Northwest Atlantic Fisheries Organization

Serial No. N5721 **SCR Doc. 09/ 060**

NAFO/ICES *PANDALUS* ASSESSMENT GROUP MEETING – OCTOBER 2009

Precisions for biomass-index series in fitting a stock-production model of the dynamics of the West Greenland shrimp stock by Bayesian methods.

by

Michael C. S. Kingsley
Pinngortitaleriffik, Greenland Institute of Natural Resources
Box 570, DK-3900 Nuuk, Greenland

Abstract

The West Greenland Stock of *Pandalus borealis* is assessed from indices of biomass and biomass density based on catch and effort data from commercial fishing fleets and a research trawl survey, catch data, and information on stock demographics and on the distribution of the stock as revealed by fishery logbooks. The assessment framework incorporates a logistic stock-recruitment model, fitted by Bayesian methods, that uses CPUE, survey and catch data and includes a term for predation by Atlantic cod, using available series of cod biomass. Results include a modelled trajectory for the stock biomass.

The modelled trajectory has had a tendency to follow the standardised CPUE series closely and to weight the survey series more lightly. There was a possibility that this occurred by carryover from run to run, as converged values from one run are usually input as starting values to another.

This was investigated by forcing the model to follow the survey series more and more closely, recording the effects of this constraint on the values of other precision settings. Finally, the constraint was removed and the model again allowed to choose its own settings for precisions. It then returned to following the CPUE series. This showed that this behaviour could not be ascribed to starting values.

Other precisions were little affected until the survey c.v. was reduced to below about 15%. Then the process error started to increase quite markedly, and at the same time the error associated with cod predation decreased.

Background

The stock–production model now in use for the assessment of the stock of the northern shrimp (Pandalus borealis) on the West Greenland shelf uses two series of biomass indicators. One is a series of CPUE values going back to 1976. It includes data from 4 fleets: the KGH fleet, the offshore fleet, the coastal fleet, and the Canadian fleet. The KGH fleet comprised a small number—6 or 7—of ocean-going trawlers that had a practical monopoly of the fishery in its earliest days. The offshore fleet of ocean-going trawlers has replaced it; it is restricted geographically to grounds more than 3 n.mi. outside the baseline. The coastal fleet can fish in these inshore areas, but is nominally restricted to vessels under 80 tons. The Canadian fleet fishes on the small ground where the West Greenland shelf extends west of the midline. These four CPUE series are combined into a single unified series for input to the model. (The statistical properties of the resulting series have not been investigated and are poorly known.)

The second series is a series of swept-area estimates obtained from a trawl survey that is designed expressly to estimate shrimp biomass. The survey has been executed every year since 1993. It covers waters from 150 to 600 m deep; it is stratified both geographically and by depth; it has always been semi-systematic in nature, although different methods have been used at different times to avoid the occasional closely-spaced stations that are a consequence of complete spatial randomness in sampling.

Other inputs to the model do not directly indicate biomass level, but rather biomass change: they are a series of fishery catches, assumed error-free, and series of cod (predator) biomass estimates, which are modelled as affecting the biomass through a specified predation function.

The CPUE series and the survey series had good correlation up to the end of the 1990s, both showing considerable increase from the values of the late 1980s and early 1990s. Then the survey series took a sudden jump upwards to very high values in 2003–2004, not mirrored in a more modest increase in catch rates, after which it has continuously decreased until in 2009 it is less than half the peak value; this decrease has also not been mirrored in catch rates, which have been maintained at high levels.

Problem

When the two biomass series are input to the stock-production model with equally uninformative prior distributions assigned to their precisions, the stock trajectory has always had, a tendency to track the CPUE series closely, giving it very high precision, and to assign a low precision to the survey series, essentially ignoring it. The precisions of the stock-dynamic process and of the predation function are then, apparently, also fitted round the requirement for the trajectory to track the CPUE (Table 1).

This propensity for stock-dynamic models to pick a favourite from among the data series offered and to scorn the rest is a recognised problem. In the present case the model is fitted to the data series by Bayesian methods using WinBUGS and, being large and complex, is difficult to start from scratch; there is therefore a tendency to carry over starting values from one formulation of the model, and one year's assessment, to the next. So every run has starting values that conform to the CPUE series, and outputs that do the same. This raised the question of whether, if a run were offered starting values that instead conformed to the survey series, it would switch over to producing outputs that gave the survey high precision and ignored the CPUE.

