



Serial No. N1691

NAFO SCR Doc. 89/91

SCIENTIFIC COUNCIL - 1989

WORKING GROUP ON SHRIMP AGEING - OCTOBER 1989

An Application of the MULTIFAN Method for Estimating Growth Parameters and
Age Composition from Multiple Length Frequency Samples
to the Pink Shrimp (Pandalus jordani)

by

Jim Boutillier

Department of Fisheries and Oceans, Pacific Biological Station
Nanaimo, B. C., Canada V9R 5K6

and

David A. Fournier and John R. Sibert

Otter Software, P. O. Box 265
Nanaimo, B. C., Canada V9R 5K9

We present a method for simultaneously analyzing multiple length frequency data sets. The method utilizes a robust likelihood-based estimation procedure which provides an objective criterion for hypothesis testing. The ability to simultaneously analyze multiple samples permits the method to exploit the extra information not available when analyzing samples one by one. The computer program maintains a database of fits to the data which enable the user to organize the results of the analysis. Graphical displays permit the user to view any of the fits, and an interactive graphics routine aids the user in finding good initial parameter estimates. The method is applied to length frequency data from the pink shrimp *Pandalus jordani*. The best fit to a collection of fifteen length frequency data sets spanning a three year period is presented. Several alternative fits to the data and their implications for fisheries management are explored. It is concluded that policy formulation is insensitive to the number of significant age classes assumed to be present in the data.

The MULTIFAN system of length frequency analysis was introduced in Fournier et al (in press) where it was applied to length frequency data from southern bluefin data (*Thunnus macoyii*). Southern bluefin tuna are a slow growing relatively long-lived species. MULTIFAN is a robust likelihood based method which simultaneously analyzes multiple length frequency data sets. The extra information made available by simultaneously analyzing multiple length frequency data sets enables the model to perform hypothesis tests to distinguish between different models for fitting the data. In particular it is possible to estimate the number of significant age classes which are present in the data. To illustrate the flexibility of MULTIFAN it is interesting to apply it to a fast growing, shorter-lived species with a very different life history from southern bluefin tuna.

At the Pacific Biological Station in British Columbia Canada aging of shrimp involves analysis of length frequency data with no independent method of aging. Over the past years the length frequency analysis used was that describe by Schnute and Fournier (1980). This method basically involved determining the mean length at each age, the distribution of lengths and the distribution of ages. As is pointed out by Schnute and Fournier (1980) the principle motive for length-frequency analysis has been to determine the distribution of ages of animals which also gives information on the length at age. Using other methods of length-frequency analysis the distribution of ages is determined and then a growth curve is then constructed from this information as a separate problem. The Schnute and Fournier analysis introduced a single procedure which gives the percentage of animals at each age, mean length and standard deviation in length at each age of the stock and parameterizes the growth. Another way that this method differs from other methods is that instead of allowing arbitrary mean lengths and standard deviations, constraints are set such that the means and or standard deviations conform to a growth pattern. It was generally felt that more biological structure leads to less ambiguity. Unconstrained methods often lead to many possible solutions but by restricting population characteristics (such as mean lengths at age) so that they have some relationship to one another on biological grounds it is felt that one can perhaps eliminate a significant number of solutions.

In 1988 the MULTIFAN system of length frequency analysis was introduced. This system is described in Fournier et al (in press) where it is applied to length frequency data from southern bluefin tuna

data. The system evolved from the theory developed in Schnute and Fournier (1980) and Fournier and Breen (1983). with some significant differences. Some of these differences include: an integrated data analysis system for simultaneously analyzing multiple sets of fisheries length-frequency samples, 2) estimation of seasonal variation in growth rates, 3) correcting for bias due to selectivity, 4) estimation of total mortality, 5) a new robust likelihood estimation procedure developed by Fournier et al. (in press), and 6) the use of interactive graphics routines for obtaining initial parameter estimates and evaluating the results of the analysis.

Model Formulation

The following notation is introduced to describe the MULTIFAN model:

- i subscript indexing the length frequency intervals.
- j subscript indexing the age classes.
- α subscript indexing the length frequency data sets.
- N_j the number of age classes present in the data sets.
- N_A the number of length frequency data data sets.
- N_I the number of length intervals in each length frequency data set.
- S_α the number of animals in the α th length frequency data set.
- $f_{i\alpha}$ the number of animals whose lengths lie in the i th length interval in the α th length frequency data set.
- $p_{j\alpha}$ the proportion of age class j animals in the α th length frequency data set.
- $q_{ij\alpha}$ the probability that an age class j animals picked at random from the animals which were sampled to get the α th length frequency data set has a length lying in length interval i .
- $Q_{i\alpha}$ the probability that an animal picked at random from the animals which composed the α th length frequency data set has a length lying in length interval i .
- $\tilde{Q}_{i\alpha}$ the proportion of animals in the α th length frequency data set having a length lying in length interval i .
- $\mu_{j\alpha}$ the mean length of the age class j animals in the α th length frequency data set.
- $\sigma_{j\alpha}$ the standard deviation of the length distribution of the age class j animals in the α th length frequency data set.

- x_i the midpoint of the i th length frequency interval.
- w the width of the length frequency intervals.
- L_1 the mean length of the first age class on the von Bertalanffy curve in month 1.
- L_N the mean length of the last age class on the von Bertalanffy curve in month 1.
- K the von Bertalanffy K parameter.
- ρ the Brody growth coefficient ($K = -\log_e(\rho)$).
- λ_1, λ_2 parameters determining the standard deviations $\sigma_{j\alpha}$.
- τ_α parameters determining the overall variance of the sampling errors in the α th length frequency data set.
- $\xi_{i\alpha}$ parameters determining the relative variances of the sampling errors within the α th length frequency data set.
- δ parameter determining the variance of the sampling errors in all the length frequency data set.
- b_1 parameter determining the amount of size selectivity for the first age class.
- ϕ_1, ϕ_2 parameters determining the amount of the seasonal component of the growth curve.

Assume that the $Q_{i\alpha}$ depend on a set $\Theta = (\theta_1, \dots, \theta_r)$ of parameters. The exact form and number of the parameters depends on the specific model formulation which we do not wish to specify at this stage. We denote this dependence by writing $Q_{i\alpha}(\Theta)$.

If the length frequency data arose from random samples where the probability of each animal in the α th data set having a length lying in length interval i is $Q_{i\alpha}(\Theta)$, the probability of obtaining the observed length frequencies $f_{i\alpha}$ is equal to a constant times the expression

$$(1.0) \quad \prod_{\alpha=1}^{N_A} \prod_{i=1}^{N_I} Q_{i\alpha}(\Theta)^{f_{i\alpha}}$$

Robust Estimation

The method of maximum likelihood has long enjoyed a pre-eminent position among statistical estimation procedures. Recently it has been realized that, although maximum likelihood estimates have good properties when the model's assumptions are satisfied, their performance can degrade rapidly when departures from the model's assumptions occur. The property of a statistical estimator which permits it to behave well over a range of deviations from the statistical assumptions is called robustness.

In real applications, robustness considerations often are more important than the theoretical optimal properties of the conventional maximum likelihood estimates. Our goal in this paper is to retain the advantages of maximum likelihood estimation, but to incorporate into the model's structure the possibility of a few large deviations from the model's hypotheses and to construct a robust likelihood-based estimation scheme which can provide an adequate description of the data. In order to provide an adequate statistical model for performing hypothesis tests to analyze the length frequency data, Fournier et al. (in press) introduced a new robust likelihood function for length frequency analysis.

Two different types of deviations from a model's hypotheses can occur in length frequency analysis. A length may be recorded in a region where the model predicts that there is almost no probability of observing such an animal. An example is an animal whose length is much longer than that of any other animal. Such an observation may arise from an error in transcribing data or simply represent an unusual event. We shall call the occasional occurrence of an event of very low probability a type I deviation.

The other type of deviation occurs in a region where the probability of observing a fish length is high. In this case, the frequency observed in a length interval is either much higher or much lower than it should be considering the accuracy of the majority of the observations. This will be termed a type II deviation. Our goal is to make the model insensitive to a small number of type I or type II deviations.

If the $\tilde{Q}_{i\alpha}$ are derived from a random sample of size S_α , they would be random variables with means $Q_{i\alpha}$ and variances $(1 - Q_{i\alpha})Q_{i\alpha}/S_\alpha$. Set $\xi_{i\alpha} = (1 - Q_{i\alpha})Q_{i\alpha}$. The $\xi_{i\alpha}$ determine the relative variances of the $\tilde{Q}_{i\alpha}$ within a sample while the S_α can be regarded as a scale factor determining the overall size of all the variances of $\tilde{Q}_{i\alpha}$ in a sample.

As $Q_{i\alpha}$ tends to zero, the corresponding $\xi_{i\alpha}$ tends to zero as well. This reflects the fact that for random sampling an event with very small probability will almost never occur. For real world data, however, highly unlikely events seem to occur quite often. To render the model less sensitive to the occurrence of unlikely events, a lower bound is put on the relative variances of the $\tilde{Q}_{i\alpha}$. The scale factor S_α is replaced by a parameter τ_α to be estimated within the model.

We assume that the variance of the random variable $\tilde{Q}_{i\alpha}$ is equal to $(\xi_{i\alpha} + .1/N_I)\tau_\alpha^2$ where the τ_α are parameters to be estimated. The term $.1/N_I$ is

included so that the expression for the relative variance does not tend to zero as the corresponding $Q_{i\alpha}$ tends to zero. This renders the model insensitive to type I deviations.

