

WATER RESOURCES ACT 1991

**THE WALES ROD AND LINE (SALMON AND SEA TROUT) BYELAWS 2017
THE WALES NET FISHING (SALMON AND SEA TROUT) BYELAWS 2017**

**DOCUMENT NRW/3B
PROOF OF EVIDENCE
OF
DR JON BARRY
STATISTICIAN FOR CEFAS**

on behalf of

**CENTRE FOR ENVIRONMENT FISHERIES AND AQUACULTURE SCIENCE
(CEFAS)**

and

NATURAL RESOURCES WALES

NOVEMBER 2018

National River Classification Model - Report

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.

Current method

The status of individual river stocks in England and Wales is assessed annually against stock conservation limits (CLs) and management targets. The *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. An approach introduced in 2004 provides a way of summarizing the performance of a river's salmon stock over the last 10 years (including the current year) in relation to its CL. Regression analysis is applied to egg deposition estimates from the last 10 years, on the assumption that there might be an underlying linear trend over the period. The method fits a 20 percentile regression line to the data and evaluates the probability that this regression line is above the CL, and thus that the CL will be exceeded four years out of five (the management objective). Compliance in the future 5 years is assessed by considering the predicted regression values. When the upper bound (95 percentile) of the regression line confidence interval is below the CL line, the river is judged to be "At Risk", with a probability of <5% of meeting its management objective.

Comments

The current assessment method implements a linear regression model to describe the deposited egg counts (in millions) as a function of the year. However, the approach is inadequate for this type of data and the intent for multiple reasons:

1. Fitting a regression where the egg counts are expressed as a linear function of the year is not adequate to investigate the deposition trend over time. In the regression, the year is used to explain the variability in egg counts, however other factors may affect this quantity and time is not an appropriate variable to regress against. It is a categorical and not numerical variable and hence violates the basic requirements of this statistical model. Recoding of years to integer values as has been done does not remedy this problem.

2. A regression model is not a suitable model to describe the evolution of egg counts over time. Indeed, the regression model does not take into account the time dependence in egg counts from one year to the other, which is a crucial aspect of the process.
3. Even setting aside the fact that the regression variable is not appropriate, assuming a linear trend is too restrictive and does not reflect the behavior of the data. Once a trend is taken, the model strictly follows it and does not allow for possible variations represented by jumps in egg counts. It also does not reflect the non-linear nature of the data trend, which is decidedly quadratic or U shaped in many instances.
4. 20 percentile regression is applied. The logic behind this seems to be to infer a pure ad hoc conservative estimate and there is no statistical justification for this approach. Any such conservatism should only be conferred through setting of the CLs which in turn must be justified by conservation theory, which is a non-statistical challenge. The statistical modelling and predictions should be performed strictly on a best estimate basis, which is not currently being done.
5. Only 10 years (and thus 10 observations) are used to compute the regression line. Assessment is based only on few observations and any statement (especially related to future compliance) is likely to be not valid and subject to very high uncertainty. This may necessitate changing the underlying calculations for egg deposition to apply over shorter periods of time than annually but would be very beneficial to model building and prediction. If this is not possible and only annual assessments can be made then this further necessitates a move away from a regression-based approach to a moving averages or time series based approach.
6. 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 factors. There is no statistical justification for not including such variables, where available, in order to improve model predictive performance and to replace the use of year, which is not a valid predictor. It may prove difficult to undertake nationally but we suggest that rod effort/catch performance and weather factors are appropriate for consideration and incorporated in a valid and expanded regression model if this approach is maintained. It is notable that these were suggested improvements were taken as to be adopted in the 2004 3.1.3 Spawning Escapement Recommendations.
7. No model validation is performed. The quality of the 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 and this should be addressed through the suggested major amendments to the methodology used.

