# Presence-Absence Sampling for Fisheries Management[1]

Marc Mangel[2]

*Center for Population Biology and Zoology Department, University of California, Davis, CA 95616 USA*

and Paul E. Smith

*Southwest Fisheries Center, P.O. Box 271, La Jolla, CA 92037 USA*

It is often important to estimate the abundance of a fish stock when the stock is somewhat depleted. For pelagic species this presents great operational difficulties, because adult surveys may be prohibitively expensive and time consuming. Here we introduce a method for estimating the spawning biomass of a stock by means of egg or larval surveys. In particular, we develop a series of models for presence-absence sampling of eggs or larvae and show how presence-absence data can be used to estimate adult spawning biomass. The models are based on an underlying probabilistic description of the aggregation of eggs or larvae, a search process, and a description of habitat structure. Methodologies are given for estimating the distribution of the size of the spawning biomass from presence-absence data. A case study of sardine is used to justify a number of the assumptions. The methods are applied to a 1985 survey for sardine eggs and are compared to an alternative method based on egg production.

Dans bon nombre de cas, il est important d'estimer l'abondance d'un stock de poissons lorsque ses effectifs sont quelque peu réduits. Pour les espèces pélagiques, cela entraîne des difficultés opérationnelles majeures, car le dénombrement des adultes peut être excessivement coûteux et long. Les auteurs ont mis au point une méthode permettant d'estimer la biomasse des géniteurs d'un stock par la présence ou l'absence de leurs oeufs ou de leurs larves. À cette fin, ils ont élaboré une série de modèles et démontrent comment l'on peut se servir de données relatives à la présence ou à l'absence des oeufs ou des larves de poisson pour estimer la biomasse des géniteurs adultes. Les modèles reposent sur un calcul probabiliste du rassemblement des oeufs ou des larves, une méthode de recherche et une description de la structure de l'habitat. Les auteurs donnent des techniques pour estimer la répartition de la biomasse des géniteurs d'après les données sur la présence et l'absence des oeufs ou des larves. Ils ont recours à une étude relative aux sardines pour justifier un certain nombre d'hypothèses. Ils appliquent leurs méthodes à une étude au sujet d'oeufs de sardine datant de 1985, et les comparent à une autre méthode d'estimation reposant sur la production d'oeufs.

Very often, fisheries managers must estimate the adult biomass of pelagic stocks. When the stocks are depleted, direct assessment of adult population levels is not only difficult, but may be highly inaccurate (Hewitt et al. 1984). In such situations, an alternative is the survey of egg or larval populations as a means of estimating adult biomass (Lasker 1981). The interpretation of data from egg or larval surveys, called plankton surveys in this paper, is complicated by the patchy and highly aggregated nature of eggs and larvae (Smith 1978). A method based on egg production (Santander et al. 1982; Hewitt 1984; Hewitt et al. 1984; Lasker 1985; Wolf and Smith 1985; Wolf et al. 1987) has been proposed as a means of relating the data from egg surveys to adult biomass. The objective of this paper is to introduce another set of tools which can be used by the fisheries scientist or manager who is involved in egg or larval surveys.

The work reported here was motivated by the need to manage the recovery of the northern population of California sardine,

---

[1]Dedicated to the memory of Reuben Lasker, an exemplary scientist and individual.

[2]Author to whom correspondence should be addressed.

*Sardinops sagax caerulus* (Wolf 1986). Once a thriving population, with a peak standing biomass estimated to be $11 \times 10^6$ metric tons (t) in the California current in 1860, to about $4 \times 10^6$ t during the cannery heyday (Smith 1978), it is no longer strong. Current State law (Wolf 1986) prohibits any direct fishery on sardine (there will always be incidental catch that is essentially uncontrollable) until the spawning biomass (the fraction of standing biomass older than 2 yr) exceeds 20 000 short tons (1 short ton = 0.907 t). If the spawning biomass is determined to exceed the critical level of 20 000 short tons, then a fishery with a catch quota of about 5% of the spawning biomass is recommended.

The methodology developed here, however, should be applicable to a wide variety of other plankton sampling programs (e.g. Sette and Ahlstrom 1948; Squire 1983) and many benthic sampling programs (Downing 1979; Resh 1979). The methodology introduced here addresses the same kinds of problems as the work of Pennington (1983) but does so with very different techniques.

For the current management of sardine, the key question is whether the biomass exceeds a critical level $B_c$ (in this case 20 000 short tons). We thus do not need to estimate the mean

*Can. J. Fish. Aquat. Sci., Vol. 47, 1990*

1875

and variance of the spawning biomass directly, but instead want to assess the probability that the biomass exceeds the critical level. We do this as follows: first, we associate the critical level of biomass $B_c$ with a critical level $m_c$ of eggs per sampling station. Second, given the sampling data, we estimate the probability that the mean of the distribution of eggs per sampling station exceeds $m_c$.

In the next section, the models for egg and larval distributions are developed and discussed. In the third section, a sequence of sampling models is developed. Particular attention is paid to the justification for presence–absence sampling (PAS). In the fourth section, the methodology developed here is compared with a method based on egg production. We then develop a criterion indicating when direct counts are preferable to presence–absence sampling. A case study of sardine from 1940–65 is used to justify a number of the assumptions necessary for the analysis. We conclude with a discussion illustrating how presence–absence sampling can be placed into the overall context of management.

## Models for Egg and Larval Distribution

Eggs and larvae are commonly distributed over space in an aggregated and patchy manner. Although this aggregation is undoubtedly behavioral in nature (Vlymen 1977; Lasker 1981; Grosberg and Quinn 1986), the models introduced here are operational ones, with attempted biological justification. The highly aggregated nature of eggs and larvae means that samples of egg counts will have considerable dispersion. That is, a high proportion of the samples will have no eggs and samples that have one egg are likely to have many. A typical situation is the following data on anchovy eggs: for a survey involving 419 samples there were 208 samples with no eggs; the range for positive samples was 2–12 200 eggs per sample. The mean was 206 eggs/sample and the SD was 1022 (i.e. a coefficient of variation of about 500%). Compared to sardine, anchovy is a healthy stock, leading to many samples with eggs. For example, in the 1985 sardine cruise discussed in the fourth section, only 11 of 419 sites sampled had eggs. The actual counts of eggs per positive station were 42, 15, 10, 5, 5, 5, 3, 2, 1, 1, 1; these data have mean 7.8 and SD 11.6 for a coefficient of variation about 150%.

Some of the samples with no eggs occur because samples are taken beyond the current habitat of the population. That is, if we knew exactly the current range of the population, it would be possible to design surveys in which a much higher proportion of samples had eggs. The difficulty for pelagic species is that the spatial range of habitat can vary considerably (see, e.g. Lasker 1985, p. 18); thus the sampling must both determine the boundaries of the habitat and estimate abundance. We believe that in some cases, fully ⅔ of the samples may be used to define the boundary of the current range of the spawning population. That is, in some cases more than 50% of the samples taken may have no eggs and will be used to help delineate the boundaries of the habitat. A good, operational model should be compatible with the assumptions that (1) not all samples will be in the habitat of the spawning stock, (2) samples in the habitat may still show high levels of aggregation.

The question of spatial scale is an important one as well. For the sardines, the relevant spatial scales are:

| Unit | Spatial scale |
|---|---|
| Individual fish | ~ cm |
| School | ~ 100 m |
| Egg patches | ~ 1000 m |

| School groups | ~ 10 000 m |
| Sampling scale | ~ 10 000 m |

The actual area of the device used to sample sites in the ocean is 0.05 m². It is reasonable then to assume that samples taken at different sites (called stations) are independent random variables. Let (Table 1 summarizes symbols and their interpretations)

(1) $X_i$ = number of eggs (or larvae) taken at the $i^{th}$ station.

The properties of $X_i$ suggested by the empirical data are

(i) $X_i \geq 0$,
(ii) $Pr\{X_i = 0\}$ is considerable, and
(2) (iii) $Var\{X_i\} >> E\{X_i\}$.

In this equation, $Var\{ \}$ and $E\{ \}$ denote the variance and mean of a random variable respectively, and $Pr\{X = x\}$ is the probability that the random variable $X$ takes the value $x$. The variance greatly exceeding the mean is a commonly accepted definition of an overdispersed or aggregated population.

One model with properties (i)–(iii) is the negative binomial (NB) distribution. At the outset, we stress that using the NB is not crucial to any of the conceptual development that follows. We choose it because it is commonly used in ecological modeling. We will show below that other distributions could be used as well. For the NB distribution,

$Pr\{X_i = x \mid$ station is a habitat$\}$

(3) $= [\Gamma(k+x)/\Gamma(k)x!] \, [k/k+m]^k \, [m/k+m]^x$

In this equation, $\Gamma(x)$ is the gamma function, and $k$ and $m$ are parameters. In particular,

(4) $Pr\{X_i = 0 \mid$ station is a habitat$\} = [k/k+m]^k$

The conditional mean and variance of $X_i$ are (Johnson and Kotz 1969)

$E\{X_i \mid$ station is a habitat$\} = m$
(5) $Var\{X_i \mid$ station is a habitat$\} = m + m^2/k$

so that $m$ is the mean of the distribution and $k$ is a measure of the amount of "overdispersion" relative to the Poisson distribution.

