**Examining Jack-knifes for Diagnostics on Indian Ocean Albacore and East Atlantic Bluefin Assessments**

**By Rishi Sharma, Iago Mosqueira and Laurie Kell**

**Introduction:**

In fish stock assessment model validation is often based on a naïve adaptation of Pearson residuals, i.e. the difference between observations and posterior means, even if this approach is flawed (Thygesen, 2017). A reason for this is because statistics based on residuals from model fits are not always a good guide to how well that model will predict, since a high $R^2$ or low root mean square error (RMSE) can be obtained by over-fitting. For example, in a simple polynomial regression better fits to the data can be obtained by adding higher order terms but the predictions from the model on new data will usually get worse as higher order terms are added. This problem is compounded by often having to compare scenarios for alernative datasets and model structures with different data requirements, and so AIC cannot be used to compare models. In addition there are also a range of potential problems to identify when examining residuals, e.g. bias, drift, skewness, missing variables, and heteroscedasticity. When inspecting residual patterns, however, there is a danger of hypothesis fishing (Wasserstein and Lazar, 2016) and so it is good practice to reserve part of the data for validation. This ensures that the significance of a pattern in the data is not tested on the same data set which suggested the pattern.

Cross validation evaluates the predictive error of a model by testing it on a set of data not used in fitting. There is often insufficient data, however, in stock assessment datasets to allow some of it to be kept back for testing. A more sophisticated way to create test datasets is, like the jackkife, to leave out one (or more) observation at a time. Cross validation then allows prediction residuals to be calculated, i.e. the difference between fitted and predicted values where the later is calculated from the out-of-sample predictions.

Prediction residuals can either be for historical or future observations. In the later case for example a one-step forward prediction is where data points are made available to the model one measurement at a time, and the model is evaluated by its ability to predict the next data point. This is the general principle of frequentiststatistics (Dawid, 1984).

In this study we show how prediction residuals can be used to validate stock assessment scenarios, using as an example East Atlantic and Mediterranean bluefin. Model validation examines if the model family should be modified or extended, and is complementary to model selection and hypothesis testing. Model selection searches for the most suitable model within a family, whilst hypothesis testing examines if the model structure can be reduced.

## Methods

The method is simplistic. It leaves one variable out in the CPUE that was used in the assessment and computes the overall model fit with and without the point. Some of the key parameters, like B0, and

$B_{curr}/B_{MSY}$ are compared, and residual diagnostics on the missing data point are compared when the data is ignored with the full model, i.e. all data are used.

If the model, is indeed being over-fitted, then the residual diagnostics of the full model on all values, as compared to the missing (jack-knifed) values on the missing points will be lesser (a matric like mean squared error). If not, then the model is doing a decent job, as the prediction error is comparable to the residual error of the full model.

$$PE = \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n-1} \tag{1}$$

$$RE = \frac{\sum_{i=1}^{n}(E_i - O_i)^2}{n-1} \tag{2}$$

If RE<<PE then the model is overfitting. If PE is approximately equal to RE then the model is doing a decent job.

## Results

### Bluefin Tuna

Runs were made on the jacknife as shown below. Uncertainty on each run is indicated on some key parameters (Figure 1).

