Reprinted from

Canadian Journal of Fisheries and Aquatic Sciences

Réimpression du

Journal canadien des sciences halieutiques et aquatiques

# Sensitivities and variances of virtual population analysis as applied to the mackerel, *Scomber japonicus*

M. H. PRAGER AND A. D. MACCALL

Volume 45 • Number 1 • 1988

Pages 539–547

Canada

Fisheries and Oceans
Pêches et Océans

Printed in Canada by The Runge Press Limited

# Sensitivities and Variances of Virtual Population Analysis As Applied to the Mackerel, *Scomber japonicus*

Michael H. Prager[1] and Alec D. MacCall

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

Prager, M. H., and A. D. MacCall. 1988. Sensitivities and variances of virtual population analysis as applied to the mackerel, *Scomber japonicus*. Can. J. Fish. Aquat. Sci. 45: 539–547.

Virtual population analysis (VPA) is widely used in fish stock assessment. However, VPA results are generally presented as point estimates, without error variance. Using numerical methods, we estimated the total variance of historical (1929–65) biomass estimates of mackerel, *Scomber japonicus*, off southern California. In the years before 1940, coefficients of variation (CV's) approached 100%; later, when weights at age and the age structure of the catch were better known, the CV's were about 25%. Most of the variability derives from uncertainties in estimates of natural mortality (*M*) and of weights at age. We also developed dimensionless coefficients (sensitivities) to examine the effects of errors in the inputs on the VPA biomass estimates. The largest sensitivities were to *M* and the total catch and varied substantially from year to year. As expected, sensitivity to *M* decreased with increasing exploitation, and sensitivity to catch increased with increasing exploitation. Using such sensitivities, one could estimate the error in a biomass estimate for a past year when *M* (or any other input) was thought to be unusually high or low. Thus, retrospective corrections can be made. Also, such sensitivities form an analytic tool for examining the properties of VPA, or any quantitative model.

L'analyse de population virtuelle (APV) est largement utilisée pour l'évaluation des stocks de poissons. Les résultats de cette analyse sont généralement présentés comme des estimations ponctuelles, sans erreur de variance. Les auteurs ont estimé par méthodes numériques la variance totale, de 1929 à 1965, des estimations de biomasse de maquereau, *Scomber japonicus*, au large du sud de la Californie. Au cours des années antérieures à 1940, les coefficients de variation (CV) étaient presque de 100 % tandis que plus tard, lorsque l'on a mieux connu les poids selon l'âge et la structure des âges des prises, les valeurs de CV étaient de 25 % environ. La plus grande partie de la variabilité découle d'incertitudes liées aux estimations de la mortalité naturelle (*M*) et du poids selon l'âge. Les auteurs ont aussi déterminé des coefficients sans dimension (de sensibilité) afin d'étudier les effets des erreurs des paramètres d'entrée sur les estimations de biomasse par APV. Les éléments de sensibilité les plus importants avaient trait à *M* et aux prises totales et variaient de façon appréciable d'une année à l'autre. Tel que prévu, la sensibilité à la valeur de *M* diminuait à mesure que le taux d'exploitation augmentait et celle aux prises augmentait en même temps que le taux d'exploitation. Ces éléments de sensibilité pourraient permettre d'estimer l'erreur d'une biomasse estimée pour une année antérieure lorsque la valeur de *M* (ou de tout autre paramètre d'entrée) apparaît anormalement élevée ou faible. Il serait donc possible d'apporter des corrections a posteriori. De plus, ces éléments de sensibilité constituent un outil analytique pour l'examen des propriétés de l'APV ou de tout autre modèle quantitatif.

*Received February 21, 1987*
*Accepted November 2, 1987*
*(J9150)*

*Reçu le 21 février 1987*
*Accepté le 2 novembre 1987*

Virtual population analysis (VPA) has been applied widely to commercial fish stocks in the more than 20 yr since its development (Murphy 1965; Gulland 1965). Although several studies have examined the errors implicit in the method, VPA abundance estimates still tend to be presented as exact numbers, i.e. without error variances. Pope (1972) examined separately the effects of error in the estimate of terminal instantaneous fishing mortality (*F*) and of errors in the catch statistics. Sims (1982) examined the errors caused by assuming a constant fishing effort throughout the fishing season, in the presence of seasonal variation in effort. Saila et al. (1985) discussed the propagation of error in Pope's (1972) approximation (and therefore the basic model). However, none of the above authors discussed the effects of simultaneous errors in several parameters, nor of covariances among the parameters. Sampson (1984) overcame these limitations for analysis of a single cohort. However, the cited studies have stopped short of estimating the variances of VPA abundance estimates themselves.

We used the delta method (Seber 1973) to examine the errors caused by simultaneous errors in any number of parameters, including covariances, and have applied the method to the southern California stock of mackerel, *Scomber japonicus*. Our work provides a numerical estimate of the variance of each year's biomass estimate. In addition, we have developed a method to examine the sensitivity of VPA to its input parameters.

The California fishery on *S. japonicus*, known locally as Pacific mackerel and elsewhere as chub mackerel, supported a commercial fishery from the late 1920's to the mid-1960's (Parrish and MacCall 1978). The resource was severely depleted

---

[1] Current address: Oceanography Department, Old Dominion University, Norfolk, VA 23508, USA.

during the following decade, but recovered strongly in the late 1970's, when it reached record levels of abundance (MacCall et al. 1985). As the fishery is monitored closely by the California Department of Fish and Game, continuous landings and age composition data are available except for the years 1970–75, when landings were insufficient to support the monitoring effort. We used the data collected from 1920 to 1969 as the basis for the present study. The data collection and subsequent analysis are described in detail in Parrish and MacCall (1978), MacCall et al. (1985), and Prager and MacCall (1987).

## Methods

### Cohort Analysis

To compute population estimates, we used a variation of MacCall's (1986) approximation, which can be stated in the form

(1) $N_t \approx N_{t+1} \exp(M) + kC_t$

where $N$ is the population estimate in numbers, $M$ is natural mortality, $C$ is the catch, the subscripts $t$ and $t + 1$ represent relative time in years, and $k$ is a constant whose value depends on $M$ and on the temporal distribution of the catch throughout the year. MacCall suggested using

(1a) $k = M/\{1 - \exp(-M)\}$

as an improvement over Pope's (1965) approximation. However, to maximize the accuracy of our VPA, we calculated a value of $k$ based on an iterative monthly VPA on the unaged monthly catch data so that equation 1 holds on an annual basis. These year-specific values of $k$ were then used in the analysis of the aged annual catches. Since the Pacific mackerel fishery is highly seasonal, MacCall's approximation (MacCall 1986) is more accurate than the iterative solution by annual VPA, even when equation 1a is used without fitting $k$ to the seasonal distribution of catch.