Often $k$ is essentially independent of spawning biomass, although it depends upon sampling scale (DeBouzie and Thioulouse 1986). Smith and Richardson (1977, p. 57) found, for example, that for sardine eggs the estimate of $k$ varied between 0.08 and 0.21 as the spawning biomass varied between 0.2 and 3.9 × 10⁶ t. This is a variation of less than a factor of 3 in $k$, during variation by a factor of 19.5 in biomass; we interpret this as relative constancy of $k$.

The value of $k$ may also be tied to the life history of the species (Hewitt 1981). Table 2 shows our estimate of the life history dependence of $k$ for the early stages of anchovy (P. E. Smith, unpubl. data). Because of the depleted nature of sardine, we are unable to determine a similar life history dependence of $k$ at the present time.

The NB model has a long and distinguished history in ecology (e.g. Anscombe 1948, 1950; Taylor 1953; Sampford 1956; Bliss and Owen 1958; Bissell 1972; Pielou 1977; Taylor et al. 1979; Ripley 1981; Zweifel and Smith 1981; Diggle 1983; Dennis and Patil 1984). It also has limitations. For example, L. R. Taylor and his colleagues (Taylor et al. 1979; Perry and Taylor 1986) have sharply attacked the use of the NB distribution with

1876

*Can. J. Fish. Aquat. Sci., Vol. 47, 1990*

TABLE 1. Summary of symbols, interpretations, and values for computations.

| Symbol | Interpretation | Value used and source |
|---|---|---|
| $B, B_c$ | Spawning biomass and its critical level | 20 000 short tons (State of California Law) |
| $X_i$ | Number of eggs at the $i^{th}$ station | Random variable |
| $m$ | Mean number of eggs at a station | Estimated by the procedure |
| $k$ | Overdispersion parameter of negative binomial distribution | Ranges from 0.08 to 0.21 for sardine eggs (Smith and Richardson, 1977) |
| $p_i$ | Probability that the $i^{th}$ station is a habitat for the spawning stock | All $p_i = p = 0.05$(historical data) |
| $m_c$ | Critical value of $m$, linearly related to critical value of biomass | $m_c = 1.14$ (historical data) |
| $Y_i$ | Number of eggs detected at the $i^{th}$ station | Random variable |
| $\alpha_i$ | Maximum probability of detecting eggs at the $i^{th}$ station when eggs are overly abundant | $\alpha_i = 1$ |
| $b_i$ | Sampler efficiency | Varies between 1/6 and 6 |
| $\mathcal{P}, \mathcal{N}$ | Denote samples with positive number of eggs and no eggs | Data |
| $\mathcal{L}$ | Likelihood of an egg sample | Computed by Eq. (12) |
| $\hat{m}$ | Maximum likelihood estimator (MLE) for $m$ | Computed by methods of this paper |
| $N_n, N_p, N$ | Number of samples with no eggs or a positive number of eggs, respectively and total number of samples | Data; $N = N_n + N_p$ |
| $R(m, \hat{m})$ | Likelihood ratio of a value of $m$ relative to MLE | Computed by Eq. (16) |
| $Y_{ij}$ | Presence or absence of eggs of age group $j$ at station $i$ | Data |
| $k_j, m_j$ | Overdispersion parameter and mean number of eggs of age group $j$ | Age dependence given by Smith (1973) |
| $r_c(i)$ | Association or correlation coefficient for the presence of eggs of different age groups at station $i$ | Varies between 0 and 1. |

TABLE 2. Early life history values of the negative binomial $k$ for northern anchovy larvae (P. E. Smith, unpubl. data).

| Length (mm) | Value of $k$ | |
|---|---|---|
| | Day | Night |
| 3.25 | 0.064 | 0.073 |
| 5.25 | 0.093 | 0.104 |
| 7.25 | 0.102 | 0.121 |
| 9.25 | 0.095 | 0.126 |
| 11.25 | 0.069 | 0.120 |
| 13.25 | 0.038 | 0.085 |
| 15.25 | 0.019 | 0.053 |
| 17.25 | 0.006 | 0.036 |

constant $k$. They propose that $k$ is always a function $k(m)$ of the mean and that one should use Taylor's power law (Variance = $a \cdot m^b$, where $a$ and $b$ are constants) and a moment estimator for the variance to determine the form of the function $k(m)$. It is our opinion that, for egg sampling especially, $k$ can be treated as constant. That is, we believe that eggs are highly clumped in order to insure successful fertilization and thus the clumping should be relatively independent of spawning biomass. Mangel (1987) discusses other biological motivations for choosing the $NB$ distribution.

The distribution given in Eq. (3) is conditioned on the assumption that the $i^{th}$ station is a habitat for the spawners. Let

$$(6) \quad p_i(m) = Pr\{\text{the } i^{th} \text{ station is a habitat for the spawners}\}$$

Methods for determining the functional form of $p_i(m)$ will be described in the next section. We assume that $p_i$ may depend upon $m$ because as the stock abundance increases and decreases, the spawning habitat used by the stock may expand and contract (MacCall 1988). Since $m$ increases with biomass, the net effect is that the probability that a station is a habitat for the spawners depends upon biomass, and through that, $m$. Note, however, that we do not assume that all samples are taken at stations that are habitats.

*Can. J. Fish. Aquat. Sci., Vol. 47, 1990*

1877

The unconditional distribution of eggs at stations is then

$Pr\{X_i = 0\}$ = $Pr\{$station is not a habitat$\}$ + $Pr\{$station is a habitat and no eggs are present$\}$
$$= (1 - p_i(m)) + p_i(m) \; [k/k + m]^k$$

$Pr\{X_i > 0\}$ = $1 - Pr\{X_i = 0\}$
$$= p_i(m)\{1 - [k/k + m]^k\}$$

$Pr\{X_i = x > 0\}$ = $Pr\{$station is a habitat and $x$ eggs are present$\}$

(7)  $= p_i(m) \; \{[\Gamma(k + x)/ \; \Gamma(k)x!] \; [k/k + m]^k \; [m/k + m]^x\}$.

In the analysis that follows, we assume that $k$ is known but that $m$ is unknown and that the objective of the sampling is to make inferences about $m$. In particular, we assume that the adult biomass and mean number of eggs per station are linearly related. In general, this assumption must be verified by historical data, which we do in the case study. This means that there is a critical value of $m$, denoted by $m_c$, related to the critical level of biomass $B_c$ in the sense that

(8)  $Pr\{B > B_c\} = Pr\{m > m_c\}$.

In order to justify this assumption, we must either back-calculate from egg abundance to spawning biomass level (e.g. Lasker 1985; Wolf and Smith 1985, 1986) or use historical data that relate previous levels of spawning biomass and mean egg density to determine the value of $m_c$, as in the case study presented below.

Other types of "contagious" distributions can be used in the analysis in place of the NB distribution. Some choices are discussed in the Appendix.

## Presence–Absence Sampling

In this section, presence–absence sampling (PAS) is described as a method for estimating and making inferences about $m$. A number of extensions of the basic NB model — including imperfect detection of the plankton and multiple age groups — are described.

In presence–absence sampling, the incidence of eggs over space (time is essentially "fixed" over a short interval) is determined and used to estimate the spawning biomass. There are a number of reasons, both analytical and operational, for simply noting the incidence of eggs rather than counting them. First, for certain distinctive eggs (e.g. northern anchovy), PAS can be done on board ship whereas counting eggs requires extensive off ship effort. (In order to reduce possible errors in on-ship presence–absence sampling, one often introduces a threshold [e.g. 4 or 5 eggs]: a sample is marked as "eggs present" only if the number of observed eggs exceeds the threshold.) Thus, PAS could be done as a "real time" operation. Second, when the overdispersion parameter $k$ is small (e.g. $k < 0.1$), the eggs or larvae will be highly aggregated and thus estimating the value of $m$ with any precision from a sample mean will require a large number of samples, usually larger than the number available. In addition, the value of the sample mean will often be driven by one or two large samples; sampling for incidence avoids this problem. (As the spawning biomass increases, however, so that virtually all stations have eggs, one should switch from PAS to actual counts. Below, we describe a way to determine when the switch should be made). For problems in which one needs to determine habitat boundaries as well as spawning biomass it is likely that most stations will have no eggs. (As behavioral

information about the species, and the effects of physical variables becomes better known, it may be possible to reduce the number of stations without eggs by satellite presurvey or appropriate design of the survey.) Smith (1973) discusses how overdispersal may affect estimates of survival.