A related problem was what to do about the tendency of the model to produce precisions for the CPUE series that were much higher than was reasonable. We don't believe that CPUE values really have precisions of a few percent, as the model estimates. However, we have very little idea of whether the precisions estimated by the model for the predation function and for the stock-dynamic process itself are reasonable, having no extrinsic estimate of their variability.

Enquiry

I first investigated whether the model's habit of tracking the CPUE closely and taking little notice of the survey series was due to starting values that pointed the model in that direction. I did this by setting fixed values for the survey precision approximately equivalent to error coefficients of variation (c.v.s) of 20, 15, 10, 6, 4, and 2.5%, so as to force the model to fit stock trajectories that would stick more and more closely to the survey series. At each value of survey precision the model was run for a few tens of thousands of iterations and its state saved as starting values for the next run with a smaller c.v. for the surveys. In this way the model was driven over to the condition of being required to heed the survey series closely and to adjust its attitude to the other series accordingly, as best it could, in respect of both the precision parameters and the stock-size trajectory. After running with the survey precision equivalent to a c.v. of 2.5%, the model was returned to its first state with a free choice of precisions for all series.

The second enquiry was whether the model would run if equal precisions were specified for the two biomass index series. So a single precision parameter was set up with the same uninformative prior as the others, and the precisions of both CPUE and survey were set equal to it.

This second trial was unsatisfactory in that it neither allowed the model to look at the data, nor the user to look at the reality: it simply specified equal precision for two very different series. So a third trial was carried out in which the same single precision parameter was specified for 'biomass series', but CPUE and survey could have different parts of it. This was done by splitting the precision between CPUE and survey, and allowing the model to fit the split. Two moderately informative prior distributions were tried for the split: beta(8,8) and beta(4,4).

The principle is to provide a single precision for biomass indices, and allow the model to divide this precision between the survey and the CPUE. When the model was run with independent precisions for these two series, both were given the same uninformative prior, a gamma distribution with shape parameter 0.02 and rate 0.006. A multiple of a gamma distribution is a gamma distribution with the same shape, so allowing the model to split a single precision parameter between the two series still gave each an uninformative gamma prior. Clearly, a very uninformative prior for the split parameter would be the same as allowing the two series to have separate priors. A moderately informative prior for the split—a beta(8,8) distribution—and a moderately uninformative prior—beta(4,4) were tried. Both are symmetrical—consistent with no preconceptions about which series deserves to be considered the more precise indicator of stock size—but they don't have to be.

Results

Precisions for survey series set to fixed values.

With precisions for the survey series specified at progressively increasing values, the model ran quite orderly. The deviation of the survey and CPUE series meant that as survey c.v. was reduced, the CPUE c.v. necessarily increased. However, the survey results have at times undertaken some considerable excursions, and therefore as the survey c.v. was screwed down, the process error found itself increasing. With CPUE and process errors thus being required to take larger values, a major beneficiary was the predation function. The period of high cod biomass largely antedates that covered by the survey, and once the model could no longer insist on high precision for the CPUE fit and was also allowing a large process error, the predation fit could take advantage of the extra slack and its error c.v. was much reduced.

Fixed error c.v. for survey (%)	c.v. CPUE	c.v. process	c.v. predation	MSY	P.present
Free (21.2)	3.6	9.4	44.4	148	1.28
20	3.5	9.3	45.3	148.5	
15	4.3	9.3	42.6	146	1.35
10	10.9	11.2	20.1	135.9	1.179
6	15.9	15.2	15.3	139	1.176
4	17.6	16.6	15.2	142	1.198
2.5	18.3	17.8	13.0	137.1	1.219

However, after the survey c.v. was screwed all the way down to 2.5% and the model once again set free to decide which series to choose by being given an uninformative prior for the survey precision, it returned immediately to its first state, i.e. following the CPUE very closely. This seems to show that this is not due to starting values directing the model which state to choose.

Precision for survey series set equal to that for CPUE.

