SPICT, CMSY and BSM Comparisons for Celtic Sea and North Sea ICES stocks

Paul Bouch

11 July 2018

# CMSY, BSM and SPICT

All three assessment methodologies are surplus production models that are suited for data poor stocks. They require a catch time series for the stock. CMSY requires an estimate of resilience of the stock, and some prior ranges for biomass. If these ranges of biomass are not specified, there are default rules that use the catch data to determine a starting, intermediate and final biomass range.

CMSY then lools for viable pairs of r and k, that prevent the stock from crashing and maintain the biomass within the biomass ranges at the intermediate and end point. From the viable pairs tested, the most likely r and k estimates are calculated from the tip of the triangle where r is highest.

The BSM (Bayesian Surplus Model) is a modified version of CMSY that utilises an abundance index, either in the form of a biomass, CPUE or survey timeseries. It will use the estimate of resilience to create an r prior. It will also use the biomass ranges from CMSY to create a k prior, and will create a prior for q based on the effort timeseries. So from the data, BSM will create and use priors for r, k and q when fitting the model. It is not possible to fit BSM without these derived priors.

SPICT is similar to BSM in that it fits a surplus production model to the data, yet it is more flexible in the data requirements. SPICT does not assume a symmetrical surplus production curve. It can be fixed as n=2, which makes the curve symmetrical and analogous to the Schaefer model and identical to BSM. Priors for r, k and q can be used by the SPICT model, so we can test the model against BSM and CMSY using the same priors that are derived by the Froese methodologies.

SPICT has the additional option of using multiple abundance/survey indices, and for the 20 stocks assessed this is an option for 13 of them.

# The Data

The stocks used for the comparison are ICES data rich stocks from the Celtic, Irish and North Seas. The data is from the most recent assessment (2017 or 2018) and was obtained from the relevant working group or from stockassessment.org

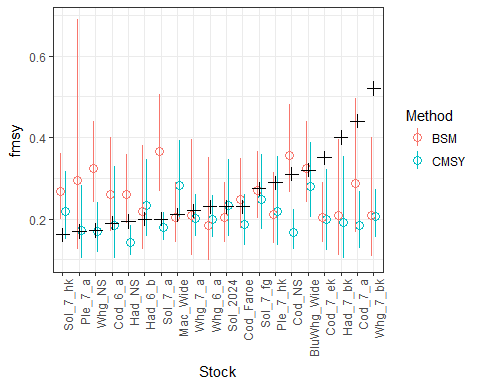
# Questions:

1. How does BSM and CMSY compare?
2. Should SPICT fix n=2?
3. Which priors of r, k and q are useful?
4. Does adding multiple indices improve SPICT?
5. Of all the methods and all the options, which is the best for the different reference points?
6. How sensitive are CMSY and BSM to incorrect starting assumptions?

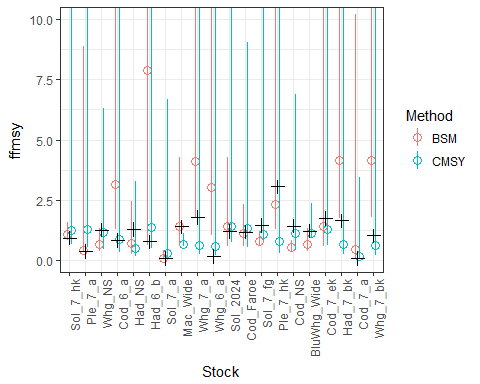
# Comparison of CMSY and BSM

CMSY uses the catch data and an estimate of resilience to perform the assessment. If not provided, as in this case, it will also make some assumptions about biomass reference points.

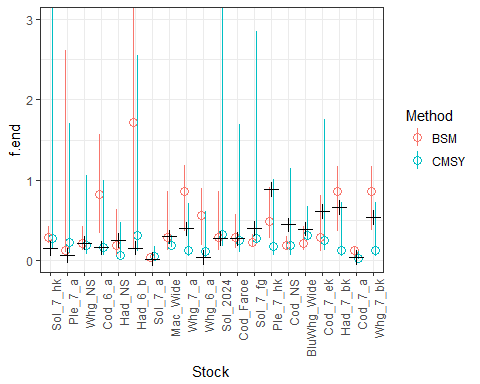
BSM uses the catch data, an estimate of resilience and also an estimate of abundance, in this case a survey index. If not provided it also makes assumptions about the reference points for r, k, and q.



