NOT TO CITED WITHOUT PRIOR
REFERENCE TO THE AUTHOR(S)

Northwest Atlantic Fisheries Organization

Serial No. N5201 **NAFO SCR Doc. 05/95**

SCIENTIFIC COUNCIL MEETING – OCTOBER/NOVEMBER 2005

Testing OGMAP as a Tool for Estimating Biomass and Abundance of Northern Shrimp

by

Katherine R. Skanes and Geoffrey T. Evans

Science Branch, Department of Fisheries and Oceans
P. O. Box 5667, St. John's, NL A1C 5X7, Canada
E-mail: SkanesK@dfo-mpo.gc.ca

Abstract

The OGive MAPping (OGMAP) method for estimating biomass (with confidence intervals) from surveys was tested on simulated northern shrimp populations designed to have statistical properties similar to those of autumn shrimp surveys in NAFO Divisions 3L, 3N and 3O. A variety of years of surveys were used both to create reference populations and to simulate samples from them. Proposed OGMAP confidence intervals (CIs) contained the reference biomass slightly less often than they should have (57% of the time for 60% CIs; 91% for 95% CIs), but were still usefully close. The CIs are also usefully narrower than those produced from stratified aerial expansion which has the false assumption of Gaussian distributions within a stratum.

Introduction

The biomass of Northern shrimp *Pandalus borealis* in NAFO Div. 3L, 3N, and 3O is estimated using data from stratified random bottom trawl surveys carried out in the autumn each year. The bottom trawl surveys randomly sample from strata designed for groundfish assessments. In the current standard NAFO analysis, the biomass and abundance of Northern shrimp are estimated using stratified aerial expansion calculations (Cochran, 1997), and confidence intervals for the integrated abundance are calculated assuming a Gaussian distribution. This can lead to confidence intervals that include negative values, particularly when there are one or two large catches during a survey.

OGive MAPping (OGMAP) (Evans *et al.*, 2000; Evans, 2000) is a proposed non-parametric method for estimating abundance or biomass with Monte Carlo confidence intervals. The probability distribution (pdf) is estimated at a set of points covering the whole of Div. 3LNO with high resolution and the expected value of the distribution is integrated to provide a biomass estimate in the region. Like the standard analysis, OGMAP assumes that each trawl set is an independent random sample from an underlying probability distribution at a given point. Unlike the standard analysis, it assumes that probability distribution for catch varies smoothly over space, so that observations have information about distributions at nearby points even in other strata. Also unlike the standard analysis, it does not assume that the pdf is Gaussian. The biomass, abundance and length frequencies are similar to those of stratified aerial expansion results but OGMAP can give tighter confidence intervals, especially when there are one or more exceptionally large catches. Details of the differences between stratified aerial expansion and OGMAP, as well as an example on deriving TACs from OGMAP, can be found in (Orr, *et al.*, 2004).

NAFO's Scientific Council meeting on shrimp in 2004 requested that sensitivity analysis be conducted on OGMAP. Is OGMAP an appropriate tool for calculating the biomass and abundance of Northern Shrimp in Div. 3L, 3N, and 3O? We address this question with the following steps:

- Create a reference population with known (stochastic) characteristics.
- Simulate repeated surveys from the reference population.
- For each simulated survey, compute confidence intervals for integrated biomass using OGMAP.
- Note what proportion of the computed confidence intervals contain the known total biomass of the reference population. It is hoped that approximately 60% of the computed 60% confidence intervals would contain the true value.

Reference Populations and Simulated Surveys

To try to capture the characteristics of a real population, we use the OGMAP analysis of a survey to define a reference population. This has exactly the form we need: it specifies the probability distribution for biomass concentration at every point in the area at which horizontal position and bottom depth are known. Bandwidths (characteristic distances for the influence of nearby points on the pdf) suitable for Div. 3LNO were taken to be 200 km horizontally, and 0.1 vertically on a scale of the square root of depth in meters. Reference populations were created for each Div. 3LNO autumn survey from 1995 to 2003.

Simulating (or indeed conducting) a survey has two steps: selecting the points at which the population will be sampled (the "design points") and taking a random sample at each design point. Variability and inaccuracy can enter at each step: design points can be unrepresentative of the region, and random samples from them can be unrepresentative of the pdf. In this study we use an existing survey (in general different from that used to construct the reference population) to provide the design points. This means that we do not investigate the full range of variability possible due to the choice of design points. From the nine years there were 81 combinations of reference population and survey design.

For each combination of reference population and survey design points, 1 000 surveys were simulated by sampling at random from the pdfs at the design points. Each simulated survey was analyzed with OGMAP (using the same bandwidths), and we noted whether the estimated 60% and 95% confidence intervals contained the true value for the reference population. When all 1 000 simulations were done we noted the fraction of, say, 60% confidence intervals that did contain the true value, and whether the fraction was indeed close to 60%.

