Fisheries and Oceans
Science

Pêches et Océans
Sciences

C S A S

Canadian Science Advisory Secretariat

S C C S

Secrétariat canadien de consultation scientifique

Research Document 2001/082

Document de recherche 2001/082

Not to be cited without
permission of the authors [1]

Ne pas citer sans
autorisation des auteurs [1]

A simple matrix based analysis of multiple area tagging data applied to the northern cod stock off Newfoundland

Application au stock de morue du nord de Terre-Neuve d'une analyse matricielle simple des données de marquage obtenues dans plusieurs zones

J. Pope and J. Brattey*

Norwegian College of Fisheries Science, Tromsø, Norway.
Correspondence to NRC (Europe) Ltd, The Old Rectory, Burgh St. Peter, Norfolk, UK NR34OBT

*Science, Oceans and Environment Branch
Department of Fisheries and Oceans
Box 5667, St. John's, NF Canada
A1C 5X1

ISSN 1480-4883
Ottawa, 2001

Canada

Abstract

A simple model for the analysis of tagging data is developed. This accounts for both harvest within stock components and migration between components, as well as initial tagging mortality, tag loss, tag reporting rates, and assumed natural mortality. An exact matrix solution is provided for the simple case where all components are tagged in a year and extensions of this are constructed to handle the singular matrix situations where not all data are available for all years. This model is applied to inshore components of the Newfoundland northern (NAFO Divs. 2J3KL) cod (*Gadus morhua*) stock and to inshore and offshore components of the southern Newfoundland NAFO Subdiv. 3Ps cod stock. Post-moratorium fisheries being conducted only in the inshore areas (Lilly et al 2001) complicate assessment of the northern cod stock. These inshore areas are not fully covered by ground-fish surveys which provide the main link to the pre-moratorium population estimates. Hence, assessments of abundance of these inshore sub-populations must rest heavily on the results of tagging experiments. Results indicate that the biomass of inshore components of the northern cod have remained at about 40Kt for the past three years and that harvest rates have typically been of the order of 10% on the major 3L inshore North component and rather higher on the inshore 3K component during this period.

Résumé

Nous avons mis au point un modèle simple pour analyser des données de marquage. Le modèle tient compte de la pêche des composantes du stock, de la migration entre les composantes, de la mortalité initiale due au marquage, des étiquettes perdues, des taux de déclaration des poissons marqués et de la mortalité naturelle présumée. Nous présentons une solution matricielle exacte pour le cas simple d'une année où toutes les composantes sont marquées et élaborons des extensions de cette solution pour accommoder les situations de matrice singulière où les données complètes ne sont pas disponibles pour toutes les années. Nous appliquons ce modèle aux composantes côtières du stock de morue (*Gadus morhua*) du nord de Terre-Neuve (divisions de l'OPANO 2J3KL) et aux composantes côtières et hauturières du stock de morue de la sous-division 3Ps, au sud de Terre-Neuve. Le fait que la morue ne soit pêchée que dans les zones côtières depuis le moratoire (Lilly et al., 2001) complique l'évaluation du stock de morue nordique. Ces zones côtières ne sont pas entièrement couvertes par les relevés du poisson de fond, lesquels fournissent le principal lien avec les estimations de la population avant le moratoire. Par conséquent, l'évaluation de l'abondance de ces sous-populations côtières dépend largement des résultats des expériences de marquage. Les résultats montrent que, depuis trois ans, la biomasse des composantes côtières du stock nordique se maintient à environ 40 kt, le taux d'exploitation est généralement de l'ordre de 10 % pour la principale composante côtière du stock nordique dans 3L et ce taux est plus élevé pour la composante côtière de 3K.

Introduction

The analysis of tagging experiments can become complex when tagged fish are not fully mixed in the overall population. Where tagging is conducted in a series of areas and recaptures are from the same closed group of areas it is possible to propose a simple model. This is based on the assumptions:

- that fish tagged in each area (or possibly each tagging experiment) have a probability of occurrence in each area (approximately the average residence time).
- that each recapture area has a specific harvest rate such that fish from any experiment have a harvest rate of the area rate times their probability of occurrence in the area.