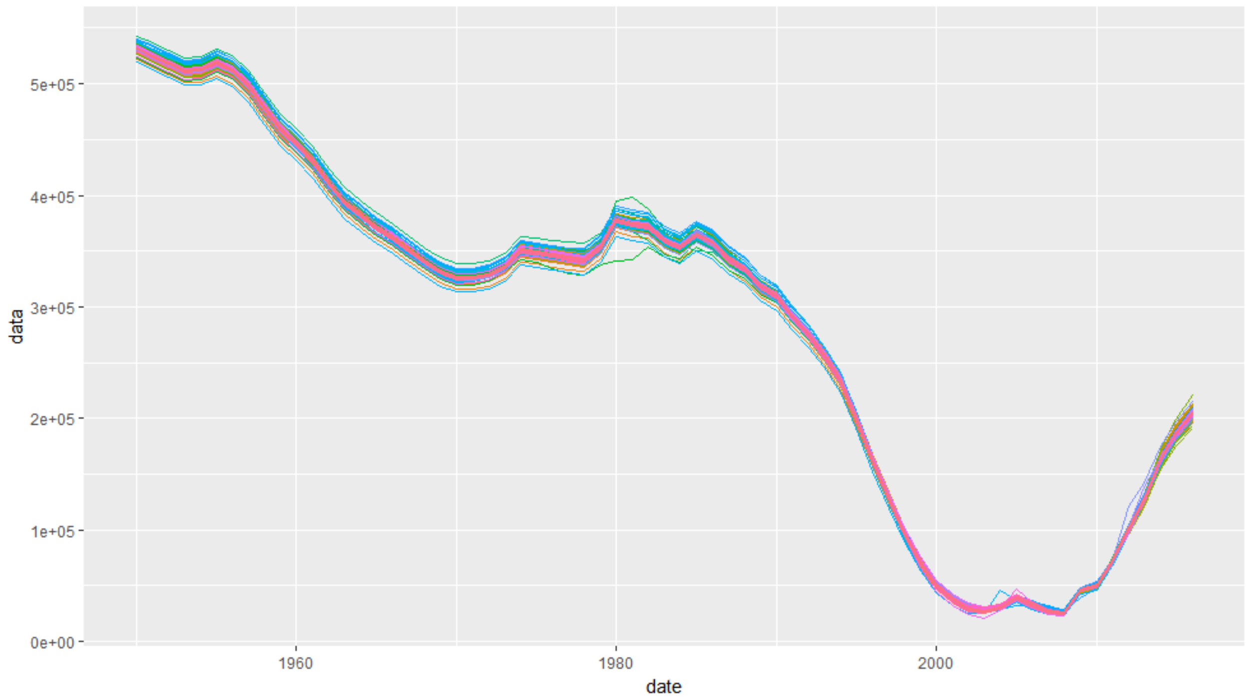


Figure 1: Jack-knife based on EBFT assessment on taking one point out at a time

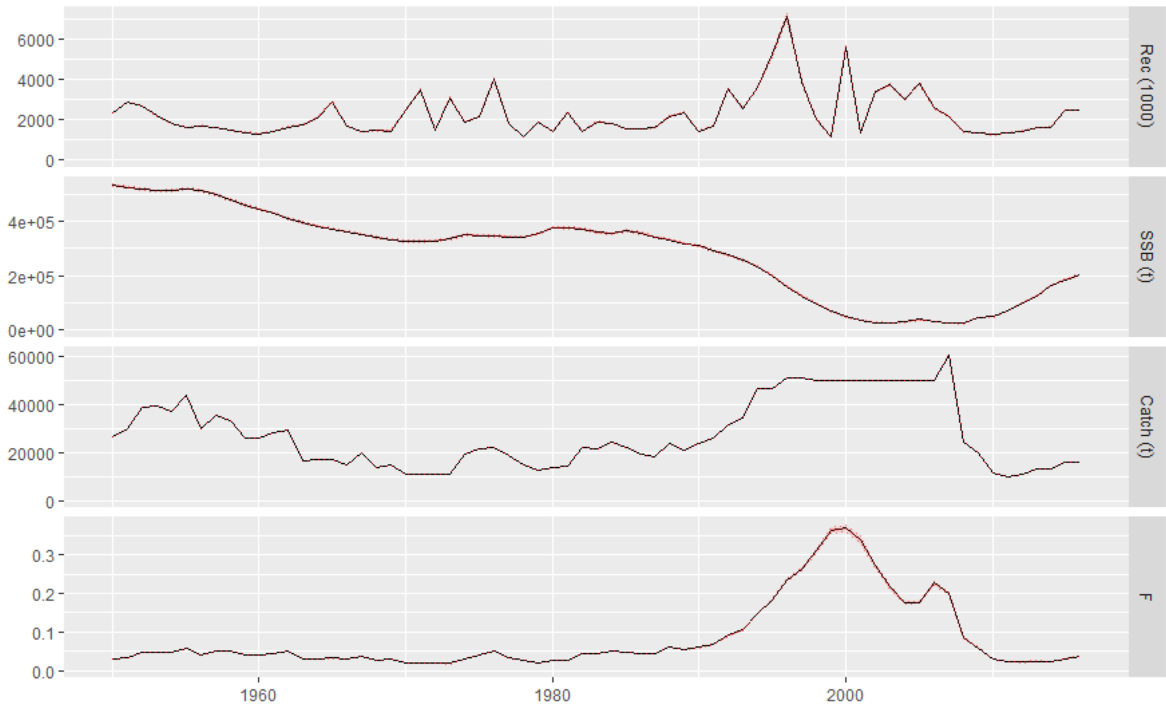


Figure 2: results across the runs in Jackknife runs evaluated for EBFT.

