CONF-770501--2

Estimation of Age Structure of Fish Populations from Length-Frequency Data*

NOTICE
This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Energy Research and Development Administration, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

K. Deva Kumar and S. Marshall Adams

Environmental Sciences Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee

ABSTRACT

A probability model is presented to determine the age structure of a fish population from length-frequency data. It is shown that when the age-length key is available, maximum-likelihood estimates of the age structure can be obtained. When the key is not available, approximate estimates of the age structure can be obtained. The model is used for determination of the age structure of populations of channel catfish and white crappie. Practical applications of the model to impact assessment are discussed.

Key words: age-length key, age structure, channel catfish, length-frequency distribution, maximum likelihood estimates, mixture of normals, probability model, simulation study, white crappie

INTRODUCTION

Fish monitoring programs at nuclear power plants are designed to detect impacts and to ensure that these plants operate without causing unacceptable environmental damage. However, these monitoring programs, in general, are not comprehensive enough to detect impacts on fish populations due to power plant operations. Problems in the design of meaningful monitoring programs occur primarily because of the difficulties in sampling and interpreting monitoring data collected in open aquatic systems under highly variable environmental conditions. The inherent problems in sampling fish populations, such as size selectivity of sampling gear, mobility of fishes, and some of the behavioral characteristics of fish such as schooling, all tend to increase the effort and cost of an effective monitoring program.

Several approaches have been tried to evaluate the extent of changes in fish populations due to perturbations. One approach attempts to detect changes in density but is very difficult to implement due to inherent problems in quantitative sampling. Tag and release methods seem best for

*Research sponsored in part by the Energy Research and Development Administration under contract with Union Carbide Corporation and in part by the Nuclear Regulatory Commission Office of Nuclear Reactor Regulation under Interagency Agreement ERDA 40-554-75. ESD Publication No. 1040, Environmental Sciences Division, Oak Ridge National Laboratory.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

estimating the size of a fish population, but these programs are expensive and time consuming. Methods currently employed to monitor potential impacts of power plant operations on the density of fish populations include comparing the catch per unit effort between control and discharge areas and between preoperational and operational periods. However, interpretation of catch-per-unit-effort data and extrapolation to effects on a population are difficult. For example, higher catches of fish in a thermal area compared to catches in a control area or variations in catches from year to year may mean either that some redistribution of the population has occurred, or that a change in population structure has occurred, or both. For meaningful interpretation, this method requires obtaining estimates of the fish density of the entire aquatic system, which is difficult.

Another approach to monitoring fish population dynamics involves the species structure. The best example of this is the comparison of fish diversity indices and similarity coefficients between discharge and control areas or between preoperational and operational periods. It is difficult, however, to relate changes in these indices to changes in the functional aspects of populations such as population growth, production, mortality, and age structure. Functional parameters, particularly production, can be used to quantify the ecological success of a species and to provide a measure of stress in aquatic ecosystems (Chadwick 1976). Functional parameters for evaluation of potential perturbations are also desirable because they reflect the dynamics of the entire system rather than of a single component. A change in density may be reflected in changes in mortality and growth rates, unless mortality and growth are density independent. If operation of a power plant is having a major impact on a fish population, then significant changes could possibly be expected in the mortality rate, growth rate, and production rate which in turn will be reflected in the age structure of the population.

In this paper we are interested in estimating the age structure of a fish population. Knowledge of the age structure is useful for the estimation of natural mortality, fishing mortality, and production (Ricker 1975, Seber 1973). The age-length key* is usually based on a subsample of the catch (Allen 1966), and the age distribution in the subsample is assumed to be an estimate of the age structure in the catch and, therefore, of the age structure in the population. The length-frequency distribution of a fish population is usually derived from either the total catch or a very large subsample. However, due to the cost and effort involved, the age-length key is based on a much smaller sample. If the subsample chosen for aging is "optimal" (based on some criteria), then the age-length key method will yield reliable estimates of the age structure.