We extend Eq. (7) to include the failure to detect eggs present at a site. Let $Y_i$ denote the number of eggs in the sample at the $i^{th}$ station. There are three ways to obtain $Y_i = 0$: (1) the station is not a habitat for the spawning species; (2) the station is a habitat and there are no eggs present; and (3) the station is a habitat, there are eggs present, but the sampler misses the eggs. The models developed thus far account for (1) and (2) but not for (3), which is a likely possibility since the CalCOFI stations (Fig. 1) represent about 102 km$^2$ and the sampler area is 0.05 m$^2$. Introduce a detection function

(9)  $Pr$ {detect eggs | number of eggs present $X_i = x$}
$$= \alpha_i \; (1 - e^{-b_i x})$$

where $0 \leq \alpha_i \leq 1$ is the maximum probability of detecting eggs as $X_i \rightarrow \infty$ and $b_i$ is a measure of sampler efficiency when eggs are present. Eq. (9) represents a random search formula (Mangel 1985a, b, 1987). We thus obtain

$Pr\{Y_i = 0\}$
$= Pr\{$station is not a habitat$\}$ +
  $Pr\{$station is a habitat$\} \; \Sigma_x \; Pr\{x$ eggs present$\}Pr\{$no detection at the level $x\}$

$= (1 - p_i) +$

(10)  $p_i\Sigma_x \; [\Gamma(k + x)/\Gamma(k)x!] \; [k/k + m]^k[m/k + m]^x$
$$\{1 - \alpha_i(1 - e^{-b_i x})\}$$

Eq. (10) can be simplified by the use of the moment generating function for the NB distribution and becomes

(11)  $Pr\{Y_i = 0\} = (1 - \alpha_i p_i) + \alpha_i p_i \; [k/(k + m(1 - e^{-b_i}))]^k$

Eq. (11) is the sarting point for the rest of the analysis in this paper.

To recap: we want to estimate $m$ from values of $Y_i$ and assume that the parameters $\alpha_i$, $p_i$, $b_i$, and $k$ are known. The first three are essentially operational parameters, fixed by the operational situation and equipment. The last, $k$, is a biological and operational parameter, fixed by the species and the scale of the sampling procedure. The objective is to sample $N$ stations for the incidence of eggs and to then make inferences about the properties of the mean $m$. In this paper, a Maximum Likelihood Estimation (MLE) approach is adopted (see Edwards 1972 for general discussion); Mangel (1987) discusses Bayesian methods for presence–absence sampling.

Assume that a total of N stations are sampled and that there are $N_p$ stations (indexed by $i \; \varepsilon \; \mathcal{P}$) at which eggs are discovered and $N_n$ stations (indexed by $j \; \varepsilon \; \mathcal{N}$) at which no eggs are discovered. The likelihood $\mathcal{L}$ of such a data set is constructed from the basic model Eq. (11). Begin by setting

(12)  $w_i = 1 - e^{-b_i}$

The likelihood of the data set is then

(13)  $\mathcal{L} = \Pi_{j \varepsilon \mathcal{N}} \; [(1 - \alpha_j p_j + \alpha_j p_j \; \{k/(k + mw_j)\}^k] \; \Pi_{i \varepsilon \mathcal{P}} \; [\alpha_i p_i - \alpha_i p_i \{k/(k + mw_i)\}^k]$

From this likelihood function, we want to compute the maximum likelihood estimate (MLE) for $m$. The MLE, which we denote by $\hat{m}$, must generally be computed numerically. There
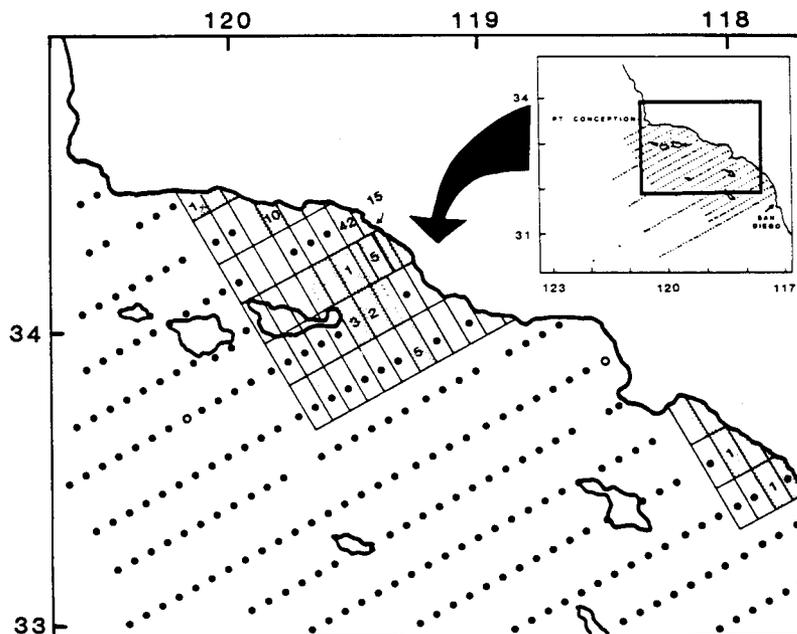
1878

FIG. 1. Sampling stations with eggs in the May, 1985 sardine survey.

is at least one case, however, in which $\hat{m}$ can be found analytically. This is the special case in which for all values of $i$ $p_i = p$, $\alpha_i = \alpha$, and $w_i = w$. This assumption must be somewhat restrictive, since it is based on the assumption of uniformity of oceanic habitat over space. On the other hand, for a survey done over a relatively small region and done quickly, these assumptions might be valid. Assuming the constancy of $\alpha$, $p$, and $w$, leads to the likelihood

(14) $\quad \mathscr{L} = [1 - \alpha p + \alpha p\{k/(k+mw)\}^k]^{N_n}$
$\qquad \times [\alpha p - \alpha p\{k/(k+mw)\}^k]^{N_p}.$

In this equation, $N_n$ and $N_p$ are the number of stations with no eggs and with eggs, respectively. Set $N = N_n + N_p$, so that $N$ is the total number of stations sampled. Direct computation shows that the MLE is

(15) $\quad \hat{m} = (k/w)[\{N\alpha p/(N\alpha p - N_p)\}^{(1/k)} - 1]$

Thus, for fixed $k$, $w$, $N$, and $\alpha p$, the MLE estimate is a function of the number of positive samples $\hat{m} = \hat{m}(N_p)$. Eq. (15) represents a generalization of a formula due to D. J Gerrard (Southwood 1976, p. 55).

The likelihood can be viewed as a function of $m$, $\mathscr{L} = \mathscr{L}(m)$ so that the likelihood ratio

(16) $\quad R(m, \hat{m}) = \mathscr{L}(m)/\mathscr{L}(\hat{m})$

provides a measure of the relative likelihood of any value of $m$ in comparison to the MLE. Alternatively, $R(m, \hat{m})$ is approximately the posterior distribution of $m$ under Bayesian analysis when a uniform prior is assumed. Figure 2 shows one such likelihood ratio, for parameters corresponding to the 1985 sardine egg cruise shown in Fig. 1. The likelihod ratio is highly skewed with a long tail (indicating the chance, albeit small, of some very large values of $m$). If $m$ is proportional to the spawn-

ing biomass, then we can view $R(m, \hat{m})$ as a likelihood distribution for the biomass. The probability that $m$ exceeds $m_c$ and aproximate confidence intervals for the biomass can then be constructed from this likelihood function (Hudson 1971).

Since the number of positive samples is a random variable, the function $\hat{m}(N_p)$ is also a random variable. Its mean can be computed by the application of Seber's delta method (Seber 1973) and is

(17) $\quad E\{\hat{m}(N_p)\} \sim \hat{m}(E\{N_p\} + (1/2) [\hat{m}''(N_p) \mid_{E\{N_p\}}] \text{Var}\{N_p\}$

where $E\{N_p\}$ is the expected value of $N_p$ and is

(18) $\quad E\{N_p\} = N\alpha p[1 - \{k/(k+mw)\}^k]$

and $\hat{m}''(N_p)$ is the second derivative of $\hat{m}(N_p)$; in Eq. (17) it is evaluated at the expected value of $N_p$.
The following properties are verified

1) $\hat{m}(E\{N_p\}) = m$
2) $\hat{m}''(N_p) =$
$\quad (1/w) [(N\alpha p)^{(1/k)} (1 + (1/k)) (N\alpha p - N_p)^{-(2+(1/k))}]$
3) $\text{Var}\{N_p\} = N\alpha p[1 - \{k/(k+mw)\}^k] [1 - \alpha p + \alpha p\{k/(k+mw)\}^k]$.

Thus, although we find that $\hat{m}(E\{N_p\}) = m$, it is not in general true that $E\{\hat{m}(N_p)\} = m$ and the MLE is in general a biased estimator. Table 3 provides numerical information on the size of the bias. Note the trend that the bias decreases with decreasing $b$. The other trend — that the bias decreases with increasing $\alpha p$ — is easily understood since we are dealing with less uncertainty as $\alpha p$ increases.