This paper describes the solution of the model by matrix algebra given these assumptions. This model is applied to inshore components of the Newfoundland northern (NAFO subdivisions 2J3KL) cod stock and to inshore and offshore components of the NAFO Subdiv. 3Ps cod stock. Post moratorium fisheries on the Newfoundland northern cod stock have been conducted only on its inshore sub populations (Lilly et al 2001). These sub-populations are not fully covered in the ground-fish surveys that provide the main link to the pre-moratorium population estimates. Hence, assessments of the exploitable biomass of these sub-populations must rest heavily on the results of the tagging experiments described in Brattey et al. (2001). For these data sets the model is used to make estimates of migration rate and local harvest rates and hence to convert local catches to estimates of local population biomass.

Materials and Methods

Mathematical model

The model assumes that the recaptures (CAP(k, j)) in area j from a tagging experiment conducted in area k is given by the equation

$$CAP(k, j) = T(k)*OC(k, j)*H(j) \qquad (1)$$

Where T(k) is the effective numbers of fish that were tagged in area k at the beginning of a period, where OC(k, j) is the probability of fish tagged in area k being subject to the area k harvest rate and where H(j) is the harvest rate in recapture area j.

If the recapture areas designated by k match the tagging areas designated by j and if these are the only areas occupied by the tagged fish then the estimation problem may be written in matrix form as follows.

$$\mathbf{CAP} = \mathbf{T}*\mathbf{OC}*\mathbf{H} \qquad (2)$$

$$\mathbf{Unit} = \mathbf{OC}*\mathbf{Unit} \qquad (3)$$

Where **CAP, T , OC** and **H** are the square matrix forms of the terms shown in equation 1 (note **T** and **H** are diagonal matrices) and **Unit** is a column vector of 1's of the same length. Equation 2 thus describes recaptures and equation 3 expresses the fact that the sum of OC(k,j) for all j equals 1 (i.e. the fish have to be available somewhere in the model).

We may rewrite 2 as

$$\mathbf{(T)^{-1}*CAP*(H)^{-1} = OC} \qquad (4)$$

Where $\mathbf{(\ )^{-1}}$ indicates the matrix inverse.

Multiplying both sides by **Unit**

Gives $\mathbf{(T)^{-1} *CAP*(H)^{-1} *Unit= OC*Unit=Unit}$ (5)

Hence,

$$\mathbf{(H)^{-1} *Unit = \{(T)^{-1} *CAP\}^{-1} *Unit} \qquad (6)$$

provided $\mathbf{\{(T)^{-1} *CAP\}^{-1}}$ is not singular or ill conditioned (the matrix equivalent of dividing by zero or a very small number). Equation 6 thus can be solved for H (recall H is a diagonal matrix). **OC** can then be solved using equation 4. These matrices equations can be readily solved (for example in an EXCEL spreadsheet using the MMULT and MINVERSE functions).

In the simple form proposed the method uses all the degrees of freedom of the data and thus produces no residuals. However, since tag returns under the model are rare events they might thus be expected to have a Poisson distribution. Hence, some idea of the models statistical properties may be obtained by simulating sampling error by replacing data with random numbers drawn from Poisson distributions with means equal to the data values.

Data

Returns from cod tagging conducted in 1997, 1998, 1999 and 2000 in the Newfoundland area are summarized in Brattey et al. 2001. The experiments included single tagged, double-tagged and high value tagged fish in order to estimate tag loss and reporting rates. Tagging experiments were conducted for a number of inshore areas of subdivisions 2J3KL (designated 3K inshore, 3L In- North, 3L In-South) and inshore areas of subdivision 3Ps (Placentia Bay (3Ps-PB), western Bays (i.e. west of the Burin Peninsula, 3PS-WB) and offshore areas (3Ps Off) of Subdiv. 3Ps and also in Divs. 3NO. For 1999 and 2000 in particular these broadly accord with the requirements of the theory developed in the previous section. Results for 1997 and 1998 do not have tagging experiments in some areas and can thus only be analyzed in this fashion in conjunction with 1999 and 2000 data. Only a very few returns, from fish tagged in the 3K In, 3L In- North, 3L In-South, 3Ps-PB and 3Ps Off, were reported from outside these areas. Hence, the analysis was restricted to these 5 areas with minor reassignment of the few outside recaptures to the 5 areas.

