thesis, I developed an extension of the stratified model using Bayesian P-splines (Lang and Brezger, 2004) to model the expected counts in each time period as a smooth, but flexible, function of time. This explicitly accounts for the temporal structure, and simulations based on real data demonstrated greatly increased precision of abundance estimates without loss of accuracy.

I presented this work at the 2007 WNAR/IMS Annual Meeting and the 2008 Meeting of the Statistical Society of Canada, for which I received the student presentation award, and am currently preparing a manuscript based on this work. These methods are also being adopted by the Trinity River Restoration Program and in October I will be assisting my PhD supervisor, Dr. Carl Schwarz, in a training workshop for the program's members.

The second problem in CR methodology I have studied is to incorporate information from covariates that are unique to each individual and vary over time, like body mass or other indicators of an individual's fitness, in the modelling of capture probabilities and other parameters. The challenges with these types of variables are that they can only be observed if and when an individual is captured and that it is not reasonable to assume they are missing at random. In my MSc, I developed a Bayesian solution to make inference concerning the effects of such covariates on capture and survival probabilities. Using Bayesian methods, my solution explicitly models the observed covariate values, accounting for the dependence in the observations, and uses this to define a distribution for the missing data. I presented this work at the 2004 EURING Analytical Meeting (the

methodological conference for the European Bird Ringing Union and the primary forum for capturing-recapture methods) and a corresponding paper was published in *Biometrics* in 2006.

During my PhD I spent four months in the UK working with Drs. Byron Morgan and Ruth King to compare this method with a competing approach, known as the trinomial method, developed by Dr. Morgan and his collaborators (Catchpole et al., 2008). In our tests, the Bayesian method outperformed the trinomial method except when capture probabilities were very high or the extra assumptions required in defining the full Bayesian method did not fit the data. A paper based on this work has been conditionally accepted for publication in *Biometrics* (Bonner et al., 2009).

I later extended the Bayesian method to the problem of estimating the size of the population – the added challenge being that valid inference about capture probabilities can be obtained by modelling only the captured individuals, while abundance estimates require modelling the entire population, including the individuals that were never observed. My approach is based on an approximation to the posterior distribution obtained for a specific, non-informative prior distribution, and leads to an intuitive, two step procedure for sampling from the posterior distribution with reduced computational effort. I believe that this work has wider applications in CR methodology and general sampling problems and will be presenting this material at the 2009 EURING Analytical Meeting in September. I am currently finalizing a manuscript based on this work for submission this fall.

## Functional Data Analysis

Through my work with spline models in CR analysis I became interested in spline smoothing methods and the related area of functional data analysis (FDA). In January 2009, I joined the Department of Statistics at the University of British Columbia as an NSERC/PIMS post-doctoral fellow with Dr. Nancy Heckman, an expert on semi-parametric methods and FDA. One challenge I encountered in fitting the spline models in the CR context was that the choice of smoothing parameter often had a dramatic effect on the fit of the model, essentially because the spline was smoothing values that had not actually been observed and had to be estimated from the data themselves. The Bayesian P-spline approach incorporates this uncertainty, but the Markov chain Monte Carlo methods required to sample from the posterior distribution can take a long time to produce a representative sample. Together with Dr. Jiguo Cao at Simon Fraser University, Dr. Heckman and I have been working to adapt the parameter cascading method recently developed by Dr. Cao and Dr. Jim Ramsay to simplify inference from hierarchical functional models (Cao and Ramsay, 2007).

A second project I have been working on as part of my post-doctoral fellowship concerns the application of FDA to study changes in climate over small regions. In conjunction with Dr. Nathaniel Newlands at Agriculture and Agri-Foods Canada, I have been working to implement functional principal components methods to describe the major components of climate variation from data gathered at several weather stations in a small area of British Columbia, and then to model the variation as functions of predictors like time, elevation, and indices of the of local and global climate cycles. Challenges in this project include the sparsity of data caused by inconsistent sampling and the high correlation between observations at the closely spaced stations.

## Future Work

One future project I am very excited about will combine my CR spline models with what I am currently learning about FDA to produce more powerful methods for studying changes in fisheries populations. Capture-recapture studies of migrating fish populations have been conducted for many years at different locations along the Pacific coast of North America and around the world. By applying my spline method to each data set, I intend to construct a set of smooth curves which summarizes variation in the dynamics of these populations over time and between the different populations. The tools of FDA will then allow me to study the differences in these curves in relation to species, time, location, and global climate predictors. Because the estimates of the population size from my spline method are more precise than those of previous models, the

FDA approach should increase the power to detect factors affecting the populations. Moreover, FDA methods will provide information about more complicated changes in the migrations, like the timing of peak abundance, than can be obtained from simple abundance estimates.

Another area of CR methodology that has opened in recent years and that I am very interested in concerns the use of natural rather than man-made marks, like patterns in skin pigmentation or genetic markers obtained from hair or fecal samples. Experiments that rely on natural marks do not require animals to be physically captured, which may affect their behaviour or survival, and have the potential to generate much more data because every individual in the population is effectively marked from birth. In 2008 I began working with Jason Holmberg of Ecocean, a non-profit organization founded to study the global population of whale sharks, a species about which little is known despite their massive size. Using a custom designed web-interface, Ecocean collects photographs of whale sharks taken by tourists on scuba-diving trips and matches the sharks' skin patterning with other photographs in the database to construct an individual observation history. The ultimate goal of this work is to understand human impacts on the shark population, but there are several statistical issues that must be resolved first. One problem I have begun working on concerns how to combine data from multiple marks, e.g. patterns on different areas of the sharks' bodies, that may vary in their visibility or their reliability in matching.

Along with my research on CR methodology, I also plan to continue my work on functional climate modelling, adapting the methods I am developing with Dr. Heckman and Dr. Newlands for application to data from moderate scales at which the correlations between stations depends on their proximity (e.g., provincial data). To account for this, we will incorporate explicit models of spatial dependence into the procedures for computing the principal components and modelling their variability over time and space. Through Dr. Newlands, I have recently joined a group applying for a grant from Agriculture and Agri-Foods Canada's Sustainable Agriculture Environmental Systems (SAGES) initiative which we hope will provide funding for further work on this project.

## References

- S. J. Bonner and C. J. Schwarz. An extension of the Cormack-Jolly-Seber model for continuous covariates with application to *Microtus pennsylvanicus*. *Biometrics*, 62(1):142–149, March 2006.
- S. J. Bonner, B. J. T. Morgan, and R. King. Continuous, time-varying covariates in mark-recapture-recovery analyses: A comparison of methods. *Submitted to Biometrics.*, 2009.
- J. Cao and J. O. Ramsay. Parameter cascades and profiling in functional data analysis. *Computational Statistics*, 22(3):335–351, 2007.
- E. A. Catchpole, B. J. T. Morgan, and G. Tavecchia. A new method for analysing discrete life-history data with missing values. *Journal of the Royal Statistical Society – Series B*, 70:445–460, 2008.
- S. Lang and A. Brezger. Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13:183–212, 2004.