ANNUAL MEETING - JUNE 1971

AN INVESTIGATION OF THE ACCURACY OF VIRTUAL POPULATION ANALYSIS

by

J. G. Pope
Ministry of Agriculture, Fisheries and Food,
Fisheries Laboratory, Lowestoft

Gulland's virtual population analysis (Gulland 1965) is an extremely useful technique when assessing a fishery, because it enables estimates of population at age and fishing mortality to be made independently of the measurement of effort. These estimates are however subject to various errors which might adversely affect an assessment. What causes these errors and how can their magnitude be calculated?

1. Cohort analysis as an approximation to virtual population analysis

Definition of symbols used:

M is the instantaneous coefficient of Natural Mortality;

F is the instantaneous coefficient of Fishing Mortality;

Z is the instantaneous coefficient of Total Mortality;

N_i is the population of a year-class at the ith birthday;

C_i is the catch of a year-class at age i;

t is the last age of a year-class for which catch data are available;

exp is the exponential function.

Cohort analysis is a new form of virtual population analysis developed by the author. It is in fact an approximation to Gulland's virtual population analysis which is usable at least up to values of M = 0.3 and F = 1.2. A detailed explanation of the method will be the subject of a later publication - this research document is intended only to give some indications of the results relating to errors. The method is based n the approximate formula

$$N_i = C_i \exp\{M/2\} + N_{i+1} \exp\{M\} \quad \ldots\ldots\ldots\ldots \quad 1.1$$

Thus, using 1.1 as a recurrence relationship,

$$N_i = C_i \exp\{M/2\} + C_{i+1} \exp\{3M/2\} + C_{i+2} \exp\{5M/2\} \ldots$$
$$\ldots\ldots N_t \exp\{(t-i)M\} \quad \ldots\ldots\ldots\ldots \quad 1.2$$

As with Gulland's virtual population analysis N_t has two possible forms. The first form is when C_t refers to the catch in year t only, which is the case with the last year's catch of a year-class which is still being fished.
In this case

$$N_t = \frac{C_t Z_t}{F_t \left(1 - \exp\{-Z_t\}\right)} \quad \ldots\ldots \quad 1.3$$

and consequently

$$N_i = C_i \exp\{M/2\} + C_{i+1} \exp\{3M/2\} + C_{i+2} \exp\{5M/2\} +$$

$$+ \frac{C_t Z_t \exp\{(t-i)M\}}{F_t \left(1 - \exp\{-Z_t\}\right)} . \quad \ldots\ldots \quad 1.4$$

The second form is when C_t refers to the catch in year t and all subsequent years. This is usually the case with a completely fished year-class. In this case

$$N_t = \frac{C_t Z_t}{F_t} \quad \ldots\ldots \quad 1.5$$

and consequently

$$N_i = C_i \exp\{M/2\} + C_{i+1} \exp\{3M/2\} + C_{i+2} \exp\{5M/2\} +$$

$$+ \frac{C_t Z_t \exp\{(t-i)M\}}{F_t} \quad \ldots\ldots \quad 1.6$$

In either case

$$F_i = \log_e \{N_i/N_{i+1}\} - M. \quad \ldots\ldots \quad 1.7$$

The closeness with which these formulae approximate the results of virtual population analysis can be judged from Table 1 where results of both methods are compared. It can be seen that in no case do the estimates given by the two methods differ by more than 2%. Consequently an investigation of the errors of cohort analysis is an approximate investigation of the errors of Gulland's virtual population analysis. It can be seen from equations 1.4 and 1.6 that errors in N_i, and consequently errors in F_i, can be introduced by the incorrect choice of F_t and by the sampling errors in the C_i. These two sources of error are investigated in the next two sections. Errors in M can also cause errors in N_i and F_i, but for the purpose of this document M will be considered as fixed.

