**Efficient Estimation of Electrofishing Capture Efficiency from Large Datasets.**

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.
* Estimating abundance from electrofishing data requires a reliable estimate of capture probability.
* Often capture probability is estimated using multi-pass electrofishing, although there are other possibilities.
* Hierarchical Bayesian models have been developed to model capture probability (and jointly fish density) but these approaches do not scale well to larger datasets.
* Consequently, the use of these models appears to be restricted to smaller case studies and recent development in capture probability modelling do not seem to have been taken up by the applied fisheries community.
* Single pass electrofishing is widely used as it allows for a greater number of sites to be covered for a given sampling effort.
* Modelling capture probability allows the appropriate use of single pass electrofishing data in quantitative analyses, which would otherwise be based on assumptions of constant capture probability.
* 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
* The use of conditional likelihood provides a simple way to investigate complex relationships.
* Capture probability has been reported to vary with habitat, time of sampling and sampling procedure, including personnel and equipment. We consider proxies for these processes based on covariates that require minimal site specific habitat data by using site location to derive map based covariates.
* 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.
* Sampling effects were the most important, followed by Fish characteristics (age class, species) and sampling date. Physical characteristics such as width and gradient affected capture probability to a lesser extent.
* A simulation study was conducted to show the effect of using a simple conditional likelihood approach over a full joint model for capture probability and density.
* The results reinforce the dangers of assuming constant capture probability when interpreting spatial trends in fish density derived from single pass electrofishing data.
* Furthermore, we show how our model improves density estimates from single pass electrofishing data.

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 habitat, sampling procedure and physical characteristics that can vary spatially and temporally.
* This has negative implications for the use of single pass electrofishing (widely used) for quantitative advice.
* Capture probability estimation requires multipass fishing.
* Often single pass is used for practical reasons.
* When capture probability and density are 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.
* The use of a model for capture probability would allow the incorporation of single pass fishing data (in conjunction with sufficiently representative multipass data) into quantitative advice.
* We present an analysis of capture probability for xxxx sites covering Scotland for the years 1996 to 2014.
* We consider covariates that require minimal site data collection to make the most use of available data and allow predictions to be made for new sites.
* Physical Covariates reported to affect capture probability are: velocity (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 also…
* 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
* A simulation study is used to support the conclusions and to investigate the implications of simplifying the modelling assumptions.
* In addition, because it is still common to ignore the effects of variable capture probability we also use the simulation study to show the effects of assuming it constant.

# Materials and Methods

## Study site

* We modelled the capture efficiency of electrofishing in 208 catchments in Scotland between 1980 and 2014. **Figure 1**
* Scotland consists of XX catchments greater than 500 m2 draining a total of xxx m2.
* There is spatial variation in catchment characteristics e.g., west coast of Scotland is characterised by many catchments, while the east coast comprised several large catchments.
* Reference NASCO FAR.

## 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, and fishing area were retained.

## Fish sampling

* A variety of sampling methods were used across the organisations who supplied data: with and without stopnets, backpack electrofishing, bankside generators, boat sampling etc. IMPORTANT FOR DISCUSSION.
* 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. IMPORTANT FOR DISCUSSION.
* We assume that the sampling protocols are consistent within organisation, although this is not strictly true it does allow for some aspect of sampling procedure to be included in modelling.
* Electrofishing samples varied between 2 and 6 passes, however the majority of samples had 3 (84%) or 2 (14%) passes.

## Physical site characteristics

* 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.
* Table describing what variables are supposed to be proxies of.

## Spatial information

* It is often the case that spatial terms in models represent unmodelled variation that is not captured by the chosen covariates.
* The use of habitat and sampling covariates should allow a complete model of capture probability; however as mention there are aspects of sampling and habitat that could not be included that are likely to be spatial in nature.
* To account for this additional variability two possibilities were considered: a smooth process over latitude and longitude, and a regional smoother based on groups of catchments. We use catchment groupings based on SEPA hydrometric area (CAN WE DO BETTER) to define spatial strata of roughly similar areas.

## 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 manner. Linear terms can be lines (ie with a slope and intercept), factor level means, unpenalised 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 than 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.
* Discuss potential correlations
* Plot marginal scatter plots

## Capture efficiency modelling

* The best working model / best approximating model 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

## General discussion/recommendations

* We present an approach for estimating and modelling electrofishing capture probability for large datasets that provides a balance of complexity and computational efficiency.
* An issue identified is the lack of uptake of recent developments for modelling capture probability into the applied fisheries community. The software developed for the present paper is freely available for use in the popular R environment for statistical computing.
* Tie back to introduction: augment single pass fishing with p estimates. This brings in extra observations. Looking forward consider value of approach for increasing spatial coverage for limited costs with 1 pass EF - also a potential for 1 pass EF use in catchment wide stock assessments.
* 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.
* Management consequences when assuming constant capture prob / site by site capture prob.

## 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

* 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 attempt to develop penalised likelihood approaches.

## 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.

# References

# Tables

# Figures