Canadian Atlantic Fisheries
Scientific Advisory Committee

Comité scientifique consultatif des pêches
canadiennes dans l'Atlantique

CAFSAC Research Document 91/62

CSCPCA Document de recherche 91/62

The effects of set to strata allocation adjustments

by

Pierre Gagnon
Science Branch
Department of Fisheries and Oceans
Maurice-Lamontagne Institute
P.O. Box 1000, Mont-Joli (Québec)
G5H 3Z4

RÉSUMÉ

On étudie par simulation l'effet de différentes stratégies d'ajustement de l'allocation des traits dans les strates sur la justesse et la fidélité (précision) des valeurs estimées. Les modèles de populations utilisés sont la biomasse de la morue dans la division 3Pn-4RS de l'OPANO l'hiver et l'abondance de la morue dans 2J. Les lois de Poisson, log-normale et négative-binomiale sont utilisées pour modéliser la variabilité des populations dans chaque strate.

Les résultats démontrent que l'ajustement optimal des taux d'échantillonnage permet une augmentation significative de la fidélité des valeurs estimées. Par contre, les stratégies qui consistent à modifier l'allocation des traits en tenant compte de l'échantillonnage courant introduisent un biais dans les estimations du même ordre de grandeur que l'écart-type des valeurs.

ABSTRACT

The effect of different set to strata allocation adjustment strategies on the precision and accuracy of estimates is studied by simulation. The population models used are the cod biomass in NAFO division 3Pn-4RS and the cod abundance in 2J. The Poisson, log-normal and negative-binomial distributions are used to model the population variability within each stratum.

The results show that the optimal adjustment of sampling fractions gives significant increase in the precision of estimated quantities. They also show that the strategies that involve adjusting the allocation of sets based on the results of the current survey introduce a bias in the estimates of the same order of magnitude as their standard errors.

Introduction

CAFSAC stock assessments often rely on catch per unit effort data from research surveys to estimate (tune) missing parameters in age-structured population models. The research survey CPUE data have a better consistency over the years than commercial data because the fishing methods are well documented, closely monitored and kept constant. However, the high cost of conducting sampling surveys at sea put severe restrictions on the amount of data available.

The high variances of the sampling survey results have limited their usability for stock evaluation purposes. This has prompted many research initiatives that have tried to replace the traditionnal set to strata allocation strategy which was proportionnal to stratum area by some more optimal allocation strategy.

The aim of these *optimal* strategies is to avoid wasting in uninteresting strata, valuable sets that could be allocated to some other strata where there is more variability to be observed. Various strategies have been suggested (Francis 1984, Gagnon 1990) to adjust the set to strata allocation to avoid such waste.

In this paper, I look at the effects of different set to strata allocation adjustment strategies on the precision and bias of abundance estimates.

Types of set to strata adjustments strategies

Set to strata allocation can be either *fixed* or *adjusted annually* before the sampling is done or *adjusted progressively* during the sampling process. The purpose of making adjustments to the allocation of sets to strata is to improve the precision of abundance estimates. More adaptive strategies will be more responsive to changes in the spatial distributions but the search for a minimum variance might cause a negative bias to the mean when the sample means increase with the sample variances. On the other hand, more precise biased estimators can be preferred to imprecise unbiased estimators when the incurred bias is small.

Directed surveys

The number of species that a survey must target dictates the type of allocation strategy that can be used. For generic surveys where all species are considered equally important, the allocation of sets to strata should be proportional to strata area. For directed surveys, where a limited number (usually one or two) of species are considered more important than the others, allocation strategies can be optimized by making the allocation density in every stratum an increasing function of the variance of the abundance of these more important species.

The simulations

Each simulation covered a combination of a population model and a distribution with a strategy for set to strata allocation adjustment. The populations considered are simplified versions of observed 3Pn-4RS cod biomasses and 2J cod abundance. The parameters of the simulations for each population are summarized in Table 1. The *number of strata*, the *number of sets*, the *minimum allocation* and the *allocation to phases* (for the 2J population) were chosen to approximate closely the survey designs that provided the population data.

Population	3Pn-4RS	2J
number of replicates	2000*	4000
number of strata	20	30
strata areas	all equal	all equal
number of sets	200	145
minimum allocation	3	2
allocation to phases	100/100	108/37

*increased to 4000 in certain cases

Table 1: Simulation parameters

For each simulated survey, the stratified mean and its variance were calculated. I compared the average of estimated means to the true mean from the models and the average of the stratified variances between simulations.

Population models

There is no agreement on the statistical distribution of fish abundance in a typical marine environment. The model that I used assumes that the population can be divided into homogeneous strata. The set results within a stratum are assumed to be independent and identically distributed with a common mean and variance. The distribution is the same in all strata but the means and variances vary from stratum to stratum.