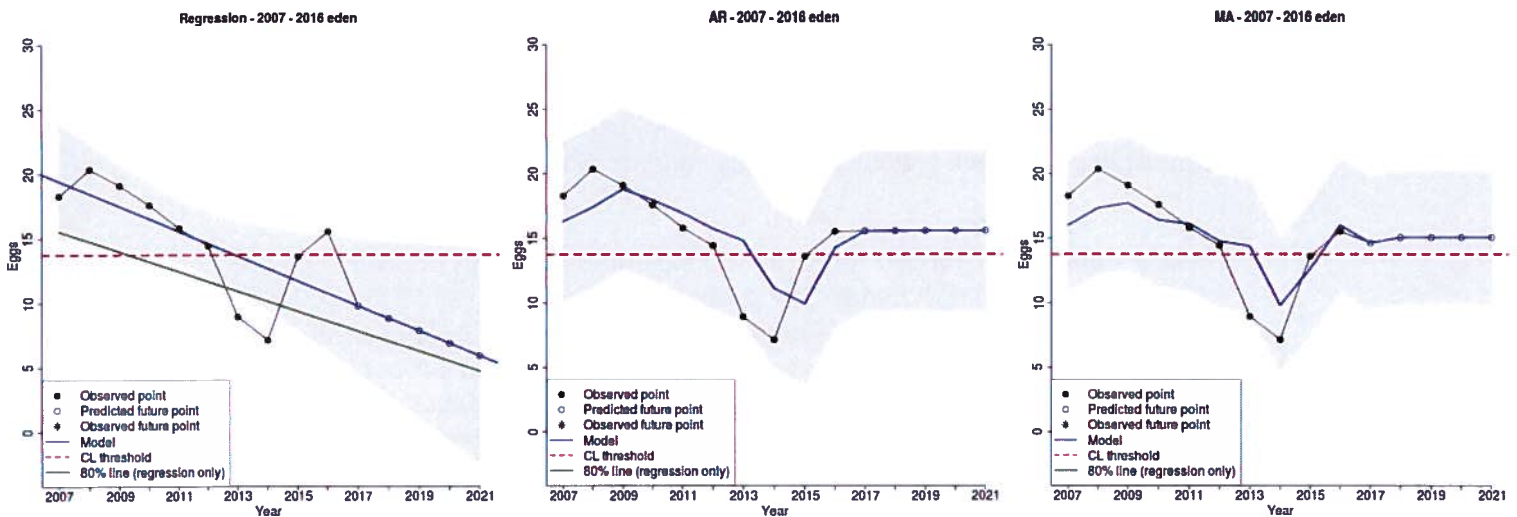
8. The objective of the approach seems to identify the likelihood of the eggs deposition failing to meet the management objective by evaluating the probability that the regression line will be above/below the CL. The confidence intervals reported are misinterpreted and do not measure this quantity. Probabilities 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.
9. 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 percentages attached to confidence intervals relate to the percentage of the time an individual year's interval will contain the true value under repeated sampling, they do not map to probabilities of a particular outcome being true or not true. The current classification of rivers into AR ,PaR, PNaR and NaR statuses based on this reasoning is deeply flawed.
10. 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 inevitably necessitate further such ad hoc adjustments/corrections since the underlying approach is not statistically valid or robust.
11. There appear to be further fundamental misunderstandings in producing the AR ,PaR, PNaR and NaR bands. The probability of meeting the MO in 2004 is described as "exceeding CL in 4 out of 5 years", which makes logical sense. An amendment to include "on average" came in 2008 when the probabilities (which are themselves flawed) were used to produce the bands. This idea of something being true "four out of five years on average" is not logically sound. An outcome may be true four out of five years, or the average value of a quantity over five years may be above a threshold, but there is no logical meaning in saying that something is true on average over four out of five years.

12. The effects of such misunderstandings are evident throughout the observed data versus banded predictions/decisions. Two rivers (Ribble & Lune) have clear CL achievement over the last 10 years and yet are both AR in 2016 and in 2021 have a specified <5% probability of meeting MO. This contrasts with the Wye for example, where it has failed CL in 9 of the last 10 years but is PaR in 2016 (specified to have 5% to 50% probability of meeting MO) and PNaR in 2021 (50% to 95% probability of meeting MO). The reasons for such issues arising are the fundamental flaws in the underlying regression-based approach and use of probabilities derived incorrectly from confidence bands.

Discussion

The regression-based approach used so far 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 following figure shows a simple comparison between regression, Auto Regressive (AR) model and a Moving Averages (MA) model for river Eden based on data for the period 2007-2016. The grey shaded area shows the 95% confidence interval calculated around the ensuing predictions over predictions made 5 years into the future.



Note that the AR model is a type of time series model where the egg counts at year (t) depend on the previous year (t-1) egg counts and a component of random error. Formally, the equation for egg counts at year (t) would be:

$$Egg_{(t)} = aEgg_{(t-1)} + Error_{(t)} \quad \text{where } a \text{ is a coefficient.}$$

The AR model applied here is a simplified version for illustrative purposes only, with only dependence on the previous year conferred. If selected as a predictive model, it should be adapted to confer appropriate dependence structure for egg counts based on a four to five-year historical data linkage window. However, the underlying principle and the nature of the observed outcome cited below will persist.