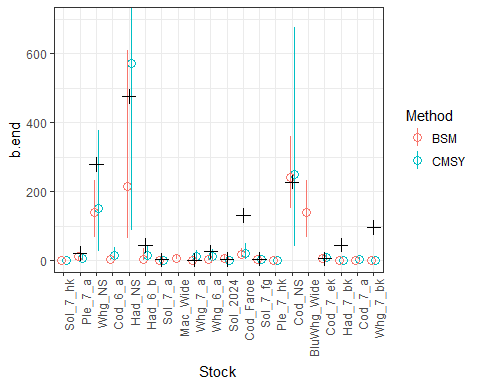
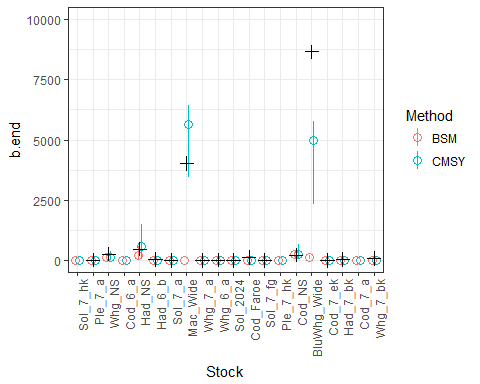
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | Fmsy\_ARE\_mean | Fmsy\_ARE\_median | Fmsy\_MSE\_mean | Fmsy\_MSE\_median |
| BSM | 14 | 33.549 | 30.634 | 0.014 | 0.005 |
| CMSY | 14 | 23.787 | 17.372 | 0.014 | 0.002 |



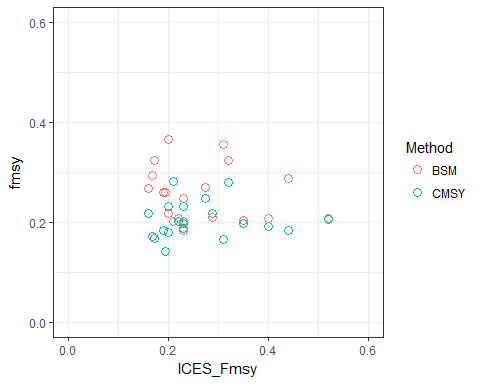
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | F\_Fmsy\_ARE\_mean | F\_Fmsy\_ARE\_median | F\_Fmsy\_MSE\_mean | F\_Fmsy\_MSE\_median |
| BSM | 14 | 134.285 | 50.688 | 4.409 | 0.359 |
| CMSY | 16 | 32.709 | 31.801 | 0.537 | 0.145 |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | F\_ARE\_mean | F\_ARE\_median | F\_MSE\_mean | F\_MSE\_median |
| BSM | 13 | 185.778 | 55.145 | 0.198 | 0.032 |
| CMSY | 17 | 71.966 | 62.486 | 0.068 | 0.016 |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | B\_ARE\_mean | B\_ARE\_median | B\_MSE\_mean | B\_MSE\_median |
| BSM | 5 | 68.114 | 74.957 | 5545899 | 1129.902 |
| CMSY | 6 | 84.982 | 58.025 | 1011252 | 625.597 |

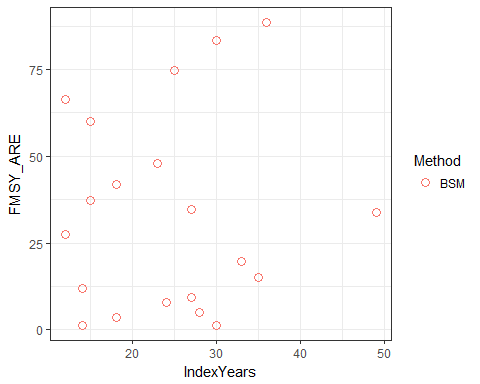


##   
## Pearson's product-moment correlation  
##   
## data: BSM$fmsy and BSM$ICES\_Fmsy  
## t = -0.72038, df = 18, p-value = 0.4805  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.567855 0.297132  
## sample estimates:  
## cor   
## -0.1673997

##   
## Pearson's product-moment correlation  
##   
## data: CMSY$fmsy and CMSY$ICES\_Fmsy  
## t = 0.25746, df = 18, p-value = 0.7997  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.3924692 0.4899592  
## sample estimates:  
## cor   
## 0.06057133

So the results show that the BSM method doesn’t really do that well when the data provided is from a survey index. The BSM can only use one survey index, so there isn’t too much to fit the model with. The default assumptions, that the CMSY method uses, seem to work pretty well and for most reference points give reasonable results.

Is there a relationship between timeseries length of the abundance index and the Fmsy error?



##   
## Pearson's product-moment correlation  
##   
## data: BSM$fmsy and BSM$IndexYears  
## t = 1.4383, df = 18, p-value = 0.1675  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.1415776 0.6685890  
## sample estimates:  
## cor   
## 0.3210578

So no relationship between the absolute relative errors of the reference points and the length of the index timeseries. So even with a great survey index timeseries, would you ever use BSM rather than CMSY???