In order to test the estimation procedures for $m$ further, we used a Monte Carlo simulation to generate data according to

(19) $\quad Pr\{\text{positive sample}\} = \alpha p[1 - \{k/(k+mw)\}^k]$

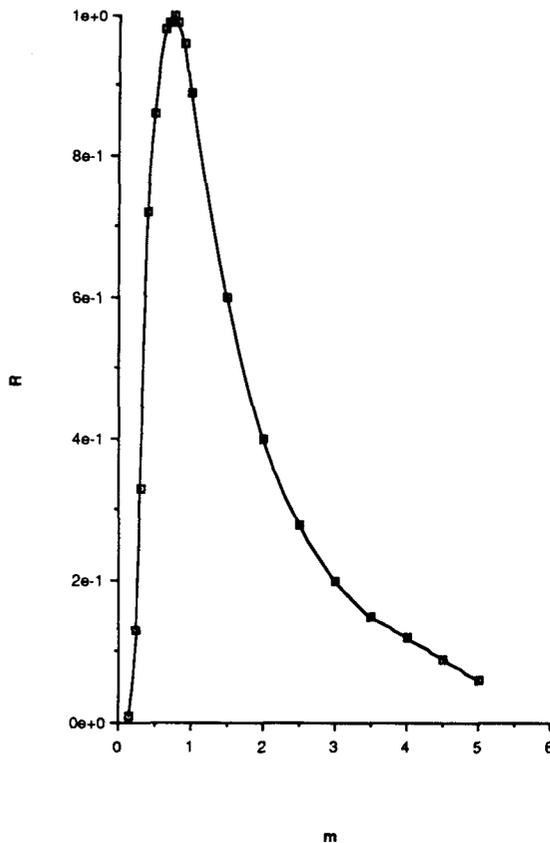Can. J. Fish. Aquat. Sci., Vol. 47, 1990

1879

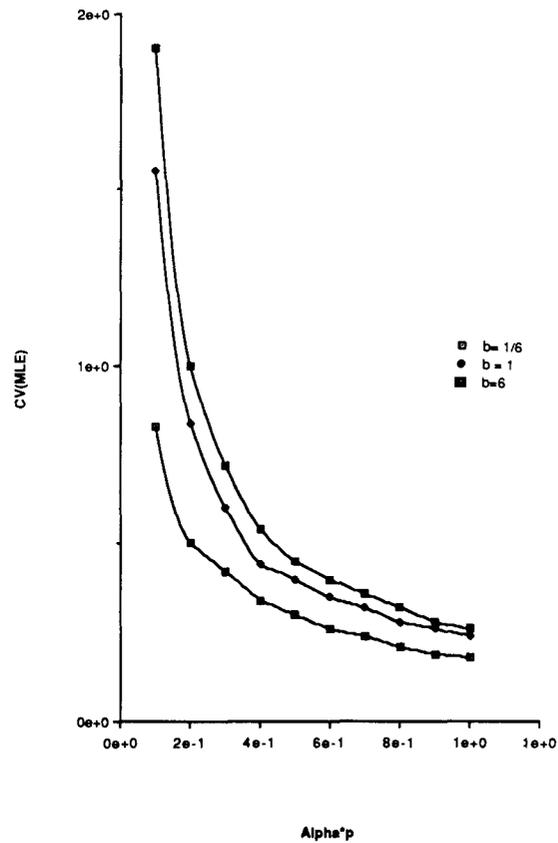FIG. 2. Likelihood ratio $R(m, \hat{m})$ (Parameters are $\alpha p = 0.1$, $b = 6$, $k = 0.2$, $N = 419$, $N_n = 408$).

FIG. 3. Coefficient of variation (CV) of the MLE for $m$, from the simulations, as a function of $\alpha p$ and $b$. (Parameters are $N = 419$, $m = 5$, $k = 0.2$).

TABLE 3. Bias in the MLE $\hat{m}$ $(N = 419, m = 5, k = 0.2)$.

| | Bias for | | |
|---|---|---|---|
| $\alpha p$ | $b = 6$ | $b = 1$ | $b = 1/6$ |
| 1 | 0.36 | 0.30 | 0.18 |
| 0.9 | 0.43 | 0.36 | 0.20 |
| 0.8 | 0.53 | 0.44 | 0.24 |
| 0.7 | 0.65 | 0.53 | 0.28 |
| 0.6 | 0.81 | 0.66 | 0.34 |
| 0.5 | 1.04 | 0.83 | 0.42 |
| 0.4 | 1.39 | 1.09 | 0.54 |
| 0.3 | 1.96 | 1.54 | 0.74 |
| 0.2 | 3.10 | 2.43 | 1.14 |
| 0.1 | 6.54 | 5.09 | 2.34 |

and in each run of the simulation, the value of $\hat{m}$ was determined. Figure 3 shows the coefficient of variation CV($\hat{m}$) of $\hat{m}$ obtained using the simulations. The same trends with $b$ and $\alpha p$ are observed for the coefficient of variation as for the bias.

In many instances, the eggs or larvae sampled will include different age-classes that involve different levels of dispersion and different means. For example, sardine eggs are usually classified as $S$ (eggs encountered during spawning), $A$, $B$, or $C$

(1-, 2-, or 3-d-old). We explicitly extend the model to include age-classes of eggs as follows. Let

$$(20) \quad Y_{ij} = \begin{cases} 1 & \text{if there are eggs of the } j^{\text{th}} \text{ age-group} \\ & \text{at the } i^{\text{th}} \text{ station} \\ 0 & \text{otherwise}. \end{cases}$$

The natural extension of the basic model is

$$(21) \quad Pr\{Y_{ij} = 0\} = 1 - \alpha_i p_i + \alpha_i p_i \{k_j/(k_j + m_j w_i)\}^{k_j}$$

where $k_j$ and $m_j$ are the values of $k$ and $m$ for the $j^{\text{th}}$ age-group. For sardine, we choose the parametrization (Smith 1973; Lasker 1985, p. 17 ff) $k_j = 0.1 j$ and $m_j = \{1 - 0.3(j-1)\}\, m$ with $j = 1, 2, 3,$ or 4 for $S, A, B,$ or $C$ eggs respectively. The interpretation of this parametrization is the following. The $k_j$ increase with time, indicating a more even (less patchy) distribution of older eggs. We presume that this is due to oceanic mixing after spawning. The $m_j$ decrease, with time, indicating mortality at about 30% per day for the first 3 d. This 30% per day mortality is true for sardine eggs; for anchovy eggs it is about 20%. Detailed justification of these values, along with variance estimates, is given by Smith (1973) and Smith and Hewitt (1985).

To construct the likelihood function, we now need to include the interaction between eggs of different age-groups. It is con-

1880

ceivable that the presence or absence of eggs of a certain age-group is not completely independent of the presence or absence of eggs of other age-groups. On the other hand, we do not know the exact correlation between eggs or different age-groups. As a way around this difficulty, we introduce a parameter $r_c(i)$ that can be interpreted as a simple "association coefficient" or "presence–absence correlation coefficient": when $r_c(i)$ is set equal to 0, the eggs at the $i^{th}$ station are treated as completely independent, random variables and when $r_c(i)$ is set equal to 1, the eggs at the $i^{th}$ station are treated as completely dependent random variables so that if eggs of one age-class are present, eggs of all age-classes are present. With this model,

$Pr$\{no eggs of any age group at the $i^{th}$ station\}

$$= (1 - r_c(i)) \prod_{j=1}^{4} Pr\{Y_{ij} = 0\}$$

$$(22) \quad + r_c(i) \min_j Pr\{Y_{ij} = 0\}$$

This model is used in the next section for the analysis of the 1985 sardine egg survey.

## Application to the 1985 Sardine Egg Survey

The methodology developed in the previous section was applied to the data for the 1985 sardine egg survey, conducted in May 1985 by P. Wolf (California Department of Fish and Game) and one of us (P. E. Smith). In that survey, 419 CalCOFI stations were sampled and 11 of those stations (Fig. 1) had sardine eggs of any age group. The total number of eggs discovered was 86. The stated objective of the survey was to determine whether the spawning biomass exceeded 20 000 short tons. The data thus consist of $N_n = 408$ and $N_p = 11$. From those data, the objective was to estimate the maximum likelihood value of $m$ and the probability that $m$ exceeded the critical value $m_c$.

The multiple age group model was used for calculations, with the $\{m_j, k_j\}$ given above. Other parameters were set at $\alpha p = 1/20$, $b = 0.6$, and $m_c = 1.14$. These are justified as follows. Based on historical records (from the period of the cannery heyday), a sardine spawning biomass of 20 000 short tons covered about 5% of the long-term spawning area (Kramer and Smith 1971; Lasker 1985, p. 18). We thus set $p = 0.05$ and $\alpha = 1$ (note, of course, that only the product $\alpha p$ enters into the computations); the choice $\alpha = 1$ means that if eggs are abundant enough they will be detected. We choose $b = 0.6$ as an upper bound on the effectiveness of the sampler. The value of $m_c$ was also determined by comparison with historical data from the cannery period. We performed computations for $r_c(i)$ constant for all $i$ and in the range of 0.4–1.0. The MLEs obtained in this manner ranged between 1.2 (for $r_c(i) = 0.4$) and 12.5 (for $r_c(i) = 1.0$); the probability that $m$ exceeded the critical level $m_c$ was high for all combinations of parameters tested.

Wolf and Smith (1986) used a method based on egg production and concluded, using the same data, that the spawning biomass exceeded 20 000 short tons and recommended opening a small, directed fishery on sardine in 1986. The "inverse egg production method" used by Wolf and Smith is motivated as follows. The egg production method (Parker 1980; Picquelle and Hewitt 1983; Hewitt 1984; Lasker 1985) is used to estimate spawning biomass $B$ by

$$(23) \quad B = P_0 A \ kW/RFS$$

where $P_0$ is daily egg production (eggs/0.05 m²/d), $W$ is average female weight, $R$ is sex ratio of the population, $F$ is batch fecundity (number of eggs spawned/female/batch), $S$ is the proportion of females spawning per day, $A$ is the total area of the survey, and $k$ is a conversion factor from grams to short tons.