In the following sections we discuss a probability model for the length of a fish which takes into consideration the age-structure of the population and

*A double frequency table, usually with age in the columns and lengths in the rows (Ricker 1975, p. 206).

is estimated from length-frequency data. If an estimate of the age-length key is available, reliable estimates of the age structure can be obtained, and even when this key is not available, the model gives a first approximation for the age distribution. While the Allen estimator (see above) for the age structure is a straight extrapolation from a subsample, this model attempts to use the larger sample associated with the length-frequency data to estimate the age distribution.

MODELS FOR FISH LENGTH

In this section we discuss a probability model for the length of fish. For a historical review of the model, see Ricker (1975).

Assume that there are k age classes in a fish population, and that ρ_i is the proportion of the population of age i. Let the lengths x of fish of age i be distributed normally with mean μ_i and variance σ_i^2. The probability density function $f(x\colon\ \Theta)$ of x for the population can be written as

$$f(x\colon\ \Theta) = \sum_{i=1}^{k} \rho_i N(x; \theta_i)\,. \tag{1}$$

where

$\Theta = \langle \theta_1, \theta_2, \ldots, \theta_k, \rho_1, \rho_2, \ldots, \rho_k \rangle$,

$\theta_i = \langle \mu_i, \sigma_i^2 \rangle$,

$N(x;\theta_i)$ = density function of a normal distribution with mean μ_i and variance σ_i^2,

ρ_i = proportion of the population of age i,

k = number of age classes in the population which is generally assumed to be known.

The parameter vector Θ is related to some well-known fishery statistics. For example, the catch curve* can be represented by the plot of $\log_e \rho_i$ against age i, and the plot of μ_i against age i is the growth curve. If the variances of the individual normal distributions are small, then the length-frequency plot will have k well-defined modes at $\mu_1, \mu_2, \ldots, \mu_k$. If there are differential growth rates for various year classes, the modes may not be well defined. MacDonald (1971) attempted to estimate modes from the length-frequency data for fast-growing pike (*Esocidae*), and he was able to identify

*A graph of the logarithm of the number of fish taken at successive ages or sizes (Ricker 1975, p. 2).

five separate age classes. For these data, the average distance between the mean lengths for successive age classes was 100 mm. Separations this distinct would not be expected for most smaller or slower-growing fish. In many cases small separations occur between age classes, especially between the middle and upper age class.

Age-length data for white crappie (*Pomoxis annularis*) presented in Table 1 show that distinct and large differences between age classes after age class 2 do not occur. For example, the mean lengths for age classes 2 and 3 differ by only 21 mm and the mean lengths for age classes 3 and 4 differ by only 28 mm. If the growth curve is modeled by the von Bertalanffy equation (Ricker 1975, p. 22), the average distance between the mean lengths of successive age classes decreases as age increases. As a result, the estimation of all the parameters in the vector Θ becomes very difficult.

Table 1. Age-length relationships for white crappie collected by trap nets in Conowingo Pond from 1966–1973[a]

Year of capture	Average length (mm) at ages 1–5				
	1	2	3	4	5
1966	117	188	220	247	278
1967	101	171	209	236	306
1968	114	183	209	230	240
1969	127	191	204	240	273
1970	116	173	196	221	229
1971	121	184	193	215	257
1972	112	168	173	202	252
1973	114	171	197	215	221
Mean	115	179	200	228	257

[a]Data from Mathur et al. (1975).