2. <u>Error in cohort analysis due to the incorrect choice of F_t</u>

If an incorrect value F_t is chosen for the terminal fishing mortality when its true value is $\bar{F}_t$, then the proportional error in N_t, $\rho(N_t)$, is given as follows in the case when C_t is the catch in year t only:

$$\rho(N_t) = \frac{Z_t \bar{F}_t \left(1 - \exp\left\{-\bar{Z}_t\right\}\right)}{Z_t F_t \left(1 - \exp\left\{-Z_t\right\}\right)} - 1 \; ; \qquad 2.1$$

since

$$\rho(N_i) = \rho(N_{i+1}) \exp\left\{-F_i\right\} \qquad 2.2$$

it follows that

$$\rho(N_i) = \left(\frac{Z_t \bar{F}_t \left(1 - \exp\left\{-Z_t\right\}\right)}{Z_t F_t \left(1 - \exp\left\{-Z_t\right\}\right)} - 1\right) \exp\left\{-F_i - F_{i+1} \dots F_{t-1}\right\} \; ; \qquad 2.3$$

for small values of Z this is approximately given by

$$\rho(N_i) \simeq \left\{\frac{\bar{F}_t - F_t}{F_t}\right\} \exp\left\{-F_i - F_{i+1} \dots F_{t-1}\right\} \qquad 2.4$$

while for larger values of Z this formula tends to overstate the error and is therefore still of some value.

A similar formula to 2.4 gives $\rho'(N_i)$, the proportional error in N_i when C_t is the catch in year t and all subsequent years. In this case

$$\rho'(N_i) = \frac{M}{Z_t} \left(\frac{\bar{F}_t - F_t}{F_t}\right) \exp\left\{-F_i \dots -F_{t-1}\right\} \qquad 2.5$$

and therefore

$$\rho'(N_i) = \frac{M}{Z_t} \rho(N_i) \, . \qquad 2.6$$

It is therefore simple to convert a table of $\rho(N_i)$ into a table of $\rho'(N_i)$. In either case the proportional error of F_i, $\rho(F_i)$, is given approximately by the formula

$$\rho(F_i) \simeq - \frac{\rho(N_i)}{1 + \rho(N_i)} \, . \qquad 2.7$$

Figures 1 and 2 show graphs of $\rho(N_i)$ and $\rho(F_i)$ plotted against the sum of the fishing mortality from year i to year t - 1 (cumulative fishing mortality). It can be seen that the underestimation of F_t results in estimates of N_i which are too large and estimates of F_i which are too small, whereas overestimating F_t has the reverse effect. It can also be seen that as the cumulative fishing mortality increases, both types of error decrease. As an example, if F_t was overestimated by 100% for a year-class and the cumulative fishing mortality from year i to year t - 1 was 2.0, then the percentage error in N_i would at most be -7% and the percentage error in F_i would be +7%. If, however, F_t was underestima by 50% and the cumulative fishing mortality was equal to 2.0, then the percentage error in N_i would be at the most 14% and the percentage error in F_i would be -12%. Thus, provided that F_t can be estimated within this range and provided that the cumulative fishing mortality is greater than 2.0, the error in the estimates of N_i and F_i should be small enough for most uses. If, however, the cumulative fishing mortality is small, which is the case when the number of recruits to a year-class is estimated from the catches of partially recruited age groups, then the accurate estimation of N_i and F_i will require the accurate choice of F_t. It should also be realised that since the cumulative fishing mortality is the sum of the fishing mortalities from age i to age t - 1 it must, for a particular year-class, be a monotonically decreasing function of age. Hence the bias in F_i caused by the incorrect choice of F_t will be greatest amongst the oldest age groups and this may upset estimates of selectivity with age. Table 2 shows the results of a cohort analysis for the 1956 year-class of the Arcto-Norwegian cod. This assumes that the true values of M and F_t are 0.3 and 0.8 respectively and shows the percentage errors in N_i and F_i when F_t is overestimated by 100% or underestimated by 50%. These errors were computed by rerunning the data with the appropriate value of F_t and are therefore precise. It can be seen that these percentage errors are similar but, in general, smaller than their estimates in Figures 1 and 2.

