

National River Classification Model – Qualified Statement

Dr. Adrian O'Hagan and Dr. Michael Fop

Aim

- Evaluate and discuss the current methodology used for the National river classification model and its application to determine a river's Stock status
- Give a brief description of some possible alternatives and their benefits.

NOTE: THIS “QUALIFIED STATEMENT” REPRESENTS A CONDENSED SUMMARY OF THE CORRESPONDING “REPORT” AND ITS ASSOCIATED APPENDICES.

Fitting a regression where the egg counts are expressed as a linear function of the year, as is currently done, is not adequate to investigate the deposition trend over time. “Year” is a categorical and not numerical variable and hence violates the basic requirements of this statistical model.

The regression model does not take into account the time dependence in egg counts across years, which is a crucial aspect of the process, another reason it is not fit for purpose.

Assuming a linear trend in the regression model is too restrictive and does not reflect the true behaviour of the data. Once a trend is taken, the model strictly follows it and does not allow for possible variations represented by more recent increases in egg counts. The true trend is generally curved and reflects these recent increases in making more optimistic projections.

20 percentile regression is used, apparently to confer prudence. However, there is no statistical justification for this approach. Any such conservatism should only be conferred through setting of the CLs. The statistical modelling and predictions should be performed strictly on a best estimate basis, which is not currently being done.

Only 10 years/observations are used to compute the regression line. Future predictions regarding compliance are hence exposed to high uncertainty. Increased observation count through more frequent observation of the process could remedy this but if not possible, there is further motivation to move away from a regression-based approach.

Key variables are not included in the Annual Assessment that would necessarily impact on estimated stock levels, such as rod effort/catch performance/success and weather information. These should be included as valid numerical regression variables, in place of year, if the regression-based approach is to be maintained.

No model validation is currently performed. The quality of the regression model should be evaluated by comparing the predicted egg counts versus the actual observed egg counts that subsequently emerge as the years pass. If done, significant error in the predicted versus actual egg counts based on the current methodology will be evident.

The confidence intervals reported are misinterpreted and do not correctly measure probabilities as they claim. Probabilities cannot be inferred from confidence bounds, even for a single year projection, but certainly not over multiple years based on lower confidence bounds. This renders the subsequent classification of rivers into AR, PaR, PNaR and NaR categories based on these probabilities invalid.

Likelihood of stocks meeting required thresholds should be based on predicted stock counts versus a conservation-justified threshold, not on relative levels of confidence bounds versus a conservation threshold. The current classification of rivers into AR, PaR, PNaR and NaR based on this reasoning is deeply flawed.

The flaws in the current regression-based methodology are particularly evident in the context of the ad hoc adjustment/correction made following 2003 when actual outcomes were markedly out of sync with those forecast from the modelling process. Persistence with the current methodology will necessitate further ad hoc adjustments in the future.

The current management objective (MO) for salmon stocks in England and Wales is that they should meet or exceed their CLs in at least four years out of five, on average. This does not make logical sense. Something cannot be true at least four years out of five “on average”. The previous definition absent the words “on average” did at least make logical sense.

The net effect of all failings cited is that there are many inconsistencies in the observed data versus banded predictions/decisions from the current methodology, failing simple common-sense reasoning. Specific examples are provided in the Report document.

The regression-based approach used, as well as being based on a non-valid predictor variable of “year” in the first instance, seems biased towards excessive protectionism of stocks via extremely pessimistic and unrealistic forecasting and does not correctly represent the uncertainty around actual and future egg counts. An alternative approach based on time series models should be used to better reflect the nature of the data, the dependence over time and the uncertainty about the actual counts. The likely outcome of such a move to an appropriate model is that some rivers will be reclassified from “At Risk” to “uncertain”, others moving between “uncertain PaR” and “PNaR” and some reclassified to “NaR” in their ultimate status.

Two simple methods to overcome these issues are an auto regressive time series model or a moving averages model, where in both instances more recent observations of egg counts provide greater weight in making predictions of future egg counts. This is vital since in most of the observed data recent egg counts exhibit an increasing trend that is not being correctly reflected in the predictions from the regression model, which remains anchored to data more distant into the past when the egg count trend was decreasing.

A corresponding Scottish stock model of this nature, based on 5 year moving averages, provides an obvious alternative to the current flawed regression approach and is already in use for this purpose in Scotland. The authors strongly advocate for movement to such a model to remedy the serious statistical failings associated with the current methodology, predictions and decisions arising that are outlined in this statement.