Estimating Abundance of Adult Salmon

May 2018

1 Introduction

Assessment of the conservation status of salmon requires knowledge of the number of salmon present in a given stock and the number required to meet the conservation objective. Here we present an updated analysis which uses information on the relationship between catch and counts for 8 sites in Scotland to produce a general method for estimating salmon numbers from reported catches.

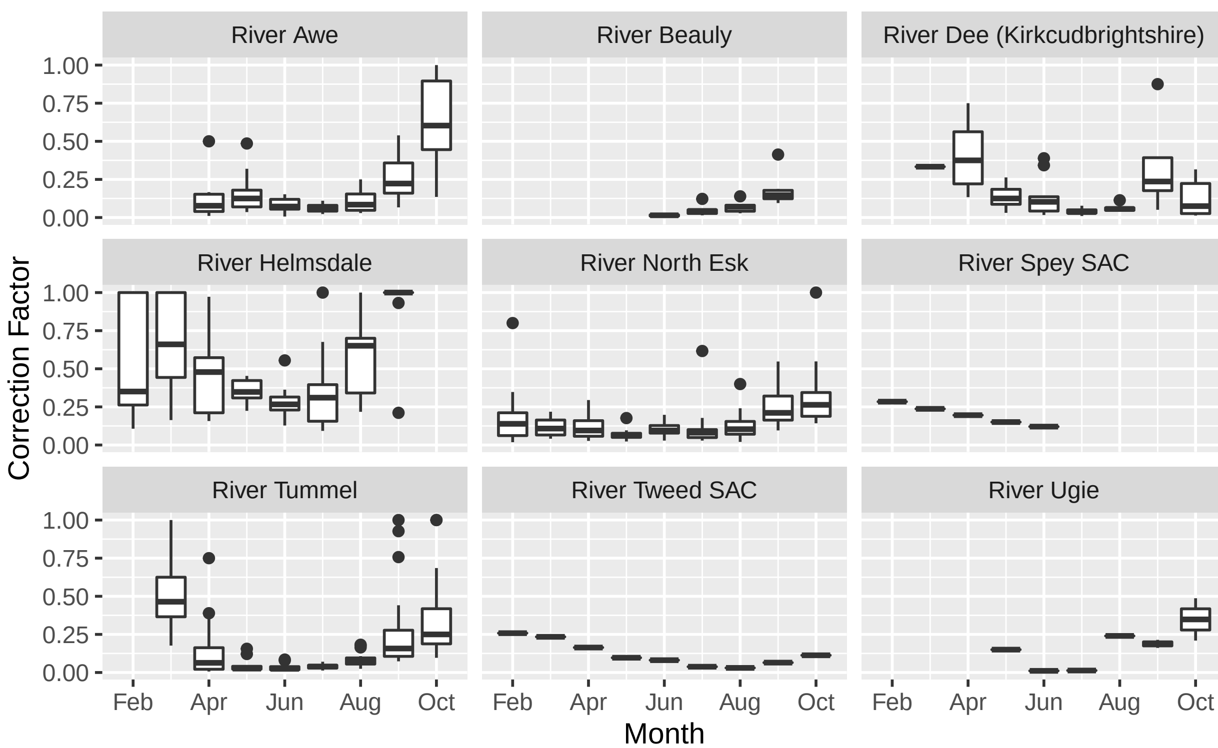

Figure 1: Boxplot of raw data used in the analysis showing monthly correction factors for the 8 sites.

2 Data

Information was available for 9 rivers in Scotland - the Awe, Beauly, Dee (Kirkcudbrightshire), Helmsdale, North Esk, Tummel, Tweed and Ugie (Figure 1). For all except the Tweed and Spey this was in the form of monthly counter from fish counters and associated rod catches from areas above the counters. Rather than directly estimating counts the model instead examined the relationship between suggested predictor variables (see below) and a correction factor (CF) which is a number bounded between 0 and 1 which can be used to convert catches to counts [1]. The CF is the catch expressed as a proportion of the count for each month (or set to 1 where catch is greater than count, ~2% of observation primarily in October). For the Tweed and Spey information was available on the recapture of salmon caught in nets at the bottom of the river, floy or radio tagged, released and recaptured by rod fishers. Monthly correction factors for each site are shown in Figure 1. There is a clear monthly trend in the correction factors. Therefore a factor for month was included in all of the models examined.