Fish tagged in or after week 21 are not included in the first years analysis of 3L IN-S, 3Ps PB, 3Ps OFF. They are included in subsequent years` analyses.

Tag loss, reporting rate and natural mortality rate

Following results reported by Brattey *et al*. 2001, initial tagging mortality is taken as 13% and tags were assumed to be lost exponentially with a loss rate of 0.2 in the first year and thereafter to remain attached. In the available data set double-tagged fish are only recorded as recaptured if both tags were attached to the fish at the time of recapture. Consequently, their tag loss was assumed to have a rate of 0.4 in the first year only (i.e. their tag losses were taken to be the square of the value used for single tagged fish).

Based upon a simple study of the summary of overall returns from the single, double and high value tagged fish, reporting rate were taken to be 100% for high value tags and to be 65% for single low value tags and 77% for double-tagged fish in all areas. Natural mortality rate was taken to be 0.2 per year for all tagged fish.

Populations of tagged fish in subsequent years (y) to the tagging year were taken to be estimated by a reverse cohort analysis such that:

Tag Population(y)
=Effective Tag population(y-1)*Exp{-(natural mortality and tag loss rates)]
– Tag recaptures(y-1)*Exp{-0.5*(natural mortality and tag loss rates)}/Reporting rate

Handling matrix singularities and other structural problems

$\mathbf{\{(T)^{-1} * CAP\}^{-1}}$ was found to be near singular for some recapture years and this seemed to result from the similarity of recaptures from some tagging areas. Where this occurred the areas were combined to give an alternative 4 or 3 area model.

Where data are missing from some cells of one experiment the model may be modified on the assumption that the migration is the same for both experiments. Where two sets of tagging (Experiment. 1 and Experiment. 2) exist where the same migration model (same OC) is postulated but where **T1, T2, CAP1, CAP2, H1** and **H2** differ we might write

$$\mathbf{(T1)^{-1}*CAP1 = OC*H1} \quad (7)$$

$$\mathbf{(T2)^{-1}*CAP2 = OC*H2} \quad (8)$$

(Note that if as postulated, CAP2 is missing a row, as a fudge we will have to replace the indeterminate numbers of $(T2)^{-1}$*CAP2 by zeros)

Therefore,

$$[(T1)^{-1}*CAP1+(T2)^{-1}*CAP2]*(H1+H2)^{-1} = OC \qquad (9)$$

Hence,

$$(H1+H2)^{-1}*Unit = \{[(T1)^{-1}*CAP1+(T2)^{-1}*CAP2]\}^{-1}*Unit. \qquad (10)$$

Having estimated $(H1+H2)^{-1}$ we may estimate a combined **OC**.

Results

Applying the model to the 5 area data set

When the model was applied to single years results estimates of migration were somewhat variable. This was particularly the case for the 5 area model where the similar occurrence ratios found in 3L In-South and 3Ps rendered matrices near singular and frequently resulted in negative OC terms and consequent estimates of negative harvest rates.

The simple model fits exactly to the data and thus cannot estimate residuals. However, estimates of confidence regions can be made by randomizing the recapture rates assuming a Poisson distribution with a mean equal to the number of tags recaptured per cell in the case of positive cells and an arbitrary small mean (say 0.25) for null cells. Each set of random realizations is then interpreted as probabilities of occurrence and as harvest rates by the matrix model.

