# Virtual Population Analysis (VPA) Equations for Nonhomogeneous Populations, and a Family of Approximations Including Improvements on Pope's Cohort Analysis

## Alec D. MacCall

*Southwest Fisheries Center, National Marine Fisheries Service, P.O. Box 271, La Jolla, CA 92038, USA*

A set of "backward" virtual population analysis (VPA) equations relates catch $(C_t)$ from continuous fishing between times $t$ and $t + 1$ to population size $(N_t, N_{t+1})$ when a portion of the stock is unavailable to fishing. The usual VPA equations become a special case where the entire stock is available (i.e. the stock is homogeneous). A close approximation to the VPA equations is $N_t = N_{t+1} \exp(M) + C_t M/(1 - \exp(-M))$, which has properties similar to Pope's "cohort analysis" and is somewhat more accurate in the case of a continuous fishery, especially if the natural mortality rate $(M)$ is large. Much closer simple approximations are possible if the seasonal pattern of catches is known.

Une série d'équations d'analyse de population virtuelle (APV) « à rebours » lie les prises $(C_t)$ tirées de la pêche continue entre les temps $t$ et $t + 1$ à la taille de la population $(N_t, N_{t+1})$ quand une partie du stock n'est pas disponible pour la pêche. Les équations habituelles d'APV deviennent un cas spécial là où le stock en entier est disponible (c.-à-d. que le stock est homogène). Une approximation étroite des équations d'APV est $N_t = N_{t+1} \exp(M) + C_t M/(1 - \exp(-M))$ qui a des propriétés semblables à l'analyse par cohortes de Pope et est un peu plus précise dans le cas d'une pêche continue, surtout si le taux de mortalité naturelle $(M)$ est important. Des approximations simples beaucoup plus étroites sont possibles si l'on connaît la tendance saisonnière des prises.

Virtual population analysis (VPA), also known generically as cohort analysis, has come to be the preeminent method of fish stock assessment since its appearance 20 yr ago (Gulland 1965; Murphy 1965). The method usually is used in the form of a "backwards solution" where historical abundances of a cohort are estimated on the basis of subsequent catches from a presumed homogeneous stock: all fish of a specified age are assumed to be equally susceptible to capture. The equations unfortunately do not allow convenient separation of the unknown quantities from the known and/or assumed quantities, so the solution must be obtained by numerical methods, e.g. Sims (1982), or by approximation, e.g. Pope's (1972) popular "cohort analysis." Computers have made numerical solution very easy, but there remain applications of approximation methods. For example, Pope's approximation has proved useful for statistical and mathematical analysis of the behavior of VPA, whereas the exact equations are intractable. In practice, good approximations and efficient algorithms remain useful, as computers may not conveniently be available in working group meetings, in field locations, and in some developing countries.

The following discussion provides two extensions to existing VPA methods. The first is a set of equations which include possible assumptions concerning homogeneity of cohorts, i.e.

part of the cohort may not be available or vulnerable to fishing. These equations also provide the basis for an easy and rapid iterative backward solution of the VPA equations. The second is description of a family of approximation formulas which are similar in concept to Pope's cohort analysis.

## VPA Equations for Nonhomogeneous Populations

Assume that the cohort consists of two segments, one fished and one unfished (cf. Widrig 1954). This division could arise under a variety of circumstances. For example, some members of the cohort may reside in habitats or regions which cannot be fished. Another possibility is differential vulnerability to the gear, depending on fish size or other attribute. The present treatment corresponds to the common model of "knife-edge" gear selection. Importantly, there must be no interchange between the two segments during the time interval. Let $P$ be the age-specific fraction of the cohort which is subject to fishing at the beginning of a time interval. The remaining notation follows that of Ricker (1975), except that $F$ refers to the instantaneous coefficient of fishing mortality *on the fishable segment of the cohort*. As usual, $F$ and $M$ are assumed to be independent, and $M$ has the same value in both segments. Both of these assumptions are open to question (e.g. Lawless 1982, p. 491; Ricker 1976), but there seems to be no practical alternative