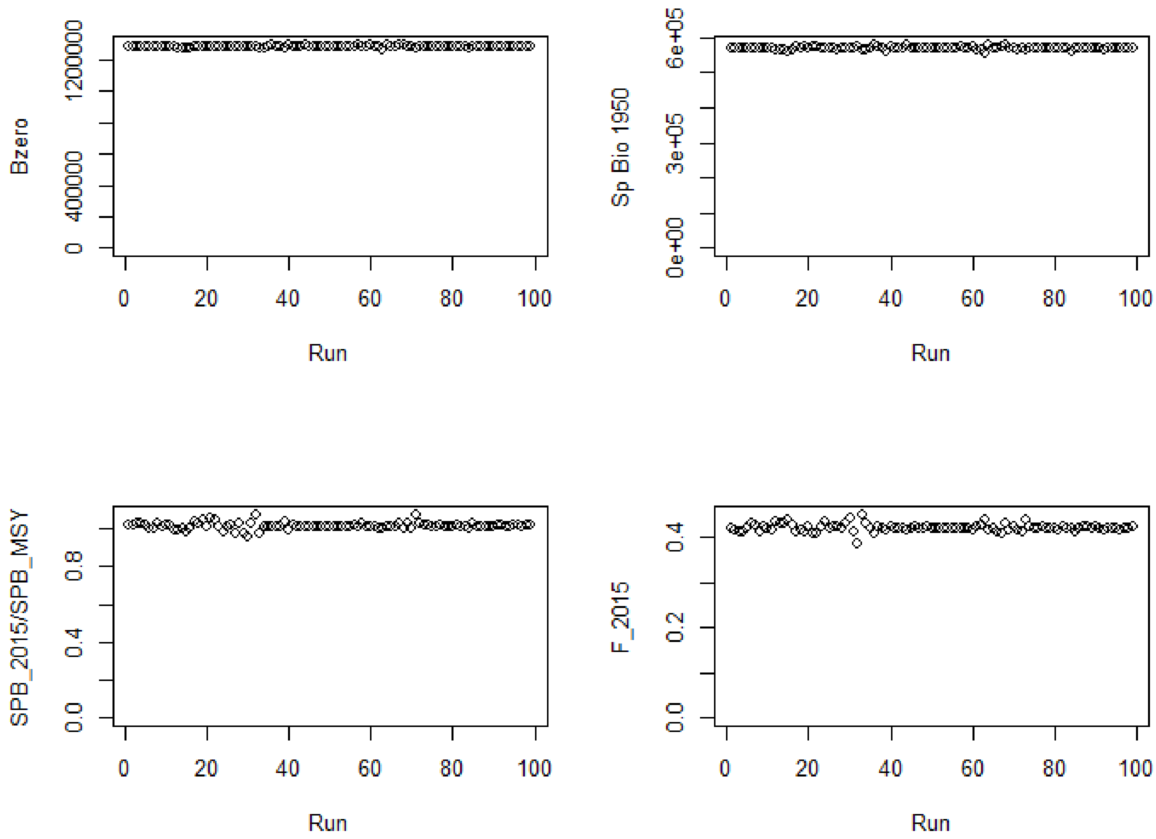


Figure 3: Model performance is stable over the runs

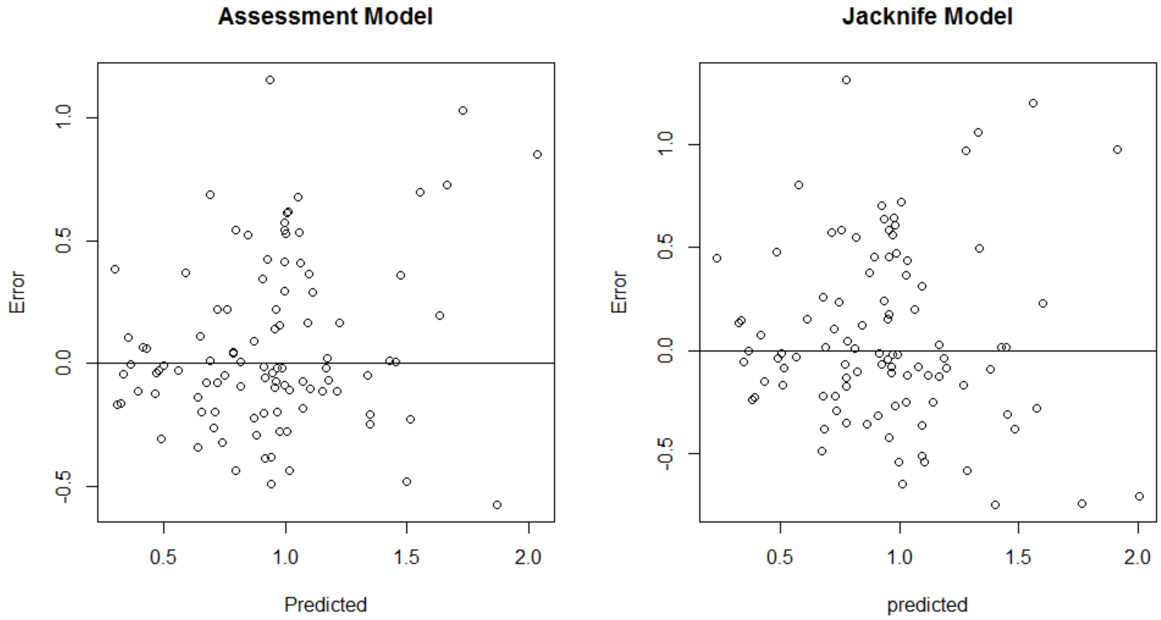


Figure 4: Prediction error compared to Assessment model error

Table 1: Mean Sq. Error values for the jack-knife and Assessment models

| | |
|---|---|
| MSE (Prediction/Jack-knife) | 0.188 |
| MSE (Assessment) | 0.121 |
| Bias | -0.09 |

Although model misspecification is possibly occurring as the assessment residuals are lesser that the jack-knife based residuals (prediction error). However, based on residual diagnostics (Figure 4) and the dynamics (Figures 1, 2 & 3) the model appears to be fairly stable.