If the $\tilde{Q}_{i\alpha}$ were independent, normally distributed random variables, their likelihood function would have the form

$$(1.1) \quad \prod_{\alpha=1}^{N_A} \prod_{i=1}^{N_I} \left[\frac{1}{\sqrt{2\pi(\xi_{i\alpha} + .1/N_I) \tau_\alpha}} \times \exp\left\{-\frac{(\tilde{Q}_{i\alpha} - Q_{i\alpha})^2}{2(\xi_{i\alpha} + .1/N_I) \tau_\alpha^2}\right\} \right]$$

This likelihood function, based on a normal distribution, is too sensitive to large deviations from the expected value $Q_{i\alpha}$. It lacks robustness. To make (1.1) more robust, we assume that there is a small uniformly distributed contamination of the normal distribution. This will render the model insensitive to type II deviations. The likelihood function becomes

$$(1.2) \quad \prod_{\alpha=1}^{N_A} \prod_{i=1}^{N_I} \left[\frac{1}{\sqrt{2\pi(\xi_{i\alpha} + .1/N_I) \tau_\alpha}} \times \left(\exp\left\{-\frac{(\tilde{Q}_{i\alpha} - Q_{i\alpha})^2}{2(\xi_{i\alpha} + .1/N_I) \tau_\alpha^2}\right\} + .01 \right) \right]$$

The reason for picking the value .01 in (1.2) is that for a normal distribution one usually begins to consider an observation to be a candidate for outlier status when it is about three standard deviations removed from the mean. For the moment let σ denote the standard deviations of a normally distributed random variable, and let $u = 3\sigma$. Then $\exp(-u^2/2\sigma^2) \approx 0.011$, so that the use of .01 in (1.2) ensures that the influence of observations on the parameter estimates decreases rapidly as their distance from the mean grows greater than three standard deviations. Taking the logarithm of (1.2) we obtain the log-likelihood function

$$(1.3) \quad -1/2 \sum_{\alpha=1}^{N_A} \sum_{i=1}^{N_I} \log_e(2\pi(\xi_{i\alpha} + .1/N_I)) - \sum_{\alpha=1}^{N_A} N_I \log_e(\tau_\alpha) + \sum_{\alpha=1}^{N_A} \sum_{i=1}^{N_I} \log_e \left[\exp\left\{-\frac{(\tilde{Q}_{i\alpha} - Q_{i\alpha})^2}{2(\xi_{i\alpha} + .1/N_I) \tau_\alpha^2}\right\} + .01 \right]$$

There are several possibilities for reparameterizing the τ_α which represent different assumptions about

the relative accuracies of the observed length frequencies composing the different length frequency data sets. In this paper we have assumed that all samples are equally accurate. Let δ be a model parameter which determines the overall size of the variances in the sampling errors. The assumption that all samples are equally accurate is implemented by setting $\tau_\alpha = \delta$ for all α .

Expression (1.3) forms the basis for the robust likelihood-based estimation scheme we have adopted, that is, the parameters are estimated by maximizing expression (1.3).

The Main Assumptions of the Model

At this point, there are far too many independent parameters in the model to obtain useful parameter estimates. Most of the model structure involves some biological assumptions which are made to reduce the number of independent parameters. The main structural assumptions are:

1. The lengths of the animals in each age class are normally distributed around their mean length (see 2.1).
2. The mean lengths at age lie on (or near) a von Bertalanffy growth curve (see 2.3).
3. The standard deviations of the actual lengths about the mean length at age are a simple function of the mean length at age (see (3.4)).

If the lengths of the age class j fish in the α th length frequency data set are normally distributed around their mean $\mu_{j\alpha}$ with standard deviations $\sigma_{j\alpha}$, the $q_{ij\alpha}$ can be expressed in terms of $\mu_{j\alpha}$ and $\sigma_{j\alpha}$ by

$$(2.1) \quad q_{ij\alpha}(\mu_{j\alpha}, \sigma_{j\alpha}) = \frac{1}{\sqrt{2\pi} \sigma_{j\alpha}} \int_{x_i-w/2}^{x_i+w/2} \exp\left\{-\frac{(x - \mu_{j\alpha})^2}{2\sigma_{j\alpha}^2}\right\} dx$$

As long as $\sigma_{j\alpha} > w$ the integral can be approximated sufficiently well by setting

$$(2.2) \quad q_{ij\alpha}(\mu_{j\alpha}, \sigma_{j\alpha}) = \frac{w}{\sqrt{2\pi} \sigma_{j\alpha}} \exp\left\{-\frac{(x_i - \mu_{j\alpha})^2}{2\sigma_{j\alpha}^2}\right\}$$

This approximation has been used in the model.

If the mean lengths $\mu_{j\alpha}$ lie on a von Bertalanffy curve then

$$(2.3) \quad \mu_{j\alpha} = L_1 + (L_{N_j} - L_1) \left(\frac{1 - \rho^{j+(m(\alpha)-1)/12}}{1 - \rho^{N_j}} \right)$$

where L_1 , the mean length of the first age class, L_{N_j} , the mean length of the last age class, and ρ ,

the Brody growth coefficient, are the three parameters which determine the form of the von Bertalanffy curve, and $m(\alpha) - 1$ is the number of months after the presumed birth month of the animals in the α th length frequency data set. This parameterization of the von Bertalanffy growth curve is from Schnute and Fournier (1980).

Including seasonal variation in growth

Seasonal growth is an important consideration for an animal like a shrimp because the animal can not grow continuously it is restricted to growth during periods of molt. As can be seen in the histograms of the prawn *Pandalus platyceros*, the females prawns begin carrying their eggs in August and September and will continue carrying these until hatching occurs in March. The version of the MULTIFAN model used in this paper allows for the estimation of a seasonal component of growth.

To include seasonal variation into the growth curve we have used the form proposed by Pauly and Gaschütz (1979).

$$(3.1) \quad \mu_{j\alpha} = L_1 + (L_{N_j} - L_1) \times \left(\frac{1 - \rho^{j+(m(\alpha)-1)/12 + \frac{\phi_1}{2} \sin(2\pi(m(\alpha)/12 - \phi_2))}}{1 - \rho^{N_j}} \right)$$

where $0 \leq \phi_1 \leq 1$ and $0 \leq \phi_2 < 1$.

The parameter ϕ_1 determines the magnitude of the seasonal effect. If $\phi_1 = 0$ there is no seasonal effect. If $\phi_1 = 1$ the animals actually stop growing at time $6 + 12\phi_2 \text{ mod } 12$ months after the month which the user has designated as month 1. (Note that mod12 means divide by 12 and take the remainder so that for example if $\phi_2 = 0.9$ then $6 + 12 \cdot 0.9 \text{ mod } 12 = 6 + 10.8 \text{ mod } 12 = 4.8$ months.)

The seasonally oscillating growth curve for the best fit to the data is shown in Figure 2.

Sampling bias

An additional complication in parameterizing the mean length at age is due to the fact that for some length frequency data sets, the sampling procedure or the fishery does not select the smallest animals in the first age class. The effect of this size selectivity is that the mean length of the animals in the first age class in the length frequency data set is larger than the mean length of the animals in the first age class in the population. If this sampling bias is not accounted for, biased parameter estimates may be produced.

With the integrated data analysis which allows simultaneous analysis of multiple sets of length-frequency samples, biases can also arise if the samples are not taken in a consistent manner, but the method can also lead to valuable insights which are not always evident when the samples are analyzed in isolation. As an example pink shrimp samples off the west coast of Vancouver Island are collected either from a synoptic research biomass survey or through sampling of the commercial fleet.

The length frequency histograms for the pinks shrimp data are shown in Figure 1. The mean lengths at age of the first age class are considerably shorter in samples 1, 7, 8, 9, 10, and 13 than they are for the other samples. Samples 1, 7, and 13 are research cruise samples where an effort was made to sample smaller animals. The other samples are derived from commercial samples where smaller shrimp are generally avoided. Although length frequencies 8, 9, and 10 were composed of commercial samples, they were taken in a year when there was great competition for shrimp. This could have led to smaller shrimp being included in the commercial catch in that year.

We assume that size selective bias only affects animals in the first age class in those length frequency samples for which it occurs, and that it decreases linearly with age until the shrimps reach the second age class. For those length frequency samples where size selectivity is assumed to occur the equation (2.3) becomes

$$(4.1) \quad \mu_{j\alpha} = L_1 + (L_{N_j} - L_1) \left(\frac{1 - \rho^{1+(m(\alpha)-1)/12}}{1 - \rho^{N_j}} \right) + b_1(12 - m(\alpha))/12$$

We have assumed that there is size selective bias operating in samples 2, 3, 4, 5, 6, 11, 12, 14, and 15.

Parameterizing the standard deviations

The standard deviations $\sigma_{j\alpha}$ are parameterized as a simple function of length involving two parameters λ_1 and λ_2 .

$$(5.1) \quad \sigma_{j\alpha} = \lambda_1 \exp \left\{ \lambda_2 \left[-1 + 2 \left(\frac{1 - \rho^{j+(m(\alpha)-1)/12}}{1 - \rho^{N_j}} \right) \right] \right\}$$

where the term enclosed in square brackets expresses the length dependency of the standard deviations independently of the numerical values of the parameters L_1 and L_{N_j} (cf. 2.3). The two coefficients, λ_1 and λ_2 , transform the rescaled length to

the standard deviations. λ_1 determines the magnitude of the standard deviations, and λ_2 determines the length-dependent trend in the standard deviations. If $\lambda_2 = 0$, the standard deviations are length-independent.

The parameters $p_{j\alpha}$, L_1 , L_{N_j} , K , λ_1 , λ_2 , e , and δ for $1 \leq j \leq N_j$ form the fundamental set of parameters in terms of which all the other parameters can be expressed and the log-likelihood function (1.3) can be calculated.

Note that the $Q_{i\alpha}$ satisfy the relationship

$$(5.2) \quad Q_{i\alpha} = \sum_{j=1}^{N_j} p_{j\alpha} q_{ij\alpha}$$

The steps in the calculation of the model parameters are:

1. Calculate the $\sigma_{j\alpha}$ from (5.1) using the fundamental parameters λ_1 , λ_2 , and K . (Recall that $\rho = \exp(-K)$).
2. Calculate the $\mu_{j\alpha}$ from (2.3) using L_1 , L_{N_j} , and ρ .
3. Calculate the $q_{ij\alpha}$ from (2.2) using $\mu_{j\alpha}$ and $\sigma_{j\alpha}$.
4. Calculate the $Q_{i\alpha}$ from (5.2) using $q_{ij\alpha}$ and $p_{j\alpha}$.

Estimating the Covariance Matrix of the Parameter Estimates

After choosing the model which best describes the data it remains to estimate how well the resulting parameter estimates are determined by the data. One method for doing this is to calculate an estimate for the covariance matrix of the parameter estimates.