Equation 1 estimates age-specific abundances in numbers; biomass estimates are obtained by multiplying equation 1 by the respective weights at age. The stock's total biomass $B$ in year $i$ is estimated as the sum of the age-specific biomasses (for ages $1 \ldots m$) at the beginning of the fishing season, i.e.

$$B_i = \sum_{j=1}^{m} w_{ij} \cdot N_{ij}.$$

### Murphy's Linkage Algorithm

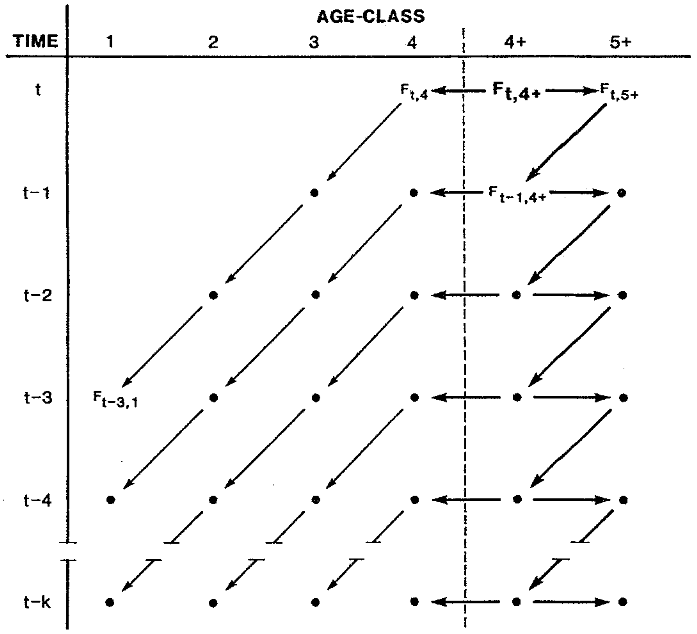

FIG. 1. Murphy's $F$-linkage algorithm and our modification for a series of $k + 1$ yr is illustrated by a matrix of $F$ values. (For clarity, most are represented by dots.) The algorithm starts at the top with $F_{t,4}^{+}$, and first follows all bold arrows diagonally and horizontally. Then, the algorithm completes all paths shown by the lighter arrows. Diagonal arrows represent calculation of $F$ values by backwards VPA. Horizontal arrows to the right represent transfer of $F$ values by the $F$ transfer coefficient (FTC, equation 3) or by equality in Murphy's original algorithm. Horizontal arrows to the left represent transfer of $F$ values by equation 5 (or by equality in the original algorithm). The bold paths yield an estimate of $F_4$ for each year. The light paths complete the biomass estimates.

Estimation of terminal fishing mortality rates is one of the largest difficulties in performing VPA. Murphy (1965) described an objective method of estimating terminal $F$ by linking cohorts. This method has been used in the previous VPA's of the Pacific mackerel stock (Parrish and MacCall 1978; MacCall et al. 1985). Murphy's linkage method requires only one estimate of terminal $F$ to start the analysis of any number of contiguous cohorts. The method as used in this fishery comprises the following steps (Fig. 1): (1) Estimate the terminal $F$ for the oldest age-class ($5^+$) in the final year ($t$), where the notation $5^+$ indicates all fish aged 5 and older. (2) Calculate $N_{t,5}^{+}$ in the final year from $C_{t,5}^{+}$ and $F_{t,5}^{+}$. (3) Use VPA to estimate $N_{t-1,4}^{+}$ and $F_{t-1,4}^{+}$. (4) Since the mackerel is considered fully recruited at age 4, assume that in each year $F_5^{+}$ is equal to $F_4^{+}$. (5) This value of $F_{t-1,5}^{+}$ supplies a new terminal fishing mortality rate for VPA estimation of $F_{t-2,4}^{+}$. By repeating steps 3, 4 and 5 for each earlier year, an estimate of $F_4^{+}$ for each year is obtained. (6) Assume that $F_4 = F_4^{+}$ in each year. From this value of $F_4$, calculate $N_4$ and perform a standard backwards VPA for the cohort from age 4 back to age 1. A key assumption of Murphy's linkage algorithm is that fishing mortality is the same among fully recruited age-classes. Thus, in each year, it is presumed that

(2) $F_4 = F_4^{+} = F_5^{+}.$

### $F$ Transfer Coefficient

We relaxed the assumption of equation 2 by postulating that $F_5^{+}$ is proportional, but not necessarily equal, to $F_4^{+}$ because of fluctuations in availability among ages and years. Varying differential growth, for example, could cause such fluctuations. The values of $F_4^{+}$ are assumed known from VPA. Then for each year, the estimate of $F_5^{+}$ is

(3) $F_5^{+} = F_4^{+} \cdot \text{FTC}$

where FTC is a transfer (proportionality) coefficient. In practice we consider FTC to be stochastic, with mean 1.0. In other words, Murphy's linkage (equation 2, which implies FTC of unity) is assumed to be valid only on average.

By parameterizing the FTC in this way, we can assume a nonzero variance $\sigma^2_{\text{FTC}}$ for the relationship between $F_4^{+}$ and $F_5^{+}$ (which leads to a similar relationship between $F_4^{+}$ and $F_4$) and examine the effect that this variability has on the VPA population estimates. The approach also allows exploration of mean

FTC values other than unity; in this study, we examined values of 1.0, 0.9, and 1.25.

For a given $F_4{}^+$ and FTC, $F_4$ can be calculated by noting the relationship

$$(4)\quad a_4 = \frac{C_{4^+}\cdot a_{4^+} - C_{5^+}\cdot a_{5^+}}{C_4}$$

where $a_i$ is defined as the reciprocal of the exploitation rate:

$$(4a)\quad a_i = Z_i/\{F_i\cdot[1-\exp(-Z_i)]\}.$$

Substituting equation 4a into 4 and solving for $F_4$, we obtain

$$(5)\quad F_4 = M/\{a_4\,[1 - \exp(-F_4 - M)] - 1\}.$$

Although equation 5 cannot be solved analytically, it can be solved easily by iteration. We used $F_4{}^+$ as a starting estimate of $F_4$ and obtained convergence in a few iterations.

