

Transportation of egg requirements

1 Conservation limits

In line with recommendations from NASCO, conservation limits of a given stock are set based on the estimated maximum sustainable yield (MSY) of that stock. This biological reference point is the spawning stock size that maximises the mean annual catch. It is determined by modelling the relationship between spawning stock and recruitment (the SR relationship) over multiple years, and is the value that maximises the difference between the SR relationship and the replacement line (Figure 1).

Previous analyses have demonstrated that there is considerable river to river variation in MSY across European populations, specifically related to latitude (White et al. 2016; Prévost et al. 2003). Figure 2 shows a preliminary analysis of 11 Scottish rivers and sub-catchments with point estimates for stock at MSY derived from independently fit Ricker curves, plotted against latitude. There is substantial variation from river to river within Scotland, consequently a “one size fits all” model may not be appropriate. However, defining appropriate conservation limits for rivers without data is non-trivial, motivating the development of a model that can explain the variation in MSY between Scottish rivers and predict the MSY of unseen rivers given some explanatory covariates.

2 Bayesian heirarchical modelling framework

Whilst the stock-recruitment relationships vary from river to river, it is assumed that there may exist a degree of similarity among Scottish rivers. The use of a Bayesian hierarchical modelling framework can account for this structure. In this framework parameters at the level of the river can be modelled to be drawn from a national level hyper-distribution. This not only allows the between river variation to be quantified, but also allows data from each river to inform the SR relationships of other rivers. This is particularly useful for helping to improve parameter estimation for data-poor rivers, by borrowing information from data-rich rivers. Finally, the Bayesian modelling framework provides a more informative process for modelling uncertainty than traditional statistical modelling techniques.

To model the SR relationships of available rivers, the widely used Ricker relationship is adopted:

$$R = \frac{e^{h^*}}{1 - h^*} S e^{-\frac{h^*}{S^*} S}$$

where S is the spawning stock and R is the recruits. Here the classical model is reparameterised in terms of the stock and harvest rate at MSY, S^* and h^* , following Schnute and Kronlund (1996) (Figure 1).

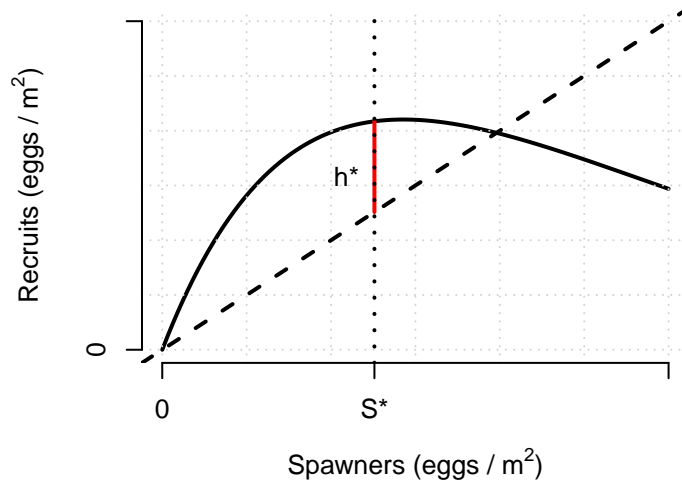


Figure 1: An example SR relationship modelled as a Ricker curve (solid black) with the replacement line $R=S$ (dashed), stock at MSY S^* (dotted) and harvest at MSY h^* (red).

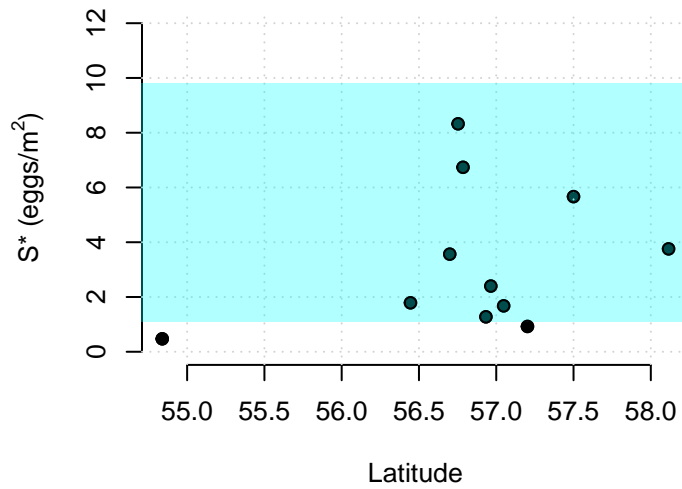


Figure 2: Preliminary point estimates of stock at MSY for 11 Scottish rivers and sub-catchments against latitude, derived from independent Ricker models. Existing egg requirement in blue.

