**A novel approach for estimating Capture Efficiency from large multi-pass electrofishing datasets: implications for density estimates.**

Colin P. Millar, Robert J. Fryer, Karen J. Millidine and Iain A. Malcolm.

# Abstract

Fisheries management and conservation decisions are often based on estimates of Juvenile salmonid abundance derived from electrofishing data In such circumstances, accurate estimates of abundance require reliable estimates of capture efficiency, which is only available from multi-pass electrofishing data. Typically, capture probability is estimated for individual sites visits, or in the case of single pass fishing, assumed to be constant. More recently hierarchical Bayesian models have been used to model capture probability (and jointly fish density) but these approaches do not scale well to larger datasets, taking considerable time to fit thereby reducing opportunities for model selection. This paper presents an approach for estimating capture probability based on classical conditional likelihood methods which allow capture probability to be modelled in terms of linear and non-linear (non-parametric) relationships with covariates that affect capture probability including habitat and personnel or equipment. The use of conditional likelihood provides a simple way to investigate complex relationships. Capture probability models were fitted to multi-pass electrofishing data for Scotland over the years 1996 to 2014. GIS covariates were included as proxies for habitat, while information on organisation allowed for variability in equipment, procedures and personnel. Organisation was the most important covariate, followed by Fish age class (fry, parr), species (salmon, trout) and day of year. Other factors such as width and gradient affected capture probability to a lesser extent. The simplifications introduced over HBMs allowed a full model selection procedure (potentially even model averaging) to be conducted. A simulation study was conducted to show the effect of these simplifying assumptions as well as the overall impact of modelling capture probability.

Keywords: Atlantic Salmon, Fry, Parr, Capture Efficiency / Probability, Electrofishing.

# Introduction

* Estimating density is important for fishery management and conservation
* To estimate density from electrofishing data you need to estimate capture probability
* Capture probability varies with x,y,z………
* Capture probability estimation requires multipass fishing
* Often single pass is used for practical reasons
* Typically capture probability and density estimated on a visit by visit basis, issues for uncertainty, especially at low numbers or zero counts.
* Recently the trend is to use HBMs to model capture probability, but these become unwieldy in terms of model selection and fitting when faced with large scale datasets
* Conditional likelihood approaches are a simplification over HBMs but provide huge savings in efficiency and allow for complex models to be applied relatively easily using standard tools in R.
* We present an analysis of capture probability for xxxx sites covering Scotland for the years 1996 to 2014.
* We consider covariates that allow predictions to be made for new sites, such as temporal, spatial and GIS derived, and sampling protocol covariates.
* A simulation study is used to support the conclusions and to investigate the implications of simplifying the modelling assumptions.

# Materials and Methods

* We modelled the capture efficiency of electrofishing in 208 catchments in Scotland between 1980 and 2014. **Figure 1**
* We use catchment groupings based on SEPA hydrometric area (CAN WE DO BETTER) to define spatial strata of similar areas. Also **Figure 1**

## Data collation

* Data was obtained from the SFCC, MSS, SEPA and Caithness DSFB. Again **Figure 1, and Figure 2** - table.
* Because many data sources do not reliably obtain ages from scale reading, electrofishing data was resolved to life-stage (fry or parr) rather than individual age class.
* Location of sample site, date, sampling protocol (see next), and fishing area were retained. Along with auxiliary site specific information such as whether the site was known to be stocked.

## Fish sampling

* A variety of sampling methods were used: with and without stopnets, backpack electrofishing, bankside generators, boat sampling etc. Only Salmon and Trout were sampled/considered.
* In some cases fish were lengthed and scaled; sometimes only salmon parr were sampled for length, while in other cases only numbers of fry and parr by species were recorded. Data was always collected for each pass.
* Electrofishing samples varied between 2 and 6 passes, however the majority of samples had 3 (84%) or 2 (14%) passes.
* [We can talk about spatial variability in sampling procedures here as we have introduced spatial strata in last section.] For MSS data (and more generally?) hard to reach areas are sampled by backpack electrofishing, while easily reached sites, or those near field stations are sampled using more powerful bankside generators.

