

# Influential Years Analysis: only past catch

## Validation of the Jul-Sep landings base models

To evaluate the base model for the Jul-Sep landings using only prior catch as covariates, three types of models were fit. The first two were GAM and linear models with Jul-Sep and Oct-Mar in the prior season only or prior season and two seasons prior as covariates:

$$\text{GAM t-1 : } S_t = \alpha + s(c_{t-1}) + e_t$$

$$\text{Linear t-1 : } S_t = \alpha + \beta c_{t-1} + e_t$$

$$\text{GAM t-1, t-2 : } S_t = \alpha + s(c_{t-1}) + s(d_{t-2}) + e_t$$

$$\text{Linear t-1, t-2 : } S_t = \alpha + \beta c_{t-1} + d_{t-2} + e_t$$

where  $c_{t-1}$  was either  $S_{t-1}$  (Jul-Sep landings in prior season) or  $N_{t-1}$  (Oct-Mar landings in prior season) and  $d_{t-2}$  was the same but 2 seasons prior.

These types of models do not allow the model parameters (the intercept  $\alpha$  and effect parameter  $\beta$ ) to vary in time. The second type of models were dynamic linear models (DLMs). DLMs allow the parameters to evolve in time. Two types of DLMs were used, an intercept only model where the intercept  $\alpha$  evolves and a linear model where the effect parameter  $\beta$  is allowed to evolve:

$$\text{DLM intercept only : } S_t = \alpha_t + e_t$$

$$\text{DLM intercept and slope : } S_t = \alpha_t + \beta_t t + e_t$$

$$\text{DLM intercept and effect : } S_t = \alpha_t + \beta_t c_{t-1} + e_t$$

In addition to the GAM, linear and DLM models, three null models were included in the tested model sets:

$$\text{intercept only : } S_t = \alpha + e_t$$

$$\text{intercept and prior catch : } S_t = \alpha_t + S_{t-1} + e_t$$

$$\text{prior catch only : } S_t = S_{t-1} + e_t$$

The ‘intercept only’ is a flat level model. The ‘prior catch only’ simply uses the prior value of the time series (in this case landings) as the prediction and is a standard null model for prediction. The ‘intercept and prior catch’ combines these two null models.

The models were fit to the 1956-2015 landings (full data) and 1984-2015 (data that overlap the environmental covariates).

The model performance is shown as the leave-one-out (LOO) one-step-ahead prediction errors. The data point  $t$  is omitted and the model is fit to the remaining data. This is the LOO model. The prediction error is the data point  $t$  minus the predicted value for data point  $t$ . This is repeated for all data points  $t$ . The AICc values for the models are also shown. AICc is based on the log-likelihood which is calculated from the one-step-ahead prediction errors for the full data (no data left out).

## Results

Figures 1D and 2D show the  $\Delta\text{AICc}$  values for the model set.

The Figure 1D shows the one step ahead predictions for the models broken out in the three categories: GAM, linear, and DLM.

that with any year left out, the set of models that has the lowest AIC was always GAM with Oct-Mar and Jul-Sep landings in the prior season or GAM with only Oct-Mar landings in the prior season. There were