Let $\hat{\Theta}$ denote the maximum likelihood estimates for the parameters Θ . As an approximation to the covariance matrix for the parameter estimates $\hat{\Theta}$ we have employed the inverse of the matrix of second derivatives of the log-likelihood function

$$(6.1) \quad \left[\frac{\partial^2}{\partial \theta_k \partial \theta_l} \left\{ -1/2 \sum_{\alpha=1}^{N_A} \sum_{i=1}^{N_I} \ln(2\pi(\xi_{i\alpha}(\hat{\Theta}) + .1/N_I)) \right. \right. \\ \left. \left. + \sum_{\alpha=1}^{N_A} \sum_{i=1}^{N_I} \ln \left(\exp \left\{ \frac{-(\tilde{Q}_{i\alpha} - Q_{i\alpha}(\hat{\Theta}))^2}{2(\xi_{i\alpha}(\hat{\Theta}) + .1/N_I)\tau_{\alpha}^2(\hat{\Theta})} \right\} + .01 \right) \right. \right. \\ \left. \left. - \sum_{\alpha=1}^{N_A} N_I \ln(\tau_{\alpha}(\hat{\Theta})) \right\} \right]^{-1}$$

It should always be kept in mind that these variance estimates are in general only accurate to an order of magnitude. They do, however, allow the user to verify that the minimization procedure has indeed found a local minimum of the log-likelihood

function. This is an important point. In a large model with many parameters it can be difficult to determine whether the putative minimum value of the parameters is indeed a minimum or whether the model has simply located a flat area of the log-likelihood function. If the matrix of second derivatives is positive definite, the point determined by the values of the parameters is located at a minimum rather than simply at a flat area.

An extensive discussion of the parameter search techniques used in the MULTIFAN model and the manner in which hypothesis testing is carried out is given in Fournier et al (in press).

The main unit of the search is a two dimensional grid search of the parameter space. The points in the grid are determined by the value of the von Bertalanffy K and the number of age classes assumed to be in the population. A systematic search is carried out for each model hypothesis, such as the hypothesis that there exists a length-dependent trend in the standard deviations of the distribution of the lengths at age or that there is a seasonal component to growth.

The accuracy of the parameter estimates which can be obtained by MULTIFAN depend in a complex fashion on many different characteristics of the data including between sample contrast in the proportions at age and modal separation. It is difficult to make any general statements about MULTIFAN's performance, but other factors being equal, the accuracy of the parameter estimates is very sensitive to the regularity of growth of the species being examined. If the mean lengths at age lie exactly on a von Bertalanffy curve, we have determined from an analysis of simulated data that it is often possible to obtain good parameter estimates without imposing constraints on the mean lengths. However, we have never encountered any actual fisheries data which exhibits this kind of regularity. For the pink shrimp data, while there are very well defined modes, the apparent mean lengths at age are somewhat irregular. It is important therefore that the user carefully inspect the data and impose bounds on the mean lengths of different age classes where appropriate.

The user should generally set quite wide bounds on the mean lengths. Fish growth is always somewhat irregular, and the mean lengths at age will not in general lie exactly on a von Bertalanffy growth curve. If the bounds are made too tight, the model may not be able to find any growth curve which satisfies the constraints. This will put the estimation scheme under considerable tension and distort the results obtained. Also, one irregular mode may exercise undue influence on the fit to the data. The

bounds are set sufficiently tight to insure that the correct age class is associated with a mode but not so tight as to require that the mean length at age lies exactly over a mode.

If the initial search for possible fits to the data does not cover a large enough region of parameter space, a local minimum of the objective function may be mistakenly accepted as the global minimum. To reduce the probability of incorrectly accepting a local minimum, the parameter space is partitioned into partially overlapping subregions and the log-likelihood function is minimized within each subregion. The variables used to set up such a systematic search are the number of age classes assumed to be present in the data sets and different user selected initial values for the von Bertalanffy K . This procedure also provides an objective method for determining the number of age classes presumed to be present in the data sets. A more detailed discussion of the systematic search estimates is presented in the description of the application to shrimp below.

Hypothesis testing

The problem of determining how many age classes are represented in the data illustrates the combined use of systematic searches and hypothesis tests in the MULTIFAN model. For each initial estimate of K , models with increasing numbers of age classes are fit. The increase in the maximum value of the log-likelihood function obtained by adding each age class is calculated. If the increase is large enough the extra age class is included in the model. This process is repeated until adding an extra age class doesn't lead to a significant increase in the maximum value of the log-likelihood function.

A χ^2 test is used to determine what constitutes a significant increase in the maximum value of the log-likelihood function. We have employed the standard theory which states that under the hypothesis that the present model is the correct one (and the additional age class is superfluous), if r parameters are added to the model, then twice the increase in the maximum value of the log-likelihood function is asymptotically distributed as a χ^2 random variable with r degrees of freedom.

Two types of errors can be made when making a decision based on a significance test. We could accept an extra age class into the model description when the present number of age classes is correct (type 1 error) or we could reject the extra age class when it is actually present in the mixture (type 2 error).

Rosenberg and Beddington (1987) have investigated the performance of Sparre's model using simulated data. They reported that better estimates of the von Bertalanffy K parameter were obtained when the number of age classes in the mixture was overestimated rather than underestimated. Therefore, when determining the number of age classes, a type 2 error is more serious than a type 1 error. To reduce the probability of occurrence of a type 2 error, we have employed the 0.90 point of the χ^2 random variable rather than the more conventional 0.95 point.

We do not recommend using the 0.90 point for testing the significance of the inclusion of other parameters such as the parameter for length-dependent standard deviations into the model. We have used the 0.95 level for all other parameters.

Setting up the Initial Systematic Search

The following discussion contains many references to the parameter estimates which are obtained by fitting the model with many different hypotheses to the data. For convenience we shall refer to the parameter estimates and associated log-likelihood function value obtained when fitting the model associated with a particular hypothesis to the data as a "fit".

Some of the most important user inputs to the MULTIFAN model occur at the initial stages of the analysis. Decisions, based on biological considerations, are made which affect the whole subsequent estimation process. In particular it is important to find the month in which the first age class enters into the data and number it month 1. The importance of this renumbering is discussed in Fournier et al (in press).

Another important user input involves setting constraints on some of the the mean lengths to insure that the model fits the obvious modes properly.

Even when the mean lengths at age are (correctly) constrained, the model can still fit the data incorrectly. It does this by estimating the proportion-at-age corresponding to a constrained mean to be very small and fitting that portion of the data with another age class. To avoid this problem, lower bounds can be placed on the proportions-at-age for the constrained age classes.

The horizontal lines over the modes in Figure 1 represent the constraints which have been placed on the mean lengths of the corresponding age class. these constraints were used in the first four length frequency data sets to "tell" the model to follow the progression

of the first and second modes in the data.

Results of the analysis

Fournier et al (in press) discuss in detail the procedures used for hypothesis testing in the model. We shall not repeat this discussion here, but note that the hypotheses tested were:

1. The existence of a length-dependent trend in the standard deviations of the distribution of the lengths-at-age. ($\lambda_2 \neq 0$)
2. The existence of a seasonal component to growth. ($\phi_1 \neq 0$)
3. The existence of size selective bias in the first age class. ($b_1 \neq 0$)
and as always
4. The number of significant age classes present in the data.

the results of the hypothesis tests were:

1. No evidence of a length-dependent trend in the standard deviations of the distribution of the lengths-at-age.
2. Evidence of a seasonal component to growth.
3. Evidence of a size selective bias in the first age class.
4. The number of significant age classes present in the data is 6.

The report of the parameter estimates of the "best" (6 age class) fit to the pink shrimp data is given in Table 1a in the form in which it is produced by the MULTIFAN program. For comparison the best fits for 5 and 4 age classes are also included in Tables 1b and 1c.

Table 1a Parameter estimates for the 6 age class fit to the shrimp data.

MultiFan v2.20.a Length-Frequency Analyzer Copyright (c) 1989 Otter Software

Fit: B1

Objective function value = 5531.74121; total penalty = 4.15326

Maximum gradient component = 0.01472

Number of non-empty length intervals: 309; Number of estimated parameters: 82

Approximate number of degrees of freedom: 227

Number of age classes: 6

Parameter Estimates:

von Bertalanffy K = 0.554 (1/year); L infinity = 22.8

First Length = 12.414; Last Length = 22.176; Brody rho = 0.575 (1/year).

Estimated age of the first age class = 1.42 years.

First month mean length sampling bias for first age class: 2.126

Mean length at age in month 1:

12.33 16.79 19.36 20.83 21.68 22.17

Standard Deviations of length at age in month 1:

0.85 0.85 0.85 0.85 0.85 0.85

Average Standard Deviation = 0.851; ratio of first to last S.D. = 1.000

Seasonal Growth Amplitude = 0.787; Seasonal Growth Phase = 0.102

Instantaneous Total Mortality Rate = 0.53 (1/year).

Selectivity Coefficients:

0.18 1.00 1.00 1.00 1.00 1.00

Table 1b Parameter estimates for the 5 age class fit to the shrimp data.

MultiFan v2.20.a Length-Frequency Analyzer Copyright (c) 1989 Otter Software
Fit: AC
Objective function value = 5503.62598; total penalty = 3.72740
Maximum gradient component = 0.00667
Number of non-empty length intervals: 309; Number of estimated parameters: 67
Approximate number of degrees of freedom: 242
Number of age classes: 5
Parameter Estimates:
von Bertalanffy K = 0.560 (1/year); L infinity = 23.4
First Length = 12.247; Last Length = 22.168; Brody rho = 0.571 (1/year).
Estimated age of the first age class = 1.33 years.
First month mean length sampling bias for first age class: 2.191
Mean length at age in month 1:
12.13 16.94 19.69 21.26 22.16
Standard Deviations of length at age in month 1:
0.88 0.88 0.88 0.88 0.88
Average Standard Deviation = 0.880; ratio of first to last S.D.= 1.000
Seasonal Growth Amplitude = 0.719; Seasonal Growth Phase = 0.110
Instantaneous Total Mortality Rate = 0.71 (1/year).
Selectivity Coefficients:
0.13 1.00 1.00 1.00 1.00

Table 1c Parameter estimates for the 4 age class fit to the shrimp data.

MultiFan v2.20.a Length-Frequency Analyzer Copyright (c) 1989 Otter Software
Fit: A5
Objective function value = 5475.91016; total penalty = 3.58871
Maximum gradient component = 0.01495
Number of non-empty length intervals: 309; Number of estimated parameters: 52
Approximate number of degrees of freedom: 257
Number of age classes: 4
Parameter Estimates:
von Bertalanffy K = 0.472 (1/year); L infinity = 24.9
First Length = 12.324; Last Length = 21.811; Brody rho = 0.624 (1/year).
Estimated age of the first age class = 1.45 years.
First month mean length sampling bias for first age class: 2.222
Mean length at age in month 1:
12.20 16.96 19.93 21.78
Standard Deviations of length at age in month 1:
0.91 0.91 0.91 0.91
Average Standard Deviation = 0.907; ratio of first to last S.D.= 1.000
Seasonal Growth Amplitude = 0.689; Seasonal Growth Phase = 0.113
Instantaneous Total Mortality Rate = 0.67 (1/year).
Selectivity Coefficients:
0.13 1.00 1.00 1.00

Table 2 Correlation matrix and estimated standard deviations of the parameters estimates for the 6 age class fit to the shrimp data.