2406

*Can. J. Fish. Aquat. Sci., Vol. 43, 1986*

given the usual state of knowledge. All parameters are assumed to be constant within the time period, but may vary among ages or time periods. A full notation is useful for fishery analysis, but requires extensive use of subscripts denoting age and time; this notation can be developed by the reader for particular applications (Schnute 1985 provides a useful structure for this purpose). The present notation is simplified for ease of exposition, and considers a single cohort and a unit time period. The time subscript $t$ is shown only where desirable for clarity. Under continuous fishing, the catch ($C_t$) for a unit time period beginning at time $t$, and ending at time $t + 1$, is given by

(1)     $C_t = N_t PFA/Z$

where $Z = F + M$, and $A = 1 - \exp(-Z)$. The abundance at the beginning of the next time period is

(2)     $N_{t+1} = N_t P\exp(-Z) + N_t(1-P)\exp(-M)$

where the two additive terms on the right side represent contributions from the fished and the unfished segments, respectively. This expression may be rearranged to give

(3)     $N_{t+1} = N_t P [\exp(-Z) - \exp(-M)] + N_t\exp(-M)$

and equation (1) allows substitution of $C_t Z/FA = N_t P$, giving

(4)     $N_{t+1} = (N_t - C_t mZ/FA)\exp(-M)$

where $m = 1 - \exp(-F)$. The backward solution therefore is

(5)     $N_t = N_{t+1} \exp(M) + C_t mZ/FA$.

Equations (4) and (5) also can be derived from the equations describing the usually assumed homogeneous case. However, because $P$ appears in equations (1) and (2), but not in (4) or (5), this derivation demonstrates that they have the somewhat surprising property of being true whether or not the homogeneity assumption ($P = 1$) is met, provided that $F$ is defined appropriately. These equations also demonstrate that for a given catch and natural mortality rate, the four variables, $N_t$, $N_{t+1}$, $P$, and $F$, possess only two degrees of freedom among them. This means that specification of any two values should determine both remaining values, but this property may be of limited use as the basis for direct parameter estimation (e.g. for a given catch, $F$ and $P$ covary strongly, and estimation errors could easily lead to misinterpretation). In practice, one value is given; for example, in the case of a backwards solution, $N_{t+1}$ may have been provided by an initial guess or by VPA of older aged catches. The second value is imposed by a simultaneous equation which may explicitly or implicitly fix $P$ (see equations (6) and (7), respectively). Equation (5) is aesthetically attractive in that it clearly expresses the sense of "virtual population" as a modified sum of subsequent catches.

There are many possible equations which give $F$ as a function of $P$, $N_t$, $N_{t+1}$, and $M$, depending on the assumed behavior and distribution of the stock and fishery. Differences among these equations become especially important when they are applied sequentially to link successive ages, as is the usual case in VPA. For example, equation (2) can be rearranged algebraically to give

(6)     $F = -\ln\left[\dfrac{(P-1)N_t + N_{t+1}\exp(M)}{PN_t}\right]$

However, if $P$ is treated as an independent population parameter, sequential use of this equation must be based on an assumed redistribution of all of the survivors at the end of each
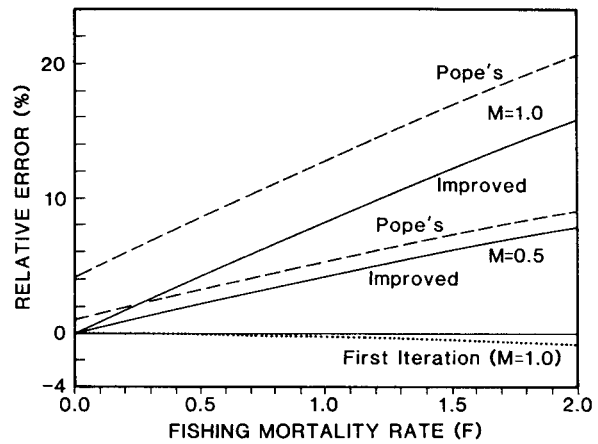
FIG. 1. Relative error (see text for definition) of Pope's and the new approximation for a homogeneous stock ($P = 1$) and two values of $M$. For comparison, the relative error of the result of the first iteration of the VPA algorithm is shown for $M = 1$, where relative error is defined as $((\hat{m}\hat{Z}/\hat{F}\hat{A})/(mZ/FA)) - 1$.