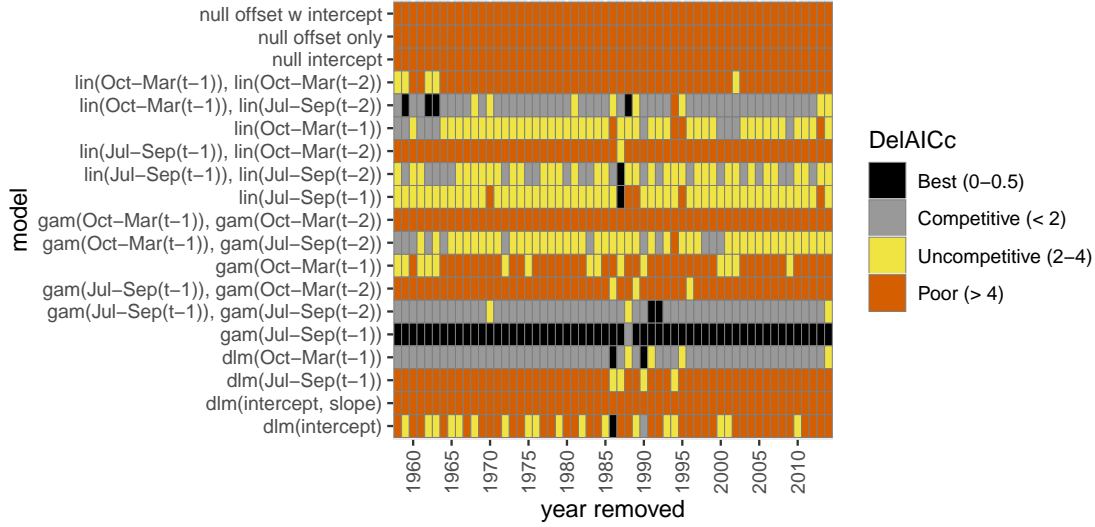


Figure 1D.  $\Delta AICc$  for the Jul-Sep landings base models with one year deleted using the full landings data set 1956-2015. Delta AIC is AIC of the model minus the AIC of the best (lowest AIC model) in the set. Black models were the best models in the set and within 0.5 AIC of each other. Green are models within 2 of the best model, thus competitive to the best models. Deleted year is shown on the x-axis; the last column, None, has no year removed.

cases where deleting a year removed one of these two from the ‘best’ category, but they were still in the ‘competitive’ category. However, when we look at the Leave-one-out predictive performance (Figure 2D), we see that the GAM with Oct-Mar and Jul-Sep landings in the prior season is better than the GAM with only Oct-Mar landings—except when 1995 or 1987-88 are left out.

This suggests that the better predictive performance is due to better prediction of 1994-95 or 1987-8. The covariate analysis also indicated that 1994-95 have a strong influence on the selected models. This is not entirely surprising since 1994 was a year of a severe decline and 1995 a year when the landings rebounded. The analysis was repeated with 1994 and 1995 removed (Figure 3D and 4D). With these 2 years removed, the GAM with Oct-Mar landings in prior season only is the model with lowest AIC and LOO predictive performance regardless of which year is removed. Thus, the GAM with Oct-Mar landings in the prior season only was chosen as the base model for Jul-Sep landings.

It should be noted that none of the models has a particularly high adjusted  $R^2$  (Figure 5D). The values are generally less than 0.3. The Jul-Sep landings tend to be highly variable and not related to the catch in prior years. Jul-Sep is during the monsoon during which fishing is not always possible due to sea-state and there is a 6-week fishing ban during this time.

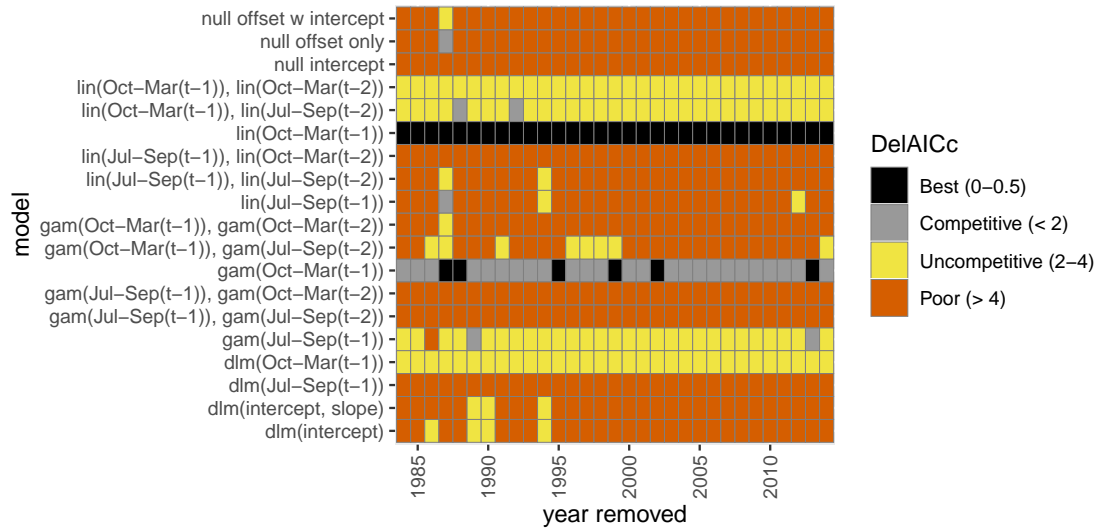


Figure 2D. Delta AICc for the Jul-Sep landings base models with one year deleted using only the landings data that overlap with the environmental data 1984-2015. See Figure 1D for an explanation of the figure.