MultiFan v2.20.a Length-Frequency Analyzer Copyright (c) 1989 Otter Software

Standard Deviations of Parameter Estimates:

Page 5

L[1]	L[2]	Rho	b1	S_Ampl	S_Phase	Ave SD	Age 1	L_inf	K	Z
0.0536	0.0536	0.0055	0.1026	0.0581	0.0111	0.0213	0.0048	0.0816	0.0096	0.0370

Correlation Coefficients between Parameter Estimates:

L[1]	L[2]	Rho	b1	S_Ampl	S_Phase	Ave SD	Age 1	L_inf	K	Z
L[2]	-0.076									
Rho	-0.000	1.000								
b1	0.654	0.476	0.527							
S_Ampl	0.186	-0.726	0.076	-0.224						
S_Phase	0.068	0.224	-0.126	0.289	-0.476					
Ave SD	-0.144	0.912	0.020	0.382	-0.676	0.127				
Age 1	-0.049	-0.083	-0.152	-0.421	-0.028	-0.020	-0.074			
L_inf	0.094	0.442	-0.710	0.222	-0.311	0.383	0.339	-0.163		
K	-0.593	-0.440	0.851	0.441	0.418	-0.104	-0.410	-0.216	-0.641	
Z	0.540	-0.476	-0.527	-1.000	0.224	-0.289	-0.382	0.421	-0.222	-0.441

As stated above, the covariance estimate for the parameter estimates (equation 6.1) should always be calculated to insure that the parameter estimates really do determine a maximum of the log-likelihood function. The estimated correlation matrix and the estimated standard deviation of the parameter estimates are shown in Table 2 in the form in which they are produced by MULTIFAN.

An interesting comparison can be made between the hypotheses that there are four, five and six age classes. Although the six age class fit to the data produces the "best" fit according to the significance tests, biological considerations suggest that there may be as few as four significant age classes in the data. From the point of view of scientific fisheries management the real question of interest is not how many significant age classes are present in the population, but rather whether the parameter estimates for the four and six age class fits to the data lead to significantly different management regimes.

Yield per recruit estimates were calculated for the best fit of the model for

each of the different age class categories in Table 3. If one looks at a annual instantaneous natural mortality value of .43 for four, five and six age classes, one sees that maximum yield per recruit values occur between relatively equal Fishing mortality values (between 1.6 and 1.8).

When one also looks at the relative proportions of the age classes in the different samples for best calculated fits of four, five and six age classes (Table 4), one sees that the proportions of the first two age classes are relatively stable. In British Columbia these are the major two year classes used in predicting the relative strength of the next year fishable shrimp stock off the west coast of Vancouver Island.

Since there are not any independent methods of age validation, the next step to pursue would be to combine the results in a catch at age model and see which if any results can provide a consistent explanation of the survey results and commercial sampling information.

Table 3a Yield per recruit table for 6 age class solution

MultiFan v2.20.a Length-Frequency Analyzer Copyright (c) 1989 Otter Software
 Yield Per Recruit Using estimated selectivities WL coff=0.0007 Page 4

Fishing	Annual Instantaneous Natural Mortality									
Mort.	0.08	0.16	0.25	0.34	0.43	0.51	0.60	0.69	0.77	0.86
0.03	0.686	0.508	0.382	0.292	0.227	0.179	0.142	0.115	0.094	0.078
0.14	2.389	1.806	1.386	1.080	0.853	0.683	0.554	0.454	0.377	0.315
0.26	3.264	2.515	1.967	1.560	1.253	1.019	0.838	0.697	0.584	0.495
0.37	3.684	2.892	2.302	1.855	1.513	1.248	1.039	0.874	0.741	0.633
0.48	3.860	3.082	2.492	2.038	1.685	1.407	1.185	1.006	0.861	0.741
0.59	3.906	3.167	2.598	2.153	1.801	1.519	1.292	1.107	0.954	0.828
0.71	3.884	3.194	2.653	2.224	1.879	1.601	1.373	1.185	1.029	0.899
0.82	3.830	3.188	2.678	2.267	1.934	1.661	1.435	1.247	1.090	0.957
0.93	3.762	3.165	2.684	2.293	1.971	1.705	1.483	1.297	1.139	1.005
1.05	3.689	3.134	2.680	2.307	1.997	1.739	1.521	1.337	1.180	1.046
1.16	3.617	3.098	2.670	2.313	2.015	1.764	1.551	1.370	1.215	1.081
1.27	3.548	3.061	2.655	2.315	2.027	1.784	1.576	1.397	1.244	1.111
1.38	3.484	3.024	2.638	2.312	2.035	1.798	1.595	1.420	1.269	1.137
1.50	3.423	2.988	2.620	2.307	2.040	1.810	1.611	1.439	1.290	1.160
1.61	3.367	2.954	2.602	2.301	2.042	1.818	1.624	1.455	1.308	1.179
1.72	3.314	2.921	2.583	2.293	2.042	1.824	1.635	1.469	1.324	1.197
1.83	3.265	2.889	2.565	2.284	2.041	1.829	1.644	1.481	1.338	1.213
1.95	3.220	2.859	2.547	2.275	2.039	1.832	1.651	1.491	1.351	1.227
2.06	3.177	2.830	2.529	2.266	2.036	1.834	1.656	1.500	1.362	1.239
2.17	3.137	2.803	2.511	2.256	2.032	1.835	1.661	1.507	1.371	1.250

Table 3b Yield per recruit table for 5 age class solution

MultiFan v2.20.a Length-Frequency Analyzer Copyright (c) 1989 Otter Software
 Yield Per Recruit Using estimated selectivities WL coff=0.0007 Page 4

Fishing	Annual Instantaneous Natural Mortality									
Mort.	0.08	0.17	0.26	0.34	0.43	0.52	0.61	0.70	0.78	0.87
0.03	0.596	0.457	0.354	0.278	0.220	0.176	0.142	0.116	0.095	0.079
0.15	2.136	1.661	1.305	1.036	0.830	0.672	0.548	0.452	0.375	0.314
0.26	3.036	2.391	1.903	1.529	1.240	1.015	0.838	0.697	0.585	0.494
0.37	3.542	2.826	2.276	1.851	1.518	1.256	1.047	0.879	0.744	0.633
0.49	3.809	3.076	2.507	2.061	1.708	1.426	1.200	1.017	0.867	0.743
0.60	3.933	3.212	2.646	2.197	1.838	1.549	1.314	1.122	0.963	0.831
0.72	3.971	3.278	2.727	2.286	1.929	1.639	1.400	1.203	1.040	0.903
0.83	3.961	3.301	2.771	2.342	1.992	1.704	1.466	1.268	1.101	0.961
0.94	3.923	3.299	2.792	2.378	2.036	1.753	1.517	1.319	1.152	1.010
1.06	3.871	3.281	2.797	2.398	2.067	1.790	1.557	1.361	1.194	1.051
1.17	3.812	3.254	2.793	2.410	2.088	1.818	1.589	1.395	1.228	1.086
1.29	3.752	3.223	2.783	2.414	2.103	1.839	1.615	1.423	1.258	1.116
1.40	3.692	3.191	2.769	2.414	2.112	1.855	1.635	1.446	1.283	1.142
1.51	3.635	3.157	2.754	2.411	2.118	1.867	1.652	1.466	1.304	1.164
1.63	3.580	3.124	2.737	2.405	2.121	1.877	1.665	1.482	1.323	1.184
1.74	3.529	3.092	2.719	2.399	2.122	1.884	1.676	1.496	1.339	1.201
1.86	3.480	3.061	2.701	2.391	2.122	1.889	1.686	1.508	1.353	1.216
1.97	3.434	3.031	2.684	2.382	2.120	1.892	1.693	1.518	1.365	1.230
2.08	3.392	3.003	2.666	2.373	2.118	1.895	1.699	1.527	1.376	1.242
2.20	3.351	2.976	2.649	2.364	2.115	1.896	1.704	1.535	1.385	1.253

Table 3c Yield per recruit table for 4 age class solution

MultiFan v2.20.a Length-Frequency Analyzer Copyright (c) 1989 Otter Software
 Yield Per Recruit Using estimated selectivities WL coff=0.0007 Page 4

Fishing Mort.	Annual Instantaneous Natural Mortality									
0.02	0.05	0.13	0.20	0.28	0.35	0.43	0.51	0.58	0.66	0.74
0.12	0.321	0.264	0.218	0.180	0.150	0.125	0.105	0.089	0.075	0.064
0.21	1.589	1.312	1.088	0.906	0.759	0.638	0.539	0.458	0.390	0.334
0.31	2.477	2.057	1.716	1.438	1.210	1.024	0.870	0.742	0.636	0.548
0.41	3.092	2.582	2.166	1.825	1.545	1.314	1.122	0.963	0.829	0.717
0.50	3.508	2.946	2.485	2.106	1.792	1.532	1.315	1.134	0.982	0.853
0.60	3.781	3.193	2.708	2.307	1.974	1.697	1.464	1.268	1.103	0.963
0.70	3.953	3.356	2.862	2.451	2.108	1.820	1.578	1.373	1.200	1.052
0.79	4.052	3.458	2.965	2.552	2.205	1.913	1.666	1.456	1.278	1.125
0.89	4.100	3.517	3.030	2.621	2.275	1.983	1.734	1.522	1.340	1.184
0.99	4.113	3.546	3.069	2.666	2.325	2.035	1.787	1.574	1.392	1.234
1.08	4.101	3.552	3.088	2.695	2.360	2.073	1.827	1.616	1.433	1.275
1.18	4.073	3.544	3.094	2.711	2.383	2.102	1.859	1.649	1.468	1.310
1.28	4.035	3.525	3.090	2.718	2.398	2.122	1.883	1.676	1.496	1.339
1.38	3.990	3.500	3.080	2.718	2.406	2.137	1.902	1.698	1.520	1.364
1.47	3.942	3.470	3.064	2.714	2.410	2.146	1.917	1.716	1.540	1.385
1.57	3.892	3.438	3.046	2.706	2.411	2.153	1.928	1.730	1.556	1.403
1.67	3.842	3.405	3.026	2.696	2.408	2.157	1.936	1.742	1.570	1.419
1.76	3.792	3.371	3.005	2.685	2.404	2.158	1.942	1.751	1.582	1.432
1.86	3.744	3.338	2.983	2.672	2.399	2.158	1.946	1.759	1.592	1.444
	3.697	3.305	2.961	2.659	2.392	2.157	1.949	1.765	1.601	1.455

Table 4a Estimated proportions at age for the 6 age class fit to the shrimp data.