With equal precisions, the model ran quite orderly and produced reasonable statistics.

c.v. CPUE and survey	c.v. process	c.v. predation	MSY	P.present
12.6 (2.9)	10.7	18.8	139	1.21

Median estimates of c.v.s (reciprocal of square root of precision): survey and CPUE series fitted with equal precisions

The median estimates of c.v. for both survey and CPUE were 12.6%. Process error was not much higher than when these precisions were free, while the predation error was much lower. The model cannot fit the stock trajectory closely to the survey and doesn't try very hard to do so, so the CPUE fit can also be somewhat relaxed. This allows the process error to stay in the same region as with free c.v.s for survey and CPUE, and the predation function to be fitted more closely.

Survey and CPUE precisions based on a split of a single biomass-index precision.

Both prior distributions tried for the split parameter were symmetrical, implying no preconceptions as to which biomass-index series was the more reliable; but both gave some push in the direction of equality of c.v. However, both, and more especially the more uninformative, have tails that extend to zero and to unity, so if the model had a very strong view that one of the biomass-index series was much more precise than the other it would be able to find posterior distributions for the split parameter and consequently for the two separate precisions that reflected that. However, it did not do that, but found in both cases fairly moderate posteriors. The survey precision was uncorrelated with the CPUE precision. As for the cases in which the survey c.v. was set to slightly lower values, the process error increased slightly, and for the informative prior on the split, which reduced the survey c.v. to a median value of 15.2%, the predation c.v. was substantially reduced. The posterior distributions of the cvs were strongly updated and well defined with interquartile ranges of only about 3 percentage points.

This method of handling survey and CPUE precisions has the clear advantage of allowing, or requiring, lots of discussion of the parameters of the prior distribution for the split parameter—which series, if either, should initially be considered more accurate, and by how much, and how definitely. This is clearly has more entertainment value than, as at present, allowing the model a free hand to make judgements that are then seen as questionable.

Prior for split	c.v. CPUE (%) (i.q.r.)	c.v. survey	c.v. process	c.v. predation	MSY	P.present
Beta(4,4)	7.6 (3.1)	17.1 (3.6)	9.1	34.4	144	1.25
Beta(8,8)	9.7 (2.6)	15.2 (3.2)	9.6	26.4	143	1.24

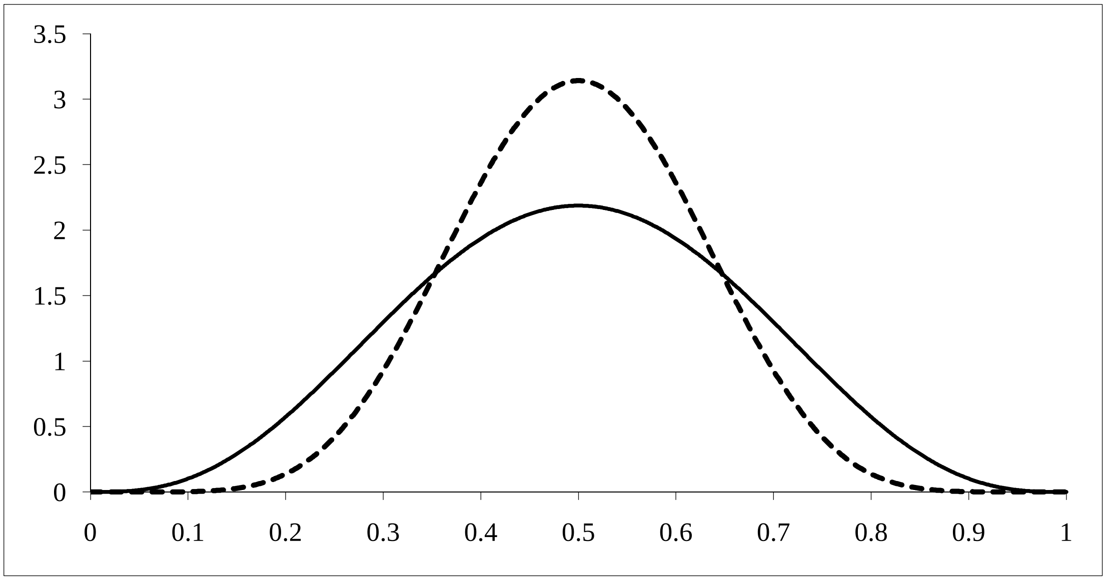

Figure 1: Beta(4,4) (solid line) and beta(8,8) (broken line) distributions.