In Table 2, the average yearly increment in length for one-year-olds (two years old at the end of the given growing season) is presented for white crappie. Identification of the good and poor recruitment years is taken from Mathur et al. (1975). The average increase in length when recruitment was poor was about 75 mm and the average increase when recruitment was good was only 48 mm. This inverse relationship between growth rate and recruitment success is probably due to inter- and intra-age-class competition for the same food source in the pond. Young white crappie (age classes 0–3) in the pond consume mainly zooplankton (Euston 1976). When there is a large recruitment of young into the pond, individuals in age class 0 are competing for zooplankton among themselves and also with age classes 1 and 2. Consequently, during a good recruitment year, growth of these three age classes is slow, whereas during a poor recruitment year, growth of these age classes is greater due to reduced competition for food. Also, during high recruitment

Table 2. The effect of recruitment success on the growth rate of age class 1 white crappie in Conowingo Pond[a]

Year	Recruitment success	Length of age class 1, mm	Length of age class 2, mm	Growth increment,[b] mm
1966	Good	117	188	
1967	Poor	101	171	54
1968	Poor	114	183	82
1969	Good	127	191	77
1970	Poor	116	173	46
1971	Good	121	184	68
1972	Poor	112	168	47
1973	Poor	114	171	59
1974	Average	129	202	88
1975	Poor	86	175	46

[a]Data from Mathur et al. (1975).

[b]Calculated as length of age class 2 in year t minus length of age class 1 in year t minus length of age class 1 in year $t - 1$.

years for white crappie, the predator-prey ratio in this system probably decreases. Therefore, predation on young-of-the-year white crappie would be reduced and probably would not be intensive enough to ultimately reduce the competition for food. Some large predators such as walleye, largemouth bass, and smallmouth bass, are present in the system, but seemingly in small numbers (Mathur et al. 1975).

The previous discussion has indicated that there is not always a direct correspondence between the modes in length-frequency data and the ages present in the catch. Because the age-length key is based on a subsample, it would be, in general, more efficient to use the larger sample associated with the length-frequency data to estimate the age-structure. This can be accomplished by obtaining the mean lengths of the various age classes (μ_i's). Once the μ_i's are known, the reduced parameter vector $\Omega = \langle \sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2, p_1, p_2, \ldots, p_{k-1} \rangle$ can be estimated; that is, instead of estimating k means, k variances, and $k - 1$ proportions, we estimate k variances and $k - 1$ proportions.

Additionally, if the variances σ_i^2 can be estimated from the subsample employed for obtaining the age-length key, then the only unknowns are the $k - 1$ proportions. The estimate of Ω will be referred to as the conditional maximum likelihood (CML) estimate, and the estimate of Θ defined in Eq. (1) will be referred to as the maximum likelihood (ML) estimate. This method will be applied to some fish catch data.

ESTIMATION OF THE PARAMETERS OF THE MODEL

The estimation of the parameters of a mixture of two normal distributions (also referred to as a compound normal distribution) has been investigated by Cohen (1967), Pearson (1894), and Rao (1952). Pearson (1894) derived the moment estimator for $k = 2$ and Cohen (1966) found that for sample sizes less than 400 the moment estimators were not very reliable.

Hasselblad (1966) has investigated the properties of the maximum likelihood estimators for a mixture of k normal distributions. He reported that (1) as $d_{ij} = |\mu_i - \mu_j|$ decreases, the variances of the proportions p increase rapidly; (2) as d_{ij} decreases, the variances of the location (i.e., μ) and scale (i.e., σ) estimators increase; and (3) as p_i approaches 0, the variances of the location and scale estimates increase. Hasselblad suggested an iterative method for the maximum likelihood estimator. Dick and Bowden (1973) used the Newton iterative scheme for a mixture of two normal distributions.

MacDonald (1971) obtained estimates that minimized the Cramer–von Mises statistic* (Cramer 1974, p. 451), and Kumar and Nicklin (1976) obtained estimates that minimized the mean squared distance between the empirical and theoretical characteristic function. In this paper, we restrict ourselves to the maximum likelihood estimator. It is beyond the scope of this paper to compare the various estimators.