## Physical site characteristics

* [TABLE THESE? Along with all covariates used] Physical Covariates reported to affect capture probability are: velocoty (Bayley and Dowling, 1990; Price and Peterson 2010), cross sectional area (Price and Peterson, 2010), fish length (Price and peterson, 2010; etc.) Wood density, gradient (Hedger et al, 2013), site width (Hedger et al, 2013), total fish captured (Hedger et al, 2013; pregler 2015), temperature, conductivity, undercut bank (pregler 2015, Rodtka, 2015). Turbidity…
* Many of these do not lend themselves well to predicting capture probability at historic single pass fishing sites, or at new sites where it could be impractical to measure all these covariates for every new site. So we sought to find GIS covariates that could act as proxies to the above features.
* The habitat covariates considered are: altitude, upstream catchment area, distance to sea, gradient, landuse and channel width. See Millar et al. for a detailed description of how these covariates were calculated / derived.

## Capture efficiency modelling

* The model for capture efficiency follows that of Huggins and Yip (1997).
* Various people have extended this model to multiple sites: Conroy et al. (2008), Wyatt, Rivot. However all of these approaches also consider density either as a random effect or in terms of covariates and in order to model capture probability end up with a hierarchical Bayesian model.
* Although HBMs are a very useful tool in ecological modelling (cressie), they can be difficult to extend to large datasets such as the Scotland wide data being analysed here.

## Covariate models

* ‘Linear’ models can be used in a straightforward maner. Linear terms can be lines (ie with a slope and intercept), factor level means, un penalised splines and the related reduced rank GMRFs models which contain among other things regional spatial models.
* If fitting penalised splines then AIC or GCV can be used to estimate the smoothing parameters. It is important when using AIC to estimate the appropriate degrees of freedom of the smoothing terms which reduce as penalisation increases.
* Models are fitted using minimum AIC, which when there are no penalised terms is equivalent to maximum likelihood.

## Simulation testing

* In order to assess the effect of conditioning on site-wise density estimates rather then an optimal approach in which densities are modelled a simulation test was run.
* Simply, one simulation is set up in which the density model is known and is of a simple form. The conditional model is fitted then the full model (with correct density model) and the results compared.
* An alternative simulation in which densities vary independently following a uniform distribution is also run to test the situation where the conditional model would be appropriate.

# Results

## Covariates

* Data was collated for 2749 discrete sites and 6049 site visits.
* Table all considered covariates, 3 headings: name; data provenance; proxy for.
* Discuss potential correlations
* Plot marginal scatter plots

## Capture efficiency modelling

* The best working model / best approximating model [I like these terms] was: [and state model]
* List effects in order of importance
* Plot effects and describe them

## Simulation testing

* Plot box-plots for results of simulations
* Describe what is there

# Discussion

## Covariates

* Discuss how GIS was at generating proxies for the reported influencers on catchability
* Issues with spatial confounding and equipment use
* Can we model river width? And other covariates? What about better landuse metrics?

## Capture efficiency modelling

* Iss
* Spatial terms can be thought of as capture unmodeled variation – advisable to attempt to find covariates that can describe the likely cause of the variation – ie average autumn temperature? Or some other integrated water temperature metric.
* If unpenalised splines not used, then

## Simulation testing

* Quick paragraph on when the conditional model is likely to be better or worse. Presumably when there is lots of structure in the density model, then the conditional model will be less good.
* Potentially one could run a range of model selection procedures conditional on different density models, and choose the best one … by some criteria.

## General discussion/recommendations

* Tie back to introduction: augment single pass fishing with p estimates. This brings in an extra xxx observations.
* Although the model can be used to estimate p for new sites, it is still advisable to maintain a level of multipass fishing to validate and improve the p model for Scotland.

# References

# Tables

# Figures