Model formulations that reduced the survey c.v. and caused, or forced, the model to take more notice of that series reduced the estimated MSY. The fall in survey biomass index from its 2003 peak is associated with a period of high catches averaging over 150 000 t/yr from 2004 to 2008. If the assessment pays most attention to the CPUE index, which showed high catch rates to be maintained from 2005 through 2008, it concludes that the MSY is high enough to support such catches. If it pays more attention to the survey index—which casts doubt on whether such high catches could be sustained—it takes a less rosy view of the MSY.

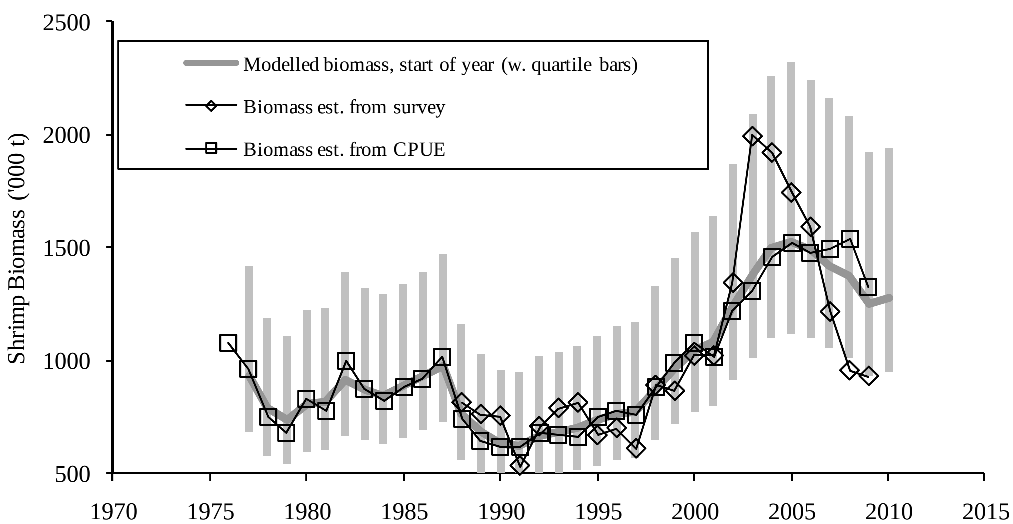

Figure 2: estimated stock trajectory with biomass-index precision split with beta(4,4) prior on the split.

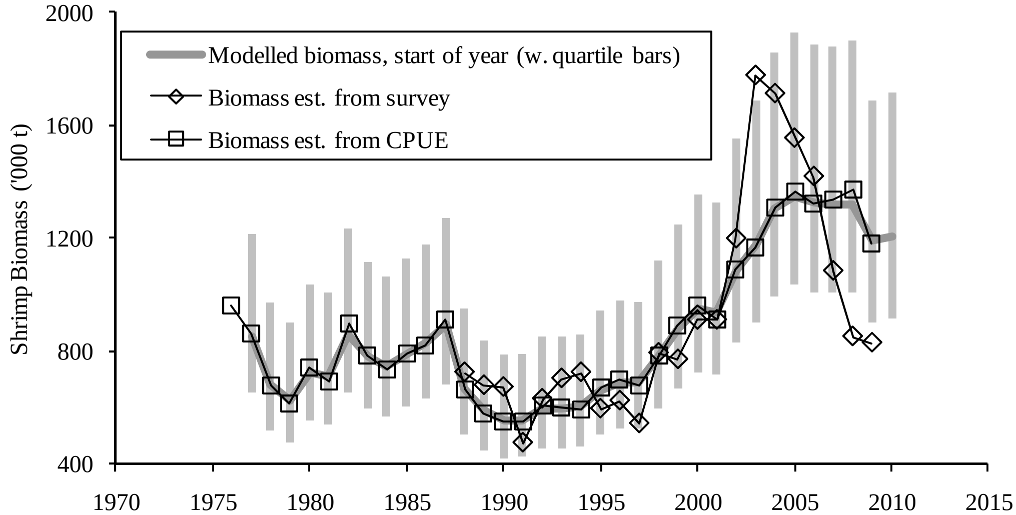

Figure 3: estimated stock trajectory when the model is free to estimate precisions for all series.