time period, so that the individuals in the fished and unfished segments have no differential tendency to remain in their previous states. Alternatively, if $P$ can be estimated independently for individual ages and times, perhaps from surveys of habitat distribution or of size composition, equation (6) allows exploitation rates to be separated into component sources of variation: those associated with changes in $P$ and those associated with changes in $F$ on the available segment. This form of "separable VPA" (sensu Doubleday 1976) is being developed by Daniel Kimura (NMFS-NWAFC, BIN C15700, 7600 Sandpoint Way N.E., Seattle, WA 98115, USA, manuscript in revision). Of course, in the case of homogeneity ($P = 1$), equation (6) simplifies to

(7)     $F = \ln(N_t/N_{t+1}) - M$

as was suggested by Pope (1972).

Equations (5) and (7) also provide a convenient rapid algorithm for solving the usual VPA equations for a homogeneous stock. The procedure is as follows: (1) given $C_t$, $N_{t+1}$, and $M$, use Pope's approximation or the improved approximation described below to obtain a first estimate of $N_t$; (2) use this value of $N_t$ in equation (7) to obtain an estimate of $F$; (3) use this value of $F$ in equation (5) to get an improved estimate of $N_t$; (4) repeat steps 2 and 3 until the desired level of precision is reached, judging by the change from the estimate in the previous iteration; and (5) repeat the procedure on the next younger age of the cohort, letting the previous $N_t$ now be $N_{t+1}$, and go to step 1. The rate of convergence is very high (Fig. 1), approaching that of Newton's method as described by Sims (1982); two iterations often are sufficient. Also, this method does not converge to extraneous solutions.

## A Family of Approximations

The independence of $C_t$ in equation (5) indicates that there exists a family of VPA approximations of the form

(8)     $\hat{N}_t = N_{t+1}\exp(M) + C_t V(F,M,\ldots)$

where $\hat{N}_t$ is an approximation of $N_t$ and $V(F,M,\ldots)$ is a

function approximating the value of $mZ/FA$. These approximations are applicable to nonhomogeneous fisheries ($P \neq 1$) as well as the more commonly assumed homogeneous fisheries. In the case of nonhomogeneous fisheries, these equations need to be coupled with an equation such as (6).

Pope's (1972) cohort analysis is a member of this family of approximations, where $V = \exp(M/2)$:

(9) $\quad \hat{N}_t = N_{t+1} \exp(M) + C_t \exp(M/2)$.

Another approximation is provided by a first-order Taylor series approximation of the exponentials in the left side of the expression

(10) $\quad mn/A \approx FM/Z$

where $n = 1 - \exp(-M)$. Rearrangement of (10) gives

(11) $\quad V = mZ/FA \approx Mn$

which may be substituted into equation (8) to give

(12) $\quad \hat{N}_t = N_{t+1} \exp(M) + C_t M/n$.

The relative accuracy of approximations (9) and (12) can be expressed as the relative error in $V$, the coefficient of $C_t$; the remaining terms of the equations are identical and tend to reduce clarity of the comparisons. Also, because equations (9) and (12) are linear in $C$, errors in the coefficients propagate in a straightforward manner. Let the "relative error" ($r_x$) of an approximation, $\hat{X}$, be defined as $r_x = (\hat{X} - X)/X$, or equivalently, $r_x = (\hat{X}/X) - 1$, where $X$ is the true value. In the present case, the approximation is the coefficient applied to $C_t$ in equations (9) and (12), and the true value is the coefficient $mZ/FA$ in equation (5)). Thus, the relative error for Pope's approximation ($r_{(11)}$) is

(13) $\quad r_{(9)} = \dfrac{\exp(M/2)}{mZ/FA} - 1$

where $\exp(M/2)$ is Pope's cohort analysis coefficient of $C_t$. Similarly, the relative error of the new approximation in equation (12) is

(14) $\quad r_{(12)} = \dfrac{M/n}{mZ/FA} - 1$.

Relative errors for these two approximations are compared in Fig. 1; generally, the differences are not large. The relative error of the new approximation in equation (12) is slightly but consistently smaller than that of Pope's approximation. The new approximation becomes considerably better than Pope's as $M$ becomes large, thus extending the range of mortality rates under which approximation is useful. Taylor series expansions of the relative errors, equations (13) and (14), clarify the nature of the difference between the two approximations. If we discard terms higher than quadratic, the relative error for Pope's approximation is $(2FM + M^2)/24$, whereas that of the new approximation is $(2FM)/24$. The constant factor of $M^2/24$ in Pope's relative error explains both its nonzero intercept at $F = 0$ and the parallel appearance of the plots in Fig. 1.