Wolf and Smith (1985) invert Eq. (22) (hence the name inverse egg production method IEPM) and argue as follows: if the spawning biomass exceeds $B_c$ short tons, then the area with eggs should exceed a critical area $A_c$ given by

$$(24) \quad A_c = B_c \ RFS/P_0 kW$$

To employ the IEPM, one multiplies the number of stations with positive egg samples by the area associated with each station (in this case, the area is 40 nm²) and compares that area with $A_c$. In particular, using the inverse egg production method, Wolf and Smith estimated that a population of 20 000 short tons spawning biomass would cover about 500 nm²; they estimated that in 1985 the population covered about 670 nm². Thus, for the first time since 1974 a directed fishery for sardine opened on 1 January 1986 (Wolf and Smith 1986). It should be noted that many of the parameters in Eq. (23) have considerable uncertainty (Wolf and Smith 1986; Mangel 1988) and this affects the confidence that one has in the estimate.

The IEPM and the methods developed in this paper are similar in spirit, since the objective of each method is to use presence/absence data to determine the spawning biomass of the stock. Although the two methods use the same kind of data, they are different in methodology (the IEPM is fundamentally deterministic while the methods of this paper are fundamentally stochastic). Both methods are easily employed using a small, desktop microcomputer. In a sense then, one has the best possible situation: the IEPM and the methods developed in this paper are two relatively independent analytical procedures for assessing the size of the spawning biomass from the same data. They thus complement each other and can be used to check one another.

Finally, both methods are ideally suited to the situation in which the spawning biomass is sufficiently small that using actual counts is not advisable. For example, analogous egg surveys are currently performed for anchovy, but use counts instead of presence–absence.

## When are Counts Preferable to Presence–Absence Sampling?

The alternative to presence–absence sampling is the use of counts, that is records of the actual values of $X_i$, rather than whether or not $X_i > 0$. In this section, we develop a method that can be used to determine when using counts is preferable to using presence–absence sampling, in the sense that one has a smaller error than the other.

We consider the situation in which all stations have the same values for the detection parameters and the probability of the station being a habitat. In that case, with $Y_i$ denoting the number of eggs in the sample

$$(25) \quad Pr\{Y_i = 0\} = 1 - \alpha p + \alpha p \{k/(k+mw)\}^k$$

The MLE for presence–absence sample, henceforth denoted by $m_{m/c}$ is given by Eq. (15). Another commonly used estimate of $m$ from presence–absence sampling (Southwood 1976) uses the fraction of zeros $f_o = N_n/N$ in the sample. This is essentially a moment estimator. Assuming that $k$ is constant, equating $f_0$ to
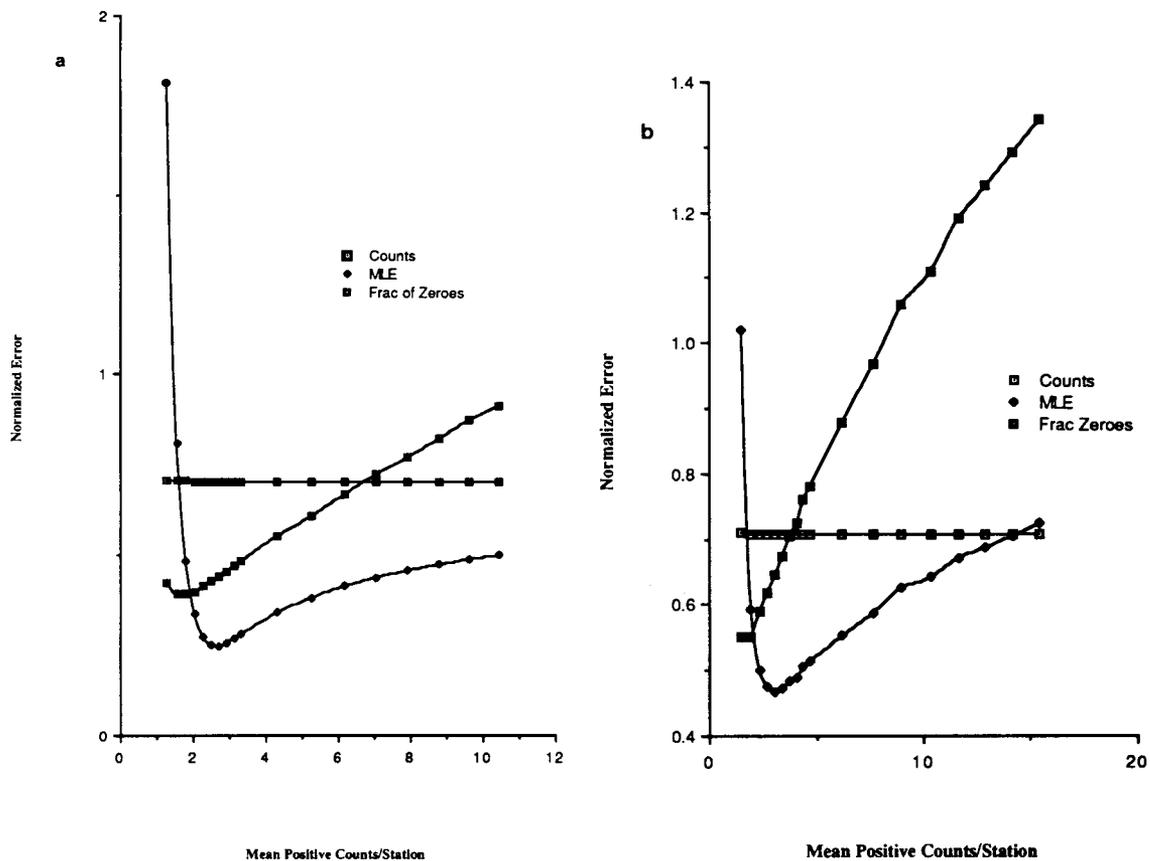
*Can. J. Fish. Aquat. Sci., Vol. 47, 1990*

1881

FIG. 4. The normalized errors for the three estimation methods as a function of the number of eggs per positive station. (a) $N = 400$, $\alpha p = 0.3$, and $k = 0.1$, (b) $N = 400$, $\alpha p = 0.3$, and $k = 0.2$.

the right hand side of Eq. (25) and solving for $m$ gives the estimate

$$(26) \quad m_{frac} = (k/w)\{\alpha p/(f_o - 1 + \alpha p)\}^{1/k} - k$$

Finally, one can estimate $m$ by the mean number of eggs in the samples. Denote this estimate by $m_c$ for count; it is

$$(27) \quad m_c = (1/N) \Sigma Y_i$$

We consider the situation in which $w = 1$, but $\alpha p < 1$. In light of Eq. (25) we write that

$$(28) \quad Y_i = \begin{cases} 0 \text{ with probability } 1 - \alpha p + \alpha p(k/k+m)^k \\ x \text{ with probability } \alpha p NB(x;m,k) \end{cases}$$

Here $NB(x;m,k)$ denotes the negative binomial probability of $x$ events when the parameters are $m$ and $k$, so that with probability $1 - \alpha p + \alpha p(k/k + m)^k$ no eggs are detected in the sample but when eggs are detected they follow the negative binomial distribution. Using Eq. (27), we see that $E\{m_c\} = \alpha pm$.

To compare these three estimators, we propose using the following normalized error

$$(29) \quad S_i = \{E(m_i - m)^2\}^{1/2}/m$$

in which $m_i$ denotes any of the three estimators and $E\{ \}$ denotes the expectation over all possible samples. If the $m_i$ were unbiased estimators, then the $S_i$ would be associated coefficients of variation.

When counts are used, $S_c$ can be computed directly and is

$$(30) \quad E\{m_c - m^2\} = (1/N^2)\{N\alpha p(\text{m}^2 + m + \text{m}^2/k) + N(N-1)\alpha^2 p^2 \text{m}^2\} + m^2 - 2\alpha pm^2$$

We were unable to develop an analytical formula for $m_{mle}$ or $m_{frac}$. They are, however, easily computed numerically according to the following scheme. First note that the number of positive samples follows a binomial distribution with number of trials $N$ and success probability $p' = \alpha p - \alpha p(k/k+m)^k$. Let $\mathcal{B}(n;N,p')$ denote the probability of $n$ positive samples for the binomial distribution with parameters $N$ and $p'$. Then

$$(31) \quad E\{(m_i - m)^2\} = \sum_{n=0}^{N} (m_i(n) - m)^2 \mathcal{B}(n;N,p')$$

In this equation, $m_i(n)$ is the appropriate estimator when the number of positive samples is $n$. The right hand side of Eq. (31) is most easily evaluated by noting that Prob $\{N_p = 0\} = (1 - p')^N$ and that for $n > 0$.
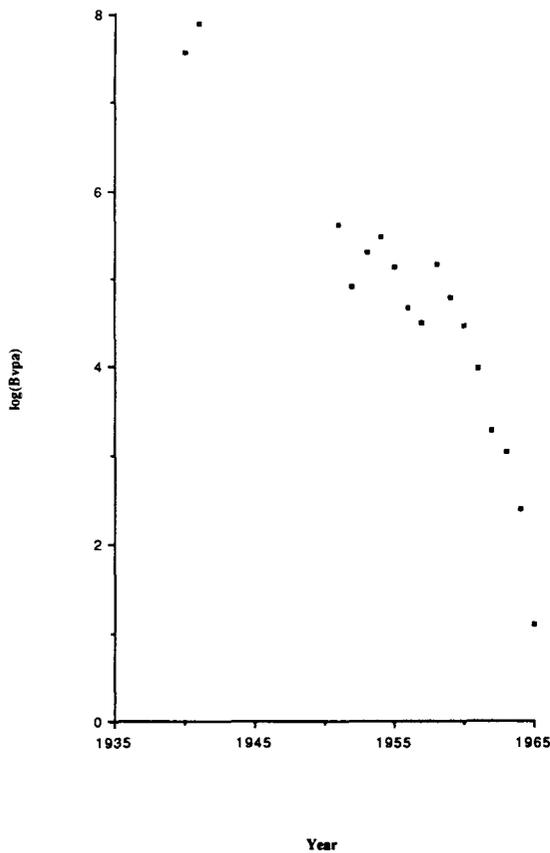
1882

Year



Year

FIG. 5. Estimated spawning biomass $B_{vpa}$ (short tons) of Pacific sardine, 1940–41 and 1951–65. The spawning biomass was estimated by Virtual Population analysis (Smith 1990).