# SPICT

# Should SPICT fix n=2?

Spict gives the option to fix the shape of the production curve or it can either be estimated by the model. By fixing n=2, the surplus production model is equivalent to the Schaefer model, and matches with CMSY, which assumes n=2. Should we force n=2 or should we allow SPICT to parameterise it?

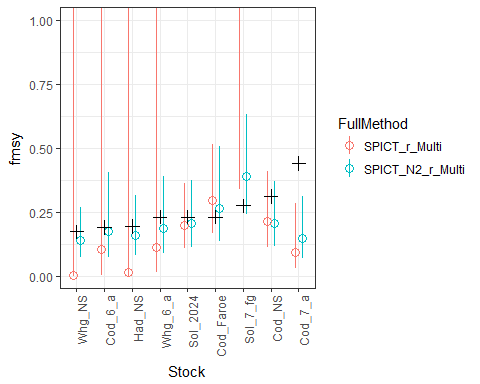
We can compare all the converged models and see the accuracy for the reference points, and also how succesful the models were at converging.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| FixN2 | CI\_Match | Fmsy\_mean | Fmsy\_median | F\_Fmsy\_mean | F\_Fmsy\_median | F\_mean | F\_median | B\_mean | B\_median | Model\_Convergence | Percent |
| N | 80 | 72.53 | 42.21 | 95.01 | 88.39 | 292.51 | 85.29 | 778.20 | 86.13 | 112 | 67.88 |
| Y | 73 | 235.41 | 32.44 | 80.75 | 88.25 | 184.32 | 89.39 | 844.18 | 87.91 | 143 | 86.67 |

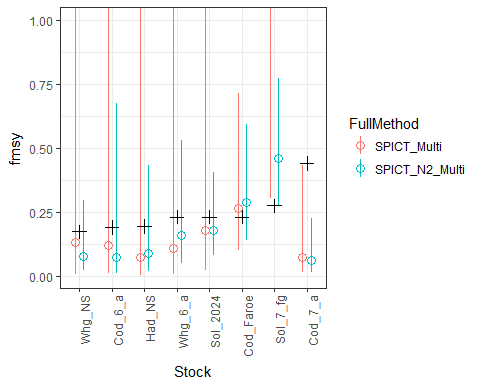
By fixing n=2, the results are closer to ICES reference points and also the model is more likely to converge. Considering that when we stipulate n=2, we are more likely to get convergence of the model when the data is poor and the results inaccurate, the benefit of fixing n=2 is greater than shown by these figures.

To get around that, we can look at pairs of scenarios and can exclude stocks where one or other model doesn’t have succesful convergence.

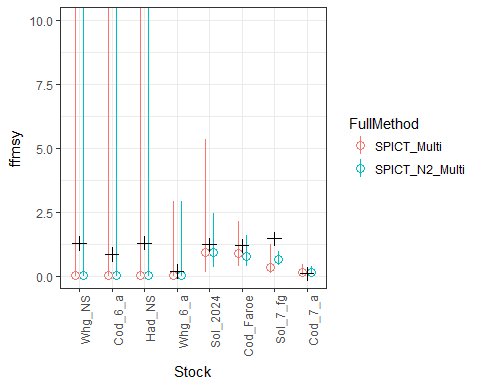
For example…



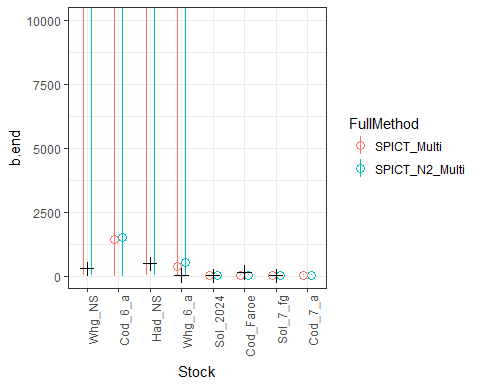
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | Fmsy\_ARE\_mean | Fmsy\_ARE\_median | Fmsy\_MSE\_mean | Fmsy\_MSE\_median | Model\_Convergence | Percent |
| SPICT\_N2\_r\_Multi | 8 | 25.841 | 18.531 | 0.013 | 0.001 | 13 | 100.000 |
| SPICT\_r\_Multi | 7 | 84.137 | 51.298 | 0.108 | 0.014 | 9 | 69.231 |