MultiFan v2.20.a Length-Frequency Analyzer Copyright (c) 1989 Otter Software
 Proportions at Age: Page 2

Age	Sample														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.06	0.20	0.27	0.45	0.25	0.21	0.03	0.05	0.15	0.26	0.07	0.14	0.01	0.02	0.02
2	0.24	0.18	0.06	0.06	0.06	0.18	0.52	0.67	0.41	0.55	0.27	0.36	0.44	0.48	0.54
3	0.24	0.32	0.26	0.19	0.19	0.09	0.16	0.16	0.24	0.13	0.49	0.41	0.44	0.29	0.17
4	0.34	0.02	0.02	0.07	0.22	0.34	0.15	0.06	0.04	0.00	0.11	0.04	0.08	0.19	0.26
5	0.02	0.02	0.03	0.13	0.17	0.12	0.09	0.05	0.09	0.00	0.04	0.02	0.02	0.02	0.02
6	0.10	0.26	0.35	0.09	0.10	0.06	0.05	0.02	0.08	0.05	0.03	0.03	0.02	0.00	0.00

Table 4b Estimated proportions at age for the 5 age class fit to the shrimp data.

MultiFan v2.20.a Length-Frequency Analyzer Copyright (c) 1989 Otter Software
 Proportions at Age: Page 2

Age	Sample														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.06	0.19	0.27	0.45	0.25	0.22	0.03	0.05	0.15	0.29	0.06	0.13	0.01	0.02	0.03
2	0.26	0.19	0.09	0.07	0.08	0.17	0.54	0.69	0.46	0.59	0.62	0.53	0.47	0.49	0.49
3	0.38	0.33	0.26	0.22	0.27	0.22	0.18	0.15	0.21	0.06	0.18	0.29	0.46	0.38	0.33
4	0.18	0.07	0.05	0.16	0.32	0.37	0.18	0.08	0.09	0.01	0.13	0.04	0.04	0.10	0.15
5	0.11	0.21	0.32	0.09	0.07	0.02	0.06	0.03	0.09	0.05	0.01	0.01	0.02	0.01	0.00

Table 4c Estimated proportions at age for the 4 age class fit to the shrimp data.

MultiFan v2.20.a Length-Frequency Analyzer Copyright (c) 1989 Otter Software
Proportions at Age: Page 2

Age	Sample														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.06	0.19	0.28	0.45	0.25	0.23	0.03	0.05	0.15	0.29	0.07	0.13	0.01	0.02	0.03
2	0.26	0.23	0.10	0.09	0.10	0.17	0.55	0.71	0.49	0.61	0.69	0.62	0.48	0.51	0.50
3	0.47	0.34	0.26	0.30	0.47	0.47	0.23	0.17	0.22	0.04	0.20	0.24	0.47	0.42	0.42
4	0.21	0.24	0.35	0.16	0.18	0.13	0.19	0.07	0.14	0.05	0.04	0.01	0.04	0.04	0.05

References

- Fournier, D. A. and P. A. Breen. 1983. Estimation of Abalone mortality rates with growth analysis. *Trans. Amer. Fish. Soc.* 112:403-411.
- Fournier, D. A., J. R. Sibert, Jacek Majkowski, and John Hampton. MULTIFAN a likelihood-based method for estimating growth parameters and age composition from multiple length frequency data sets with an application to southern bluefin tuna (*Thunnus macoyii*). To appear in *Can. J. fish. Aquat. Sci.*
- Pauly, D. and G. Gaschütz. 1979. A simple method for fitting oscillating length growth data, with a program for pocket calculator. ICES CM 1979/G: 24. Demersal Fish Ctee., 26p.
- Rosenberg, A. A. and J. R. Beddington. 1987. Monte-Carlo Testing of two methods for estimating growth from length-frequency data with general conditions for their applicability, p. 283-298. In D. Pauly and G. R. Morgan (eds.) *Length-based methods in fisheries research. ICLARM conference Proceedings 13*, 468 p. International center for Living and Aquatic Resources Management, Manila, Philippines, and Kuwait Institute for Scientific Research, Safat, Kuwait.
- Schnute, J., and D. Fournier. 1980. A new approach to length frequency analysis: growth structure. *Journal of the Fisheries Research Board of Canada* 37:1337-1351.
- Sparre, P. 1987. A method for the estimation of growth, mortality, and gear selection/recruitment parameters from length frequency samples weighted by catch per effort, p. 75-102. In D. Pauly and G. R. Morgan (eds.) *Length-based methods in fisheries research. ICLARM conference Proceedings 13*, 468 p. International center for Living and Aquatic Resources Management, Manila, Philippines, and Kuwait Institute for Scientific Research, Safat, Kuwait.