3. Error in cohort analysis due to the sampling error of C_i

Unlike the estimate of F_t, which is usually an arbitrary choice, each estimate of catch at age can be assigned a variance, although this is seldom available, due to the heavy work involved in its computation (see Gulland 1955). Assuming such variances to be available it is a simple matter to compute the resulting variance of N_i and F_i, since

$$\text{variance}\ (N_i) = \text{variance}\ (C_i)\ \exp\{M\} + \text{variance}\ (N_{i+1})\ \exp\{2M\}\ , \quad \ldots\ 3.1$$

and this may be used as a recurrence relationship to obtain

$$\text{variance}\ (N_i) = \text{variance}\ (C_i)\ \exp\{M\} + \text{variance}\ (C_{i+1})\ \exp\{3M\} + \ldots$$

$$\ldots + \text{variance}\ (C_t)\ \frac{\exp\{2(t-i)M\}\ (F_t + M)^2}{F_t^2\ (1 - \exp\{-F_t - M\})^2} \quad \ldots\ldots\ldots\ldots\ 3.2$$

which is a very similar formula to 1.4.

The equivalent variance of F_i can be approximated, since

$$F_i = \log_e \left\{ N_i / N_{i+1} \right\} - M \qquad 3.3$$

which yields approximately

$$\text{variance } (F_i) \simeq \frac{\text{variance } (N_i)}{N_i^2} - \frac{2 \text{ variance } (N_{i+1}) \exp\{M\}}{N_i N_{i+1}} + $$

$$+ \frac{\text{variance } (N_{i+1})}{N_{i+1}^2} . \qquad 3.4$$

Equations 3.2 and 3.4 should be used to calculate the respective variances of N_i and F_i in a particular case, but in order to appreciate the approximate magnitude of these variances the following approximate formulae are useful:

$$(\text{variance ratio } N_i)^2 \simeq (\text{variance ratio } C_i)^2 \left(1 - \exp\{-F_i\}\right)^2 +$$

$$+ (\text{variance ratio } N_{i+1})^2 \left(\exp\{-F_i\}\right)^2 ; \qquad 3.5$$

$$(\text{variance ratio } F_i)^2 \simeq \frac{\left(1 - \exp\{-F_i\}\right)^2}{F_i^2} \Big((\text{variance ratio } C_i)^2 +$$

$$+ (\text{variance ratio } N_{i+1})^2 \Big) . \qquad 3.6$$

Figure 3 shows graphs of these formulae for each year from the final year, that is for the number of years from the estimate in question to the final year. The graphs are given for the case when the variance ratio of the catch-at-age data is constant, and when the fishing mortality is constant throughout the life of the fish. Although these conditions are unrealistic, the rapid convergence of the graphs to asymptotic values does suggest that the graphs would indicate the approximate value of the variance ratio of the estimates of N_i and F_i, even when F_i is not constant from year to year. As an example of the use of the graph, the estimate of N_5 (for a year-class with an oldest age group of 12 years old, experiencing a fishing mortality of 0.6 per year) would have a variance ratio of approximately 54% of the variance ratio of the catch data. Similarly the estimate of F_5 would have a variance ratio of approximately 85% of the variance ratio of the catch data. Hence, if the variance ratio of the catch data was 10%, then the variance ratios of N_5 and F_5 would be 5.4% and 8.5% respectively. As a result the approximate 95% confidence limits for the estimates would be $\pm$ 10.8% of the estimate of N_5 and 17.0% of the estimate of F_5.

Table 3 shows the 1956 Arcto-Norwegian cod results, together with the standard deviations and variance ratios of N_i and F_i. These were computed from equations 3.2 and 3.4 on the assumption that the variance ratio of the catch data at each age was 10%. It can be seen that the variance ratios of these estimates are not very different from those which would have been predicted by entering the graphs of Figure 3 with appropriate values of F_1 at the asymptotic parts of the graphs. Thus Figure 3 should prove to be of some value in providing quick estimates of the variance ratios of N_1 and F_1 for any year-class which has catch data which have approximately constant variance ratios.

4. Summary

This document provides formulae for calculating the error introduced in cohort analysis (and therefore virtual population analysis) by errors in F_t and by the sampling error of catch data. It also provides some quick estimates of the likely size of such errors. These estimates suggest that such errors converge to fairly small values, but they also suggest that a knowledge of the approximate value of these errors will always be a safeguard against misinterpretation of data!