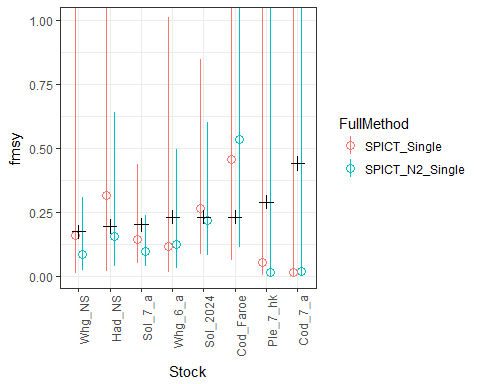
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | Fmsy\_ARE\_mean | Fmsy\_ARE\_median | Fmsy\_MSE\_mean | Fmsy\_MSE\_median | Model\_Convergence | Percent |
| SPICT\_Multi | 6 | 82.59 | 45.753 | 0.144 | 0.01 | 8 | 61.538 |
| SPICT\_N2\_Multi | 7 | 50.31 | 54.960 | 0.028 | 0.01 | 11 | 84.615 |



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | FFmsy\_ARE\_mean | FFmsy\_ARE\_median | FFmsy\_MSE\_mean | FFmsy\_MSE\_median | Model\_Convergence | Percent |
| SPICT\_Multi | 7 | 71.086 | 92.051 | 0.674 | 0.386 | 8 | 61.538 |
| SPICT\_N2\_Multi | 7 | 68.352 | 91.503 | 0.603 | 0.420 | 11 | 84.615 |



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | B\_ARE\_mean | B\_ARE\_median | B\_MSE\_mean | B\_MSE\_median | Model\_Convergence | Percent |
| SPICT\_Multi | 4 | 2445.917 | 679.039 | 229320613 | 59518.14 | 8 | 61.538 |
| SPICT\_N2\_Multi | 5 | 2599.944 | 1033.370 | 235292043 | 135793.99 | 11 | 84.615 |

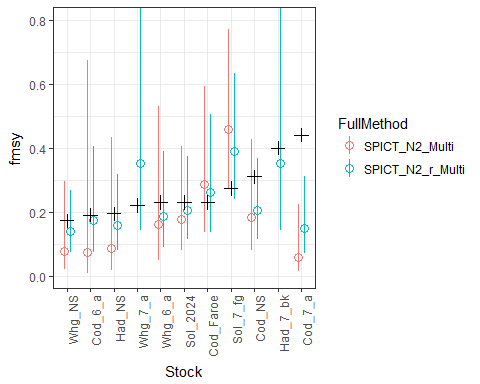
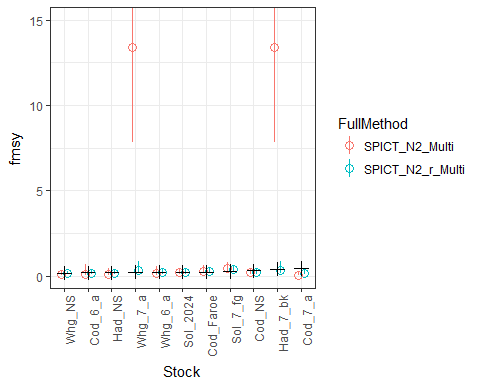


|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | Fmsy\_ARE\_mean | Fmsy\_ARE\_median | Fmsy\_MSE\_mean | Fmsy\_MSE\_median | Model\_Convergence | Percent |
| SPICT\_N2\_Single | 8 | 62.326 | 51.749 | 0.047 | 0.011 | 16 | 80 |
| SPICT\_Single | 8 | 55.309 | 56.768 | 0.040 | 0.014 | 8 | 40 |

Adding in the n=2, improves the convergence success of the models, as well as improving the accuracy and size of the confidence limits. Let’s assume that we should alway fix the surplus production curve with n=2.

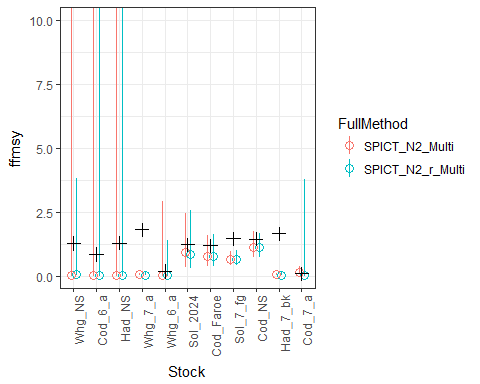
# SPICT - Effects of using an r prior

The CMSY and BSM methods require an estimate of resilience which is converted into an r prior. If we use the same prior r prior for SPICT does it improve the estimates?