**Indian Ocean Albacore**

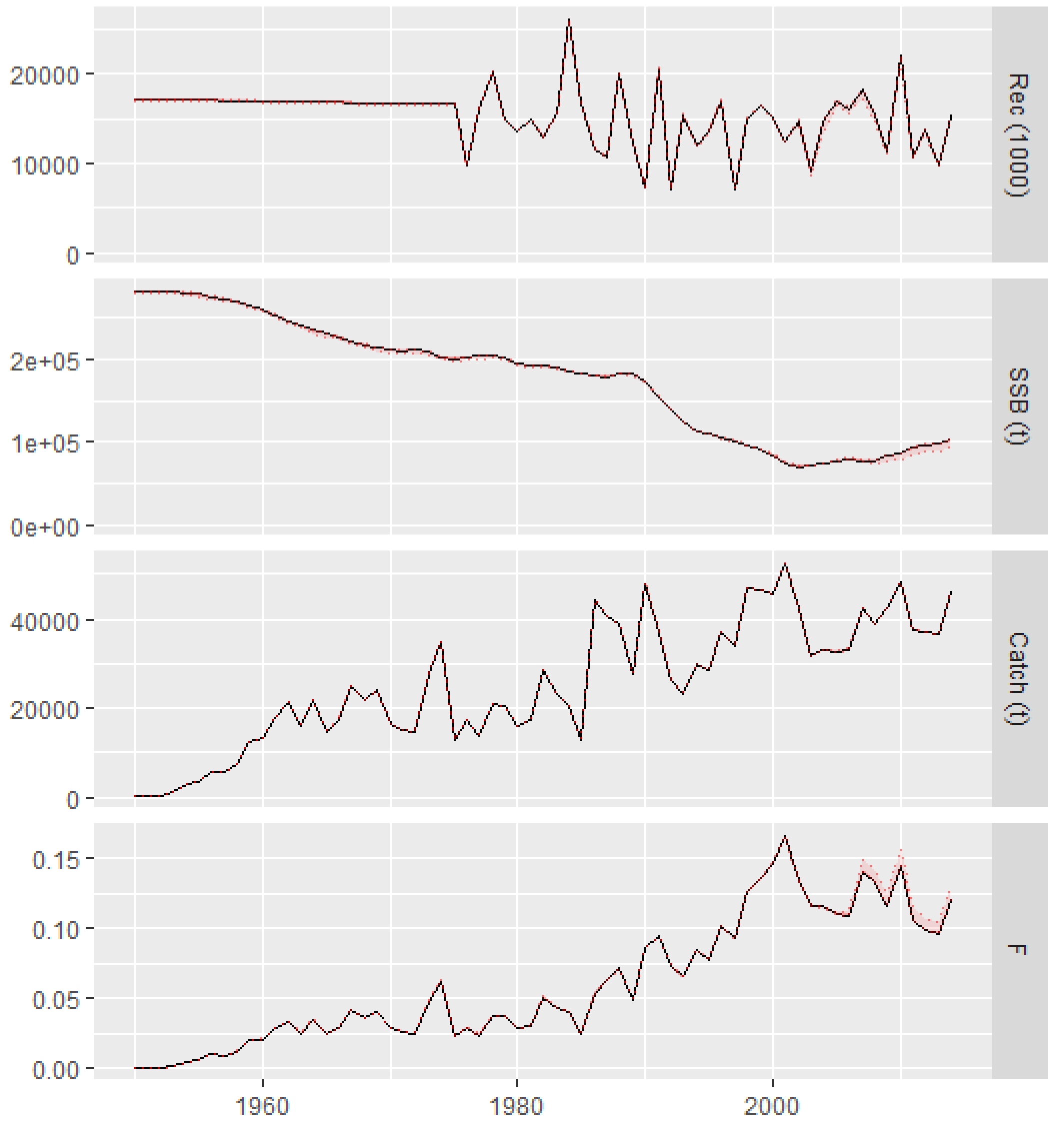


Figure 5: Runs across all jacknifes.

Summary of runs across seasons (Figure 5 is shown) below. A similar analysis was performed on IO Albacore. The results once again show similarities in the 2 models though the model appears to be be doing a better job but in terms of prediction may not perform as well (Table 2, Figure 7). However residual diagnostics again (Figure 6) show that the 2 models are quite comparable.