The following covariates were also included in the modelling process:

2.1 Flow rate

Deviation from the median monthly flow at each of the sites. Flow rates taken from SEPA gauging stations.

2.2 Land cover

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 available 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.

The geographic spread of the different types of land cover is in Figure 2.

2.3 Latitude

Previous work had shown a significant relationship between latitude and the correction factor.

2.4 Catch profile

Salmon entering throughout the year may become catchable again after September. This may impact on the estimated CF with sites with greater numbers of pre-September fish potentially having a higher cf than those with few fish entering the river before September. While the count data could

[1] count = catch/CF

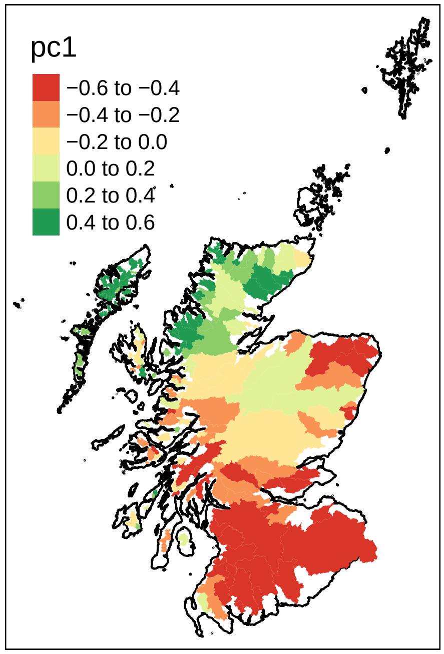

Figure 2: Map showing the geographic pattern shown by the land cover principle components analysis. Each assessment area is coloured according to the value of the first principle component.

be used to categorise rivers a method is required that can be applied to sites without counter data. Therefore, the patterns shown by catch data between March and February were examined and were used to allocate rivers to one of two groups, those with catches spread throughout the year ("spread") and those where catches indicate most salmon being caught at the end of the season ("late").

Data from 10 rivers with known differences in seasonal catches were used as the input data in a linear discriminant analysis. The pattern of monthly catches between March and September, expressed as the proportion of the total catch within this period, was calculated for the period 2011 to 2017 for the following sites:

- Rivers with late catches
  - Alness River, River Bladnoch, River Deveron, River Don, River Earn.
- Rivers with catches spread throughout the year
  - River Awe, River Dee (Aberdeenshire), River Helmsdale, River North Esk, River Spey

There are clear differences between the two groups of sites and this data was used as the training data to produce the model used for classification in the discriminant analysis (Figure 3). Monthly catch information associated with the counter sites, and the River Tweed, was then used to predict which catch profile group they belonged to. The River Awe, River Helmsdale, River North Esk, River Spey and River Tummel sites belonged to the spread group, with the River Beauly, River Dee, River Tweed and River Ugie belonging to the late catch profile group.