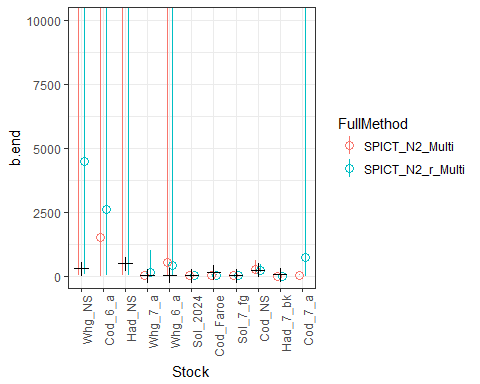


|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | Fmsy\_ARE\_mean | Fmsy\_ARE\_median | Fmsy\_MSE\_mean | Fmsy\_MSE\_median | Model\_Convergence | Percent |
| SPICT\_N2\_Multi | 8 | 878.910 | 55.260 | 31.067 | 0.014 | 11 | 84.615 |
| SPICT\_N2\_r\_Multi | 10 | 27.676 | 18.531 | 0.012 | 0.002 | 13 | 100.000 |

Using an r prior improves the estimates of Fmsy, and prevents some Fmsy estimates that are very large (>10). How does setting an r prior affect the other reference points?



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | FFmsy\_ARE\_mean | FFmsy\_ARE\_median | FFmsy\_MSE\_mean | FFmsy\_MSE\_median | Model\_Convergence | Percent |
| SPICT\_N2\_Multi | 8 | 67.963 | 94.374 | 0.955 | 0.665 | 11 | 84.615 |
| SPICT\_N2\_r\_Multi | 8 | 69.885 | 96.702 | 0.969 | 0.655 | 13 | 100.000 |

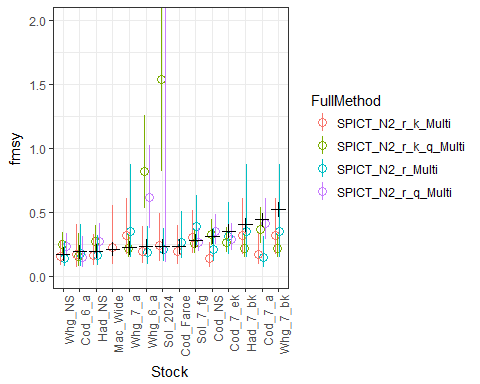


|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | B\_ARE\_mean | B\_ARE\_median | B\_MSE\_mean | B\_MSE\_median | Model\_Convergence | Percent |
| SPICT\_N2\_Multi | 7 | 1746.471 | 80.467 | 156861671 | 2021.941 | 11 | 84.615 |
| SPICT\_N2\_r\_Multi | 6 | 1723.357 | 100.000 | 67975073 | 10761.514 | 13 | 100.000 |

The broad r prior can easily be estimated, and the benefits seem clear.

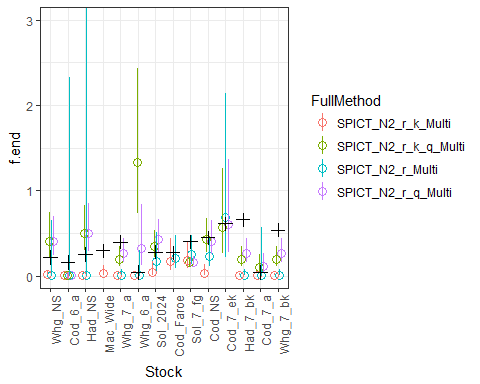
# SPICT - Effect of priors for k and q

By default, CMSY and BSM also use the data sets to estimate priors for k and q. Does using these, or a combination of them improve the reference points derived from SPICT.

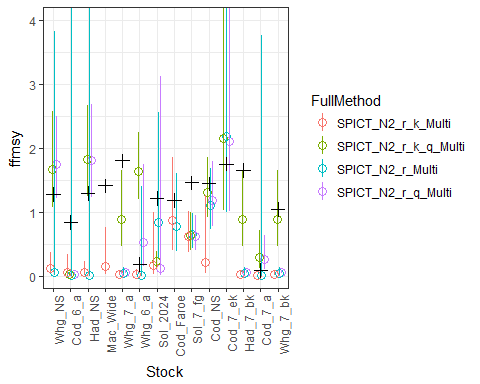


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | Fmsy\_ARE\_mean | Fmsy\_ARE\_median | Fmsy\_MSE\_mean | Fmsy\_MSE\_median |
| SPICT\_N2\_r\_k\_Multi | 11 | 24.039 | 15.517 | 0.013 | 0.001 |
| SPICT\_N2\_r\_k\_q\_Multi | 7 | 90.027 | 32.272 | 0.183 | 0.006 |
| SPICT\_N2\_r\_Multi | 12 | 26.748 | 18.531 | 0.013 | 0.002 |
| SPICT\_N2\_r\_q\_Multi | 8 | 458.781 | 38.548 | 5.140 | 0.005 |

Looking at Fmsy, adding in the q prior makes assessments less accurate. Depending on how you judge accuracy, adding in the k prior sometimes improves the Fmsy accuracy, but it’s not conclusive.