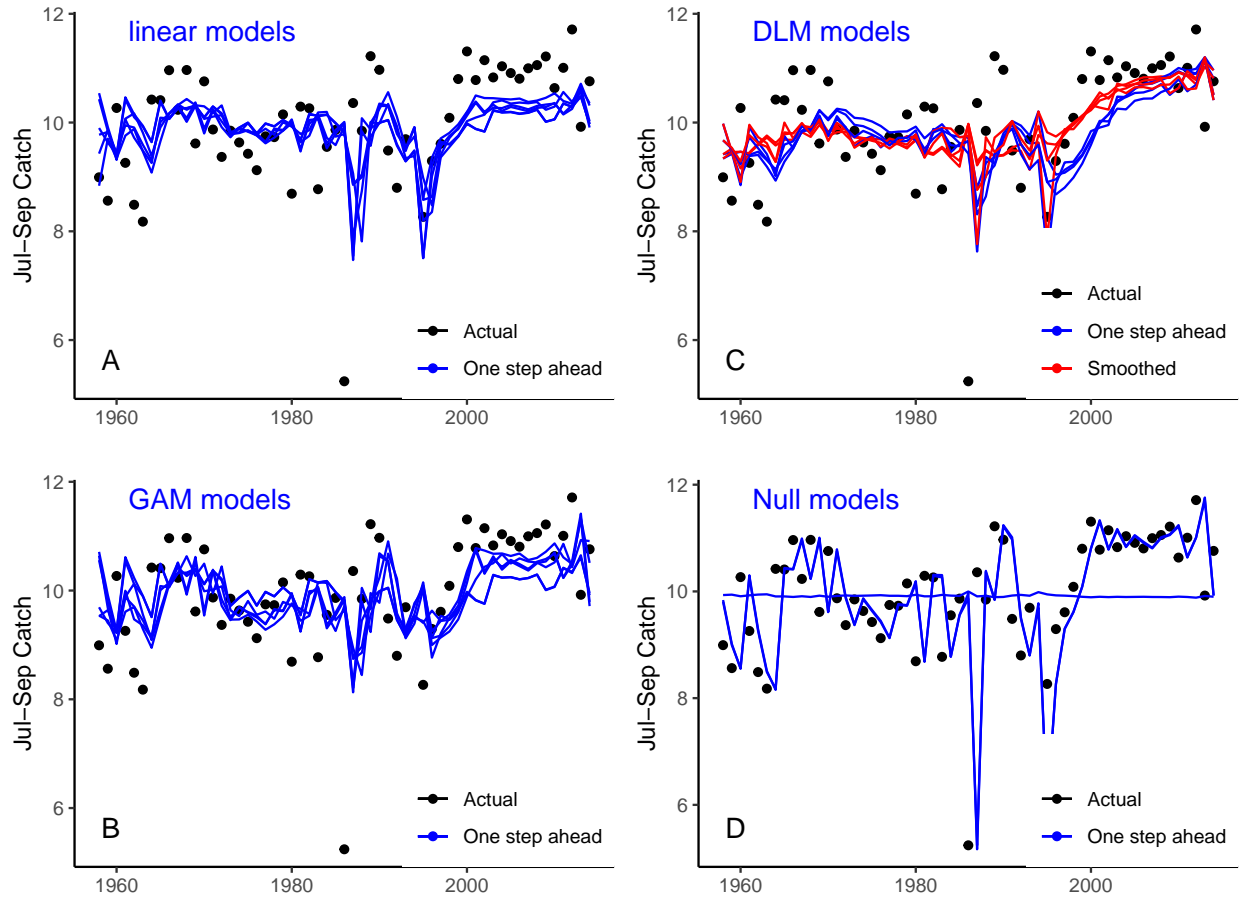


Figure 1D. Leave one out (LOO) one step ahead prediction errors for the linear, GAM, and DLM models of Jul-Sep landings. In panels A-C, the data point at year  $t$  on the x-axis is predicted from the data up to year  $t-1$ , and the model is fit to all data except year  $t$ . In panel D, the predictions use all the data; the model is still fit without year  $t$ . The predictions in panel D are called the smoothed predictions. The dashed line in the DLM plots is the best overall model in terms of root mean squared error.