Murphy's linkage algorithm resembles schemes of "separable VPA" (Doubleday 1976; Deriso et al. 1985). Separable VPA estimates relative availabilities at age, which are used to predict age-specific estimates of fishing mortality as a combination of independent age- and year-specific factors. Terminal $F$ values for adjacent cohorts are linked by optimizing the fit to this model. Murphy's algorithm does not attempt to extend this rationale to partially available ages, but focuses only on fully available ages. Murphy's use of combined cohorts increases sample sizes, and thereby reduces this source of random variability. However (as in Gulland's VPA), no degrees of freedom remain, so his algorithm leaves no direct means of accounting for stochastic variability in the model, unlike separable VPA. Our treatment attempts to reintroduce considerations of stochasticity into the linkage. At the same time, it provides a method of extending Murphy's linkage algorithm to the case of unequal availability.

### Variance Estimation

To estimate the variances of the VPA biomass estimates, we used the delta method (Seber 1973). The delta method numerically estimates the variance of a function $B(x_1, x_2,\ldots,x_r)$ by means of a Taylor expansion. The expression for the estimated variance of $B$ is

$$(6)\quad \mathrm{Var}[B(x)] \approx \sum_{k=1}^{r} \mathrm{Var}[x_k]\cdot(\partial B/\partial x_k)^2 + 2\sum_{k<l}\sum \mathrm{cov}[x_k,x_l]\cdot(\partial B/\partial x_k)\cdot(\partial B/\partial x_l).$$

In this case, the function $B$ is the population biomass estimate and the $x_k$ (and $x_l$) are the usual data needed for a VPA estimate of stock biomass: the catches and weights at age in each year, the natural mortality coefficient $M$, an estimate of the fishing mortality coefficient in the final year (terminal $F$), and, in our application, the FTC.

To solve equation 6 for Var[$B$], one must estimate the variances of each $x_k$, the covariances of each pair of $x_k$, $x_l$, $k \neq l$, and the partial derivatives of $B$ with respect to each $x_k$. Where possible, we made these estimates based on data or theoretical grounds, but were forced to make ad hoc assumptions in several cases because of lack of data.

### Partial Derivatives

We used numerical derivatives for the $\partial B/\partial x_k$. To calculate them, we made a cohort analysis estimate of stock biomass $B_1$ for a given year. Then the $x_k$ being considered was temporarily increased by 1% and a new stock biomass estimate $B_2$ was made. By definition:

$$(7)\quad \partial B/\partial x_k \approx (B_2 - B_1)/0.01x_k.$$

### Variances of Catches at Age

For most of the period under consideration, samples of mackerel from the catch were aged by examining the otoliths, and the age composition of the total catch was computed using a stratified estimate with monthly strata (Fitch 1951). As an example, let us assume that of the $n_i$ fish aged in year $i$, the proportion falling into age-class $j$ was $p_{ij}$. As the proportions in the different age-classes are distributed multinomially, the sampling variance of each $p_{ij}$ is given by

$$(8)\quad \mathrm{Var}[p_{ij}] = p_{ij}\cdot q_{ij}/n_i$$

where $q = (1 - p)$. Since the *monthly* age composition data had been lost, we assumed that the age composition of the total catch was the same as that of the sample, i.e. that

$$(9)\quad C_{ij} = C_i\cdot p_{ij}$$

where $C_i$ is the total catch in year $i$. Then the variance of $C_{ij}$ is given by

$$(10)\quad \begin{aligned}\mathrm{Var}[C_{ij}] &= C^2{}_i\cdot \mathrm{Var}[p_{ij}]\\ &= C^2{}_i\cdot p_{ij}\cdot q_{ij}/n_i.\end{aligned}$$

In the years before 1933 and between 1935 and 1938, age composition of the catch was determined by modal analysis of the length–frequency histogram, rather than by otolith ageing (Parrish and MacCall 1978). This introduced two problems. First, the numbers of fish measured in these years are not known. And second, because the method is not particularly accurate, the ageing variance Var[$C_{ij}$] for these years is higher than the nominal estimated value (equation 10) by an unknown amount. We incorporated both of these phenomena in an admittedly ad hoc way. First, we found the smallest proportion ($n_i/C_i$) of fish sampled for otolith analysis in any year. Then, for the years of modal analysis, we used for each $n_i$ (the number of fish sampled) in equation 10 this proportion times the year's catch divided by 2. For given proportions at age, this resulted in twice the estimated variances for $C_{ij}$ in those years.

### Variance of Total Catch

Equation 10 estimates the variance in each $C_{ij}$ due to sampling error of the proportions. However, the variance of $C_{ij}$ is further increased by any uncertainty in the estimate of total catch for the year, $C_i$. We treated this source of variability separately by assuming for each $C_i$ a coefficient of variation of 5%, a value believed to approximate the uncertainty in the official catch records (R. Klingbeil, California Department of Fish and Game, 245 W. Broadway, Long Beach, CA 90802, USA, pers. comm.).

### Variances and Covariances of Weights at Age

As the weight at age data are derived by two different methodologies, we used different approximations to estimate the variances and covariances of the estimated mean weights at age.

#### *1940 to 1968*

Mean weights at age for the fishing season were known (Fitch 1951), but no information on variance was available. Estimated

May 1 weights were estimated by Prager and MacCall (1987) from the published mean weights by linear interpolation based on the mean dates of each year's catch.

Estimation of the variances associated with those weight estimates posed several problems. First, variance arises from sampling errors and changes in timing of the fishery, as well as from estimation errors in the interpolations. Also, one expects covariance among weights of different ages within a season, as all are likely to be high or low together. These covariances must be known to compute the variance of the biomass estimate (equation 6).

To estimate the variance for each weight at age, a 3-yr centered moving average $W^*_{ij}$ of the estimated May 1 weights was used as a plausible estimate of the true weight at age for that year. The deviation of the year's estimate from the 3-yr mean is called $d_{ij}$. Then, $d_{ij}$ divided by the 3-yr mean provides a standardized deviate (STD) for that year:

$$\mathrm{STD}_{ij} = W^*_{ij}/d_{ij}.$$

The variance–covariance (STDCOV) matrix of these standardized deviates was used to estimate the variances and covariances of the estimated weights at age. The estimated covariance for the two ages $j$ and $k$ in year $i$ is

$$(11)\quad \mathrm{Cov}[W_{ij},W_{ik}] = \mathrm{STDCOV}_{jk}\cdot W_{ij}\cdot W_{ik}$$

where variance is the special case of $j = k$. In effect, the values in the STDCOV matrix are squared coefficients of variation.