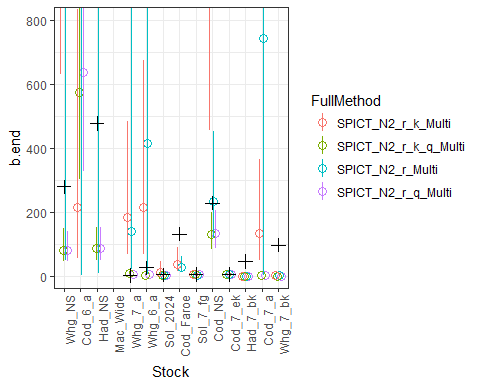


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | F\_ARE\_mean | F\_ARE\_median | F\_MSE\_mean | F\_MSE\_median |
| SPICT\_N2\_r\_k\_Multi | 2 | 86.360 | 93.546 | 0.103 | 0.058 |
| SPICT\_N2\_r\_k\_q\_Multi | 3 | 323.263 | 67.629 | 0.186 | 0.038 |
| SPICT\_N2\_r\_Multi | 9 | 72.404 | 96.741 | 0.082 | 0.025 |
| SPICT\_N2\_r\_q\_Multi | 4 | 117.088 | 60.064 | 0.045 | 0.030 |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | FFmsy\_ARE\_mean | FFmsy\_ARE\_median | FFmsy\_MSE\_mean | FFmsy\_MSE\_median |
| SPICT\_N2\_r\_k\_Multi | 1 | 82.025 | 94.415 | 1.178 | 1.104 |
| SPICT\_N2\_r\_k\_q\_Multi | 6 | 56.448 | 65.021 | 0.547 | 0.435 |
| SPICT\_N2\_r\_Multi | 9 | 75.680 | 96.702 | 0.912 | 0.655 |
| SPICT\_N2\_r\_q\_Multi | 6 | 76.195 | 84.932 | 0.829 | 0.469 |

When assessing current fishing effort and the relative F/Fmsy all methods generally perform similiarly well. Without using the q prior, fishing mortality often seems to be low, compared to high when the q prior is utilised. There are issues using a q prior for SPICT when using multiple abundance indices because it uses that same prior for all of them, when that might not be appropriate.

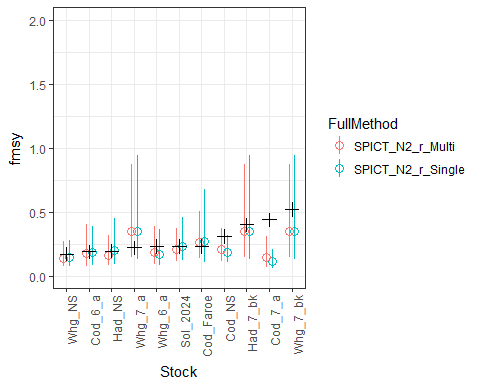


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | B\_ARE\_mean | B\_ARE\_median | B\_MSE\_mean | B\_MSE\_median |
| SPICT\_N2\_r\_k\_Multi | 2 | 1234.694 | 538.636 | 93518176.03 | 32605.536 |
| SPICT\_N2\_r\_k\_q\_Multi | 2 | 91.492 | 76.420 | 21300.32 | 1266.089 |
| SPICT\_N2\_r\_Multi | 7 | 1421.726 | 99.854 | 55616806.03 | 9204.476 |
| SPICT\_N2\_r\_q\_Multi | 2 | 79.754 | 76.871 | 21253.63 | 1253.445 |

The addition of the q prior makes the estimates worse. Adding in k, does not improve the estimates either, so in my opinion the best all round SPICT variant uses n=2 and had a broad r prior. The k and q priors generated by CMSY\_BSM are not useful, and in the case of q is a negative.

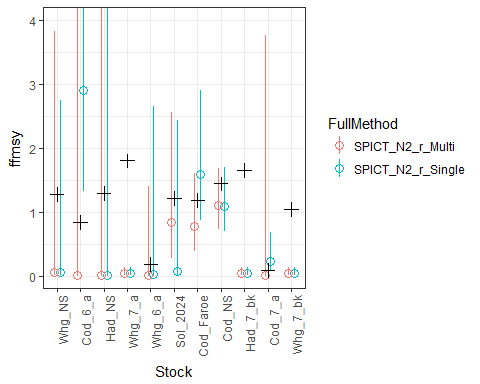
# Multiple indices?

Of the 20 stocks, 13 of them have multiple indices. Does using all the available data improve the assessment? This is just comparing the difference when we use n=2 and the r prior.