References

GULLAND, J. A., 1955. Estimation of growth and mortality in commercial fish populations. Fishery Invest., London, Ser. 2, 18(9).

GULLAND, J. A., 1965. Estimation of mortality rates. Annex to Arctic Fisheries Working Group Report (meeting in Hamburg, January 1965); ICES CM 1965, Gadoid Fish, Doc. No. 3 (mimeo).

Table 1 Comparison of the results of virtual population analysis and cohort analysis

Arcto-Norwegian cod, 1956 year-class

M = 0.3

1. Virtual population analysis
2. Cohort analysis

Age (years)	Fishing mortality, F_n			Population $N_i \times 10^{-6}$		
	(1)	(2)	% error	(1)	(2)	% error
12	0.8000*	0.8000*		0.2	0.2	
11	1.3400	1.3670	2	1.1	1.1	0
10	0.7826	0.7806	-	3.1	3.2	2
9	0.6768	0.6747	-	8.3	8.5	2
8	0.6582	0.6570	-	21.7	22.2	2
7	0.8636	0.8657	-	69.6	71.2	2
6	0.7341	0.7333	-	195.6	200.1	2
5	0.4289	0.4261	1	405.5	413.6	2
4	0.1874	0.1854	1	660.2	672.0	2
3	0.0411	0.0405	1	928.5	944.7	2
2	0.0024	0.0024	-	1256.4	1278.2	2
1	0.0007	0.0007	-	1697.1	1726.6	2

*assumed

Table 2 The percentage error in N_i and F_i when F_t is overestimated by 100% and when F_t is underestimated by 50% for the 1956 year-class of the Arcto-Norwegian cod, with $M = 0.3$ and when the true value of $F_t = 0.8$

Age (years)	$N_i \times 10^{-6}$	F_i	Cumulative F_i	% error when F_t is taken as 0.4		% error when F_t is taken as 1.6	
				in N_i	in F_i	in N_i	in F_i
12	0.2	0.8000*	-	+ 68.66		- 32.27	
11	1.1	1.3670	1.3670	+ 17.50	- 26.44	- 8.22	+22.21
10	3.2	0.7806	2.1475	+ 8.02	- 10.77	- 3.77	+ 6.07
9	8.5	0.6747	2.8222	+ 4.08	- 5.50	- 1.92	+ 2.82
8	22.2	0.6570	3.4792	+ 2.12	- 2.91	- 0.99	+ 1.45
7	71.2	0.8657	4.3449	+ 0.89	- 1.40	- 0.42	+ 0.67
6	200.1	0.7333	5.0782	+ 0.43	- 0.63	- 0.20	+ 0.29
5	413.6	0.4261	5.5043	+ 0.28	- 0.35	- 0.13	+ 0.16
4	672.0	0.1854	5.6897	+ 0.23	- 0.22	- 0.11	+ 0.16
3	944.7	0.0405	5.7302	+ 0.22	- 0.25	- 0.10	+ 0.25
2	1278.2	0.0024	5.7326	+ 0.22	- 0.00	- 0.10	+ 0.00
1	1726.6	0.0007	5.7333	+ 0.22	- 0.00	- 0.10	+ 0.00

*assumed

Table 3 Standard deviations and variance ratios of N_i and F_i calculated for the 1956 Arcto-Norwegian cod, assuming that the variance ratio for the catch at each age was 10% and that $M = 0.3$ and $F_t = 0.8$

Age (years)	$N_i \times 10^{-6}$	F_i	Standard deviation		Variance ratio (%)	
			$N_i \times 10^{-6}$	F_i	N_i	F_i
12	0.2	0.8000*	0.01449			
11	1.1	1.3670	0.08361	0.09192	7.60	6.72
10	3.2	0.7806	0.20763	0.06801	6.49	8.71
9	8.5	0.6747	0.50349	0.05865	5.92	7.51
8	22.2	0.6570	1.26666	0.05601	5.71	8.53
7	71.2	0.8657	4.46487	0.06669	6.27	7.70
6	200.1	0.7333	12.01878	0.06134	6.01	8.36
5	413.6	0.4261	21.65851	0.04047	5.24	9.50
4	672.0	0.1854	31.37061	0.01910	4.67	10.30
3	944.7	0.0405	42.51186	0.00439	4.50	10.83
2	1278.2	0.0024	57.38580	0.00026	4.49	10.77
1	1726.6	0.0007	77.46281	0.00007	4.49	10.59