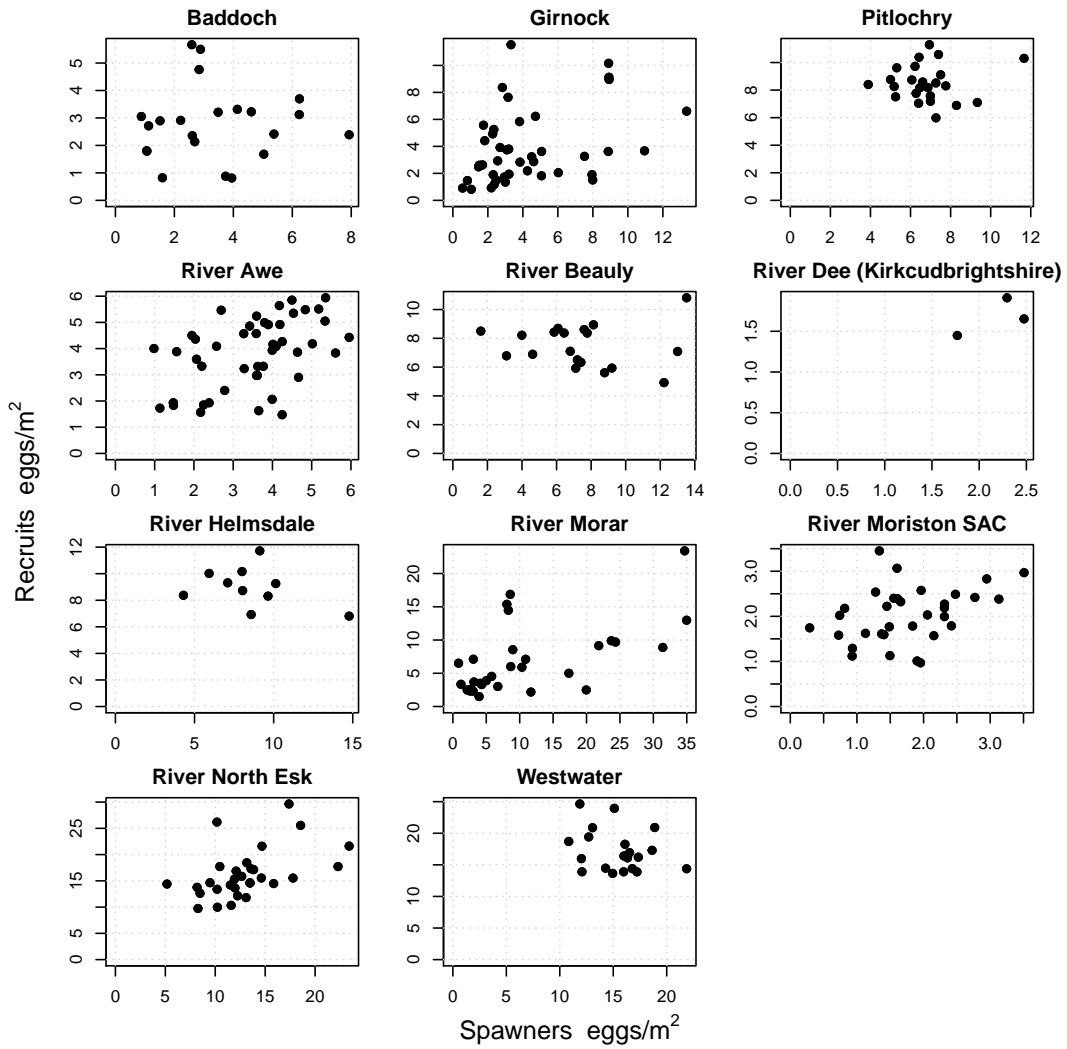


Figure 3: Stock recruitment data for the 11 rivers and sub-catchments used in model fitting. Note that x and y axis differ between rivers.