(32) $\quad \text{Prob } \{N_p = n+1\} = [(N-n)/(x+1)][p'/(1-p')]$
$$\text{Prob } \{N_p = n\}$$

One complication, however, remains. In real surveys, one does not know the value of $m$. What will be known is the mean number of eggs per positive station. From the definition of conditional probability and Eq. (29)

(33) $\quad Pr\{Y_i = x \mid Y_i > 0\} = Pr\{Y_i=x, Y_i>0\}/Pr\{Y_i>0\}$
$$= NB(x;m,k)/(1 - (k/k+m)^k)$$

Thus, the mean number of counts per positive station is estimated by

(34) $\quad E\{Y_i \mid Y_i > 0\} = \sum_{x=1} x \, NB(x;mk,)/(1 - (k/k+m)^k)$
$$= m/(1 - (k/k+m)^k)$$

Note that this mean is independent of $\alpha p$; this happens because one is only considering stations with positive counts.

Figure 4 shows the normalized errors as a function of the expected mean number of counts per positive station and suggests that when $m$ is very small, presence–absence sampling using the estimate based on the fraction of zeroes is most preferable, that for intermediate values of $m$ presence–absence sampling using the MLE is best, and that for larger $m$, counts may
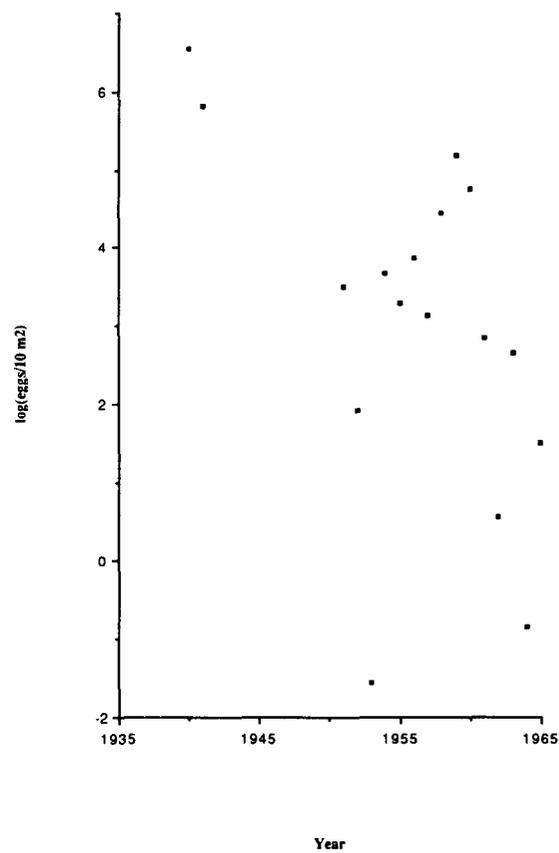
FIG. 6. Egg density (number of eggs/ 10 m² sea surface) for the same period as in Fig. 7. Note that there are years in which spawning biomass is very low but in which egg density nearly equals the egg density when the biomass was high.

be preferable (e.g. Fig. 4b). We stress that these results depend upon parameter values, and that one should repeat the computations for his or her own particular parameter values. The computational schemes outlined in Eq. (25)–(33) are, however, easily implemented on desk-top microcomputers.

## A Case Study of Sardine

To justify a number of the assumptions used in this paper, we summarize the results of a study by one of us (Smith 1990) on a long time series for sardine in the California current.

The time series tracked sardine population size as it declined from about $2 \times 10^6$ t in 1940 (Murphy 1966) to about $3 \times 10^3$ t in the mid-1960s (MacCall 1979). For the period of interest, we show the biomass $B_{vpa}$ estimated by virtual population analysis (Fig. 5), the estimated number of eggs per unit area (Fig. 6), and the percent positive stations (Fig. 7). Note the relatively large estimates for the number of eggs per unit area and the considerable fluctuations in percent positive stations during a period in which the biomass consistently declined. For example, in 1941, when the biomass $B_{vpa} = 2709$ t, the egg estimate is 337 eggs/10 m² while in 1959, when $B_{vpa} = 122$ t,

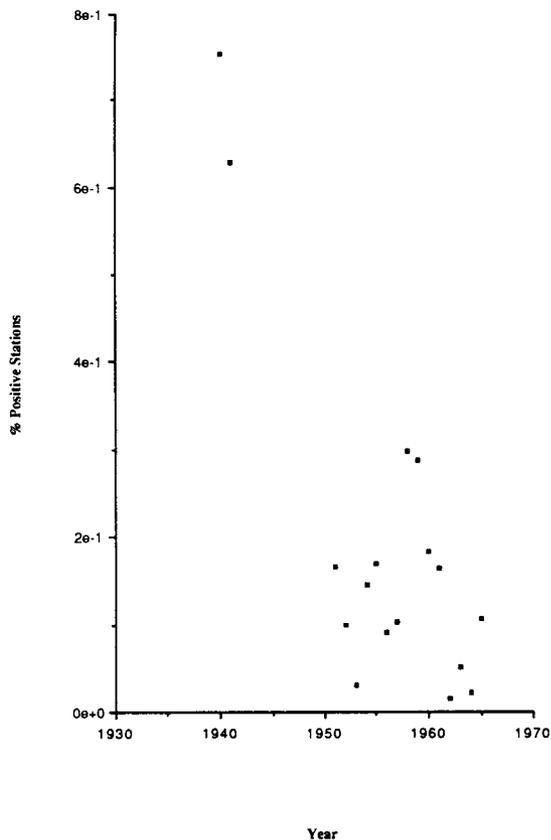Can. J. Fish. Aquat. Sci., Vol. 47, 1990

1883

FIG. 7. Percent positive stations for the same period as in Fig. 6. The difficulties identified in Fig. 6 do not occur for this measure of presence–absence sampling.
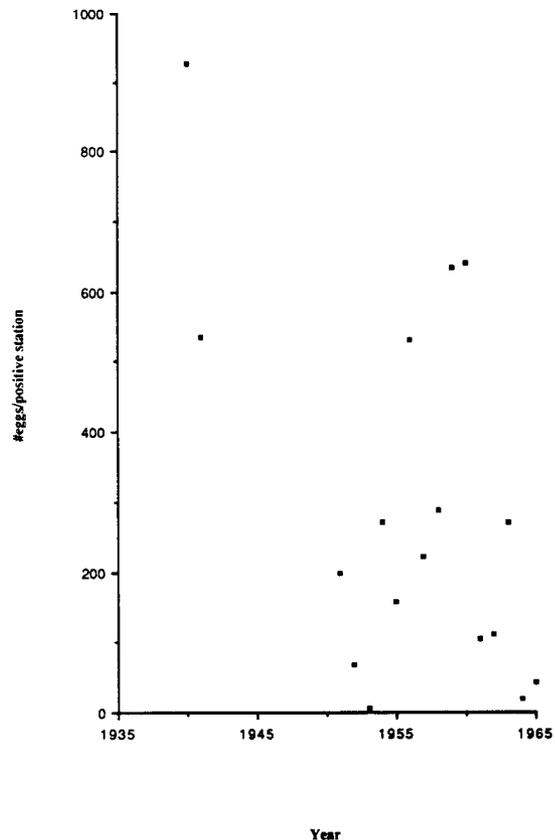


FIG. 8. Number of eggs per positive station for the same period as in Fig. 7. Note that the difficulties identified in Fig. 6 reoccur here, when actual counts are used.

the egg estimate is 182 eggs/10 m$^2$. That is, the biomass dropped by a factor of about 20 and the average estimate of eggs per unit area dropped by 2. We believe that this is caused by the clumped distribution of the eggs. The generally high values of eggs per positive station (Fig. 8) between 1955 and 1960 are an example of how counts of eggs per positive station can bias the estimation. For example, in 1941 ($B_{vpa}$ = 2709 t) the number of eggs per positive station was 536 and in 1956 ($B_{vpa}$ = 108 t) the number of eggs per positive station was 532. These results suggest that counting eggs may give incredibly misleading estimates of biomass under circumstances that could be reasonably expected to occur and that practicable alternatives such as presence–absence sampling are important.