### *Pre-1940*

Because no weight data exist, the weights at age before 1940 were estimated by regression analyses on population size. In some cases the weights of adjacent ages were included in the regression models (Prager and MacCall 1987). The STDCOV matrix developed for the 1940–68 weights at age was assumed to apply to these estimates as well. The regression models used to reconstruct the weights introduced additional error, and by the sequential nature of the reconstruction, these errors propagated from younger to older ages. On the other hand, weight estimates for the older ages used information from adjacent ages or cohorts, thus averaging out some of the error. The error of the regression model for weight of age 1 fish suggests a coefficient of variation (CV) of about 20%. The mean-squared error is 1218 $g^2$, and the mean weight is 177 g; thus the coefficient of variation is

$$(12)\quad \mathrm{CV} = (1218)^{1/2}/177 = 0.197$$

or about 20%. Although the models of older ages appear to have higher $R^2$, this is misleading, since those models rely upon the estimated weights at previous ages.

To accommodate these facts, we assumed that each reconstructed weight should have its variance inflated by 20% to reflecting a constant and independent 20% reconstruction error. This was accomplished by increasing all values in the STDCOV matrix by 0.04 before applying equation 11 to compute variance for the pre-1940 data. Because each age's variance is assumed to be independent, and the biomass estimates were composed of five ages, some of this additional variance canceled out in the computations.

## Variance of the Natural Mortality Coefficient, $M$

We know of no publication in which an estimate of the variance of $M$ has been derived for any fish stock. We regrettably cannot provide the exception and have estimated this quantity in an ad hoc manner, although the estimate is strengthened by theoretical considerations. The estimate of $M$ (0.50/yr) for *S. japonicus* is taken from Parrish and MacCall (1978).

Before estimating the variance of $M$, we considered that it might arise from two sources. The first source is *measurement error*, our inability to measure the mean of $M$ precisely. As our VPA uses a single estimate of $M$ for all years (as do all VPA methods we are aware of), any measurement error affects all years of the analysis simultaneously. This means that, to model measurement error, only one partial derivative with respect to $M$ must be computed for each year's biomass estimate ($B_i$).

The second source of variance in $M$ is *process error*, the year-to-year variability in a quantity ($M$) not estimated every year. Although process error can be autocorrelated, we assumed the simplest case, i.e. that process error would affect the value of $M$ in each year of the analysis independently. To model such process error, one must compute for each year's biomass estimate a separate partial derivative with respect to $M$ for each year whose data contribute to the biomass estimate.

We assumed for simplicity that the contributions of the two sources to the variance of $M$ were equal. For computational purposes, then, we split $M$ into two equal components: a "low-frequency" component (subject to measurement error) and a "high-frequency" component (subject to process error). The variance of $M$ was divided between the two components, also. Thus, we write

$$(13)\quad M = M_0 + M_1.$$

Here, $M_0$ is the low-frequency component of natural mortality, i.e. the component whose variance, arising from measurement error, affects all years at once. In contrast, $M_1$ is the high-frequency component, i.e. the component whose variance, arising from process error, affects each year separately.

We estimated $M$ to have a coefficient of variation of about 25%, i.e. we believed a 95% confidence interval from $M = 0.25$ to $M = 0.75$ to be approximately correct. This gives an estimated variance of 0.0156, which we rounded to 0.016. To explore the implications of the two sources of variance in $M$, we repeated our analysis for three scenarios: the first had all the variance (0.16) in $M_0$, the second had all the variance in $M_1$, and the third had the variance divided equally between them.

## Variance of Terminal $F$

The required estimate of the terminal $F$ ($F_4^+$ in 1969) was set at 1.0 (Parrish and MacCall 1978), with estimated variance of 0.5. The estimated variance, while somewhat arbitrary, does reflect the variance found in the distribution of fishing mortality values observed over the years (Parrish and MacCall 1978; MacCall et al. 1985); it also reflects our opinion of the uncertainty in this estimate of $F$. Because we used Murphy's linkage method, the value of terminal $F$ was much less important than in nonlinked cohort analyses.

## Covariances

Except as explained above for the weights at age, covariances among the VPA input quantities were not calculated directly. We assumed that the covariances between different quantities (e.g. $\mathrm{Cov}[F,M]$ or $\mathrm{Cov}[w_i, C_i]$) were zero. Since the catches at age were treated as proportions, their covariances within a given year would be negative. This fact was treated numerically, while

TABLE 1. Index to VPA biomass and variance estimates, identified by computer run number. $M_0$ is the component of natural mortality whose variance acts on all years together; $M_1$ is the component of natural mortality whose variance acts on each year independently (equation 13). Run 22 is used in the text as the base for comparisons.

| Run number | $M_0$ mean | $M_0$ variance | $M_1$ mean | $M_1$ variance | FTC mean |
|---|---|---|---|---|---|
| 20 | 0.5 | 0.016 | 0.0 | 0.0 | 1.0 |
| 21 | 0.0 | 0.0 | 0.5 | 0.016 | 1.0 |
| 22 | 0.25 | 0.008 | 0.25 | 0.008 | 1.0 |
| 23 | 0.25 | 0.008 | 0.25 | 0.008 | 1.25 |
| 24 | 0.25 | 0.008 | 0.25 | 0.008 | 0.9 |

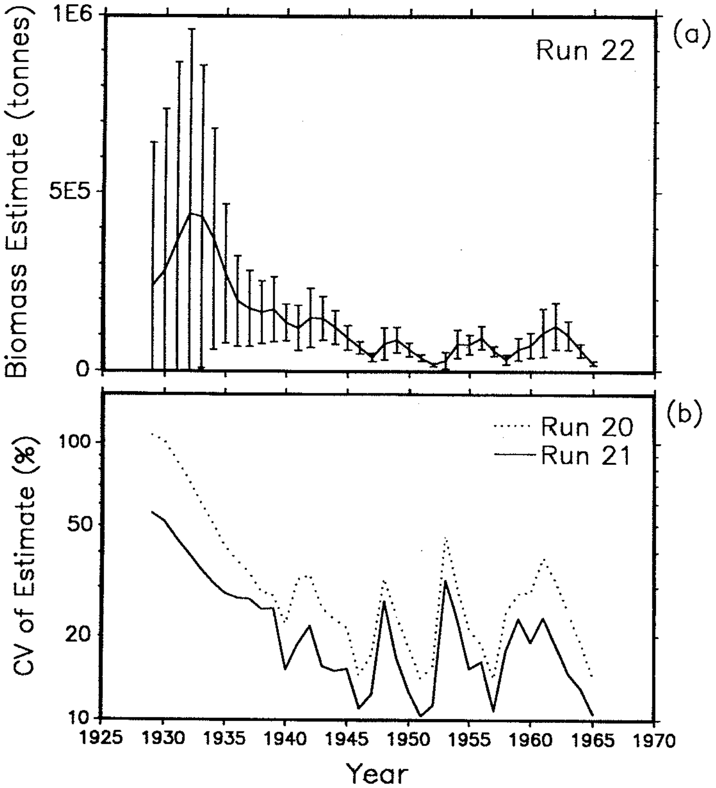

FIG. 2. Biomass and variance estimates for the mackerel, *Scomber japonicus*. (a) Biomass estimates with approximate 95% confidence intervals from run 22; (b) coefficients of variation (CV's) of biomass estimates from runs 20 and 21. CV's from run 22 were in between, somewhat closer to run 20. Runs differ in their assumptions about the variance of natural mortality (Table 1).

estimating the partial derivatives (equation 7) of catch at age, by taking the temporary 1% increase in each $C_{ij}$ from the other catches in the same year. When $C_{ij}$ was temporarily increased by 1%, the other catches were temporarily reduced (in proportion to their contribution to the total catch for the year) so that the total catch remained the same. Thus, although few of the covariance terms in equation 1 were computed explicitly, most of the covariances were taken into account.