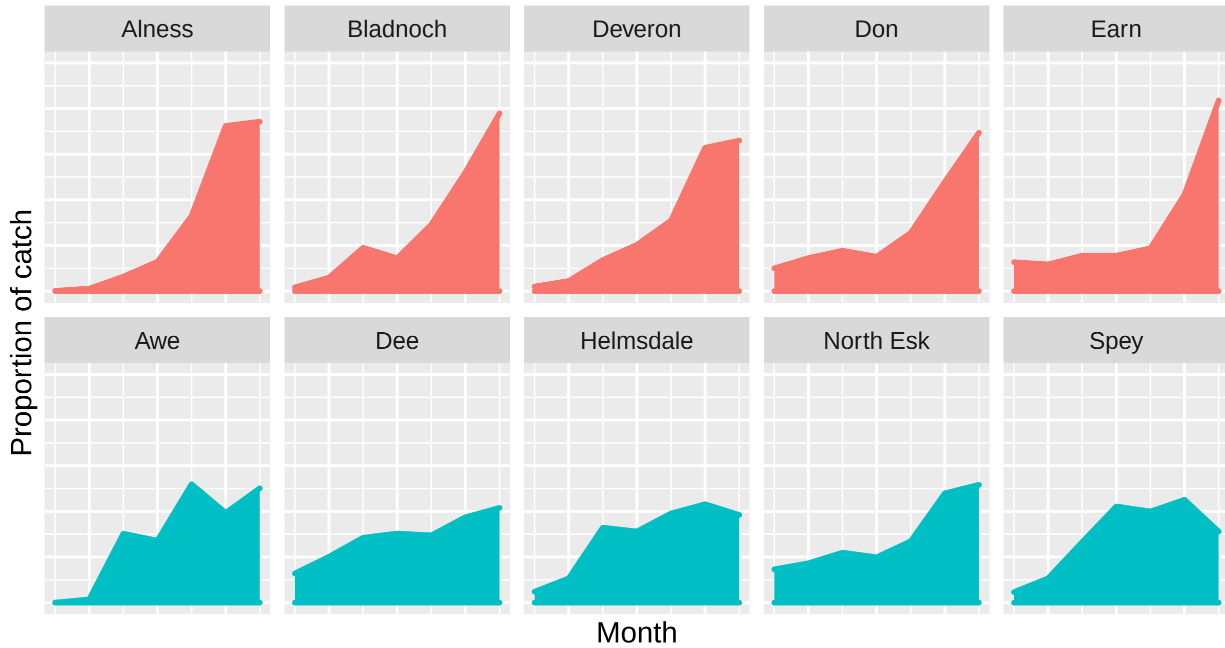

Figure 3: Monthly catch expressed as a proportion of toal Feb-Sep catch (2011-2017) for the two catch timing groups: red = late, blue = spread.

3 Analysis

Models were fitted using generalised linear mixed models with binomial errors as this allowed the incorporation of fixed and random effects. Models were fitted using the lme4 package in R. All models included the following random effects: site, site:year and an identity random effect to account for overdispersion in the data. With the exception of the null model, which contained no fixed effects, all other models contained a smoother for month as a fixed effect (natural cubic spline from the spines R package, version 3.5.1). The suite of candidate models were produced by adding the four additional fixed effects (flow rate, land cover, latitude and run timing) both singly and in pairs.

3.1 Model Selection

The aim of the analysis was to produce a model that would be used to predict CF for new sites. Therefore a cross-validation approach was taken to evaluate candidate models rather than examining predictor variables using significance tests or using information criteria to compare models. This process entails constructing a model using data from 7 out of the 8 sites (the training set). This model was used to estimate the cf value of the 8th site (hold out set). The fit of the model to the hold out set was assessed using a weighted sum of the likelihood of unseen data, taking into account the number of data points in the hold out set. This process was repeated 7 times so that all sites were left out once. The predictive power of the models were compared by summing the weighted likelihood over all the sites.

4 Results

4.1 Cross-Validation

The results of the cross-validation evaluation are given in Table 1. The models containing a combination of month, flow latitude and catch profile performing better than the null model, with the best model containing month and flow.

Table 1: The predictive performance (CV Score) and performance relative to the null model (Null Δ CV) for all models tested.

Covariates	CV Score	Null Δ CV
month, flow	-600.6	45.6
month, flow, catch profile	-604.5	41.7
month	-617.6	28.6
month, flow, latitude	-621.6	24.6
month, catch profile	-626.6	19.6
month, latitude	-642.0	4.2
null	-646.2	-
month, latitude, catch profile	-653.7	-7.5
month, flow, landcover	-655.6	-9.4
month, landcover	-680.9	-34.7
month, landcover, latitude	-725.6	-79.4
month, landcover, catch profile	-731.9	-85.7

4.2 AICc

Results of a comparison of small sample AIC (AICc) scores for the candidate models is presented in Table 2. In contrast with the cross-validation approach all models perform better than the null model. However, the poor performance of many of these models in the cross validation analysis suggests over-fitting of the covariates generating poor predictions for out of sample rivers.