Figure 1. Pinks shrimp 6 age class fit

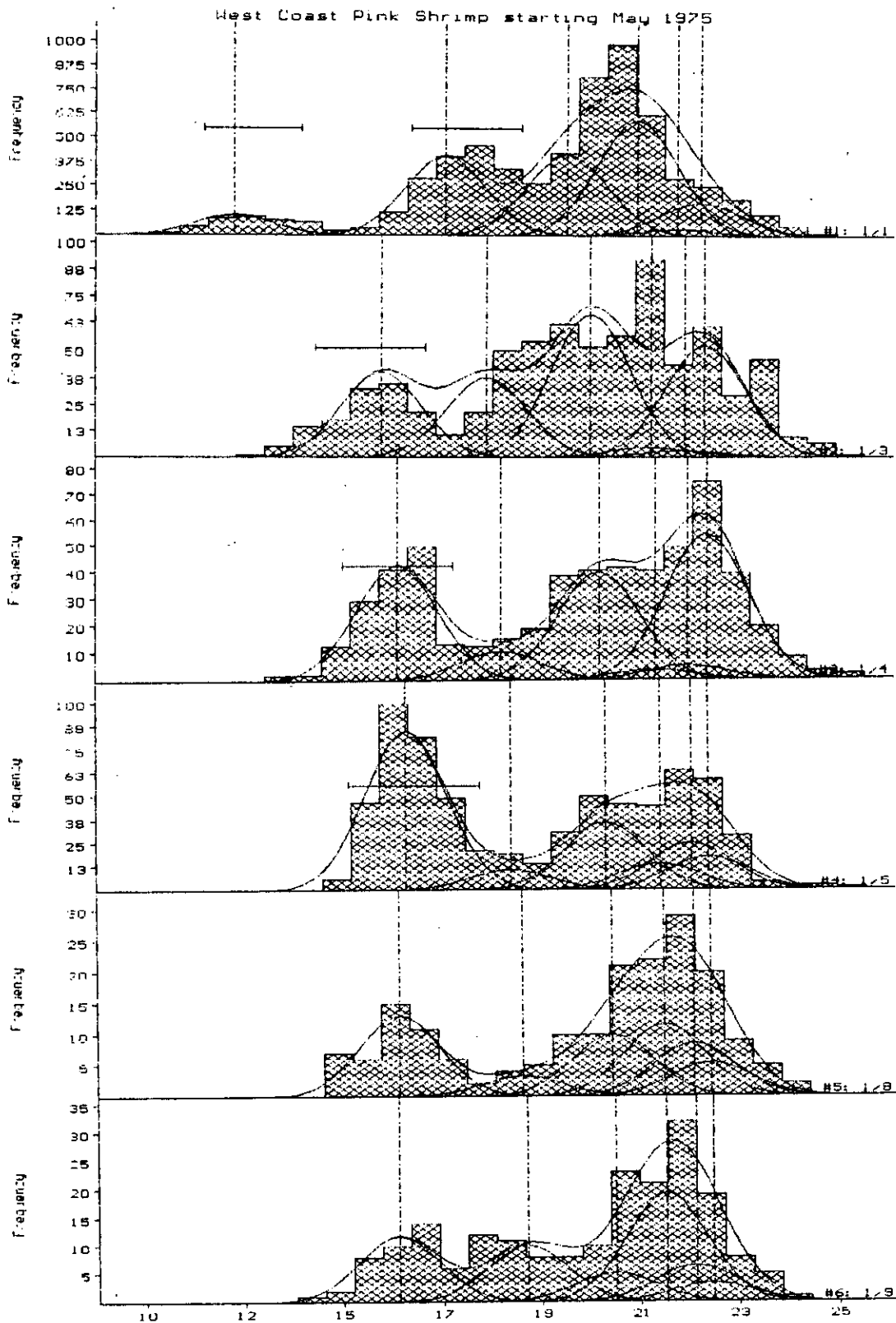


Figure 1. Pinks shrimp 6 age class fit continued

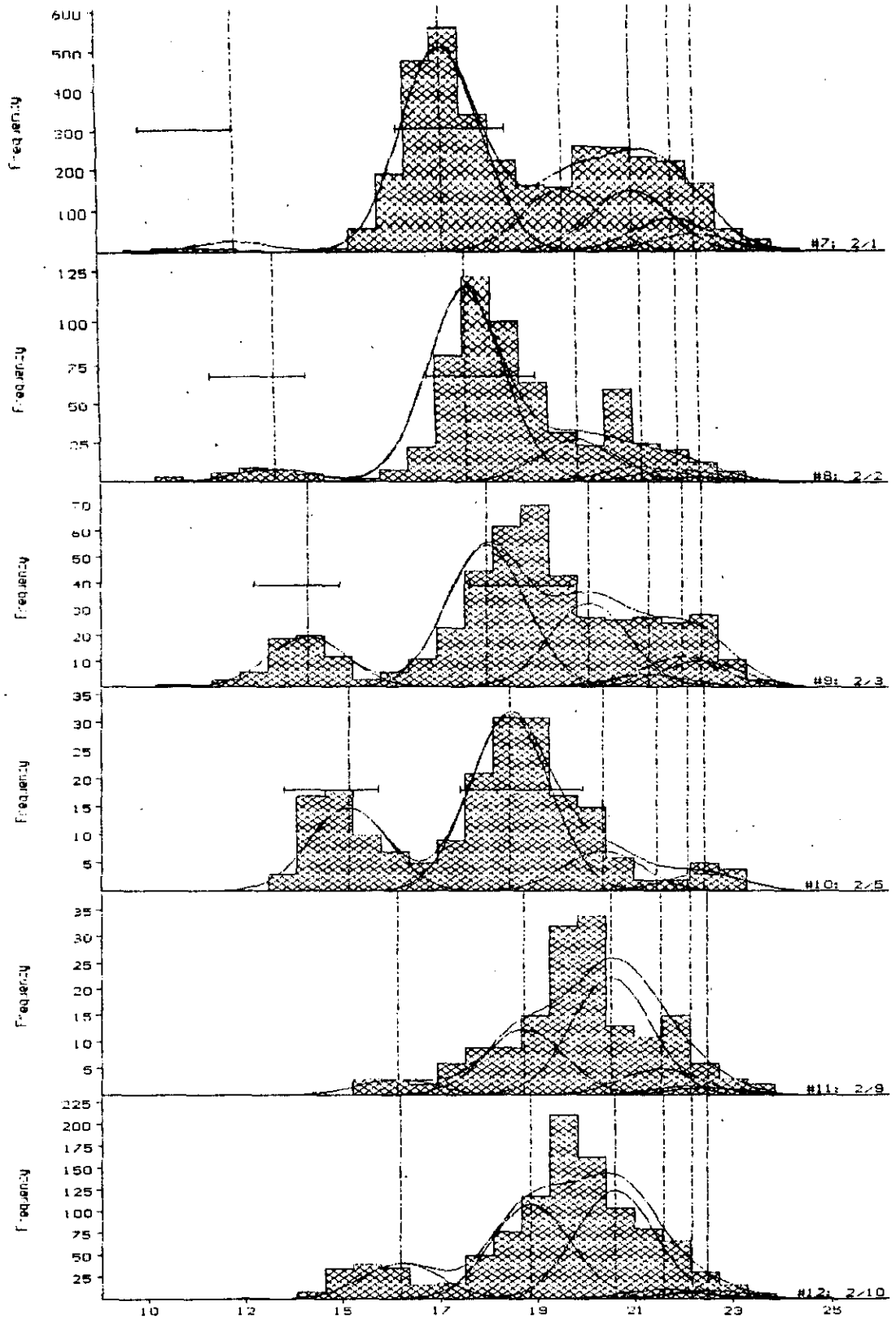


Figure 1. Pinks shrimp 6 age class fit continued

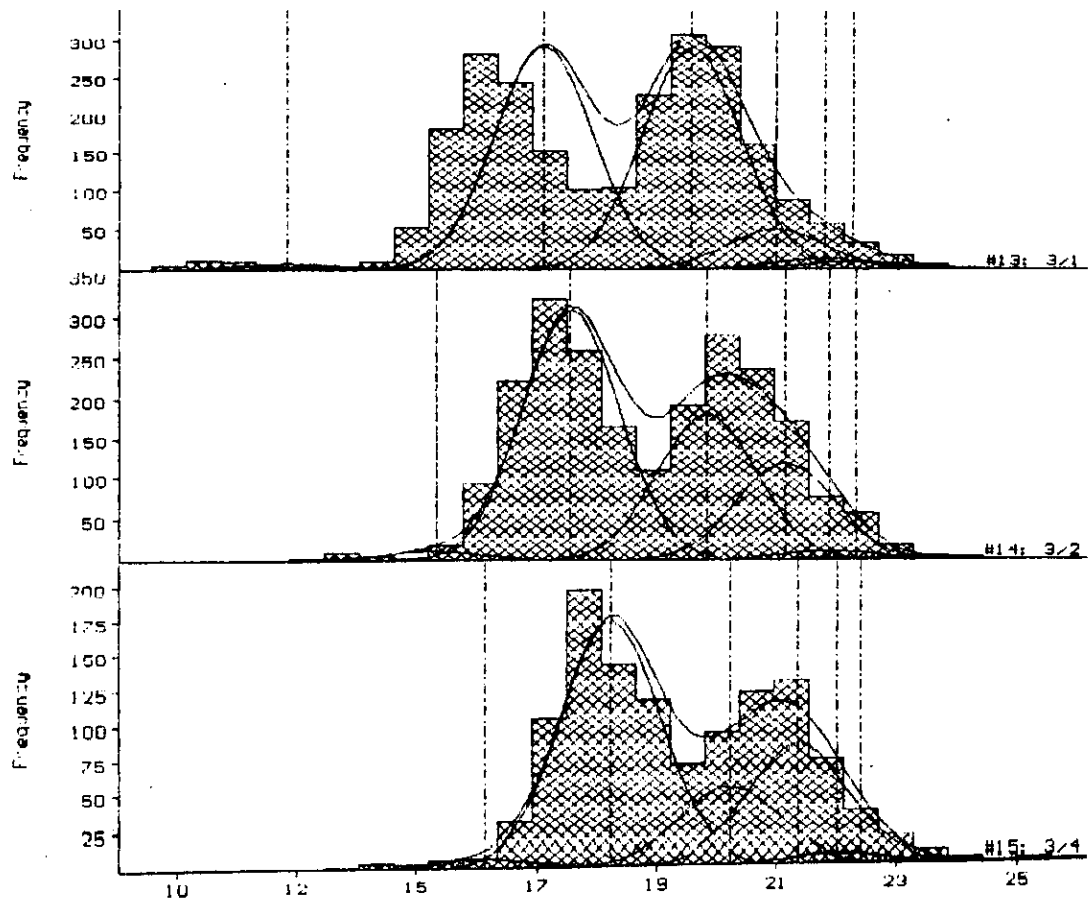
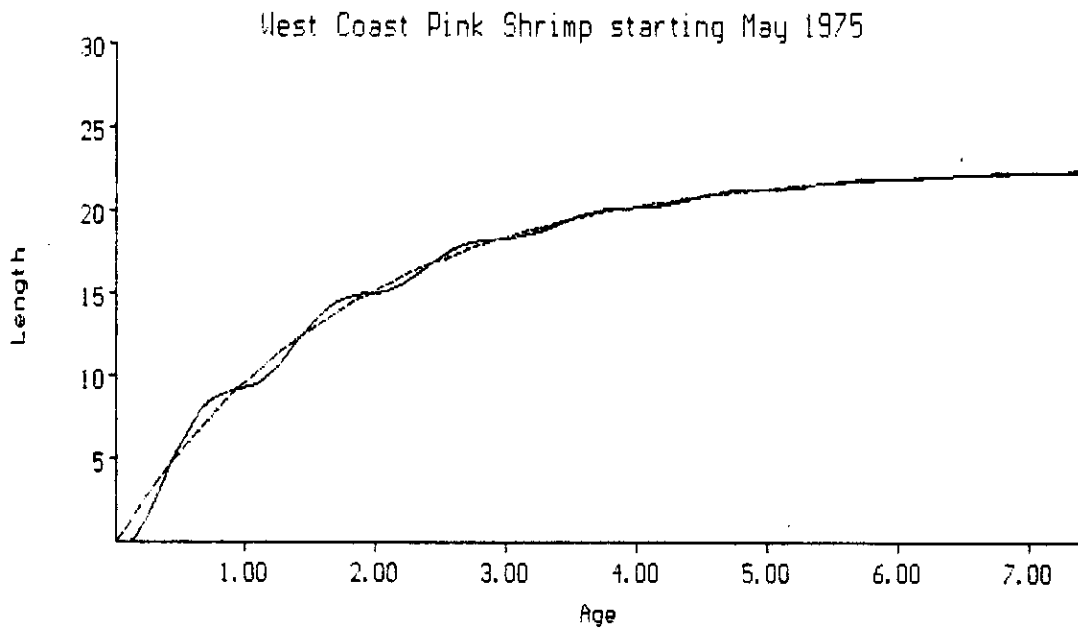