### Sensitivity Factors

During the analyses, it became apparent that the partial derivatives of the biomass estimate with respect to the input quantities ($\partial B_i/\partial x_k$), which were computed to accomplish the delta method, were of interest as indicators of the sensitivity of the VPA biomass estimates to the values of the input quantities. Furthermore, the derivatives can be transformed into dimensionless sensitivity factors by scaling them:

$$(14) \quad S_{ik} = \frac{x_k}{B_i} \cdot \frac{\partial B_i}{\partial x_k}$$

where $S_{ik}$ is our notation for the sensitivity in year $i$ of the biomass estimate $B_i$ to the input factor $x_k$. The interpretation of the $S_{ik}$ is that for a given percentage change in the corresponding $x_k$, the biomass estimate will change by that percentage multiplied by $S_{ik}$. Such sensitivities are similar to the quantities defined by economists as *elasticities*. (Clark 1976, p. 128). Sensitivities were computed for all input factors except the weights at age.

### Multiple Biomass Estimates for Each Year

We made several runs of our VPA algorithm with different assumptions about the variance in $M$. Also, runs were made to examine the effects of values of the FTC other than 1.0. Table 1 lists the various runs and their parameters.

## Results

### VPA and Variance Estimates

Figure 2a is a time-series plot of the VPA biomass estimates from run 22, with approximate 95% confidence intervals. The confidence bounds were calculated as the population estimate ±2 standard errors; the standard error is the square root of equation 6. The very large confidence intervals from 1929 to 1939 largely result from the poor precision of the estimated weights at age during these years. In later years, the CV of the estimates is much lower.

Figure 2b illustrates the CV's of the VPA biomass estimates from runs 20 and 21. (Since runs 20–22 vary only in their treatment of the variance of natural mortality, the biomass estimates are identical.) The CV's of run 20 are largest and those from run 21 smallest. Run 22 falls in between, but somewhat closer to run 20.

### Composition of Variance

Figures 3a, 3b, and 3c show the time-averaged compositions of the variance estimates from runs 20, 21, and 22. In preparing this figure, we have omitted the highest variance years, 1929–39. In each case, the largest fraction of the mean variance comes from the estimates of $M$, for which we assumed a CV of approximately 25%. The other large fraction comes from the estimates of weights at age. Smaller fractions come from the catch at age, total catch, and the FTC. All other fractions of the variance were less than 0.05%. The results for runs 23 and 24 were essentially identical to those of run 22.

### Effects of $F$ Transfer Coefficient

We found that the variance of the FTC was relatively unimportant in establishing the total variance (Fig. 3), contributing at most 2.3% of the total in years after 1935. The mean value of $F_4$ for these years was 1.01. In earlier years, however, when the mean value of $F_4$ was only 0.16, the FTC contributed up to 12% of the total variance. Changing the value of the FTC had little effect on the biomass estimates except in the initial years with low $F$ (Fig. 4). This small effect is in agreement with the low sensitivities to FTC (<1%) that we found when $F$ was moderate or high.