Table 2: Comparison of AICc scores for the candidate models. Change from the best model (Δ) and AICc weights (ω) are presented for all models tested.

Covariates	AICc	Δ	df	ω
month, flow, landcover	5,457.7	-	8	0.431
month, flow, catch profile	5,458.5	0.7	8	0.299
month, flow, latitude	5,459.6	1.8	8	0.172
month, flow	5,460.7	3.0	7	0.098
month, landcover, catch profile	5,472.2	14.5	8	<0.001
month, latitude, catch profile	5,473.3	15.6	8	<0.001
month, catch profile	5,474.5	16.7	7	<0.001

Covariates	AICc	Δ	df	ω
month, landcover	5,475.6	17.8	7	<0.001
month, latitude	5,477.3	19.5	7	<0.001
month, landcover, latitude	5,477.5	19.8	8	<0.001
month	5,478.6	20.9	6	<0.001
null	5,697.1	239.4	4	<0.001

4.3 *Final Model*

The final model chosen was the model with month and flow as the cross-validation analysis indicated that this was the best performing model when estimating for new sites. The monthly pattern was such that the correction factor was higher during the early and late months (Figure 4). The correction factor also increases with flow rate, suggesting that, all other things being equal, salmon may be less catchable during low flow conditions (Figure 5).

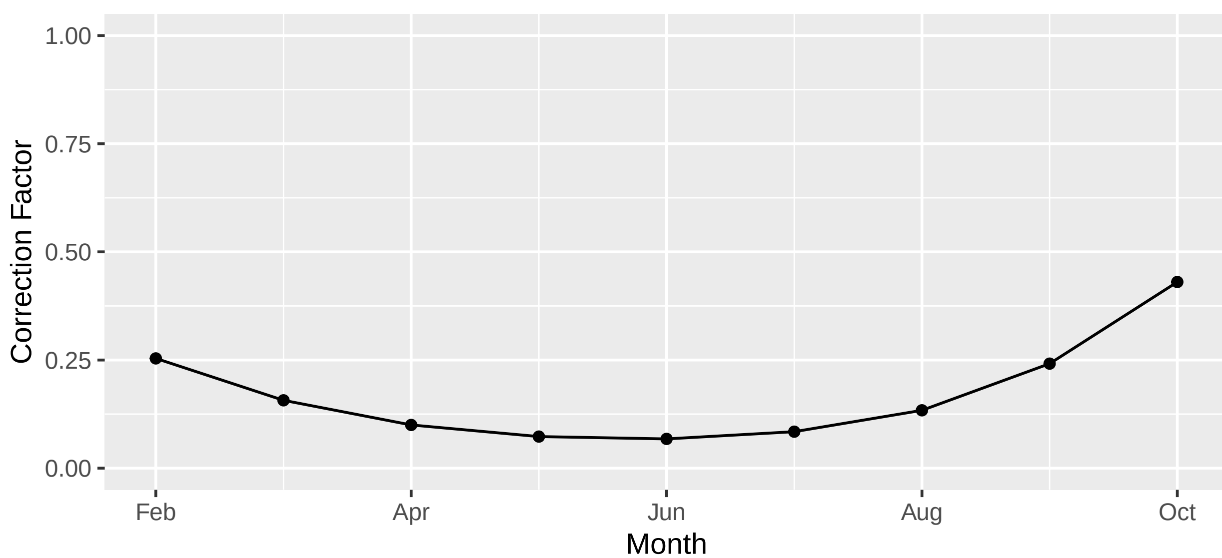

Figure 4: Monthly correction factors estimated from the final model. CF estimated under median monthly flows.