*assumed

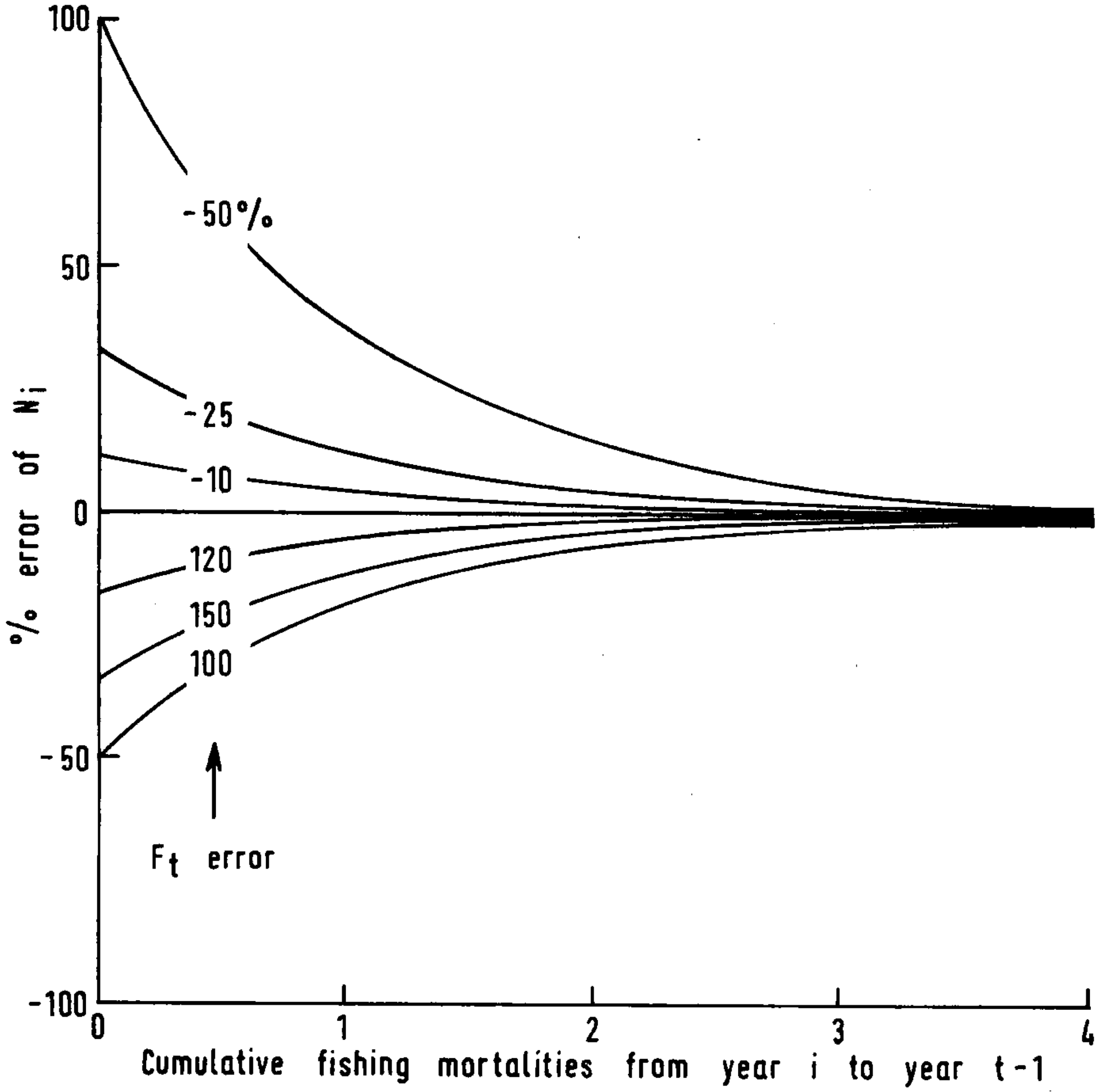

Figure 1 Graphs of the percentage error in N_i due to incorrect values of F_t plotted against the cumulative fishing mortalities from year i to year t-1.