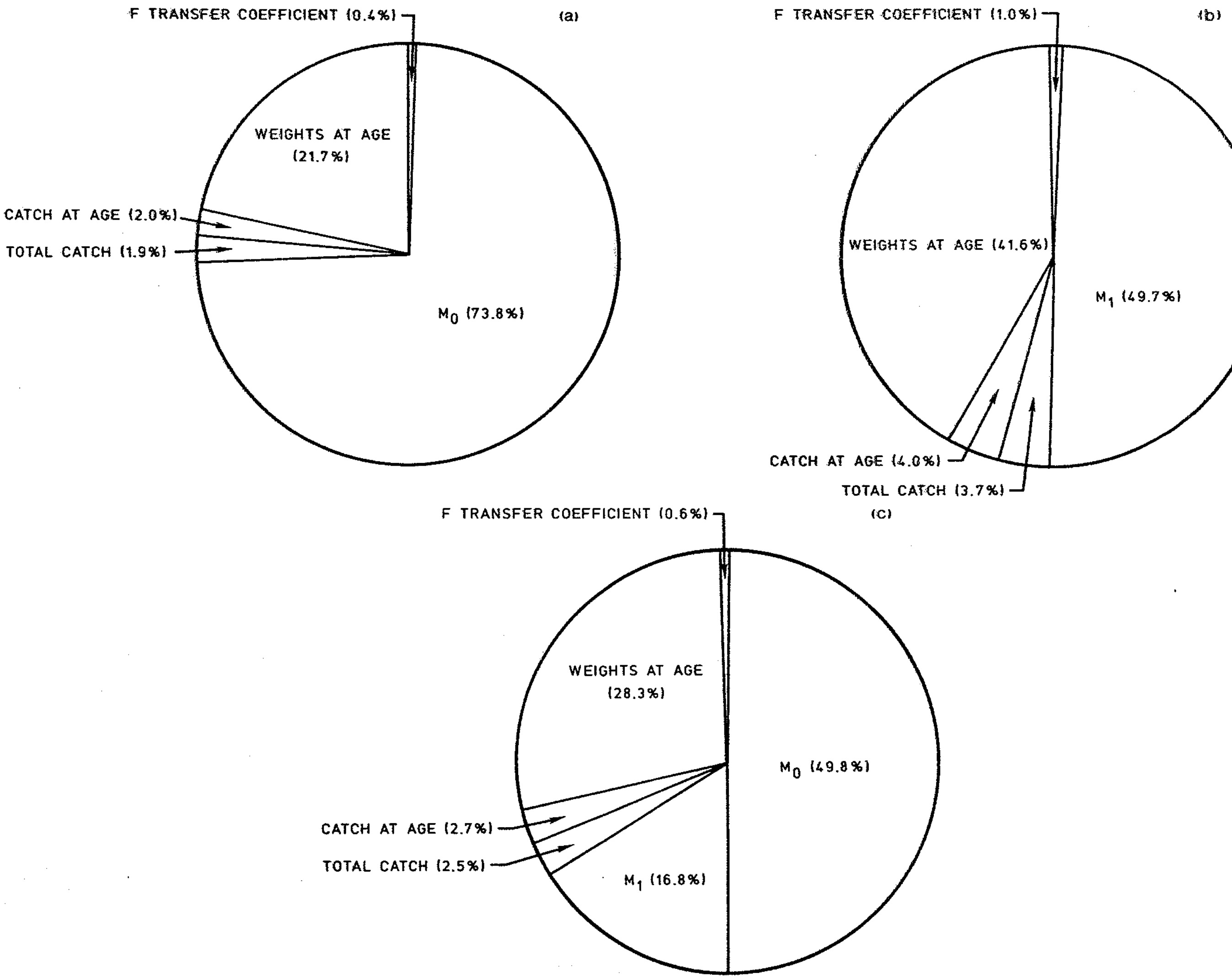

FIG. 3. Variance composition for runs 20, 21, and 22. Results of runs 23 and 24 are extremely similar to those of run 22.

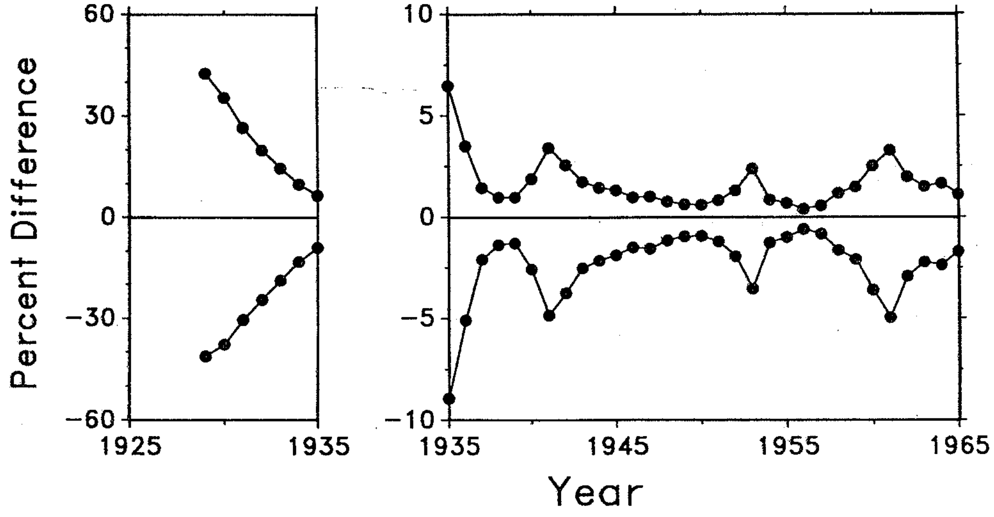

FIG. 4. Biomass estimates from runs 23 (lower curves) and 24 (upper curves), which have FTC values of 1.25 and 0.90, respectively. Plots are of percentage differences from run 22. To illustrate the different scales, the 1935 values are shown in both panels.

TABLE 2. Dimensionless sensitivities ($S_{ik}$, equation 14) of VPA biomass estimate in 1955 to various input quantities. Sensitivities are for run 22 and vary in other years and runs. Factors less than 0.01 are not shown.

| Input quantity, $x_k$ | Sensitivity of 1955 biomass estimate to this quantity |
|---|---|
| Low-frequency $M$ ($M_0$) | 0.359 |
| High-frequency $M$ ($M_1$) in 1955 | 0.178 |
| $M_1$, 1956 | 0.109 |
| $M_1$, 1957 | 0.045 |
| $M_1$, 1958 | 0.016 |
| Total catch, 1955 | 0.291 |
| Total catch, 1956 | 0.332 |
| Total catch, 1957 | 0.269 |
| Total catch, 1958 | 0.079 |
| Total catch, 1959 | 0.015 |
| Total catch, 1960 | 0.010 |
| Catch (1955, age 3) | 0.033 |
| Catch (1956, age 2) | 0.056 |
| Catch (1956, age 3) | 0.128 |
| Catch (1956, age 4) | 0.047 |
| Catch (1957, age 3) | 0.060 |
| Catch (1957, age 4) | 0.070 |
| Catch (1957, age 5) | 0.060 |
| Catch (1958, age 4) | 0.032 |
| Catch (1958, age 5) | 0.037 |
| Catch (1959, age 5) | 0.024 |
| Catch (1960, age 5) | 0.012 |

TABLE 3. Dimensionless sensitivities ($S_{ik}$, equation 14) of VPA biomass estimates to various input quantities. Sensitivities are means for the years 1940–65; those less than 0.01 or related to catch at age are not shown.

| Input quantity, $x_k$ | Mean sensitivity of biomass estimates to this quantity |
|---|---|
| Low-frequency $M$ ($M_0$) | 0.360[a] |
| Low-frequency $M$ ($M_0$) | 0.821[b] |
| High-frequency $M$ ($M_1$), year $i$[c] | 0.176[a,d] |
| High-frequency $M$, year $i + 1$ | 0.108[a,d] |
| High-frequency $M$, year $i + 2$ | 0.062[a,d] |
| Total catch, year $i$ | 0.360 |
| Total catch, year $i + 1$ | 0.272 |
| Total catch, year $i + 2$ | 0.185 |

[a] Assumes that natural mortality and its variance are divided evenly between low-frequency ($M_0$) and high-frequency ($M_1$) components.

[b] Assumes that natural mortality and its variance are entirely in the low-frequency component ($M_0$).

[c] Year $i$ is the year of the estimate.

[d] If natural mortality were entirely in the high-frequency component, this sensitivity would be approximately doubled.

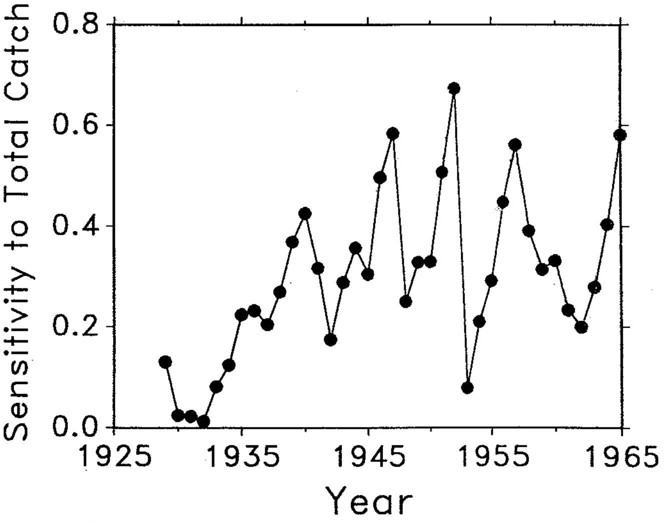

FIG. 5. Sensitivity of biomass estimates to total catch in the year of the estimate.