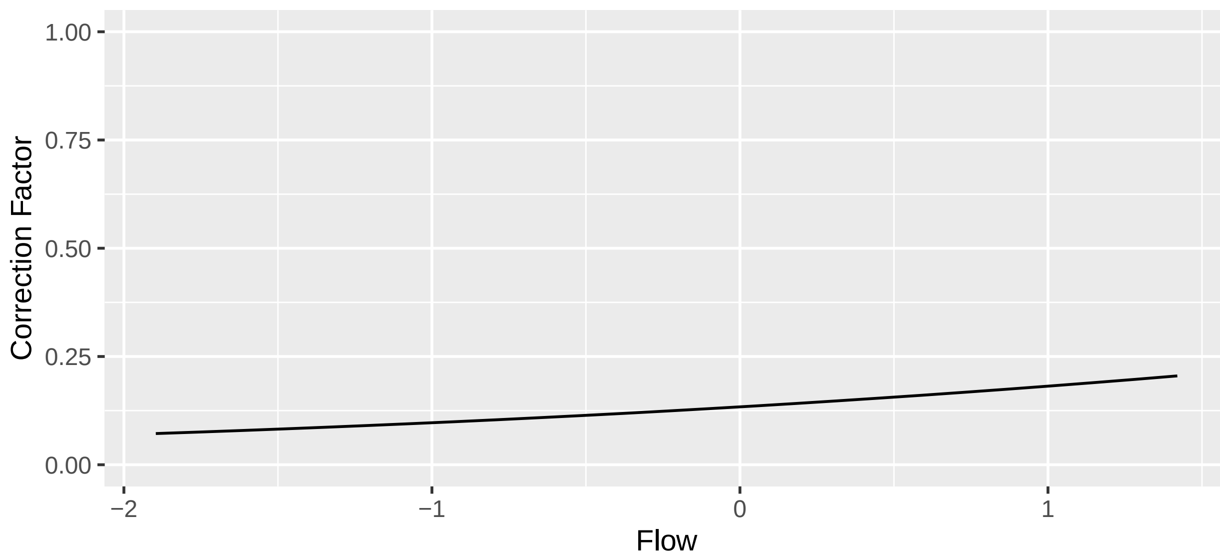

Figure 5: Relationship between flow and CF estimated from the final model. CF estimated for August. Medain flow is set at 0, with negative flows indicating lower flows, and positive values higher flows.

The model output is summarised below:

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: prop ~ ns(month, knots = KNOTS) + flow.adj + (1 | site) + (1 |
##     ident) + (1 | site:year) + (1 | site:month)
##    Data: dat
## Weights: dat$count
## Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e+05))
##
##      AIC      BIC   logLik deviance df.resid
##   5393.6   5428.4  -2688.8   5377.6      562
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.04604 -0.09393 -0.00162  0.08177  2.32428
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  ident      (Intercept) 0.8533   0.9238
##  site:year  (Intercept) 0.0849   0.2914
##  site:month (Intercept) 0.4458   0.6677
##  site       (Intercept) 0.7430   0.8620
## Number of obs: 570, groups:  ident, 570; site:year, 81; site:month, 65; site, 9
```