Figure 2. Seasonally oscillating growth curve for 6 age class fit



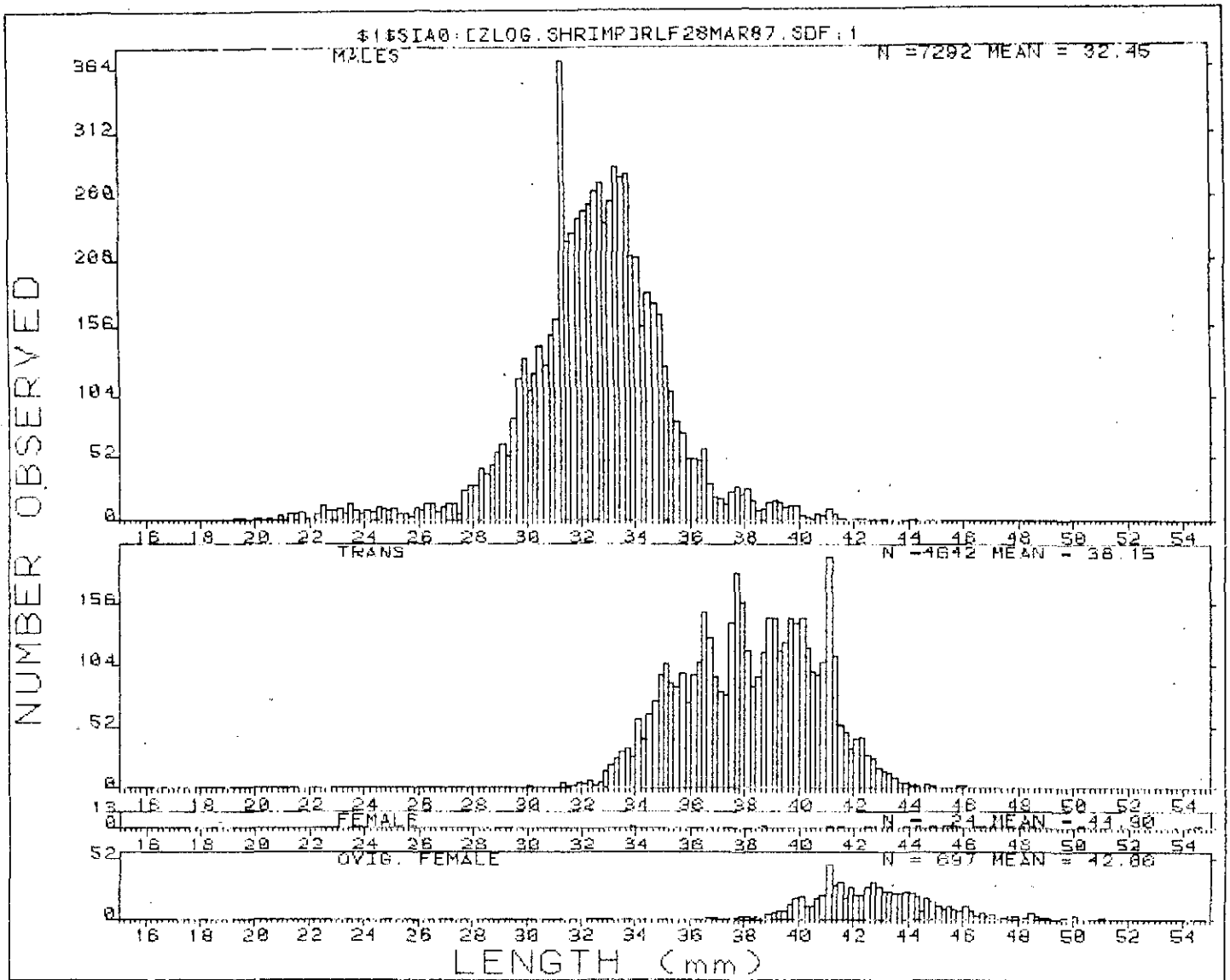


Fig 3. (a)

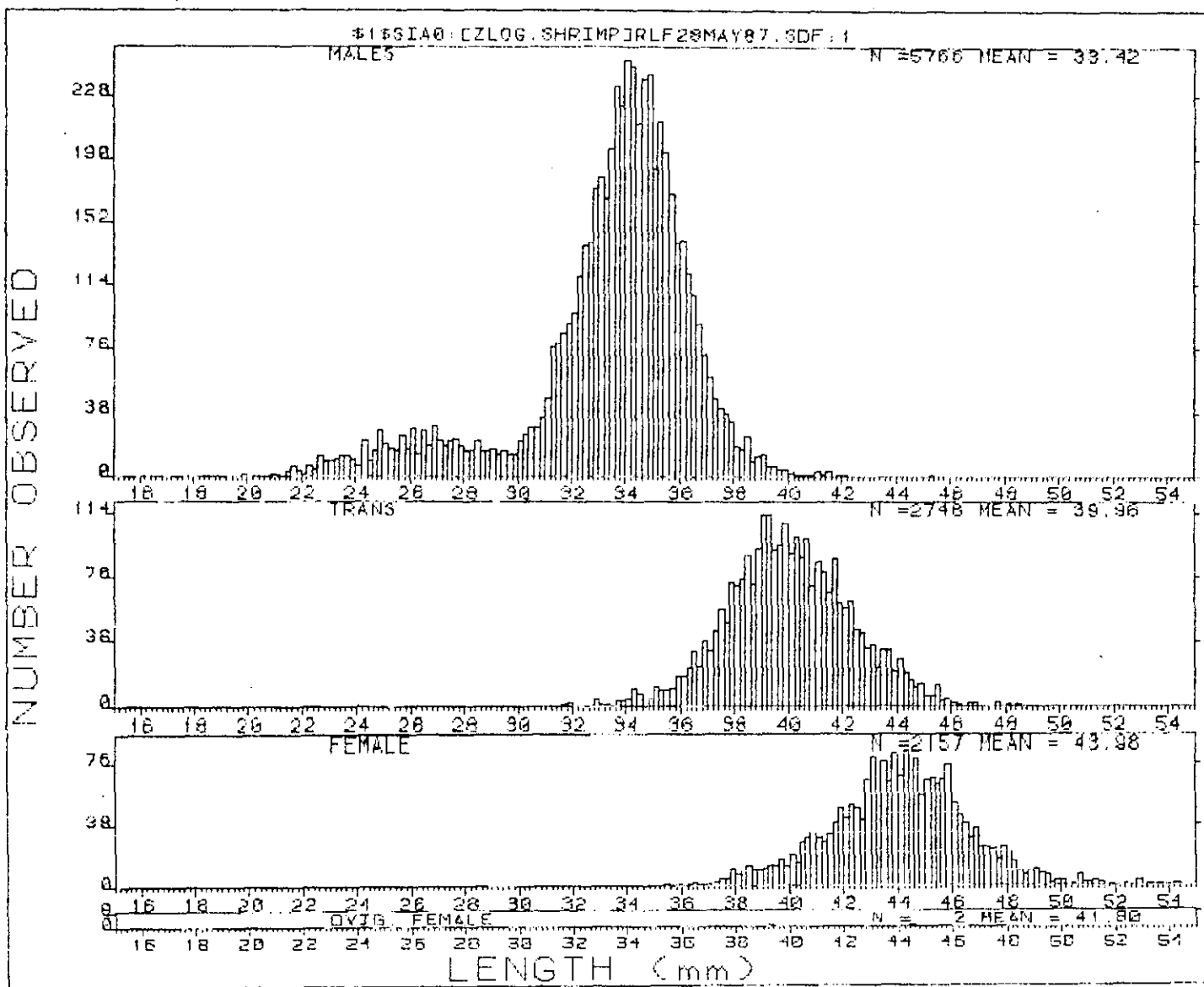


Fig 3 (B)

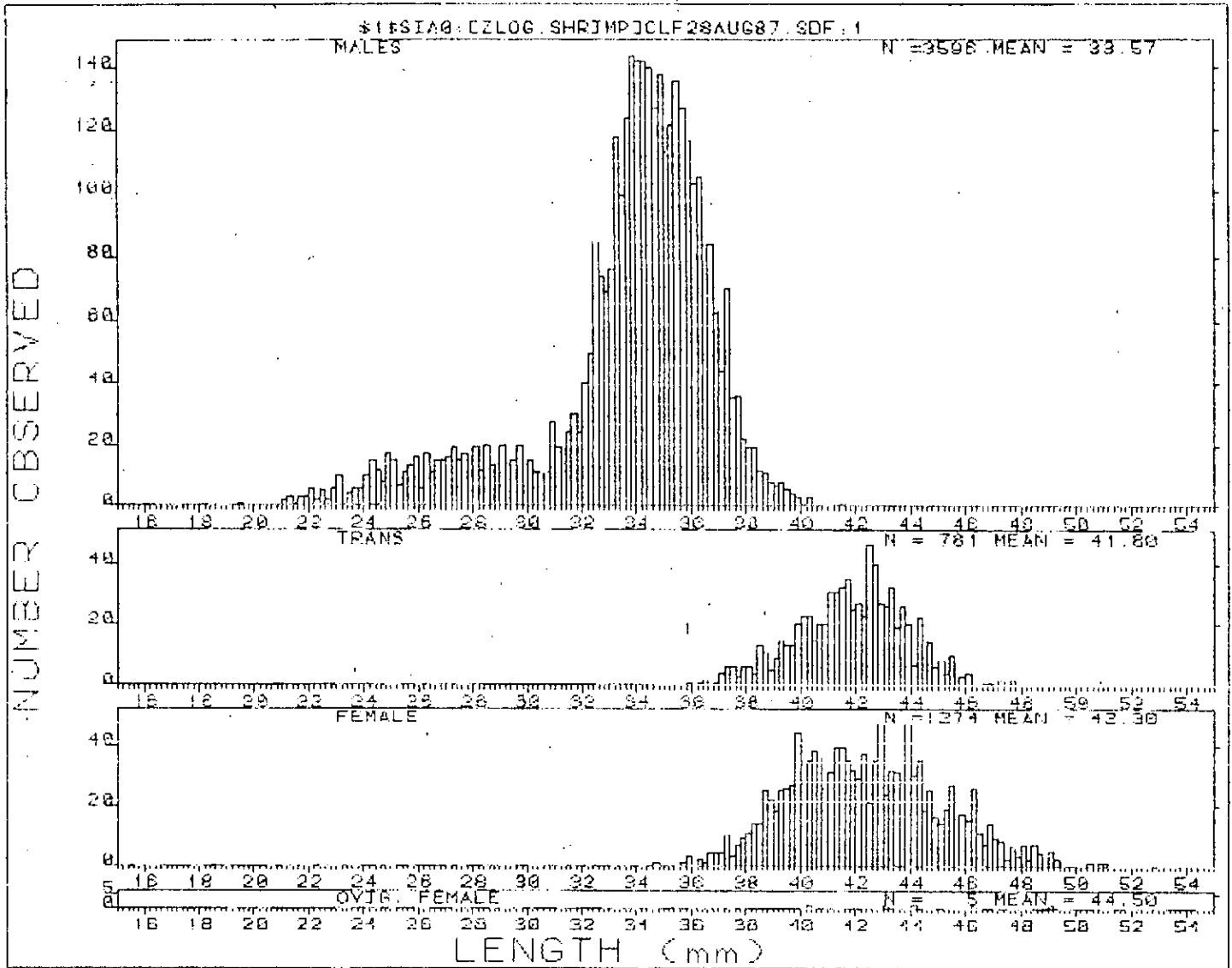


Fig 3 (c)