Results

The following two Tables present the outcomes of the sensitivity testing on OGMAP. The first Table displays the percentage of outcomes for which the biomass estimate lies within the 60% confidence intervals. The second table displays the results for the 95% confidence intervals. The column headers indicate the year from which the reference population came from, this population provided the pdf at each point and the biomass estimate. The row headers indicate the year from which the design points came from; these are the survey points at which the reference pdf was resampled.

Table 1: Percentage of proposed 60% confidence intervals that contain the reference biomass.

Design Year	Reference Year								
	1995	1996	1997	1998	1999	2000	2001	2002	2003
1995	57.4	59.2	56.5	62.0	61.6	55.2	51.3	48.8	54.8
1996	54.3	58.7	52.1	60.0	59.7	55.8	55.7	55.7	57.7
1997	56.5	59.9	59.7	56.1	54.1	53.6	56.0	59.4	56.8
1998	56.9	58.4	57.0	59.1	61.9	50.0	57.9	58.0	53.7
1999	51.4	58.3	55.3	60.0	63.5	56.9	58.4	54.5	56.3
2000	55.1	59.4	51.5	54.2	61.4	48.1	55.0	52.5	55.1
2001	61.6	57.0	57.6	60.4	60.2	45.7	55.1	55.3	54.9
2002	56.7	61.1	56.2	60.2	59.7	51.0	53.4	56.3	61.3
2003	56.2	57.0	54.1	56.9	61.5	49.7	52.4	55.0	58.3

Table 2: Percentage of proposed 95% confidence intervals that contain the reference biomass.

Design Year	Reference Year								
	1995	1996	1997	1998	1999	2000	2001	2002	2003
1995	94.8	93.9	87.2	94.6	95.0	91.2	86.6	84.4	92.2
1996	93.6	94.0	84.9	95.1	93.4	89.9	89.8	88.3	91.7
1997	90.5	93.9	89.1	91.7	90.1	89.8	91.6	92.1	91.1
1998	93.2	94.2	90.8	94.0	95.5	88.2	90.9	91.8	88.9
1999	88.7	93.1	87.8	93.0	95.2	89.9	90.5	86.2	92.2
2000	91.6	92.6	84.4	90.7	95.3	84.7	89.3	85.0	90.1
2001	93.0	91.8	89.6	94.2	94.9	82.8	90.1	91.0	90.7
2002	94.0	93.7	86.0	94.2	94.5	90.2	89.5	88.9	93.3
2003	91.8	92.1	86.2	93.0	95.1	87.9	90.3	88.3	93.1

Overall, fewer then 57% of the proposed 60% confidence intervals contain the reference biomass. The deficiency is too large to have been a sampling accident, but we feel the results are usefully close. Reference populations from 1996, 1998 and 1999 provide the best agreement, and the samples that generated them were the least dominated by single large samples (the largest catch was less then 1.7 times the fifth largest for these years, and more then 2.4 times for all other years). Almost 91% of the proposed 95% confidence intervals contain the reference biomass.

Discussion

When we use OGMAP to create the reference population, the pdfs at every point are discrete, and in general will not contain all the values seen in the original survey. It might make more sense to create the reference population with a smooth pdf: say a delta-lognormal with the same probability of a zero catch, and the same mean and variance of non-zero catches, as that estimated by OGMAP.

Instead of simulating 1000 random samples at a fixed set of design points, it might make more sense to conduct 1000 simulations each consisting of first selecting design points at random (in some way that come close to respecting the original stratified random design) and then choosing random samples from them.

Although 95% confidence intervals were estimated we would not recommend them or place a lot of confidence in them. This is because we have little reason to think we know the upper tail of the pdf that well: we have no trusted theory and very little data from which to determine it. This is a comment not about OGMAP but about the world of shrimp distributions as revealed by surveys. The only reason standard methods think they can compute 95% confidence intervals is that they insert an extra (clearly false) assertion about the Gaussian shape of the distribution, from which one can compute the tail with unjustified confidence given enough observations near the centre.

The effect of bandwidth choice, and the range of bandwidths that can give about the same statistical performance, and how the results of OGMAP (for example total biomass estimates) vary among equally acceptable bandwidth combinations, has not been investigated yet.

References

Cochran, W. G. 1997. Sampling Techniques. Third Edition. John Wiley & Sons, Toronto, 428 p.

Evans, G. T. 2000. Local estimation of probability distribution and how it depends on covariates. *Can. Stock Adv. Sec. Res. Doc.*, 2000/120. 11 p.

Evans, G. T., D. G. Parsons, P. J. Veitch, and D. C. Orr. 2000. A local-influence method of estimating biomass from trawl surveys, with Monte Carlo confidence intervals. *J. Northw. Atl. Fish. Sci.*, 27: 133-138.

Orr, D. C., P. J. Veitch, and D. J. Sullivan. 2004. Ogive MAPping (ogmap) as an alternative means of estimating indices and setting TAC's. *NAFO SCR Doc.*, No. 04/80, Serial No. N5050.