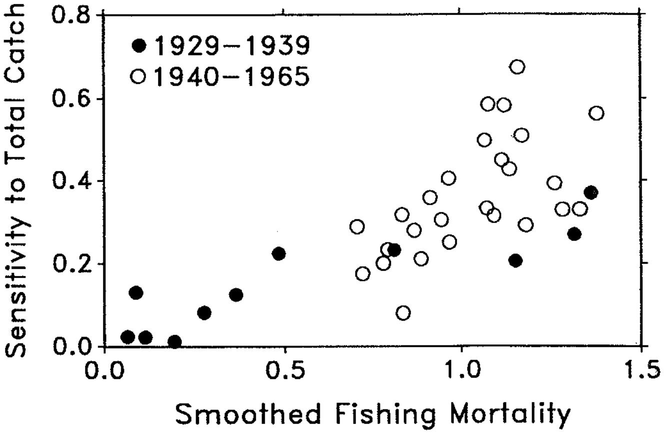

FIG. 6. Effect of fishing mortality ($F$) on the sensitivity of biomass estimates to total catch in the year of the estimate. The smoothed $F$ values on the $x$-axis are arithmetic means of $F$ in the year of the estimate and in the next three years.

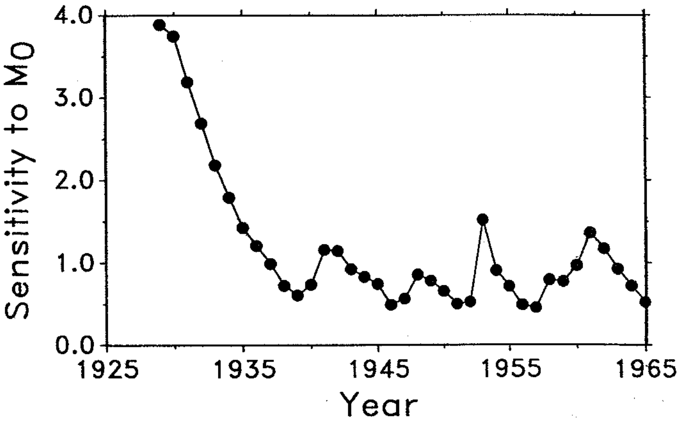

FIG. 7. Sensitivity of biomass estimates to low-frequency component of natural mortality ($M_0$). All natural mortality is assumed to be in $M_0$; if half were in the high-frequency component, sensitivities would be approximately halved.

## Sensitivity Factors

Using equation 14, we computed sensitivities to each input quantity for each year's biomass estimate. Table 2 shows the sensitivities for run 22 in 1955, which is considered typical. Factors less than 0.01 are not shown.

For run 20, the sensitivity to $M_0$ is approximately doubled from Table 2. For run 21, the sensitivity to $M_1$ is approximately doubled from Table 2.

Table 3 shows mean sensitivities for the years 1940–65. Sensitivities to specific catches at age are omitted from the table; such sensitivities are in general small (<0.05) and vary from year to year. Of them, the greatest sensitivities are to catches at age 5, probably because these are the heaviest fish and because numbers at age 5 are magnified by the backwards VPA algorithm. Table 3 reveals that, as expected, VPA biomass esti-

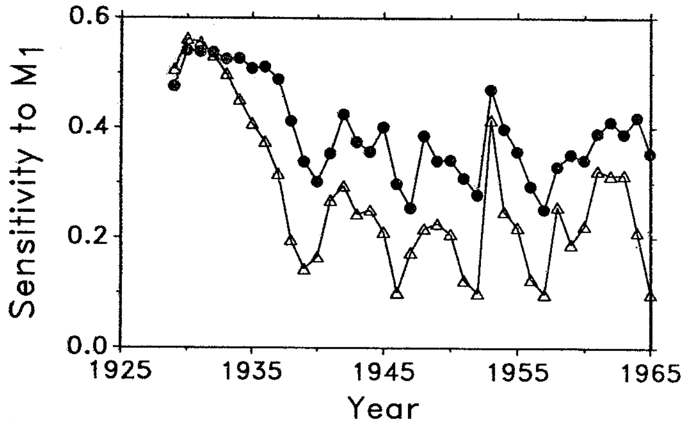

FIG. 8. Sensitivity of biomass estimates to high-frequency component of natural mortality ($M_1$). Upper curve: $M_1$ in the year of estimate; lower curve: $M_1$ in the following year. All natural mortality is assumed to be in $M_1$; if half were in the low-frequency component, sensitivities would be approximately halved.

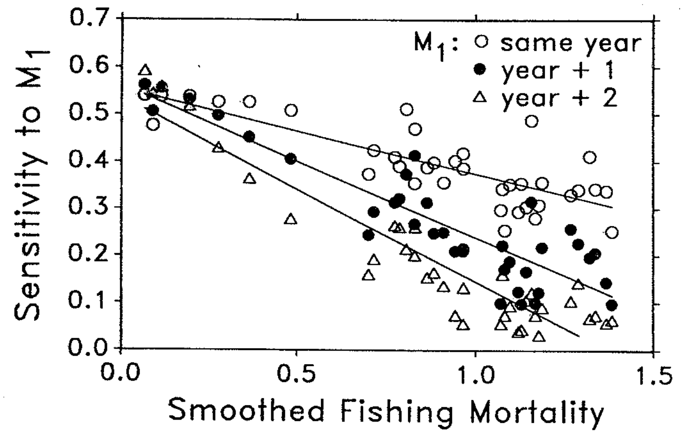

FIG. 9. Effect of fishing mortality ($F$) on the sensitivity of biomass estimates to the high-frequency component of natural mortality ($M_1$) in the year of estimate or in the following 2 yr. The smoothed $F$ values on the $x$-axis are means of $F$ in the year of the estimate and in the next 3 yr. Regression lines for each year are shown. As in Fig. 8, all natural mortality is assumed to be in $M_1$.

mates are highly sensitive to $M$; however, the highest sensitivity under the present assumptions was less than unity.

Figure 5 is a time plot of sensitivity to total catch in the year of the estimate. The sensitivity is less than 0.05 in 1930–32, when exploitation was very light, but rises to around 0.4 in later years. Figure 6 shows the same sensitivities plotted against smoothed estimated fishing mortality ($F$). The two tend to covary, even when the data are broken into pre-1940 and post-1939 groups. The years with highest estimated $F$ were most sensitive to the value of catch.