The MA model is a time series model where the egg counts at year (t) depend on an average egg counts (common to all years), the error component from the previous year (t-1) and the error component at the current year (t). The equation for the model is

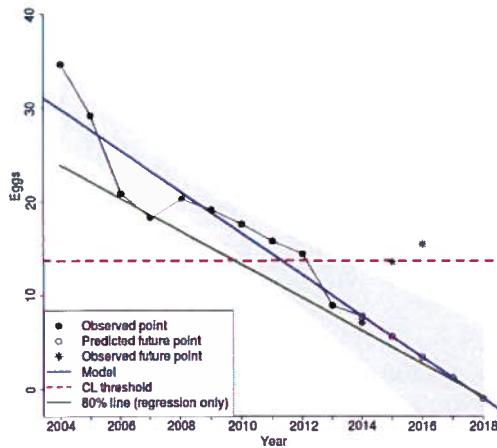
$$Egg_{(t)} = Average + bError_{(t-1)} + Error_{(t)} \quad \text{where } b \text{ is a coefficient.}$$

As we can see, the models consider the time dependence of egg counts at a given year (t) on egg counts from a previous set of years and naturally include a component of uncertainty.

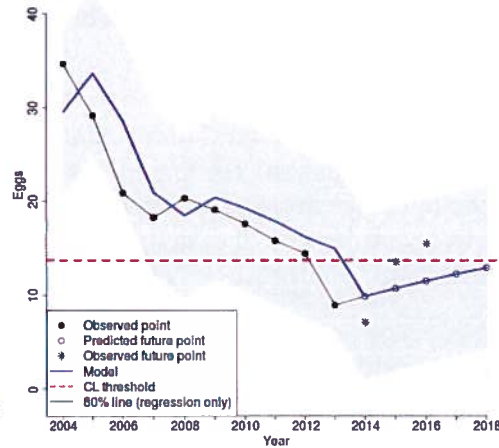
According to the regression model, the river would almost fail to meet the CL threshold over the observed period 2013-2016 and also in the future years 2017-2021. That is because the regression line is biased by the downward trend of the previous period 2007-2014, and it does not take into account the increase in the period 2015-2016. The time series models, AR and MA, better depict the egg counts trend over time and give a different picture. According to these models, the river would be judged as uncertain.

The regression approach also did not correctly represent the trend of egg counts in the past. The following figure reports data for river Eden for the observed data period 2004-2013, where the period 2014-2018 would be object of future compliance assessment. However, the actual egg counts for the years 2014, 2015 and 2016 are available, thus we can compare the modelled trend with the actual trend. Note that "asterisk" points do not enter into the model estimation process and need to be predicted.

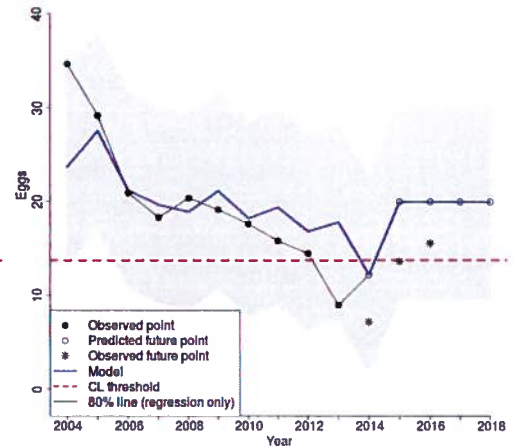
Regression - 2004 - 2013 eden



AR - 2004 - 2013 eden



MA - 2004 - 2013 eden



The regression line completely misses the actual egg deposition trend. On the other hand, the time series models consider more uncertainty and correctly take into account the possibility of an increase in egg deposition.

Four appendices are provided to this report that provide, for the Eden, Ribble, Wye and Lune, year-on-year side-by-side comparisons of the regression, MA and AR approaches, contrasting the differences in the predicted five year values produced by the valid time series approaches (MA and AR) versus the invalid regression model (current methodology).

We also wish to stress again that too few data points are used in the current model estimation process (only a 10 data-point time window is considered). All the available data should be used and should be supplemented with data collected and incorporated more than once per year if possible.

However, if this is not possible, this provides additional motivation to move away from a regression-based approach and towards a time-series based approach such as moving averages or an auto-regressive model. In the latter cases, restriction to more recent observations will generally provide a more accurate prediction of stock levels in future years. 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.

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.