Table 1 shows the results of such a Monte Carlo simulation for the full 5 area model for the various area tagging experiments conducted in 1999 using the recapture results from 1999 only. Monte Carlo simulations were made by replacing recapture data by random numbers drawn from Poisson distributions with means equal to recapture estimates. For each cell 100 random numbers were generated with a Poisson distribution with a mean equal to the cell total. These random numbers were resampled 1000 times to provide a 1000 realisations. The table shows the area harvest rates estimated by the first 20 simulations and the mean and standard deviation of the full 1000 simulations.

Table 1.
99-99 Harvest Estimates made under re-sampled Poisson distribution.
The table shows the 1st 20 of 1000 realisations and the mean and SD.

Realisation	3K IN	3L IN-N	3K IN-S	3Ps PB	3Ps OFF
1	0.281	0.039	0.127	0.121	0.010
2	0.327	0.040	0.411	0.101	0.003
3	0.278	0.036	0.199	0.116	0.009
4	0.299	0.045	0.066	0.125	0.007
5	0.276	0.048	0.326	0.114	0.006

6	0.409	0.043	-0.278	0.103	0.007
7	0.252	0.037	-0.211	0.110	0.007
8	0.313	0.044	0.067	0.118	0.003
9	0.276	0.039	0.650	0.123	0.012
10	0.297	0.043	0.072	0.120	0.010
11	0.303	0.045	0.297	0.103	0.006
12	0.244	0.047	0.437	0.115	0.009
13	0.416	0.042	0.661	0.100	0.008
14	0.379	0.041	-0.265	0.098	0.005
15	0.300	0.040	0.107	0.116	0.009
16	0.285	0.043	0.147	0.112	0.007
17	0.328	0.037	0.166	0.122	0.006
18	0.289	0.042	-0.753	0.108	0.007
19	0.317	0.045	0.493	0.111	0.007
20	0.281	0.041	0.176	0.116	0.009
Mean	0.310	0.042	0.634	0.111	0.008
S.D.	0.050	0.003	16.694	0.006	0.003

The table illustrates that while the harvest rate estimates in 3K, 3L IN-N, 3Ps PB and 3Ps OFF are rather stable to the introduction of noise, those of 3L IN-S (highlighted) are not. In the highlighted cells some results are negative (these are associated with negative occurrence probability rates) while others are large. This result is due to the migration of fish tagged in this region being very similar to that of those tagged in Placentia Bay. This renders the Inv(T)*CAP matrix near singular and results in the variable estimates. While the negatives could be trapped, the high values would remain and cause problems. Hence the solution seems to be to merge the 3L IN-S and 3Ps PB tagging results.

When applied to the 4 areas 3K IN; 3L IN-N; 3K IN-S & 3Ps PB; 3Ps OFF, full data to fit the model were not available from the 1997, 1998 and 2000 tagging. For the 1999 tagging recaptures in both 1999 and in 2000 it was possible to fit the matrix model. Tables 2a and 2b show results of the probabilities of occurrence in each area for the 1999 tagging as recaptured in 1999 (2a) and 2000 (2b).

Table 2a.

Release Area	99-99	OC matrix		
3K IN	0.71	0.28	0.01	0.00
3L IN-N	0.04	0.96	0.00	0.00
3L IN-S+PB	0.00	0.04	0.96	0.00
3Ps OF	0.00	0.02	0.06	0.92
Recaptured in	3K -IN	3L IN-N	3L IN-S +3Ps PB	3Ps OFF

Table 2b.

Release Area	99-00	OC matrix		
3K_IN	0.63	0.36	0.01	0.00
3L_INN	0.11	0.88	0.01	0.00
3L_INS+PB	0.00	0.03	0.94	0.03
3Ps_OF	0.00	0.02	0.05	0.94
Recaptured in	3K -IN	3L IN-N	3L IN-S +3Ps PB	3Ps OFF

The occurrence probabilities shown in Tables 2a and 2b are comfortingly similar. However, earlier results (not shown) suggested that low harvest rate estimates in 3Ps OFF based upon low tag returns from tagging in that area could cause the occurrence probability matrix to vary substantially. This was particularly the case with the occurrence probabilities estimated for fish tagged in 3L IN-S & 3Ps PB. Because low numbers of fish tagged in these areas were recaptured in 3Ps OFF these were sometimes interpreted as substantial migrations to a low mortality area. Reducing the model to a three-area model thus had attractions particularly in the context of the 2J+3KL assessments.