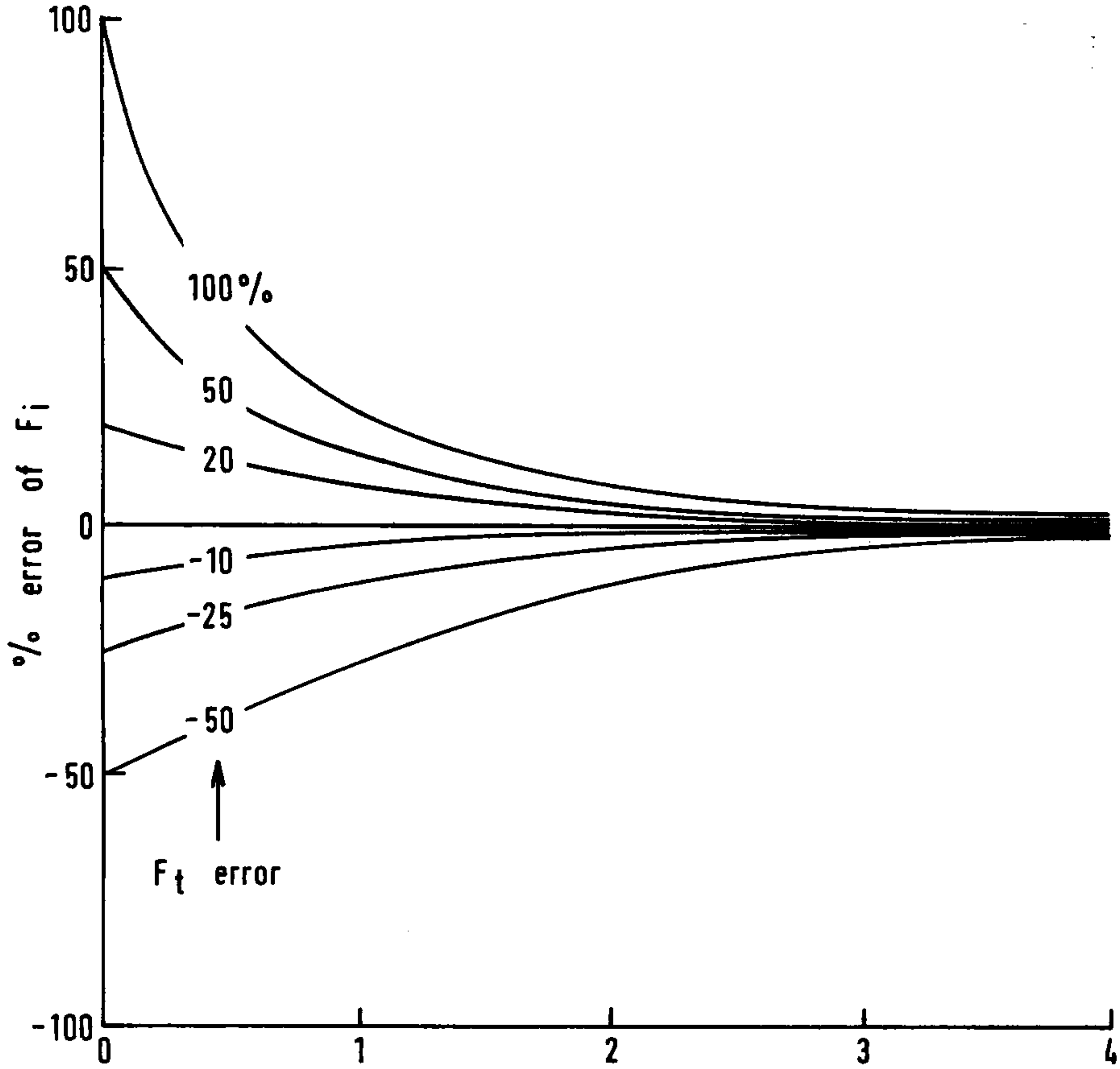

Figure 2 Graphs of the percentage error in F_i due to incorrect values of F_t plotted against the cumulative fishing mortalities from year i to year t-1.