2.1 Transportation of egg requirements with covariates

Adult to adult data on the abundance of stock and recruitment is available for only 11 monitored stocks around Scotland, consisting of 6 rivers and 5 sub-catchments (Figure 3). Consequently, modelling is required to transport conservation limits from stocks with data to those without. By including environmental and geographic covariates available for all rivers in the modelling process, conservation limits for non-monitored rivers without data may be predicted.

Covariates available for all rivers:

- a) Latitude
 - Latitude of trap/counter.
- b) Height
 - Information on the heights of the land present in each of the catchments was taken from the Ordnance Survey Terrain 50 data set (<https://data.gov.uk/dataset/0a318a38-84c1-4509-be79-bc80f71a8aad/os-terrain-50-dtm>). This contains heights for points in a 50m DTM grid. For each catchment the proportion of such points within a 100m band was estimated. Principle components analysis was used to reduce this banding into a single explanatory variable which explained 61% of the variation in the data set.
- c) Catch per area (CPA)
 - Historical information on the catches of salmon in each of the districts in Scotland was used to potentially provide information on geographic changes in the relative productive capacity of rivers across Scotland. Catches were available for each district for the period 1952-2016. Catch per area was defined as the 80th percentile of the time series divided by the area of salmon habitat.
- d) Land usage
 - Information on land cover was taken from the 2015 Land Cover Map produced by CEH (https://www.ceh.ac.uk/sites/default/files/LCM2015_Dataset_Documentation_22May2017.pdf). Land cover was broken into the proportion of 10 different land use types in each 1km grid square covering the UK. This information was used to estimate the proportion of each land use type in each of the conservation regulations assessment areas. Principle components analysis was used to reduce the 10 different variables into 1, which explained 60% of the variation in the data set.
- e) Ratio of lacustrine to fluvial habitat (Loch)
 - The ratio of still to flowing water in each catchment.
- f) Distance around coast (DAC)
 - This metric was included based on the assumption that close by rivers may be more likely to have similar values of S^* than those at a distance. In order to incorporate this information a simple metric was constructed by using latitude. For rivers on the East Coast the metric was simply the latitude minus 54.5 degrees. After the River Thurso the metric became the East coast distance ($58.62 - 54.5$) added to the difference between the most northerly latitude (58.62) and the latitude of the river mouth in question.

With the exception of the distance around coast, these covariates are included in the model through the introduction of a linear relationship between both S^* and h^* and the value of

the covariate of interest (x) for a given river. Following Prévost et al. (2003) and White et al. (2016), S^* is modeled to have a log-normal distribution and h^* , being bounded between 0 and 1, has a beta distribution. Specifically, for river i with covariate x_i , S^* and h^* are given by:

$$S_i^* \sim \text{lognormal}(\alpha_S + \beta_S x_i, \sigma)$$

and

$$\begin{aligned} \mu_i &= \text{logit}^{-1}(\alpha_h + \beta_h x_i) \\ h_i^* &\sim \text{beta}(\mu_i \phi, (1 - \mu_i) \phi) \end{aligned}$$

where σ is a variance parameter for S^* and ϕ is a precision parameter for h^* and the parameters $\alpha_S, \alpha_h, \beta_S, \beta_h$ define the linear relationship for that covariate.

The distance of the site around the coast was incorporated as a Bayesian P-spline smoother term (Lang and Brezger 2004) allowing for a non-linear spatial relationship between estimates of S^* and h^* . The smoother term was added to the model in a similar linear way to the other covariates.

In addition, the wetted area for each river was implicitly included in the model by converting recruits and spawners to egg densities.