Moreover, since the model could not be simply fitted to the incomplete tagging data from 1997, 1998 and 2000 some alternative approach was needed. As an alternative approach overall occurrence probability rates were estimated on the summed tagged and recaptured matrices of all years. The resulting all years tagging 4 stock occurrence matrix is shown in Table 3 below.

Table 3. The combined OC matrix for the four-area model.

Release Area	4-area model			
3K IN	0.65	0.34	0.01	0.00
3L IN-N	0.05	0.94	0.01	0.00
3L IN-S +3Ps-PB	0.00	0.04	0.85	0.11
3Ps OFF	0.00	0.01	0.04	0.95
Recaptured in	3K-IN	3L IN-N	3L IN-S + 3Ps-PB	3Ps-OFF

This seems similar to the 1999 tagging results shown in Tables 2a and 2b but problems associated with low recovery of fish tagged in 3Ps OFF suggested that a reduced model was advisable. An alternative three-area model was therefore also developed on the same bases of the summed recaptures and estimated tagged numbers in each year. The resulting migration matrix is shown in Table 4 below.

Table 4. The combined OC matrix for the three-area model.

Release Area	3-area model		
3K IN	0.65	0.34	0.01
3L INN	0.05	0.94	0.01
3L INS+3Ps-PB	0.00	0.04	0.96
Recaptured in	3K IN	3L IN-N	3L IN-S+ 3Ps-PB

Moreover, this 3 area model has the advantage that it covered the only three areas of concern to the 2J+3KL assessment that motivated the analysis. Hence, further work was based upon this 3-area overall matrix. This combined year, probability of occurrence matrix can be used to estimate harvest rate. This may be done either by using the inverse of the diagonal occurrence probability terms to correct the recaptures per tags at large results from the diagonal areas (i.e. fish recaptured in the area in which they were tagged) or by using the matrix equation

$(OC)^{-1}*(T)^{-1}*CAP=H$

The former approach seemed preferable in the current situation since the matrix equation is not operable for 1997,1998 and 2000 tagging results since some areas were not tagged in those years. Table 5 estimates the harvest rate for each tagging year for each recapture year for the 3 regions using only the diagonal terms.

Table 5. Harvest rate based upon diagonal terms only. Highlighted values are not based upon data but are extrapolated from earlier years results.

Recapture year		Tagged year				
	Area	1997	1998	1999	2000	Average
1997	3K In					
	3L In N	0.005				0.005
	3Ps PB	0.170				0.170
1998	3K In	0.153	0.307			0.230
	3L In N	0.061	0.064			0.062
	3Ps PB	0.142	0.134			0.138
1999	3K In	0.050	0.186	0.630		0.289
	3L In N	0.115	0.121	0.117		0.118
	3Ps PB	0.250	0.355	0.302		0.303
2000	3K In	0.032	0.141	0.148	0.156	0.119
	3L In N	0.065	0.069	0.084	0.101	0.080
	3Ps PB	0.196	0.346	0.372	0.186	0.275

Not all diagonal recapture terms were available for some of the annual tagging experiments and some means of estimating these were sought. In the tables it is noticeable that the harvest rate tends to be somewhat larger in the first year at liberty. This might result from underestimating natural and tag losses or from tagged fish gradually growing out of the selection range of gillnets. Thus to estimate missing values, the median ratio of available harvest rates in the first year at liberty to that in the same recapture year for the proceeding years tagging was used to adjust results. This ratio was used to calculate the missing values (highlighted) from the immediately proceeding harvest rate in the table row. Average estimates of harvest rate are shown in the table.

The harvest rate found in 3K IN from the 1999 recaptures of the 1999 tagging seems particularly high but in general harvest rate estimates for a particular area and year of recaptures seem coherent between the different years tagging results. While harvest levels are generally modest in 2000 they nevertheless appear to be somewhat higher than would seem wise for a stock to rebuild.