Smith (1990) used a combination of Virtual Population Analysis (VPA) and regression. The results of this work that are relevant to ours include the following. The apparent relationship between spawning biomass, $B_{vpa}$, and eggs per unit area is

(35)   $B_{vpa}$ = 44.3 + 3.43 $E$

where $B_{vpa}$ is the sardine biomass over 2 yr old and $E$ is the estimate of eggs per unit surface area (number/10 m$^2$). The $r^2$ corrected for sample size is 63.8% and the constant is not sig-

nificant, suggesting that our assumption of a linear relationship between spawning biomass and mean egg density is justified. When the analysis is done without the constant, the relationship becomes $B_{vpa}$ = 3.5419 $E$. By inverting Eq. (34), it is possible to determine the critical level of egg density associated with a critical level of biomass, as required by the analysis in previous portions of the text.

For an analysis in which the number of square miles covered ($E_a$) and the number of eggs per positive station ($E_{pos}$) are separated, the results of the regression are

(36)   $B_{vpa}$ = − 187.8 − 0.6918 $E_{pos}$ + 112.43 $E_a$

Neither the constant nor the coefficient of $E_{pos}$ are significant. The $r^2$ for the model (36) is 79.1%. Excluding both the constant and the coefficient of $E_{pos}$ gave the relationship $B_{vpa}$ = 74.367 $E_a$, justifying the assumptions of the inverse egg production method and the interpretation of our results that presence–absence sampling provides a good measure of areal coverage.

Similar results are obtained when larval data are used, with higher values of $r^2$. We believe that this is caused by a longer period in which the larvae are vulnerable to a survey (20 d for larvae vs. 3 d for eggs). In addition, there is excellent corre-

1884

TABLE 4. Mean number of sardine eggs per positive tow for Pacific sardine 1931–89.

| Sampling device | Mean number of eggs per positive station | Sardine biomass (million m-tons) |
| --- | --- | --- |
| California Dept. of Fish and Game 2 m silk ring net (1931–32) | 543 | 3.50[a] |
| One metre silk ring net (1941) | 568 | 2.50[b] |
| Half-meter silk ring net (1941) | 618 | 2.50[b] |
| High-speed net (1950) | 405 | 0.70[b] |
| CalCOFI standard net (1-m ring net) 1951–79) | 476 | <0.30[b] |
| Bongo net (1977–89) | 397 | <0.02 |
| CalCOFI vertical egg net (CalVET) 1985–86 | 932 | 0.02 |
| Simulation using a log-normal distribution with mean 3.7 and SD 2.14, based on eggs/positive tow 1951–89 | 416 | Not applicable |

[a]Murphy (1966).
[b]MacCall (1979).

lation between overall larval abundance and overall egg abundance.

Finally, we compiled essentially all of the sampling information for Pacific sardine over the last half-century (Table 4). The results of a simulation study of sampling, using the data on number of eggs per positive station from 1951–89, show that the mean number of sardine eggs per positive tow is very insensitive to the spawning biomass of the sardine and the type of net used to sample the sardine eggs (Table 4). Thus the results suggest that the area of spawning (or fraction of spawning ground covered) is most sensitive to spawning biomass in this species. All of the sample means in Table 4 could have been derived from random samples from a population in which the number of eggs per positive station was log-normally distributed with mean 3.7 and standard deviation 2.14, as in the last line of Table 4. A study of the sampling distributions, not presented in Table 4, showed that the upper limit of the number of sardine eggs per positive station was nearly uniform for all samplers and biomass levels. The lower limit is controlled by the threshold of the sampler. For example, one sardine egg in a CalVET sampler (area 0.05 m$^2$) corresponds to 200 eggs per 10 m$^2$, whereas 1 sardine egg in a Bongo net is five sardine eggs per 10 m$^2$, and one sardine egg in the CalCOFI standard net is three sardine eggs per 10 m$^2$.

## Conclusions

Our objective was to introduce a new method for the analysis of egg or larval sampling for purposes of biomass estimation.

Presence–absence sampling has a number of points that will help facilitate fisheries management. First, when compared to full fishing surveys for stock estimation, presence–absence sampling is a relatively inexpensive means of gaining information about the managed stock. For example, one can use presence–absence sampling and the methods in this paper to determine if the probability that the stock is sufficiently strong to warrant a full biomass survey. Second, presence–absence sampling provides a natural determination of the management boundaries of the stock. Third, the method has wide applicability. Fourth, the method requires a minimum amount of data and parameters and nicely complements other methods for estimating spawning biomass.

When a stock is recovering, as in the case of sardine, a variety of estimation techniques are needed for the successful management of the recovery. Early on, a relatively simple and inexpensive but cautious technique such as presence–absence sampling is valuable. During this period one wants to build confidence both in the estimation methodology and in the recovery of the stock. Afterwards, more expensive assessment techniques such as the egg production method become useful as a means of accurately quantifying the stock level. When the stock is sufficiently recovered and strong, one can use techniques that interpolate between years in which expensive assessment methods are used.

There are a number of directions for future work. The most important or promising are the following ones.

*Joint estimation of habitat boundaries and the biomass* — In general, the boundary of the geographical habitat will not be known and one simultaneously estimates the distribution of the mean of the biomass and the probability that a site is a habitat. Wolf and Smith (1985, table 2, p. 137) give an illustration of the importance of habitat exit information. As spawning population increases, the spatial spread of the stock may increase (Rosenzweig 1981; MacCall 1988). MacCall (1984, 1988) shows how to compute the spawning habitat suitability contours for the northern anchovy *Engraulis mordax*. These contours represent average phenomena and in any given year one will not know the location of the habitat contours precisely. Operationally, then, the long term average habitat contours can be used to plan the initial part of the sea survey, but then the presence–absence data are used to update (e.g. in a Bayesian sense (Mangel 1985b, 1988)) the estimation of the contours. The natural extension of these ideas is adaptive survey optimization (Walters 1986): one develops an "adaptive survey design" in which the data that have been accumulated thus far in the survey are used to determine where future sampling occurs. The sampling process {$Y_{ij}$} is now viewed as a spatial point process and one is attempting to control properties of this spatial point process.

*Egg surveys as priors and economic modeling* — If one decides that the spawning biomass does exceed the critical level, then a full biomass (i.e. sampling of juvenile and adult fish) survey may be warranted. In such a case, the egg survey can be used as a prior spatial distribution when planning the full biomass survey. In addition, the results of the presence–absence egg survey can be used to determine the economic costs and benefits of a complete biomass survey and of not allowing commercial fishing if the spawning biomass exceeds an appropriate level. Thus, the presence–absence survey fits nicely into the overall bioeconomics of fisheries management.

We believe that presence-absence sampling provides a new tool for the estimation of the biomass of pelagic fish stocks.

The great advantages of presence–absence sampling is that it can be used in situations in which other methods are either too expensive or unreliable and thus provides a technique that complements existing ones.

## Acknowledgments

## References

ANSCOMBE, F. J. 1984. The transformation of Poisson, binomial and negative binomial data. Biometrika 35: 256–254.
   1950. Sampling theory of the negative binomial and logarithmic series distributions. Biometrika 37: 358–382.
BISSELL, A. F. 1972. A negative binomial model with varying element sizes. Biometrika 59: 435–441.
BLISS, C. I., AND A. R. G. OWEN. 1958. Negative binomial distributions with a common $k$. Biometrika 45: 59–68.
DEBOUZIE, D., AND J. THIOULOUSE. 1986. Statistics to find spatial and temporal structures in populations, p. 263–282. In M. Mangel, J. R. Carey, and R. Plant [ed.] Pest Control: Operations and Systems Analysis in Fruit Fly Management. Springer–Verlag, Heidelberg.
DENNIS, B., AND G. P. PATIL. 1984. The gamma distribution and weighted multimodal gamma distributions as models of population abundance. Math. Biosci. 68: 187–212.
DIGGLE, P. J. 1983. Statistical Analysis of Spatial Point Patterns. Academic Press, New York, NY. 148 p.
DOWNING, J. A. 1979. Aggregation, transformation, and the design of benthos sampling programs. J. Fish. Res. Board Can. 36: 1454–1463.
EDWARDS, A. W. F. 1972. Likelihood. Cambridge University Press, Cambridge, UK.
GROSBERG, R. K., AND J. F. QUINN. 1986. The genetic control and consequences of kin recognition by the larvae of a colonial marine invertebrate. Nature (Lond.) 322: 456–459.
HEWITT, R. 1981. The value of pattern in the distribution of young fish. Rapp. P.-v. Cons. Int. Explor. Mer 178: 229–236.
   1984 spawning biomass of northern anchovy. Administrative Report LJ-84-18. Southwest Fisheries Center, La Jolla, CA.
HEWITT, R., A. BINDMAN, AND N. LO. 1984. Procedures for calculating the egg production estimate of spawning biomass. Admin. Rep. LJ-84-19. Southwest Fisheries Center, La Jolla, CA.
HUDSON, D. J. 1971. Interval estimation from the likelihood function. J. R. Stat. Soc. 33: 256–262.
JOHNSON, N., AND S. KOTZ. 1969. Discrete Distributions. Wiley Interscience, New York, NY. 328 p.
KRAMER, D., AND P. E. SMITH. 1971. Seasonal and geographic characteristics of fishery resources. California current region — VII. Pacific Sardine. Comm. Fish. Rev. 33: 7–11.
LASKER, R. 1981. Marine Fish Larvae. Washington Sea Grant Publications, University of Washington Press, Seattle, WA. 131 p.
   1985. An egg production method for estimating spawning biomass of pelagic fish: application to the northern anchovy, Engraulis mordax. NOAA Tech. Rep. NMFS 36, US Department of Commerce, Springfield, VA.
MACCALL, A. D. 1979. Population estimates for the waning years of the Pacific sardine fishery. Cal. Coop. Oceanic Fish. Invest. Rep. 20: 72–82.
   1984. Population models of habitat selection, with application to the northern anchovy. Administrative Report LJ-84-01. Southwest Fisheries Center, La Jolla, CA.