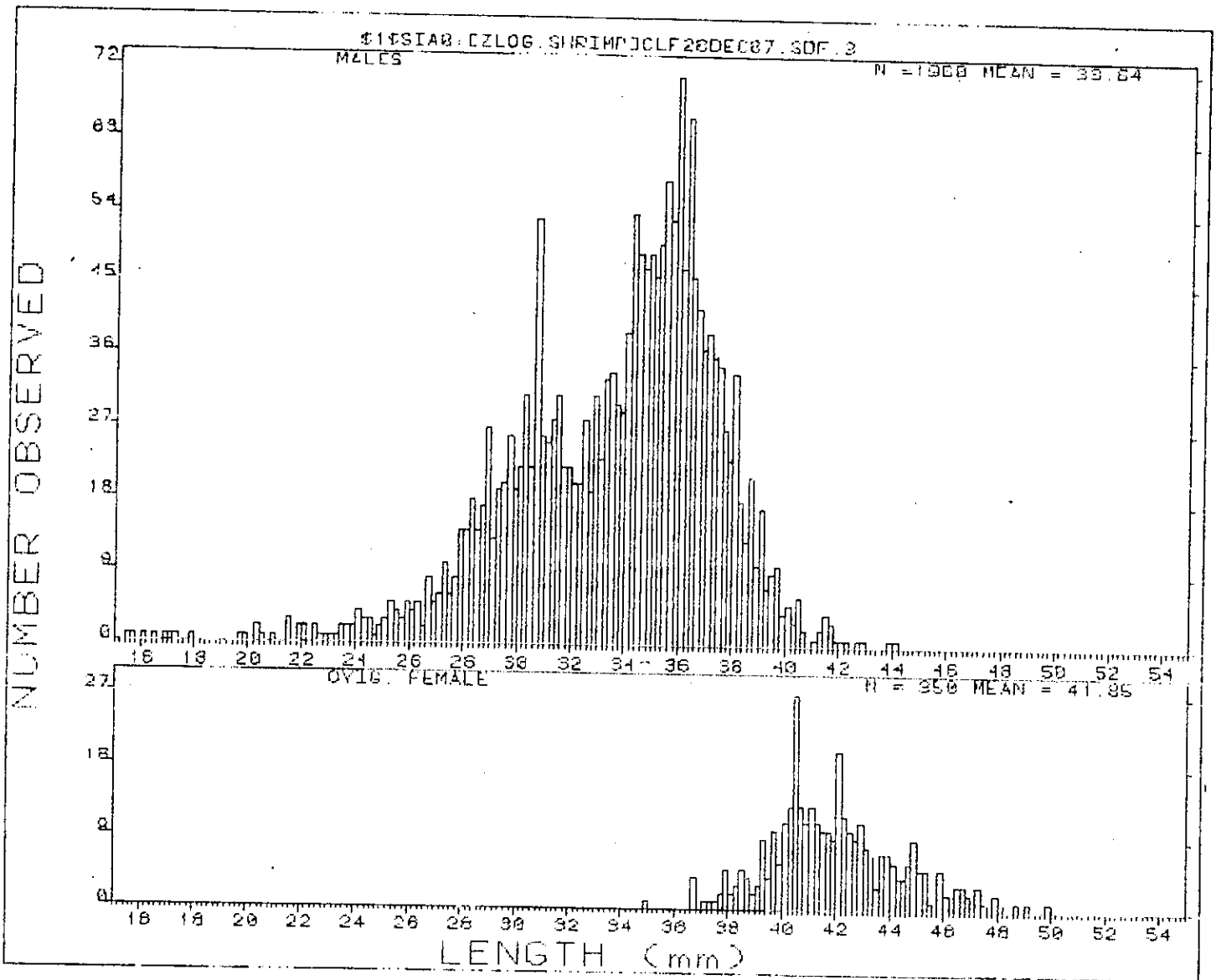


Fig 3 (d)

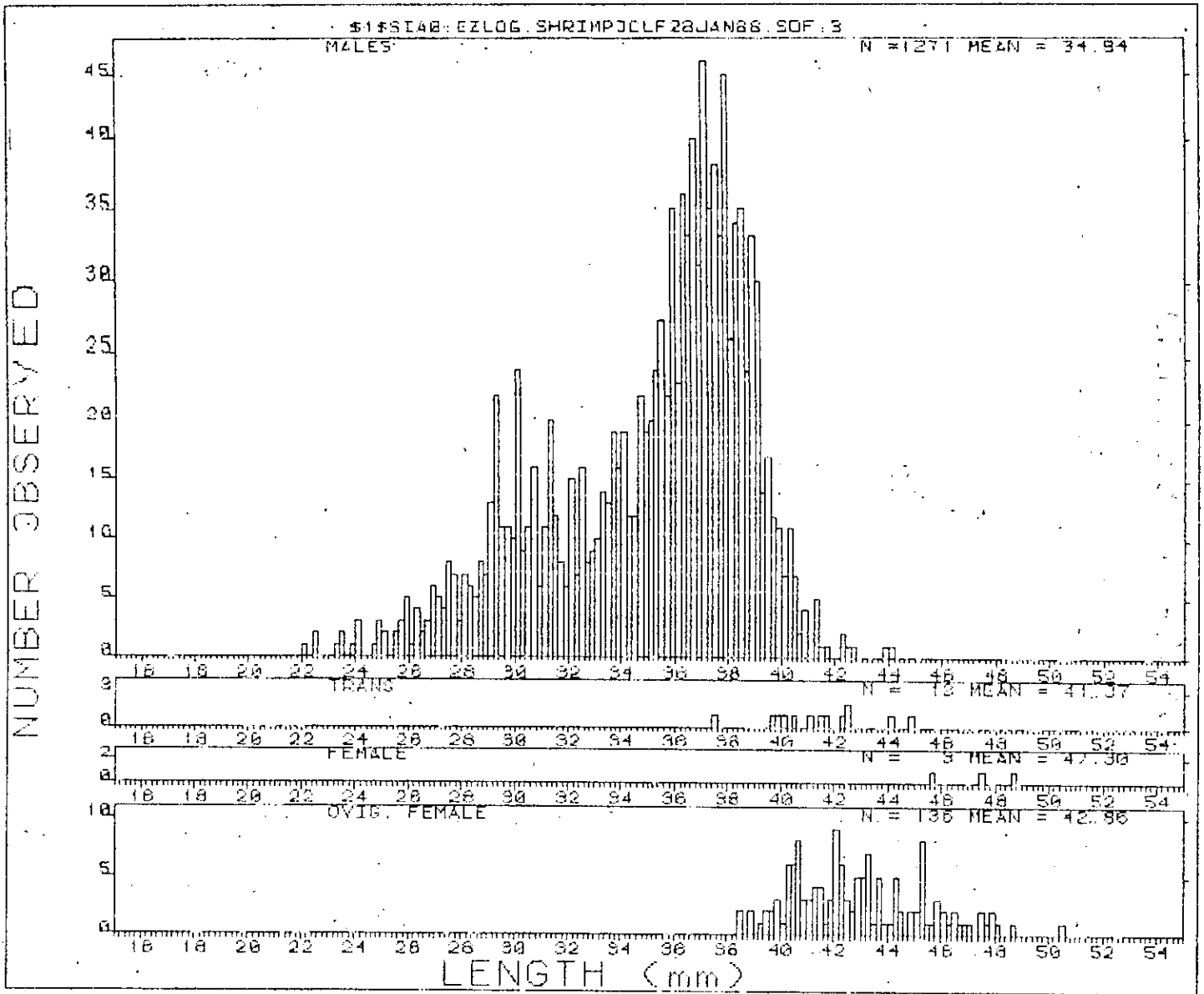


Fig 3 (e)

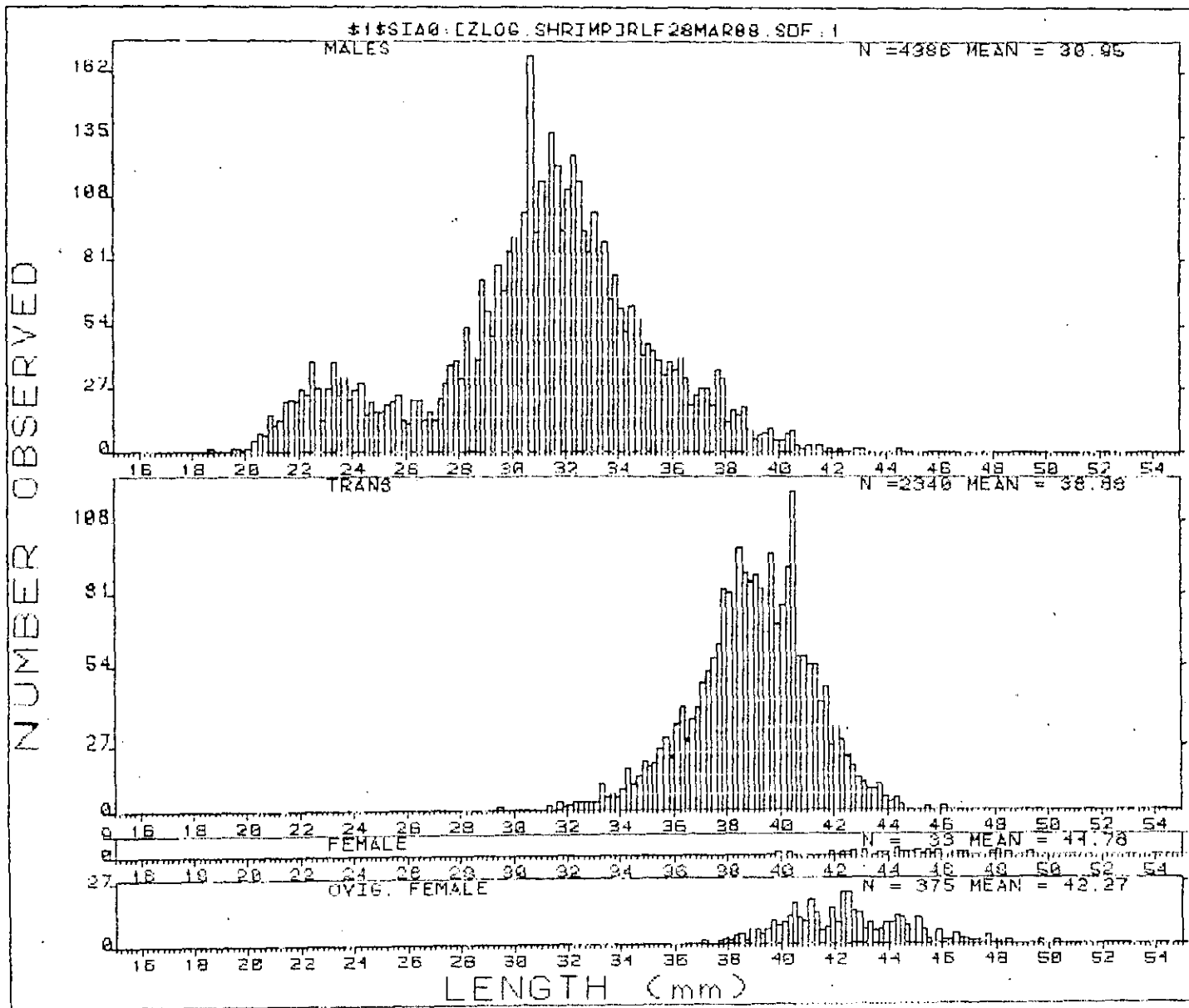


Fig 3 (f)