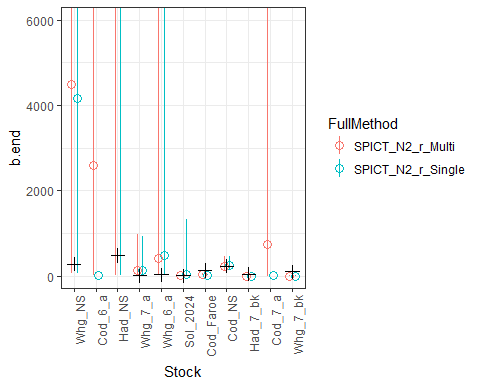


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | Fmsy\_ARE\_mean | Fmsy\_ARE\_median | Fmsy\_MSE\_mean | Fmsy\_MSE\_median |
| SPICT\_N2\_r\_Multi | 10 | 26.792 | 18.531 | 0.014 | 0.002 |
| SPICT\_N2\_r\_Single | 10 | 26.436 | 17.966 | 0.016 | 0.003 |

So it looks like the additonal indices to slighty improve the assessment.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | FFmsy\_ARE\_mean | FFmsy\_ARE\_median | FFmsy\_MSE\_mean | FFmsy\_MSE\_median |
| SPICT\_N2\_r\_Multi | 8 | 75.469 | 96.702 | 1.001 | 0.693 |
| SPICT\_N2\_r\_Single | 7 | 91.526 | 96.117 | 1.427 | 1.308 |



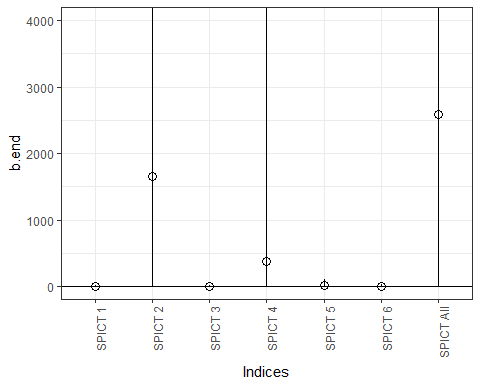
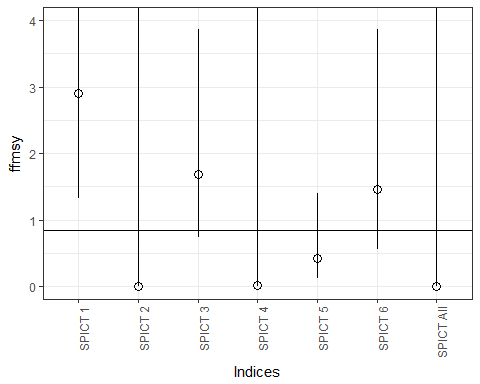
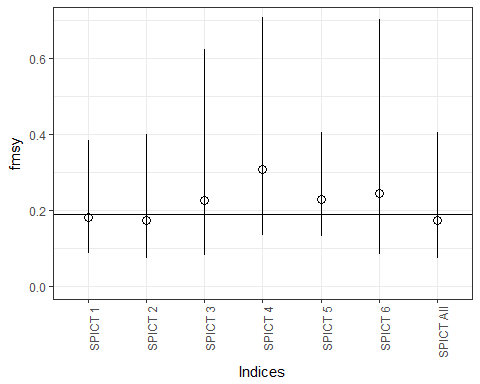
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| FullMethod | CI\_Match | B\_ARE\_mean | B\_ARE\_median | B\_MSE\_mean | B\_MSE\_median |
| SPICT\_N2\_r\_Multi | 5 | 1731.094 | 100.000 | 67976096 | 10761.51 |
| SPICT\_N2\_r\_Single | 5 | 1975.568 | 645.204 | 144906051 | 13816.58 |

Why are the f/fmsy and the f.end values so low? It seems that the abundance indices used have a big impact on the relative reference points. Perhaps more to follow in that regard!

Multiple indices can improve the assessment, but the effect can also be detrimental. Use any index with caution!

# The importance of the abundance Index

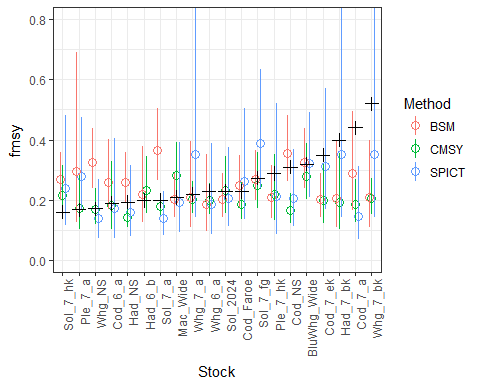
The abundance index used in the SPICT assessment will make a big difference in the calculated reference points. Cod in 6a has 6 abundance indices that are used by ICES in their assessment. In the previous assessment, all the indices are used, or if using just one index, then the longest and most recent is selected.