   1988. Dynamic Geography of Marine Fish Populations. University of Washington Press, Seattle, WA.
MANGEL, M. 1985a. Search models in fisheries and agriculture. Springer lecture notes in biomathematics. Springer-Verlag, New York, NY. 61: 105–138.
   1985b. Decision and Control in Uncertain Resource Systems. Academic Press, New York, NY. 255 p.
   1987. Sampling highly aggregated populations with application to California sardine management. Springer lecture notes in biomathematics. Springer-Verlag 72: 247–263.
   1988. From egg surveys to ecosystem models: biological assumptions in fisheries management. Springer lecture notes in coastal and estuarine studies. Springer-Verlag 28: 75–102.
MURPHY, G. I. 1966. Population biology of the Pacific sardine (Sardinops caerulea). Proc. Calif. Acad. Sci. 34(1): 1–84.
PARKER, K. 1980. A direct method for estimating northern anchovy, Engraulis mordax, spawning biomass. Fish. Bull. 78: 541–544.
PENNINGTON, M. 1983. Efficient estimators of abundance, for fish and plankton surveys. Biometrics 39: 281–286.
PERRY, J. N., AND L. R. TAYLOR. 1986. Stability of real interacting populations in space and time: implications, alternatives, and the negative binomial $k_c$. J. Anim. Ecol. 55: 1053–1068.
PICQUELLE, S. J., AND R. P. HEWITT. 1983. The northern anchovy spawning biomass for the 1982–83 California fishing season. Cal. Coop. Oceanic Fish. Invest. Rep. 24: 16–28.
PIELOU, E. C. 1977. Mathematical Ecology. John Wiley and Sons, New York, NY. 385 p.
RESH, V. H. 1979. Sampling variability and life history features: basic considerations in the design of aquatic insect studies. J. Fish. Res. Board Can. 36: 290–311.
RIPLEY, B. D. 1981. Spatial Statistics. Wiley Interscience, New York, NY. 252 p.
ROSENZWEIG, M. 1981. A theory of habitat selection. Ecology 62: 327–335.
SAMPFORD, M. R. 1956. The truncated negative binomial distribution. Biometrika 42: 58–69.
SANTANDER, H., P. E. SMITH, AND J. ALHEIT. 1982. Determination of sampling effort required for estimating egg production of anchovata, Engraulis ringens, off Peru. Inst. Mar Peru Bol. 7: 5–18.
SEBER, G. A. F. 1973. The Estimation of Animal Abundance and Related Parameters. Griffin, London, UK. 506 p.
SETTE, O. E., AND E. H. AHLSTROM. 1948. Estimation of abundance of the eggs of the Pacific pilchard (Sardinops caerulea) off southern California during 1940 and 1941. Sears Found. J. Mar. Res. 7: 511–542.
SMITH, P. E. 1973. The mortality and dispersal of sardine eggs and larvae. Rapp. P.-v. Reun. Cons. Int. Explor. Mer 164: 282–292.
   1978. Biological effects of ocean variability: time and space scales of biological response. Rapp. P-v. Cons. Int. Explor. Mer. 173: 117–127.
   1990. Monitoring of Pacific sardine (Sardinops sagax) interannual changes of spawning biomass. Cal. Coop. Oceanic Fish. Invest. Rep. (In press)
SMITH, P. E., AND R. P. HEWITT. 1985. Anchovy egg dispersal and mortality as inferred from close-interval observations. Cal. Coop. Oceanic Fish. Invest. Rep. 26: 97–110.
SMITH, P. E., AND S. L. RICHARDSON. 1977. Standard techniques for pelagic fish egg and larva studies. FAO Fish. Tech. Pap. 175, Food and Agriculture Organization of the United Nations, Rome, Italy.
SOUTHWOOD, T. R. E. 1976. Ecological Methods. Chapman and Hall, London, UK. 524 p.
SQUIRE, J. L. 1983. Abundance of pelagic resources off California, 1968–73, as measured by an airborne fish monitoring program. NOAA Tech. Rep. NMFS SSRF-762, Rockville, MD.
TAYLOR, C. C. 1953. Nature and variability in trawl catches. Fish. Bull. U.S. Fish and Wildlife Service Washington, DC. 54: 145–166.
TAYLOR, L. R., WOIWOD, I. P., AND J. N. PERRY. 1979. The negative binomial as a dynamic ecological model for aggregation and the density dependence of $k$. J. Anim. Ecol. 48: 289–304.
VLYMEN, W. J. 1977. A mathematical model of the relationship between larval anchovy (Engraulis mordax) growth, prey microdistribution, and larval behavior. Env. Biol. Fish. 2: 211–233.
WALTERS, C. 1986. Adaptive Management of Renewable Resources. MacMillan Publishing Co, New York, NY.
WOLF, P., AND P. E. SMITH. 1985. An inverse egg production method for determining the relative magnitude of Pacific sardine spawning biomass off California. Cal. Coop. Oceanic Fish. Invest. Rep. 26: 130–138.
   1986. The relative magnitude of the 1985 Pacific sardine spawning biomass off southern California. Cal. Coop. Oceanic Fish. Invest. Rep. 27: 25–31.

1886

Can. J. Fish. Aquat. Sci., Vol. 47, 1990

WOLF, P., P. E. SMITH, AND C. L. SCANNELL. 1987. The relative magnitude of the 1986 Pacific sardine biomass off California. Cal. Coop. Oceanic Fish. Invest. Rep. 28: 21–26.

ZWEIFEL, J. R., AND P. E. SMITH. 1981. Estimates of abundance and mortality of larval anchovies (1951-75): Application of a new method. Rapp. P.-v. Reun. Cons. int. Explor. Mer 178: 248–259.

## Appendix A: Alternative Choices for the Underlying Distributional Model

We briefly describe other aggregated distributional models that could be used instead of the Negative Binomial model chosen for the main part of the text. These models all have the property that, depending upon parametrization, the probability of a zero is very high and the variance may greatly exceed the mean.

For the Neyman type A distribution (Johnson and Kotz 1969)

$$Pr(X_i = x| \text{ station is a habitat}) =$$

(A.1) $\Sigma_{j=1} [e^{-\lambda}\lambda^j/j!][e^{-j\theta} (j\theta)^x/x!]$

with the properties that

$$Pr\{X_i = 0| \text{ station is a habitat}\} = \exp(-\lambda (1 - e^{-\theta}))$$
(A.2) $E\{X_i| \text{ station is a habitat}\} = \lambda\theta$
$$Var\{X_i| \text{ station is a habitat}\} = \lambda\theta(1 + \theta)$$

For the Polya–Eggenberger distribution (Johnson and Kotz 1969)

$$Pr\{X_i = x| \text{ station is a habitat}\} =$$

(A.3) $\dfrac{\binom{n}{x} \prod\limits_{j=0}^{x-1} (P + j\alpha) \prod\limits_{j=0}^{n-x-1} (Q + j\alpha)}{\prod\limits_{j=0}^{n-1} (1 + j\alpha)}$

In this equation, $P$, $n$, and $\alpha$ are parameters and $Q = 1 - P$. For this distribution

$$E\{X_i| \text{ station is a habitat}\} = nP$$
(A.4) $Var\{X_i| \text{ station is a habitat}\} = nPQ(1 + n\alpha)/(1 + \alpha)$

In the limit that $n \to \infty$, $P \to 0$, $\alpha \to 0$ such that $nP$ and $n\alpha$ are bounded and nonzero, the Polya distribution is approximated by the NB. Johnson and Kotz (1969) also note that the Polya distribution is obtained as the unconditional distribution when a variable is conditionally binomial and the parameter $p$ of the binomial distribution has a beta density.

A third model is the "zero/random" model. Suppose that a station has a positive number of eggs with probability $p$ and that if a station has a positive number of eggs, then the egg number is distributed according to the positive terms of a Poisson process. Then

$$Pr\{X_i = 0| \text{ station is a habitat}\} = 1 - p$$
(A.5) $Pr\{X_i = x > 0| \text{ station is a habitat}\} =$
$$[p/e^\lambda - 1] [\lambda^x/x!]$$

The second of these equations arises in the following manner: $X_i$ is positive with probability $p$. If it is, then its distribution is Poisson, excluding the zero term. This leads to the $\lambda^x/x!$. The normalization constant is then $\Sigma_{x=1} \lambda^x/x! = e^\lambda - 1$. The first two moments of this distribution are

$$E\{X_i| \text{ station is a habitat}\} = p\lambda e^\lambda/(e^\lambda - 1)$$
(A.6) $E\{X_i^2| \text{ station is a habitat}\} = pe^\lambda(\lambda + \lambda^2)/(e^\lambda - 1)$