```
##
## Fixed effects:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -1.07881    0.46080  -2.341   0.0192 *
## ns(month, knots = KNOTS)1 -1.61695    0.74357  -2.175   0.0297 *
## ns(month, knots = KNOTS)2  2.87207    0.42531   6.753 1.45e-11 ***
## flow.adj                  0.36232    0.07658   4.731 2.23e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) n(,k=KNOTS)1 n(,k=KNOTS)2
## n(,k=KNOTS)1 -0.732
## n(,k=KNOTS)2 -0.084  0.113
## flow.adj     -0.005  0.018        0.033
```

5 Returns Outwith Angling Season

The method detailed above (using catches to estimate the number of salmon entering rivers each month) does not allow the number of salmon entering outwith the angling season to be estimated. To do so, the relationship between counts in either the first or last month of angling season and months outside the season was examined for 6 of the counter sites (The Tummel being excluded due to its distance from the sea). There was no evidence that these relationships differed between sites of different catch profiles (Figures 6 & 7).

For each month during the post-season (up to end December) estimated stock was therefore derived using information estimated using data from all counter sites. For each month the stock level was expressed as a proportion of the stock in the last month of the new rivers fishing season . To estimate stock levels this proportion was applied to estimated stock in the last month of the fishing season for the new river. *e.g. for rivers where the fishing finishes at the end of October and the proportion of November to October counts was 0.75 a stock estimate of 100 fish in October would produce a stock estimate of 75 fish in November (100 x 0.75).*

Pre-season estimates (January onward) were undertaken using a similar procedure but using stock estimates from the first full month of the fishing season rather than the last.

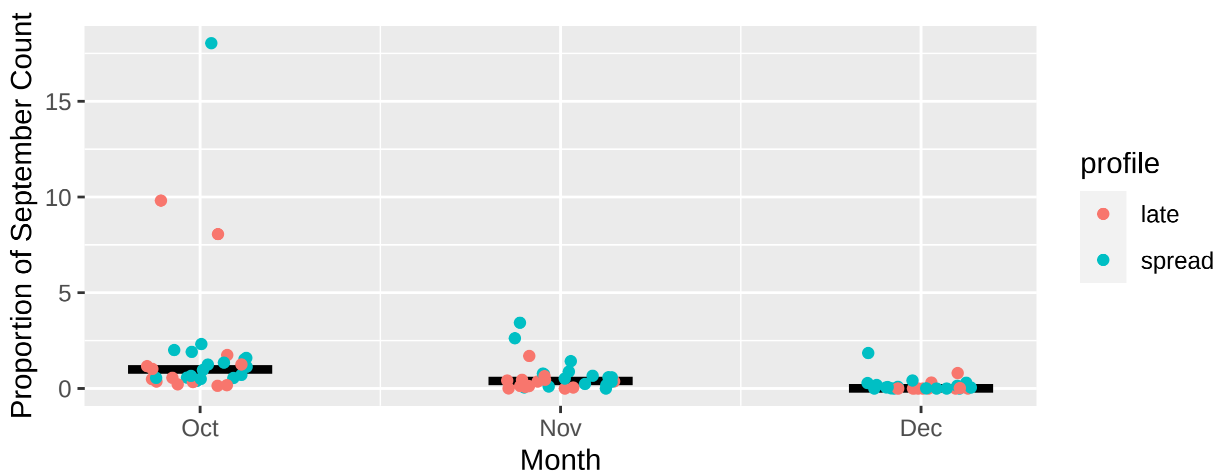

Figure 6: Graph showing counts in October-December expressed as a proportion of the September count. Dots are coloured according to the run profile of the site and horizontal lines show the median monthly values.

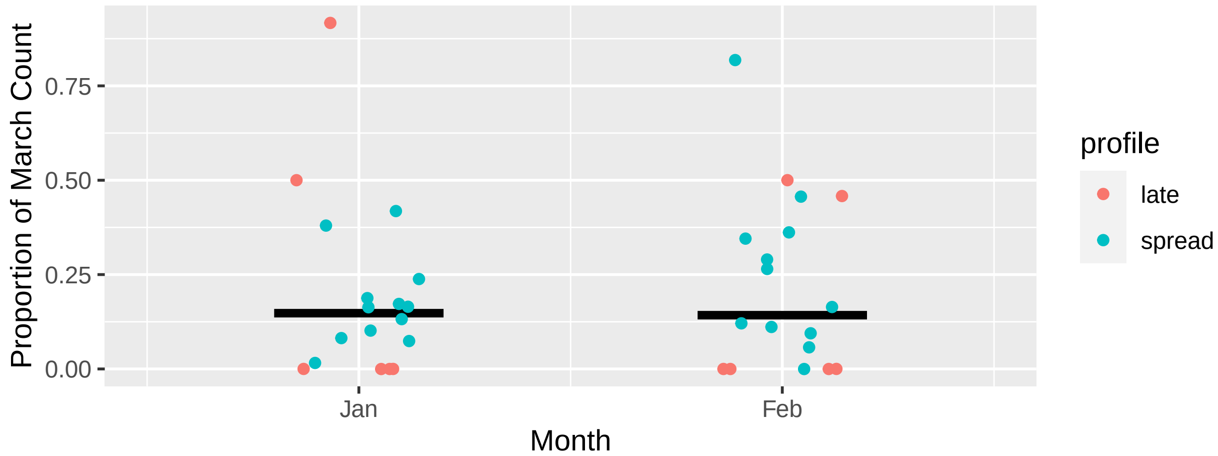

Figure 7: Graph showing counts in January and February expressed as a proportion of the March count. Dots are coloured according to the run profile of the site and horizontal lines show the median monthly values.

6 Worked Example:River Don

Here we provide an example of the estimation method applied to catches from the River Don using data from 2013-2016. The monthly catches are shown in Figure 8, highlighting a general pattern with a small peak of catches in April/May, low catches during the summer, with catches increasing again towards October.