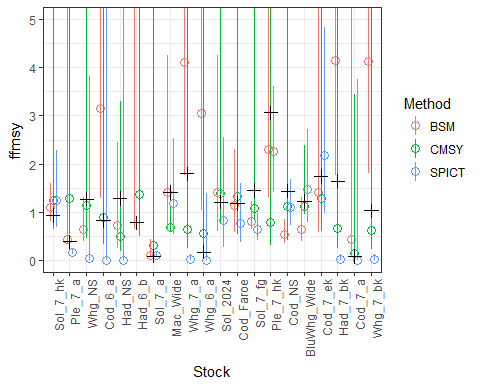
The variation in refernce points is crazy! Seems to me that the abundance indices data is too unreliable to contribute to either BSM or SPICT and by trying to use it, we are taking a really big chance with the results!

# Best Data Poor Method

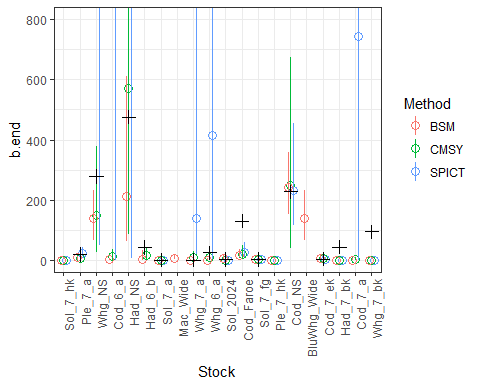
Which of CMSY, CMSY with BSM and SPICT estimations of fishieries reference points match most closely with ICES assessments?



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Method | CI\_Match | Fmsy\_ARE\_mean | Fmsy\_ARE\_median | Fmsy\_MSE\_mean | Fmsy\_MSE\_median |
| BSM | 14 | 33.549 | 30.634 | 0.014 | 0.005 |
| CMSY | 14 | 23.787 | 17.372 | 0.014 | 0.002 |
| SPICT | 18 | 27.806 | 19.503 | 0.010 | 0.002 |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Method | CI\_Match | FFmsy\_ARE\_mean | FFmsy\_ARE\_median | FFmsy\_MSE\_mean | FFmsy\_MSE\_median |
| BSM | 14 | 134.285 | 50.688 | 4.409 | 0.359 |
| CMSY | 16 | 32.709 | 31.801 | 0.537 | 0.145 |
| SPICT | 14 | 72.245 | 94.906 | 0.671 | 0.161 |



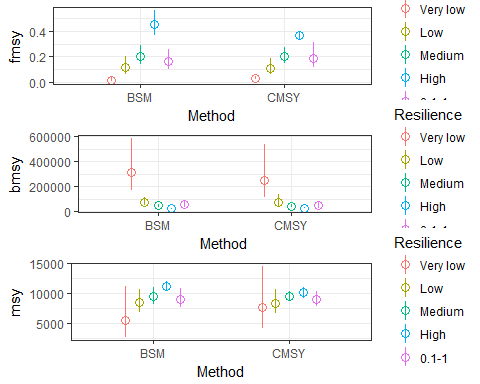
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Method | CI\_Match | B\_ARE\_mean | B\_ARE\_median | B\_MSE\_mean | B\_MSE\_median |
| BSM | 5 | 68.114 | 74.957 | 5545899 | 1129.902 |
| CMSY | 6 | 84.982 | 58.025 | 1011252 | 625.597 |
| SPICT | 10 | 1051.044 | 67.406 | 43119634 | 9204.476 |

Seems like CMSY gives the best estimates although SPICT is close. SPICT struggles way more for the reference points that are not Fmsy.

My feeling is that SPICT would perform at least as well as CMSY if the input data was good, ie that we had long time series that we had good confidence in. The reality is that often the abundance time series is less than 20 years, and when we use additional abundance indices, the time series are often irregular and showing differing trends. It is worth remembering that this data is from data rich stocks and is used as tuning data for age specific models. These are the best available data sets, yet using them rather than just CMSY really does not improve the assessments!!!

# CMSY Sensitivity to Resilience Assumption

Cod is asssumed to have a medium resilience (r between 0.2 and 0.8), but what are the effects of differing resilience assumptions for the stock? If we got the resilience assumption wrong for a stock, how much do the results vary for both the CMSY and BSM methods?



fmsy is the most sensitive to incorrect assumptions. The only real differences occurr at the extremes of the resilience scale, where it should be easier to make the correct assumption.