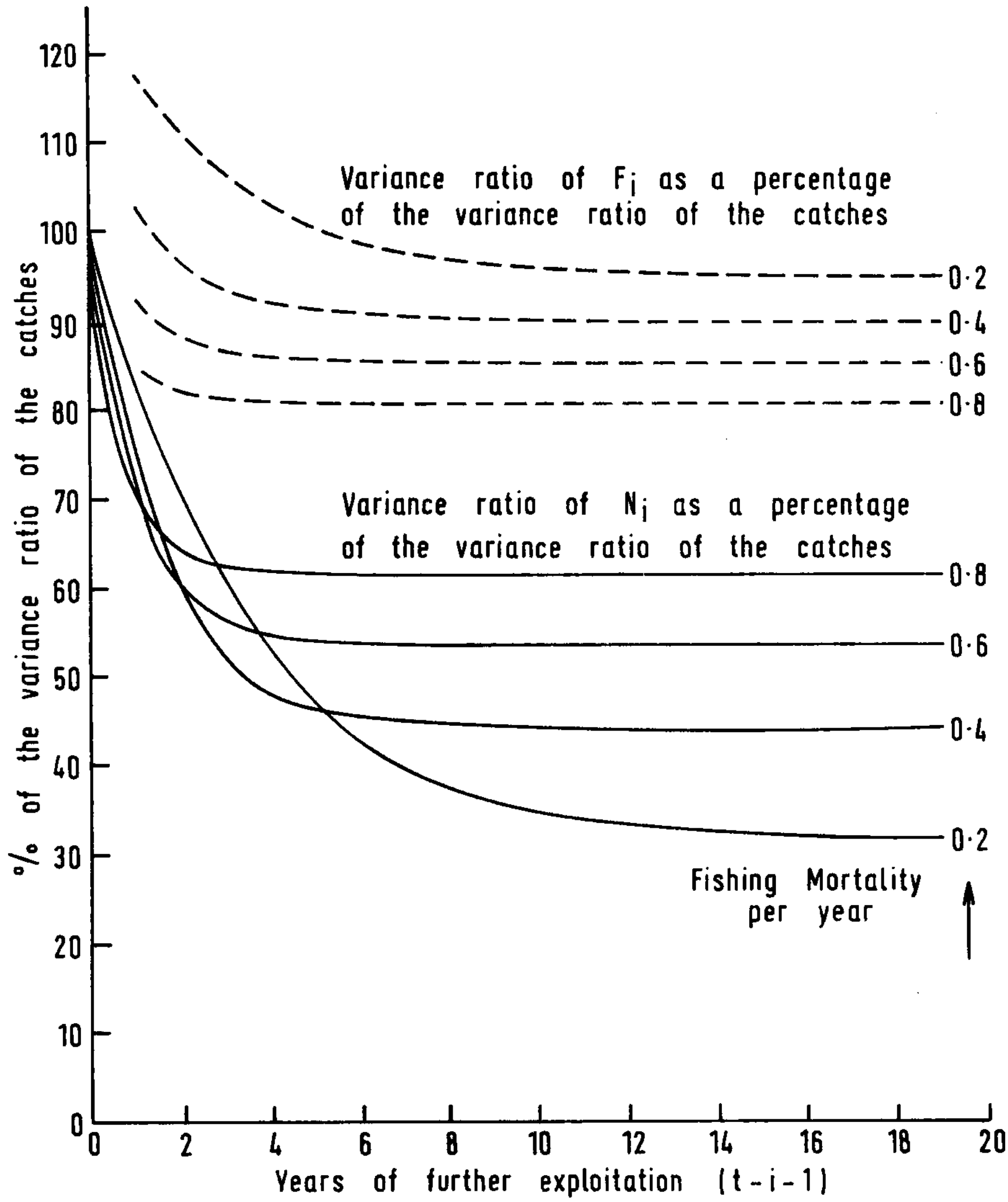

Figure 3 Graphs of the percentage variance ratio of F_i and of N_i for various constant levels of fishing mortality plotted against the years of further exploitation.