Calculation of uncertainty around abundance estimates uses a bootstrapping approach, where monthly stock estimates are made 10,000 times with each estimate using monthly CFs randomly chosen to incorporate the uncertainty around the estimates from the CF model. CFs are estimated as the average CF for an unknown river. Out of season estimates are calculated as detailed above. The outputs from the bootstrap used to estimate monthly stock levels are presented in Figure 10. These monthly estimates can be. combined to form annual estimates (Figure 11).

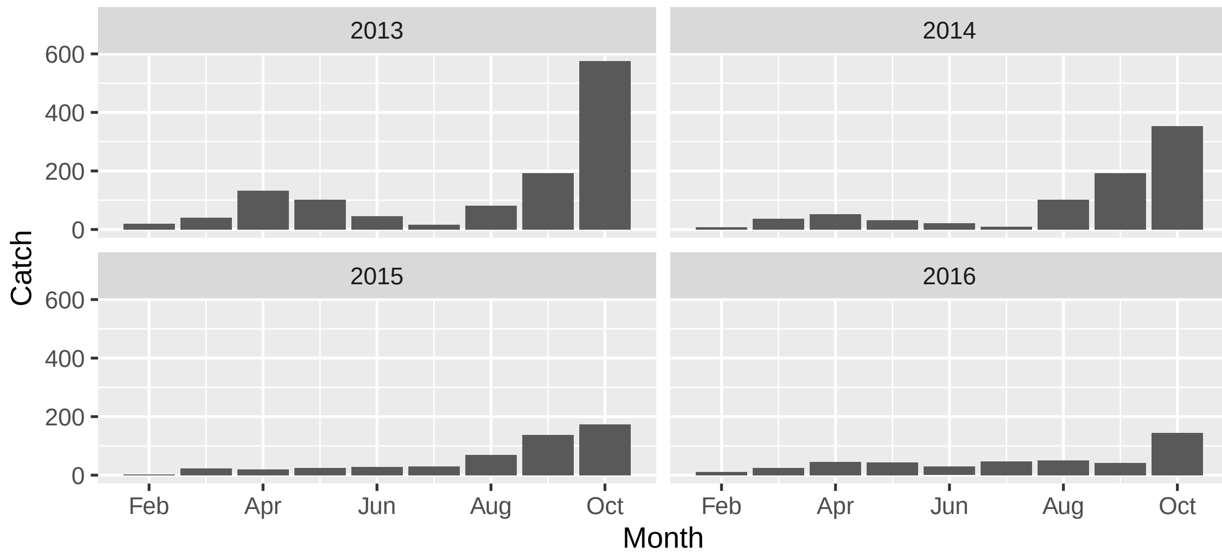

Figure 8: Monthly catches for the River Don 2012-2106.

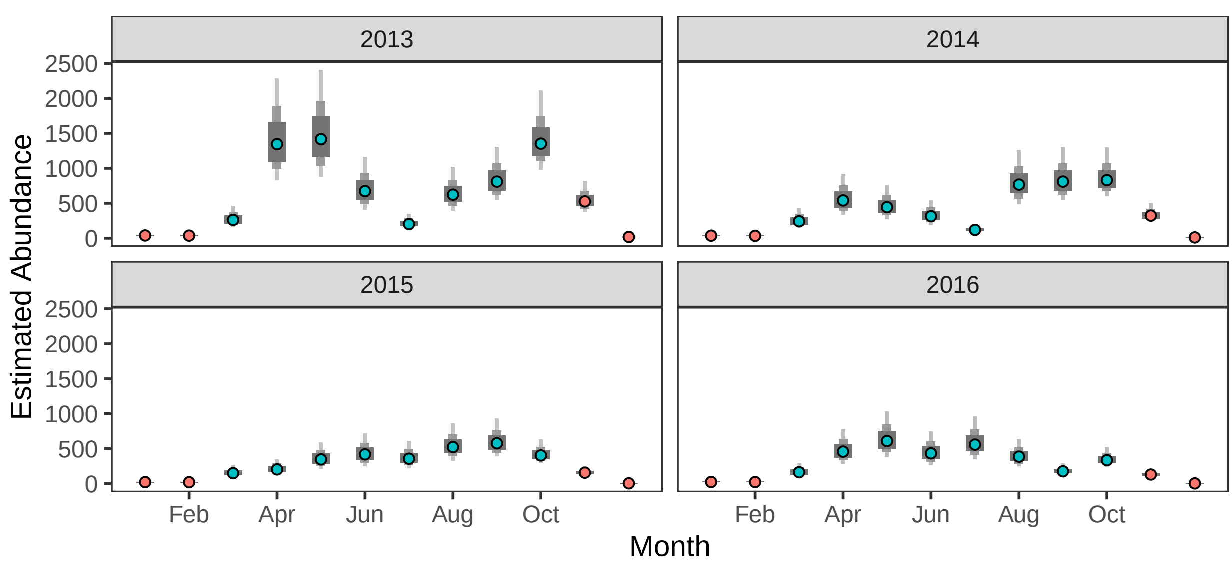

Figure 9: Estimated monthly returns of salmon to the River Don during 2013-2016. Plots show the median (filled circle), 50, 70 and 90% of the estimates (wide and dark to narrow and light regions). Blue points indicate in-season estimates, red those out of season.

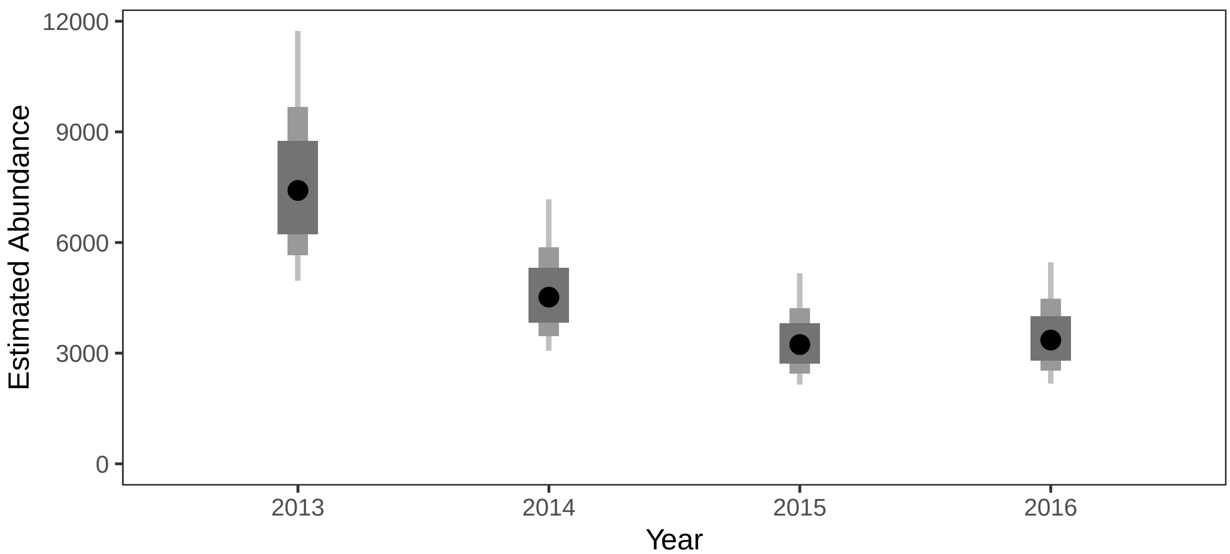

Figure 10: Estimated annual returns of salmon to the River Don during 2013-2016. Plots show the median (filled circle), 50, 70 and 90% of the estimates (wide and dark to narrow and light regions).