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/3E

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

Reply to the statistical comments made by:
Mr John Rawlinson
Chairman Ribble Fisheries Consultative Association
Our ref: P02018/09740/MH

I have read the comments made by the consultant statisticians in the above document. I generally don't agree with them. In particular, I strongly disagree with their assertion that

"A regression model using year as a predictor variable in this context is not a valid approach. This is time series data and should be modelled using time series approaches such as moving averages or ARIMA models."

I do not accept that data can be pigeon-holed into time series, regression or any other type of data and that each pigeon hole requires some particular set of solutions. For any problem, we need to find the best and most appropriate answer. In this case, the solution devised by the Environment Agency contains a fixed (regression) linear component and a correlated error structure (of the sort often used in time series modelling) around this linear component. Such trend models are routinely used in, for example, contaminant monitoring.

I won't go into each of the comments made in Mr Rawlinson's document. However, I think that the nub of the dispute is that the consultant statisticians think that the most suitable approach is to use a model that predicts future observations most closely. The Environment Agency's method is precautionary approach and assumes that what has happened in the past could happen in the future. Of the two approaches, I prefer the Environment Agency's one. This is merely a side argument, but one reason why it is difficult to use past data to see whether how well predictions performed is that remedial action is likely to have reduced the decline in fish egg numbers. This means that a predicted linear decline in egg numbers may not have happened not because the linear decline was wrong but because remedial action boosted egg numbers.

I think that it is instructive to consider the options if a linear decline in egg numbers is observed.

1. Assume that linear decline will happen in the future (the EA model approach). If you are correct, then taking remedial action will hopefully resolve the problem. If you are wrong, then your remedial actions were unnecessary.
2. Assume that linear decline will NOT happen in the future (which is essentially the time series approach proposed by the consultant statisticians). If you are wrong, then this could cause damage to the fishery because remedial action has not been taken. If you are correct, then you won't have taken remedial action unnecessarily.

Of the options above, I think that 1 is best because the adverse consequence (unnecessary remedial action) is far less serious than the adverse consequence in 2 (damage to fishery). This is clearly a value judgement and one that needs to be considered by fisheries experts, rather than statisticians.

I did the modelling work outlined in Appendix 1 in June 2018. However, I think that it is worth repeating here. Essentially, it considers the situation where a downward linear trend is observed in the past and that this downward trend continues into the future. The method used is simpler than that used by the EA, but the concept is the same. However, what it clearly shows is that, if there is a future downward trend, the EA-type models will predict it but the time series models suggested by the consultants will not.

Appendix 1: Salmon Egg Modelling

Jon Barry 5/6/18

I investigated the prediction performance of the AR(1), MA(1) and a simple 5-point moving average (models proposed by O'Hagen and Fop) for the situation where egg counts have been decreasing linearly for 10 years and then continue to decrease linearly for years 11 to 15. In our scenario, we use the first 10 years to predict the egg counts in years 11 to 15. Note that we simply predict the mean of future observations rather than considering 20th percentiles – this won't change the basic conclusions.

I simulated egg count data where the errors were independent (phi=0) and where the residuals had an AR(1) process about the mean (phi=0.3). Other values of phi were investigated, but the results were similar to those shown here.

Specifically, the models used are

$$Egg_t = aEgg_{t-1} + error_t \qquad \text{AR(1)}$$

$$Egg_t = Average + berror_{t-1} + error_t \qquad \text{MA(1)}$$

$$Egg_t = \sum_{j=1}^{5} Egg_{t-j} / 5 \qquad \text{5-point MA}$$

Prediction results are shown in Figures 1 and 2. The black circles show the actual realised values. We can see that the models do not predict the new values well. The reason for this is that these models are designed to be used for stationary processes – where the mean and variance do not change. In our situation, with a linear trend, the models can never perform well. You need to have a model which incorporates a potential trend to do well at predicting new data – as used in the current method devised by the Environment Agency.

This really crystallises the arguments that we have made already. If the first 10 years of data has a linear trend then the precautionary approach is to assume that this trend will continue in the future. This seems eminently sensible both from an environmental and a statistical point of view. Of course, the trend might not continue, but it is prudent to assume that it will.

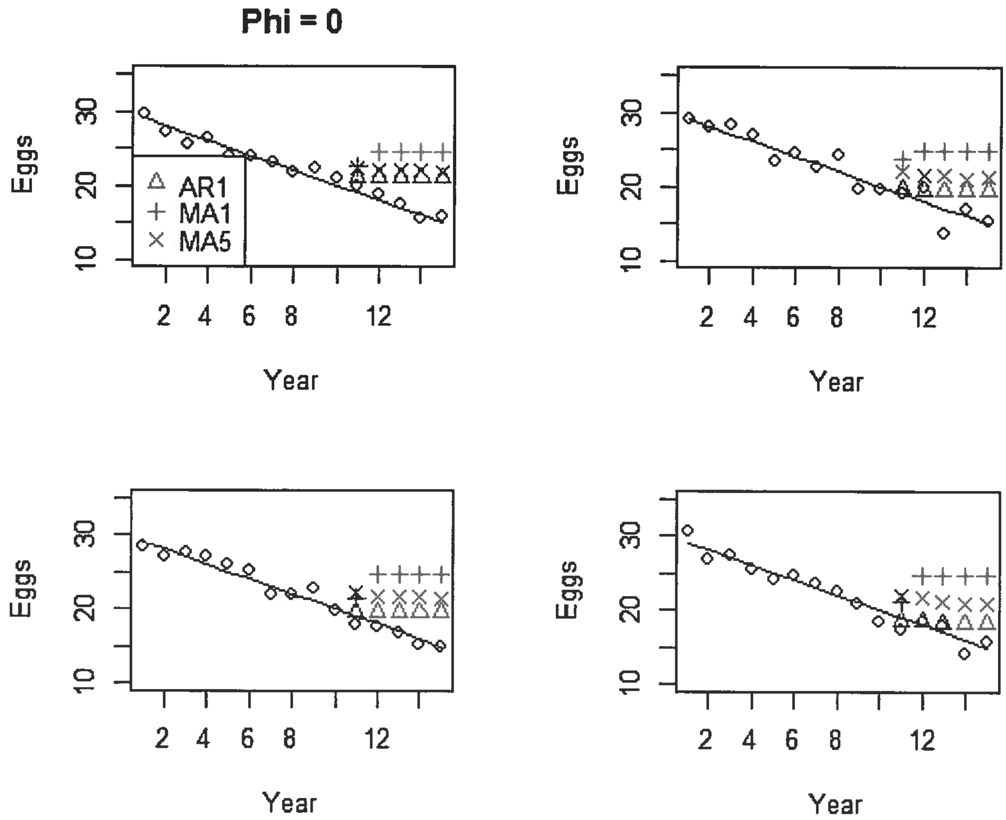

Figure 1: Four simulations of egg count data and values predicted by the models for a linear regression model with independent errors.