2.2 Assessing the predictive value of covariates

To evaluate whether a particular covariate is informative about the SR relationship for those rivers with no data, an assessment of the predictive accuracy of each covariate was made. This was performed using a leave-one-group-out cross-validation process, whereby, for each covariate combination, a model was fit to data from 10 of the 11 monitored rivers and sub-catchments, with the predictions of the resultant model compared to the SR relationship of the excluded river. This process can be summarised in the following steps:

1. Remove one river from the data set
2. Fit the hierarchical model to the remaining rivers
3. Use the fitted model to predict the data of the held out river using the covariate relationships defined above
4. Quantitatively compare these predictions to the actual data for that river
5. Repeat, excluding a different river until all rivers have been excluded
6. Determine a cross validation score for the model via a weighted sum of the likelihood of unseen data for each unseen river, taking into account the amount of data available.

The cross validation score for each model quantifies how well that model predicts unseen rivers and, though it has no meaningful absolute value, can be compared to other models fit using different covariates.

Compared to the traditional approach of looking for covariates with significant predictors, this method of assessing models based on their ability to predict unseen rivers is preferable as it replicates the intended use of the model and is much less prone to over-fitting to the available data.

2.3 Models

A suite of models were run and compared using the above method. The models tested consisted of a ‘null’ model with no covariates, a model for each of the covariates (a-f) individually, and a model for all possible pairs of covariates (a-f). The null model here represents a scenario of a “one size fits all” S^* , applying the same egg requirement to all rivers.

3 Results

Models had improved predictive performance over the null model only when either distance around the coast or catch per area were included as a covariate (Table 1). Furthermore, only the inclusion of catch per area with distance around the coast gave better predictive performance than distance around the coast alone. Whilst there is a strong correlation between these two covariates, the fact that including both in a model leads to improved predictive performance over either in isolation suggests that both covariates provide unique predictive contributions. In particular, the spline for distance around the coast allows for variable degrees of certainty in model predictions around the coast, which is not possible with the inclusion of a linear predictor for catch per area.

The poor predictive performance achieved by the use of other covariates suggests either a lack of a relationship or a degree of over-fitting, generating poor predictions for out of sample rivers. This includes latitude, which has previously been used to transport egg requirements at larger spatial scales (Prévost et al. 2003; White et al. 2016).

Figure 4 shows the results of the best predictive model with both the distance around the coast as a smooth P-spline and catch per area as a linear covariate, fit on the entire data set. The plots give the distribution of the predicted S^* for rivers in four scenarios; south east and south west coast with high and low catch per area.

There is a positive relationship between catch per area and predicted S^* . and a slight decline in the median predicted S^* for rivers in the south west compared to the south east. The uncertainty in S^* generally exceeds the range of the previous conservation limit. However, the results from the model provide more informative estimates than the previous uniform conservation limit, returning non-uniform distribution, with greater probability density around the median value.

Table 1: The predictive performance (CV Score) and performance relative to the null model (Null Δ CV) for all models tested. Covariates indicates covariates (a-f) included as predictors in the model, DAC = distance around coast (smoother), CPA = catch per area, Loch = ratio of lacustrine to fluvial habitat.

Covariates	CV Score	Null Δ CV
DAC, CPA	-24,998	555
DAC	-25,153	399
DAC, Latitude	-25,244	309
DAC, Height	-25,278	275
CPA	-25,351	202
DAC, Land usage	-25,503	50
	-25,553	-
Height, CPA	-25,620	-67
DAC, Loch	-25,720	-167
CPA, Land usage	-25,775	-222
Latitude	-25,853	-301
Latitude, CPA	-25,929	-376
Land usage	-26,063	-510
Latitude, Land usage	-26,096	-543
CPA, Loch	-26,385	-832
Height, Latitude	-26,433	-880
Height	-26,555	-1,002
Height, Land usage	-26,565	-1,012
Latitude, Loch	-26,764	-1,211
Loch	-26,778	-1,225
Loch, Land usage	-27,099	-1,547
Height, Loch	-27,711	-2,159