The means and variances that I used in the models are listed in Table 2. They are the values observed for the cod catch in the 20 strata covered by the 3Pn-4RS Gadus Atlantica winter 1991 survey (A. Fréchet, personal communication) and the cod abundance in the 30 strata of 2J in 1991 (J. W. Baird, personal communication). There are large inter-strata variations: in 3Pn-4RS, over 90% of the biomass is concentrated in strata 4 and 6; in 2J, more than half of the abundance occurs in strata 10, 13, 23 and 11. The low abundance strata have Poisson-like dispersion ($mean = variance$) but higher abundance strata show great overdispersion.

To give enough scope to the simulations, I used three distributions: Poisson, log-normal and negative-binomial. The parameter of the Poisson distribution was adjusted to the observed means but underestimated the observed variances for most strata. The log-normal and negative-binomial distributions provided two forms of overdispersion with properly adjustable means and variances.

I obtained pseudo-random numbers from the RNPOI (Poisson), RNLNL (log-normal) and RNNBN (negative-binomial) routines from IMSL (IMSL 1989). The seed was generated internally from the computer clock by IMSL. The three routines were checked to give the proper means and variances for all stratum values.

Set to strata allocation adjustment strategies

Proportional allocation For the purpose of the simulations, the strata were all given the same weight (surface). The proportional allocation of sets to these strata simply amounts to the allocation of an equal

3Pn-4RS cod biomass		
stratum	**mean**	**variance**
1	4.0	11.8
2	11.9	485.8
3	83.9	6558.3
4	1142.7	4338694.5
5	8.0	39.0
6	1914.3	28459320.0
7	31.0	1112.9
8	0.6	0.7
9	4.2	12.0
10	14.6	253.0
11	8.0	39.0
12	9.1	48.9
13	4.6	1.2
14	0.4	0.4
15	8.4	50.8
16	14.2	250.0
17	2.2	6.3
18	0.5	0.2
19	0.5	0.6
20	1.1	0.6
all	163.2	

2J cod abundance		
stratum	**mean**	**variance**
1	0.1	0.1
2	0.1	0.1
3	0.1	0.1
4	0.5	0.5
5	0.5	1.1
6	1.1	2.1
7	10.3	74.9
8	31.9	890.4
9	4.8	16.6
10	65.0	1957.0
11	125.5	13284.5
12	12.0	52.0
13	79.4	7404.0
14	41.2	4481.2
15	21.7	2112.5
16	3.0	8.0
17	0.1	0.1
18	0.1	0.1
19	0.1	0.1
20	21.5	312.5
21	5.0	50.0
22	0.1	0.1
23	86.3	15760.3
24	22.0	144.0
25	25.5	612.5
26	19.0	722.0
27	0.1	0.1
28	0.1	0.1
29	18.0	98.0
30	0.1	0.1
all	19.84	

Table 2: Means and variances of the population models

number of sets to every stratum. This fixed allocation strategy has been implemented in practice. It can be shown to give unbiased estimates of abundance.

Historical allocation adjustment In historical allocation, the allocation is a constrained optimal allocation (Gagnon 1990) based on the results of previous surveys. The first survey of the simulation is done with proportional allocation. For subsequent surveys the results from the latest surveys are pooled to estimate the strata variances from which a constrained optimal allocation is calculated. I considered the cases of allocation adjustments based on one and four previous surveys. These simulations assume that the population model does not evolve from survey to survey. This strategy can be implemented in practice if data from previous surveys is available.

Two-phase allocation adjustment In two-phase allocation, the first phase of the survey covers all strata with a certain portion (see Table 1) of the available sets using proportional sampling. The second phase implements constrained optimal allocation with the remaining sets. The first phase results are used to calculate the constrained optimal allocation of the second phase. This strategy can be implemented in practice for surveys where it is possible to cover every stratum twice.

Minimum variance allocation adjustment In minimum variance allocation, the allocation of sets to strata is determined by the results of all the previous sets of the same survey. First, all strata are sampled two or three times (see Table 1) to get a first estimate of their variance contribution. For every set thereafter, the stratum with the current maximum variance contribution is sampled. This strategy cannot be implemented in practice unless travel costs are much lower than sampling costs but it represents the most adaptive form of allocation adjustment.

Results

Table 3 gives the relative bias of the stratified mean (relative to the model mean) and the average coefficient of variation (relative to the model mean) for each combination of population model and adjustment strategy.

In the cases where the bias was found to be significantly ($* \equiv p < 0.05$) or very significantly ($** \equiv p \leq 0.0001$) different from zero, another simulation was done using a *fixed* allocation corresponding to the mean allocation used in the biased case. This was done to separate the adjustment effect from the mean allocation effect.