Figures 7–9 describe the sensitivity of the biomass estimates to the estimate of $M$. Figure 7, a time plot of sensitivity to the low-frequency component ($M_0$) of natural mortality, demonstrates the very high sensitivities (nearly 4.0) in the years when the stock was lightly exploited. The sensitivities later fall to around 0.75. Figure 8 is a similar time plot of sensitivity to the high-frequency component ($M_1$) of natural mortality. The sensitivity is higher to $M_1$ in the same year as the estimate, but also significant to $M_1$ one year later. The shape of the two curves is nearly the same. Figure 9 demonstrates that, as expected, the sensitivity to $M_1$ depends on the intensity of exploitation. With higher values of $F$, the sensitivity to the value of $M_1$ decreases.

## Discussion

We wished to estimate the variance of VPA biomass estimates so that the variances could be used in correcting a stock–recruitment model for the so-called "errors in variables" problem (Draper and Smith 1981). This problem, which arises when error is present in the predictor variable of a regression model, results in biased parameter estimates and can obscure stock–recruitment relationships (Walters and Ludwig 1981). However, knowing the variance of the biomass estimates is also valuable because it allows confidence bounds to be drawn about the population estimates (Fig. 2) and hypotheses about population sizes to be tested statistically. Since the composition of each variance estimate is also known (Fig. 3), the relative importance of better knowledge of each input quantity can be judged. This could prove useful in planning an efficient monitoring program for a stock.

Our variance estimates were slightly biased downward by the cohort analysis algorithm we used, which was selected to reflect the seasonal nature of the fishery (see Cohort Analysis above). In this algorithm, values of $k$ (which depends upon $M$) were estimated separately and not changed when computing derivatives with respect to $M$. To have done otherwise would have been computationally impractical. However, a numerical investigation of the bias revealed it to be generally quite small. In the years after 1945, the bias was approximately 1 or 2% in the CV of biomass, i.e. a CV reported as 20% was probably about 21%.

The sensitivities we computed reveal the importance of errors in the inputs in an easily interpreted way. By using the sensitivities, one could estimate the probable error in the biomass estimate for a past year when $M$ (for example) was thought to be unusually high or low. Also, such sensitivities can be used for an objective and uniform analysis of any model. Being dimensionless, they provide a scale-free expression of the relationship of a model to its input factors.

### Omission of Recent Years from Calculations

In general, the variance in VPA estimates is highest for the most recent years, when cohorts have been fished for few years and the effect of error in the estimate of terminal $F$ is not reduced by convergence. However, Fig. 2 shows no such effect. This is because we restricted our analyses to those years which permitted running the analysis with only one estimate of terminal $F$: the $F_5{}^+$ in 1969. Thus, although our data ran through 1969 (the final year of data available before the collapse of the stock), our "historical" estimates of population biomass and its variance run only through 1965. The variances of "current" VPA estimates (those nearer the final year of data) would of course be higher. Also, estimates of such variances would be much less precise; the variance of a "current" VPA estimate is almost entirely a function of the variances of $M$ and the terminal $F$ for the cohort, both typically quite high.

### Further Applications of the Method

The delta method is a valuable tool for estimating the variances of complex models. We have applied it to VPA biomass estimates, but it could be used equally well on other quantities estimated by VPA. For example, the variances of full iterative VPA could be estimated, although the computations would be more lengthy. Also, the same methodology could be used to estimate the variances of fishing mortalities or population sizes in numbers arising from VPA. A further potential application wuld be the estimation of the variances of the parameters of "separable" VPA (Doubleday 1976; Deriso et al. 1985), in which the matrix of $F$'s is estimated as a series of additive year and age effects. Finally, we expect that as fishery scientists become increasingly concerned with correcting for the effects of errors in variables, the delta method will prove an especially useful addition to the analytical toolbox.

## Acknowledgements

Financial support for this work was provided by the Ocean Assessments Division, National Ocean Service, NOAA, U.S. Department of Commerce. The manuscript was much improved by the comments of Dr. Richard Methot and of two anonymous reviewers.

## References

Clark, C. W. 1976. Mathematical bioeconomics. John Wiley & Sons, New York, NY. 362 p.

Deriso, R. B., T. J. Quinn II, and P. R. Neal. 1985. Catch-age analysis with auxiliary information. Can. J. Fish. Aquat. Sci. 42: 815–824.

Doubleday, W. G. 1976. A least-squares approach to analyzing catch at age data. Int. Comm. Northwest Atl. Fish. Res. Bull. 12: 69–81.

Draper, N. R., and H. Smith. 1981. Applied regression analysis. 2nd ed. John Wiley & Sons, New York, NY. 709 p.

Fitch, J. E. 1951. Age composition of the southern California catch of Pacific mackerel 1939–40 through 1950–51. Calif. Dep. Fish Game Fish Bull. 83: 73 p.

Gulland, J. A. 1965. Estimation of mortality rates. Annex to Arctic Fisheries Working Group Rep., ICES C.M. 1965, Doc. No. 3: 9 p. (mimeo).

MacCall, A. D. 1986. Virtual population analysis (VPA) equations for nonhomogeneous populations, and a family of approximations including improvements on Pope's cohort analysis. Can. J. Fish. Aquat. Sci. 43: 2406–2409.

MacCall, A. D., R. A. Klingbeil, and R. D. Methot. 1985. Recent increased abundance and potential productivity of Pacific mackerel (*Scomber japonicus*). Calif Coop. Ocean. Fish. Invest. Rep. 26: 119–129.

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 accuracy of virtual population analysis using cohort analysis. Int. Comm. Northwest Alt. Fish. Res. Bull. 9: 65–74.

Prager, M. H., and A. D. MacCall. 1987. Revised estimates of the historical spawning stock biomasses of the Pacific mackerel, *Scomber japonicus*. Old Dominion University Department of Oceanography Technical Report 87-7. (Available from M.H.P.)

Saila, S. B., E. Lorda, and H. A. Walker. 1985. The analysis of parameter error propagation in simple fishery models. Mar. Resour. Econ. 1: 235–246.

Sampson, D. B. 1984. Variance estimators for virtual population analysis. M.S. thesis, University of Washington, Seattle, WA. 81 p.

Seber, G. A. F. 1973. The estimation of animal abundance and related parameters. Hafner Press, New York, NY. 506 p.

Sims, S. E. 1982. The effect of unevenly distributed catches on stock-size estimates using virtual population analysis (cohort analysis). J. Cons. Int. Explor. Mer 40: 47–52.

Walters, C. J., and D. Ludwig. 1981. Effects of measurement errors on the assessment of stock–recruitment relationships. Can. J. Fish. Aquat. Sci. 38: 704–710.