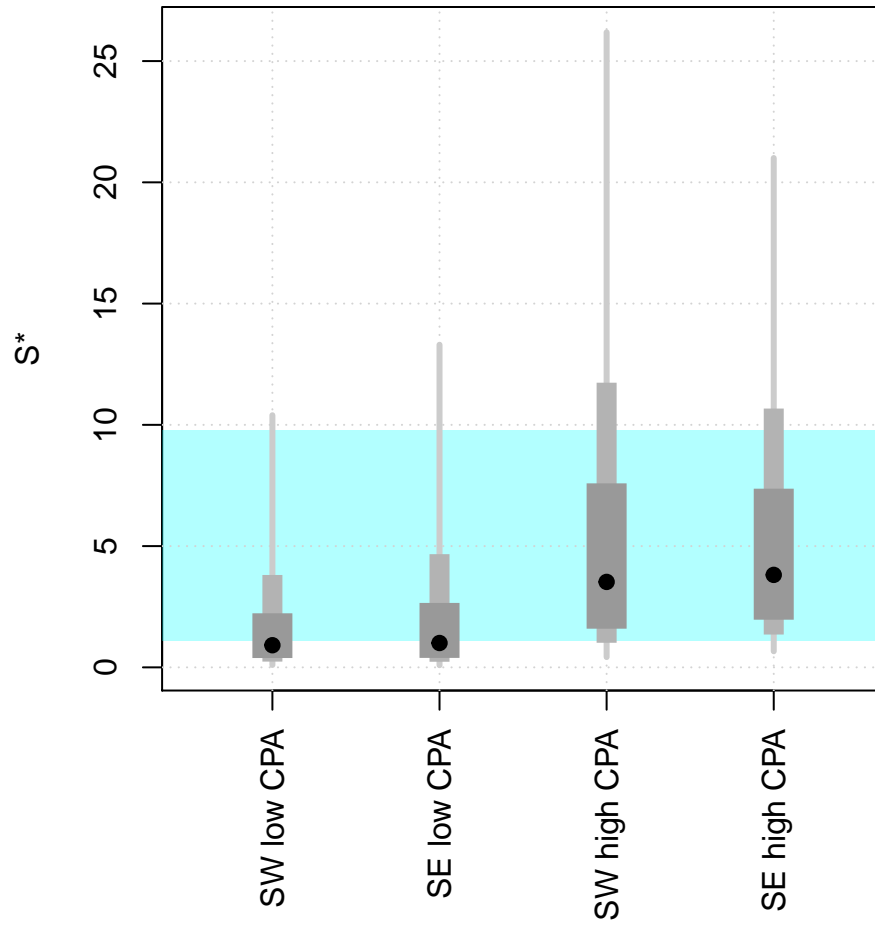


Figure 4: Median (filled circle), 50%, 70% and 90% prediction intervals (wide and dark to narrow and light regions) for S^* from the model including CPA and a distance around coast smoother for rivers with low and high CPA and on the SE coast and SW coast. Existing egg requirements shown in blue.