Pope's approximation is equivalent to assuming that the entire catch is taken instantaneously at midyear. The new approximation does not have a corresponding convenient interpretation. A tentative interpretation may be derived from the relationship $\bar{N} = N_0 n/M$, which describes average abundance ($\bar{N}$) over a unit time interval versus initial abundance ($N_0$), if only natural mortality is acting. Thus, the quantity $M/n$ used
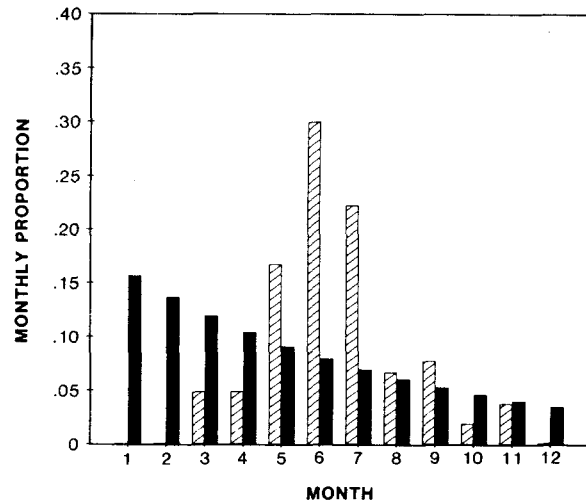


FIG. 2. Comparison of monthly catch distribution under the assumption of constant $F$ and that from a seasonal fishery (the California catch of mackerel in 1941–42, Anonymous 1944).

in the new approximation is equal to $N_0/\bar{N}$ if there is only natural mortality. Paradoxically, this means that the new approximation assumes the fishery's catch to be a continuous sampling of natural deaths in the absence of a fishery. This somewhat strange interpretation is consistent with the fact that the new approximation is identical to the expression which results from taking the limit of equation (5) for a given catch as $F \to 0$.

## VPA Approximation for Seasonal Fisheries

The usual assumption in VPA is that $F$ is constant over the time interval under consideration. This assumption requires that the catch decline exponentially over that interval. Many, if not most, fisheries are seasonal, so the pattern of catches over the year or season does not conform to this assumption (Fig. 2). A standard practice in these cases is to divide the fishing year into shorter time periods such as quarters or months, within which $F$ is more nearly constant.

If the seasonality of the fishery is predictable, a useful approximation can be developed as follows. If the exact time ($t_i$) of each catch ($C_i$) were known, VPA equation (8) would reduce to

(15) $\quad N_t = N_{t+1} \exp(M) + \sum_i C_i \exp((t_i - t_0)M)$

where $t_0$ is the beginning of the fishing year. Equivalently, if the proportion ($p_i$) of the total catch ($C_t$) taken at time $t_i$ is known, we have

(16) $\quad N_t = N_{t+1} \exp(M) + C_t \sum_i p_i \exp((t_i - t_0)M)$.

Equation (16) has the same form as equation (8), demonstrating that seasonal fisheries can be treated by the same family of approximations. Moreover, equation (16) requires information only on $M$ and the temporal distribution of catches, making function $V_{seas}$, and hence the VPA, independent of estimated $F$.

In practice, the value of $V_{seas}$ for a seasonal fishery can be

2408

obtained from sequential VPA of the typical monthly proportions of the total annual catch (see example below). By setting $N_{t+1}$ to zero, $V_{seas}$ is given directly by the virtual population of the $p_i$ values at time $t_0$, this being equivalent to the summation in equation (16).