If it is assumed that these estimates of harvest rate apply to the general fish abundance in an area as well as to the tagged fish abundance then these harvest rates may be used to make local population estimates based upon the catch weights recorded from each area.

Table 6 shows estimates of population biomass for 3K In and 3L IN-N using the average harvest rates shown in table 5. Table 6 also shows population biomass calculated on the average of harvest rate of the current years tagging and the adjusted rate of the proceeding years tagging.

Table 6. Estimates of biomass in 3K and in 3L-N for years 1998-2000 based upon two different formulations of average harvest rate.

Tag recaptured	Area	Catch (Tons)	Average Harvest Rate. (all years)	Biomass (Tons)	Average Harvest Rate (last 2 years)	Biomass (Tons)
1998	3K In	2119	0.230	9214	0.226	9368
	3L In N	1762	0.062	28207	0.061	28905
Total				37421		38274
1999	3K In	3716	0.289	12865	0.403	9211
	3L In N	3256	0.118	27633	0.116	28024
Total				40498		37234
2000	3K In	1459	0.119	12257	0.148	9849
	3L In N	2850	0.080	35769	0.090	31545
				48026		41394

Upward trends in biomass in the two regions over the three years are at best modest. An estimate of average biomass for the three years of about 10Kt in 3K-IN and about 30Kt in 3L IN-N seems reasonable. This gives a combined biomass of about 40Kt. These estimates provide a simple basis of comparison with other analyses such as acoustic surveys in the 3L In-N area.

Variance of estimates

Given the use of the diagonal terms in the estimation of harvest rate and the likely Poisson variation in the underlying data it seems likely that the coefficient of variation of harvest rate estimates would have a similar coefficient of variation to the numbers recaptured on the diagonal. Historic averages were calculated and CV's based upon the assumption that these had Poisson variation. These are shown in Table 7. It also shows the average realized CV of harvest rate estimates from the 3 stocks shown in Table 5. These are generally somewhat larger than the theoretical level postulated here, perhaps due to real changes in harvest rate as the tagged fish age.

Table 7. Theoretical and estimated coefficients of variation for the harvest rate estimates.

	Simulated CV	Estimated CV
3K In	12%	67%
3L IN-N	10%	9%
3L IN-S&PB	5%	19%

Discussion

A simple matrix solution to the migration model is developed. This is simple to use and the theory can be fairly easily modified to combine years and areas. Consideration of its application in the 2J3KL and 3Ps assessments suggests circumstances where it may have problems estimating migration and harvest due to either singularity in the tag returned matrix or to the inclusion of areas with low harvest rates. After suitable amalgamations and exclusions the method sees to give coherent results. Table 5 shows estimated harvest rates and these can be used to give approximate estimates of stock biomass (Table 6).

Various extensions to the model might be considered. The inclusion of size selection seems important and might initially be addressed by fitting the existing model separately to various size groups of tagged fish or tag returns. Consideration of how to optimally fit the common migration model to the separate years' data might also indicate ways to handle missing data.

Acknowledgements

Financial support for this study was provided by the Department of Fisheries and Oceans Strategic Science Program, Project No. 90061. We thank C. George and D. Porter, staff of the Gadoids, Sentinel Fishery, and Commercial Sampling Sections for assistance with the field work and collating recaptures. We also thank the numerous fishers and plant workers who returned tags and recapture information.

References

Brattey, J., D. Porter, and C. George. 2001. Stock structure, movements, and exploitation of Atlantic cod (*Gadus morhua*) in NAFO Divs. 2J+3KL based on tagging experiments conducted during 1999-2000. DFO Canadian Stock Assessment Secretariat Research Document 2001/072.

Lilly, G. R., P. A. Shelton, J. Brattey, N. Cadigan, B. P. Healey, E. F. Murphy, D. E. Stansbury. 2001. An assessment of the cod stock in NAFO Divisions 2J+3KL. DFO Canadian Stock Assessment Secretariat Res. Doc. 2001/044.