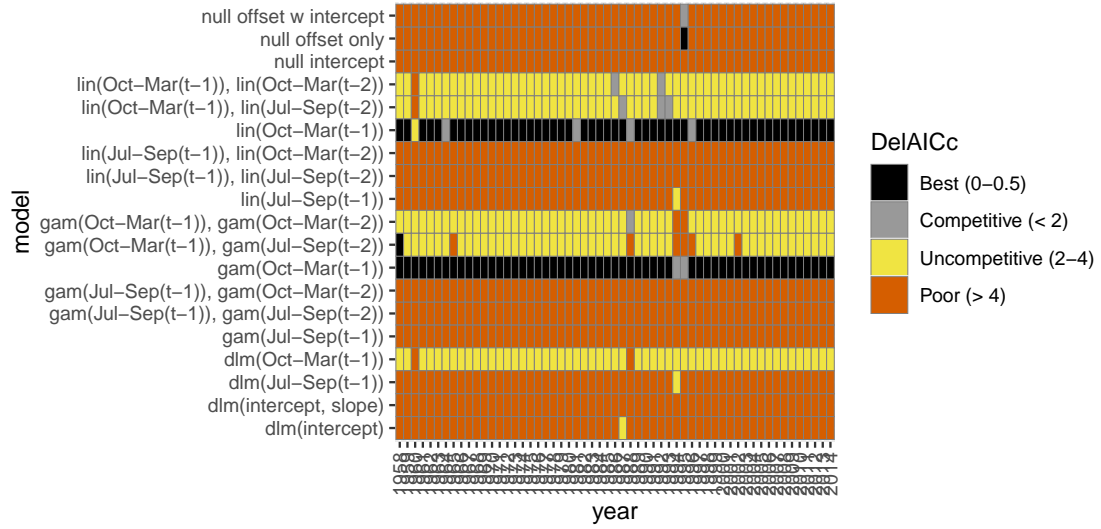


Figure 4D. Delta AIC for the Oct-Mar landings base models with one year deleted. The deleted year is shown on the x-axis.

## Validation of the Oct-Mar landings base models

The Figure 6D shows that for Oct-Mar landings the best model was always GAM with Oct-Mar in the prior season and Jul-Sep landings two seasons prior.

Again 1986 and 1994 were deleted and the leave-one-out analysis was repeated. The Figure 4D shows the best model is still always GAM with Oct-Mar in the prior season and Jul-Sep landings two seasons prior.

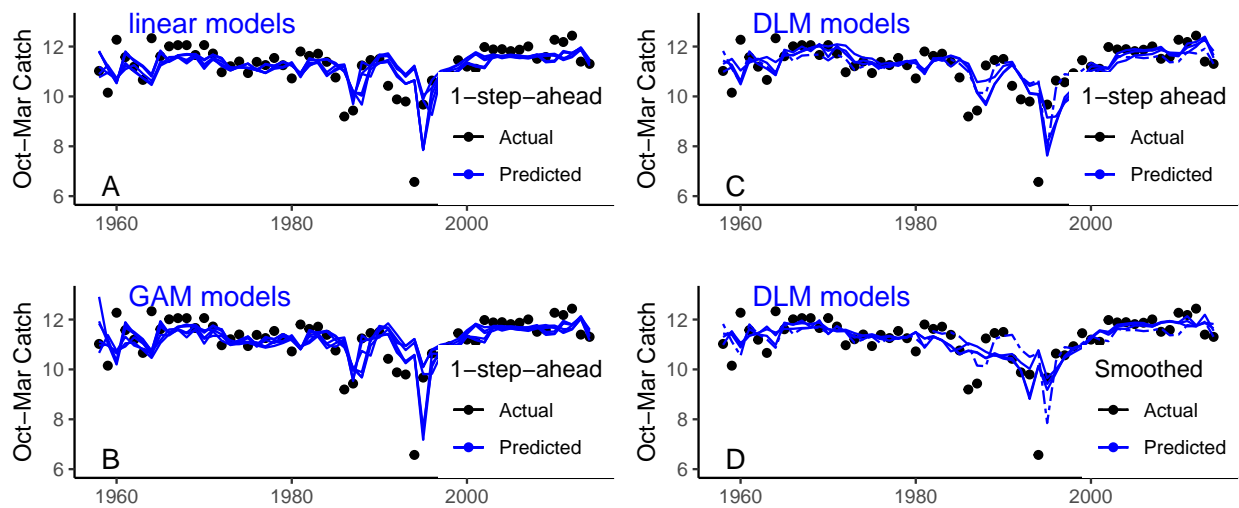


Figure 3D. Leave one out (LOO) one step ahead prediction errors for the linear, GAM, and DLM models of Oct-Mar landings. In panels A-C, the data point at year  $t$  on the x-axis is predicted from the data up to year  $t-1$ , and the model is fit to all data except year  $t$ . In panel D, the predictions use all the data; the model is still fit without year  $t$ . The predictions in panel D are called the smoothed predictions. The dashed line in the DLM plots is the best overall model in terms of root mean squared error.