## Application of Approximations

Parrish and MacCall (1978) gave VPA estimates for mackerel (*Scomber japonicus*), based on annual catch-at-age data, $M = 0.5$, and assumed constant $F$ within the fishing season. Estimated $F$ for fish aged 4 and older in the 1941−42 fishing year (May through April) was 1.115, giving a $V_{annual}$ of $mZ/FA$ = 1.2152. The monthly proportions of the total unaged annual catch are based on Anonymous (1944) and are shown in Fig. 2. These monthly catches are in weight rather than numbers, causing some error due to individual growth during the year. A value of $V_{seas}$ can be approximated by letting $V_{month}$ = $(M/12)/(1-\exp(-M/12))$. The annual value of $V$ based on the seasonal catch distribution is approximately

$$(17) \quad V_{seas} = \sum_{i=1}^{12} p_i V_{month} \exp((i - 1)M/12)$$

giving $V_{seas} = 1.2851$. Thus, the relative error of Parrish and MacCall's annually based VPA is approximately −5.4% for this year, indicating a moderate underestimate of abundance, given the assumptions. Interestingly, the new VPA approximation in equation (12) gives $V_{new} = 1.2707$, and Pope's approximation gives a nearly exact value of $V_{Pope} = 1.2840$. Pope's approximation does well because in this case the actual catch was indeed taken very near midyear. Of course, other seasonal distributions may produce values of $V_{seas}$ ranging from 1.0 (all the catch taken at year's end) to $\exp(M)$ if all of the catch is taken at the beginning of the year.

Differences in VPA estimates of abundance due to various homogeneity assumptions and/or approximations often fall within the variability due to imprecision in the value of the natural mortality rate, $M$. For example, given $M = 0.5$ and constant $F = 0.5$, Pope's approximation using $M = 0.48$ would have the same relative error as the new approximation using $M = 0.5$. The difference becomes larger for large $M$: given $M = 1.0$ and constant $F = 1.0$, Pope's approximation would require $M = 0.92$ to produce a relative error equal to that of the new approximation. Violations of the assumption of constant $F$ over the time period can produce somewhat larger errors. The relative error of Parrish and MacCall's annual VPA in the preceding example is roughly equal to the relative error which would have occurred if they had used a seasonally corrected

VPA with $M = 0.39$, which is 22% smaller than assumed $M$.

It can be argued that the choice of approximations is inconsequential because imprecision associated with alternative VPA approximations and assumptions is smaller than that associated with imprecision in assumed $M$. A counterargument is that errors due to approximations or assumptions are systematic biases and should be recognized as such. If a value of $M$ is stated, VPA using that $M$ should as nearly as possible reflect that value, rather than some systematically different effective value arising from the properties of an unnecessary assumption or approximation.

## References

ANONYMOUS. 1944. The commercial fish catch of California for the years 1941 and 1942., Calif. Div. Fish Game Fish Bull. 59: 71 p.
DOUBLEDAY, W. G. 1976. A least squares approach to analyzing catch at age data. Res. Bull. Int. Comm. Northeast Atl. Fish. 12: 68−81.
GULLAND, J. A. 1965. Estimation of mortality rates. Annex to Arctic Fisheries Working Group Rep., ICES CM 1965, Doc. No. 3: 9 p. (Mimeo.)
LAWLESS, J. F. 1982. Statistical models and methods for lifetime data. John Wiley & Sons, New York, NY. 580 p.
MURPHY, G. I. 1965. A solution of the catch equation. J. Fish. Res. Board Can. 22: 191−202.
PARRISH, R. H., AND A. D. MACCALL. 1978. Climatic variation and exploitation in the Pacific mackerel fishery. Calif. Dep. Fish Game Fish Bull. 167: 110 p.
POPE, J. G. 1972. An investigation of the accuracy of virtual population analysis using cohort analysis. Res. Bull. Int. Comm. Northwest Atl. Fish. 9: 65−74.
RICKER, W. E. 1975. Computation and interpretation of biological statistics of fish populations. Bull. Fish. Res. Board Can. 191: 382 p.
1976. Review of the rate of growth and mortality of Pacific salmon in salt water, and noncatch mortality caused by fishing. J. Fish. Res. Board Can. 33: 1483−1524.
SCHNUTE, J. 1985. A general theory for analysis of catch and effort data. Can. J. Fish. Aquat. Sci. 42: 414−429.
SIMS, S. E. 1982. Algorithms for solving the catch equation forward and backward. Can. J. Fish. Aquat. Sci. 39: 197−202.
WIDRIG, T. 1954. Method of estimating fish populations, with application to Pacific sardine. Fish. Bull. U.S. 56: 141−166.

*Can. J. Fish. Aquat. Sci., Vol. 43, 1986*

2409