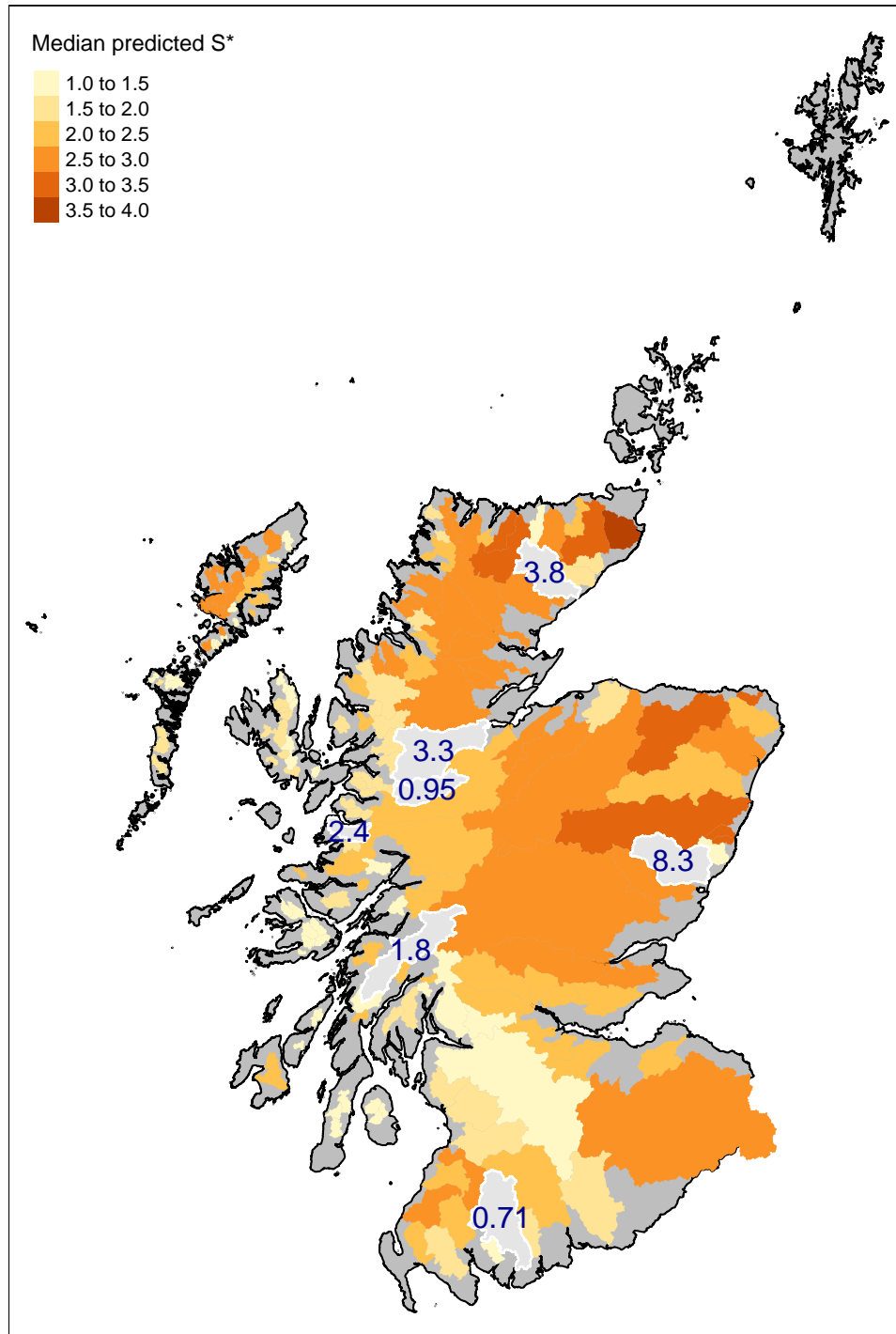


Figure 5: Median predicted egg requirements (S^*) for rivers without data from the model with both distance around the coast and catch per area as covariates mapped to assessment region. Numbers give median estimated values for rivers with data (including Moriston sub-catchment, other sub-catchments omitted).

References

- Lang, Stefan, and Andreas Brezger. 2004. *Bayesian P-Splines*. Vol. 13. 1. doi:10.1198/1061860043010.
- Prévost, Etienne, Eric Parent, Walter Crozier, Ian Davidson, Jacques Dumas, Gudni Gudbergsson, Kjetil Hindar, Phil McGinnity, Julian MacLean, and Leif M Sættem. 2003. “Setting biological reference points for Atlantic salmon stocks: Transfer of information from data-rich to sparse-data situations by Bayesian hierarchical modelling.” *ICES Journal of Marine Science* 60 (6): 1177–93. doi:10.1016/j.icesjms.2003.08.001.
- Schnute, J T, and A R Kronlund. 1996. “A management oriented approach to stock recruitment analysis.” *Canadian Journal of Fisheries and Aquatic Sciences* 53 (6): 1281–93. doi:10.1139/f96-069.
- White, Jonathan, Niall Ó Maoiléidigh, Paddy Gargan, Elvira de Eyto, Gerald Chaput, Willie Roche, Phil McGinnity, et al. 2016. “Incorporating natural variability in biological reference points and population dynamics into management of Atlantic salmon (*Salmo salar* L.) stocks returning to home waters.” *ICES Journal of Marine Science: Journal Du Conseil* 73 (6): 1513–24. doi:10.1093/icesjms/fsw015.