Allocation adjustment significantly reduces the variance in all cases but the effect is greater in the 3Pn-4RS population. This is due to the pronounced heterogeneity of the strata variances in this population that makes optimal allocation much more efficient than proportional allocation.

Proportional allocation simulation results agree well with the theoretical values. For a non-finite population in equal-sized strata, the theoretical variance of the stratified mean reduces to

$$V(\overline{y}_{st}) = \frac{1}{nN} \sum S_h^2$$

where n is the number of sets, N is the number of strata and S_h^2 is the variance in stratum h.

Pop. model	Allocation adjustment strategy	Strata abundance distribution					
		Poisson		log-normal		neg.-binomial	
		bias%	cv%	bias%	cv%	bias%	cv%
3Pn – 4RS	none (proportional)	0.006	0.55	2.1	56.8	2.9	58.7
	... theoretical	*0.000*	*0.55*	*0.0*	*55.5*	*0.0*	*55.5*
	historical (1 survey)	-0.001	0.28	0.2	19.8	0.6	19.5
	historical (4 surveys)	-0.002	0.28	0.1	19.2	-0.3	18.9
	two-phase	-0.015*	0.29	-4.4**	22.8	-10.9**	23.3
	... fixed	*-0.003*	*0.29*	*-0.3*	*21.1*	*-0.3*	*23.1*
	minimum variance	-0.005	0.29	-17.2**	15.5	-30.4**	14.1
	... fixed			*0.3*	*18.4*	*-0.5*	*19.7*
2J	none (proportional)	0.039	1.85	-0.1	16.2	-0.0	16.4
	... theoretical	*0.000*	*1.87*	*0.0*	*16.7*	*0.0*	*16.7*
	historical (4 surveys)	0.021	1.42	0.1	10.7	0.1	10.3
	two-phase	-0.157**	1.41	-9.2**	9.5	-9.3**	10.0
	... fixed	*0.035*	*1.55*	*-0.2*	*12.0*	*-0.1*	*12.0*
	minimum variance	-0.257**	1.07	-15.4**	6.1	-17.0**	6.8
	... fixed	*-0.009*	*1.48*	*-0.1*	*10.8*	*0.0*	*10.6*

* significant   ** very significant

Table 3: Bias and coefficient of variation of stratified means

Minimum variance and two-phase allocation adjustment strategies underestimate the mean. As the results of the *fixed* allocation simulations show, this is due to the adjustment strategy and not to the resulting allocation. The allocations that result from these adjustment strategies would not cause any bias if they were used consistently for sampling. It is the choice of different allocations depending on the outcome of the current random sampling that causes the bias.

> In actively seeking to reduce the variance, the sampler is indirectly reducing the mean because samples with smaller variances tend to have smaller means (for distributions of positive random variates).

Historical allocation adjustment does not give biased estimates because the estimation of the strata variances is not based on the outcome of the current sampling but on an independent sample.

Discussion

These results provide an estimate for the order of magnitude of the bias that can be caused by allocation adjustment in typical trawl sampling. The biases caused by two-phase and minimum variance adjustment strategies are of the same order of magnitude as the standard error. This is quite substantial.

For this reason, the two-phase allocation adjustment strategy should not be used any further. It is not possible to correct for the expected bias caused by this strategy because it depends on the particular allocation to the phases, the level of fish aggregation, and the true distribution of fish abundance. These may vary from year to year and so will the expected bias.

The 1989 Autumn research survey in 2J3KL used a two-phase (or 2-stage) survey design (Baird 1990). It measured a decrease of 11.4% in the overall cod abundance estimate after conducting the second phase of the survey. This is similar to the 9.3% reduction estimated by the simulation using the 2J results of 1990 with two-phase allocation adjustment.

The simulation results show similar improvements for historical and two-phase allocation adjustment strategies. The improvement in precision due to historical allocation adjustment is exaggerated in the simulations because the population model is kept constant from survey to survey. This is not true for two-phase allocation adjustment. We can thus predict that the historical strategy will not improve precision by as much as the two-phase strategy. This seems to be the price to pay for unbiased estimates.

References

Baird, J. W., C. A. Bishop and W. B. Brodie 1990. *The assessment of the cod stock in NAFO Divisions 2J, 3K and 3L.* CAFSAC Research Document 90/18.

Francis, R. I. C. C. 1984. *An adaptive strategy for stratified random trawl surveys.* New Zealand J. of Mar. and Fresh. Res. Vol. 18, p. 59-71.

Gagnon, P. 1991. *Optimisation des campagnes d'échantillonnage: les programmes REGROUPE et PARTS.* Rapp. tech. can. sci. halieut. aquat. 1818 : iii + 9 p.

IMSL 1989. *IMSL STAT/LIBRARY version 1.1 User's manual.* IMSL inc., Houston, USA, Vol. 3.