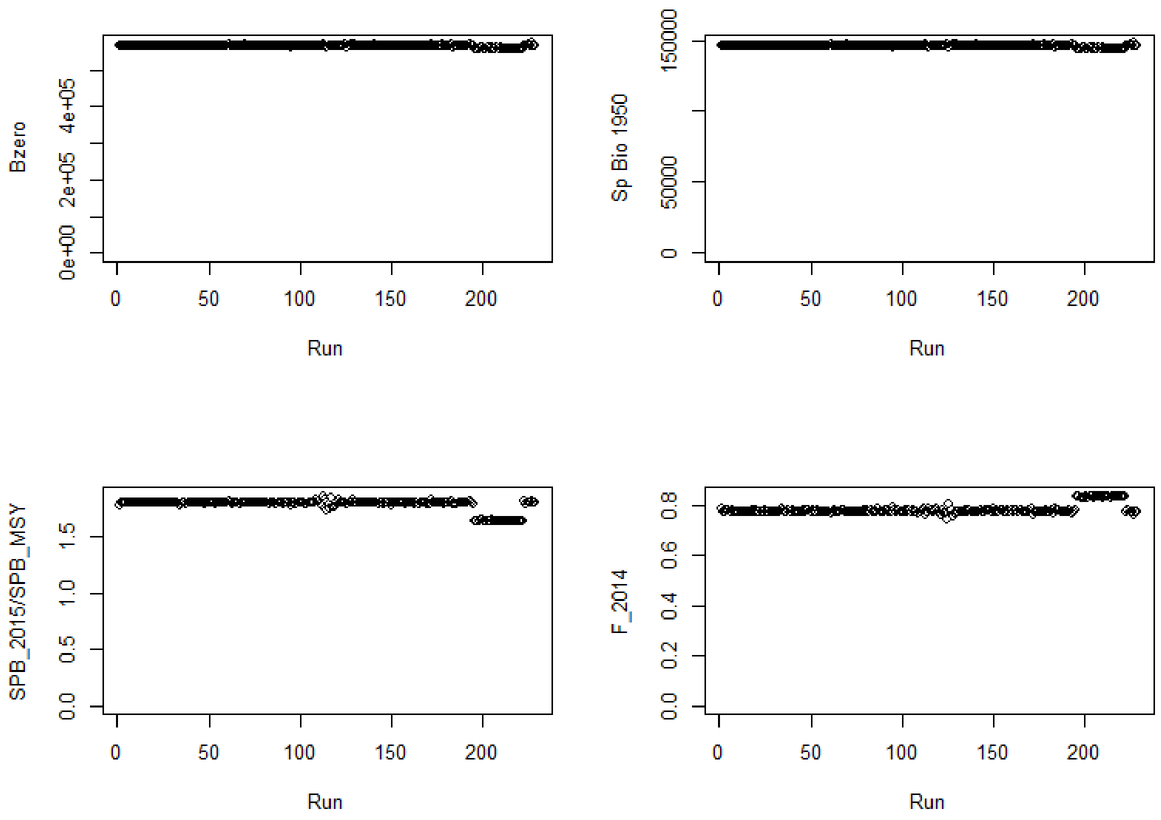


Figure 6: Run and effect on Bzero when we take one point out

Table 2: Prediction error and Bias on Albacore

| | |
|---|---|
| MSE (Prediction/Jack-knife) | 0.165 |
| MSE (Assessment) | 0.104 |
| Bias | 0.052 |

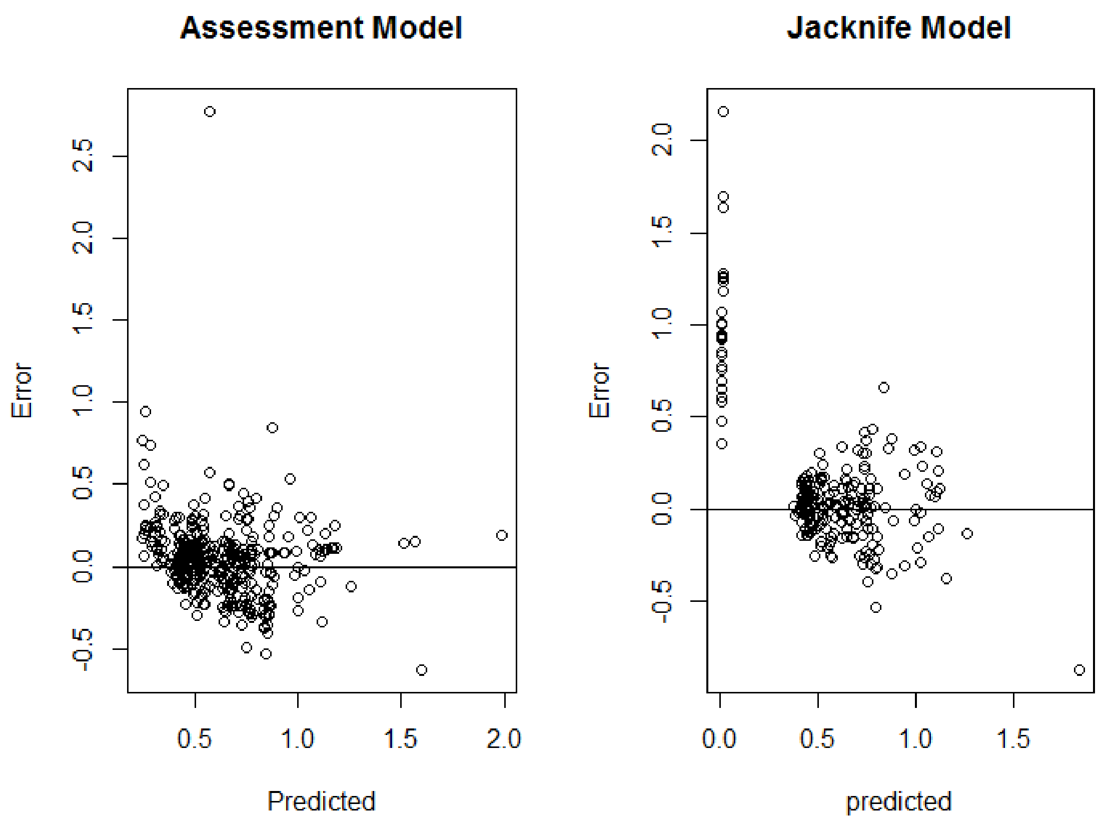


Figure 7: SS-III fitted model on residuals and jacknifed prediction residuals for the missing points.

**Conclusions**

The jacknife with prediction capability is a good indicator of model performance, and whether it tends to over-fit the data due to the high number of parameters. However, in bot cases the residual diagnostics of the predicted versus the error indicates that the model error or bias is fairly minimum. If we use another statistic to calculate bias, the Bluefin model indicates that in general the fitted model is underestimating the overall fit or has a negative bias of -9% whereas the albacore model is overestimating it by 5%.

**References**