The likelihood function is nonlinear in the parameters, Θ. It was felt that when k is greater than 2, methods based on derivatives become inefficient and the amount of computer time required tends to render this method impractical. Described below is an alternative method for estimation of the parameters (E. H. Nicklin, General Foods Corporation, White Plains, New York, *personal communication*). The negative of the natural logarithm of the likelihood function L given by

$$L = -\sum_{i=1}^{n} \ln f(x_i; \Theta) \tag{2}$$

was minimized with respect to Θ (or Ω) using the derivative-free minimization technique developed by Nelder and Mead (1965). This method is called the simplicial triangulation method. The method involves the evaluation of the function at the vertices of an $m + 1$ dimensional moving simplex (m is the number of parameters to be estimated). Based on the functional values at the vertices, the simplex moves away from the maximum calculated value. This technique has been found to be efficient for most well-behaved surfaces.

In order to keep the search for optimal values in the feasible region, transformed parameters were estimated. The location parameters (i.e., μ_i)

*The Cramer–von Mises statistic is a goodness of fit criterion for testing distributional assumptions.

were restricted to the interval $[x_{(1)}, x_{(n)}]$, where $x_{(1)}$ and $x_{(n)}$ are the observed minimum and maximum values. The transformed means are given by

$$\alpha_i = \ln[\mu_i - x_{(1)}] - \ln[x_{(n)} - \mu_i] \,. \tag{3}$$

This transformation is called the logistic transformation. The scale parameters have to be nonnegative, and this was achieved by using the transformation

$$\lambda_i = \ln(\sigma_i^2/2) \,. \tag{4}$$

Finally, since the proportions p_i must sum to unity, we need to estimate only $k - 1$ proportions. However, these proportions must also satisfy the condition

$$\sum_{i=1}^{k-1} p_i \leq 1 - b \,, \tag{5}$$

where b is some small positive constant less than 1. The constant b is introduced to prevent the pathological case of one of the proportions going to 0 or 1, and it is usually chosen to be some function of the sample size n. Inequality (5) can be rewritten as

$$\begin{aligned} & b \leq p_1 \leq (1 - b) = l_1 \,, \\ & b \leq p_i \leq \left(1 - b - \sum_{j=1}^{i-1} p_j\right) = l_i \quad 2 \leq i \leq k - 1 \,. \end{aligned} \tag{6}$$

The transformed proportion is

$$\delta_i = \ln(p_i - b) - \ln(l_i - p_i) \quad 1 \leq i \leq k - 1 \,. \tag{7}$$

Note that the ordering of i is not important.

EMPIRICAL EVALUATION OF THE ESTIMATION TECHNIQUE

In the preceding section the problems involved in the estimation of the parameters were discussed. An empirical study of a hypothetical population was conducted to determine the advantages of knowing, a priori, the mean length for each age class. The growth of a fish was assumed to follow the Bertalanffy growth curve. The average lengths for ages 1–5 were assumed to be 117, 156, 190, 220, and 245 mm, respectively, and the proportions of each of the five age classes were set to 0.13, 0.40. 0.24, 0.144, and 0.086, respectively. These proportions were assigned by assuming that the annual survival rate between ages 2 and 5 was 0.60 per year and that age 2 was the first fully recruited age class and constituted 40% of the total fish population.

The estimates of the proportions in each age class were obtained when (a) none of the parameters were known (Θ), and (b) when mean lengths were assumed known (Ω). The common variance $\sigma^2 = \sigma_i^2$; $i = 1, 2, \ldots, 5$ was varied between 100 and 400.

A sample of 1000 fish was generated from the population described above. Five sets of expected order statistics* from a normal distribution (Harter, 1952) with sample sizes 130, 400, 240, 144, and 86 were obtained, and the parameter vectors Θ and Ω were estimated with the appropriate variance. These estimates represent the "expected" behavior of the estimates, in the sense that if the methodology does not work for this sample, it is highly doubtful that it will work with a random sample. The results are summarized in Table 3 and Fig. 1 for the parameter vector Θ [case (a)] and in Table 4 and Fig. 2 for the parameter vector Ω [case (b)].

For case (a), the estimated values of the elements of the parameter vector Θ show considerable variation from the "true" values. When the variance is 400, the proportion p_2 is estimated to be 14%, whereas in fact, the true value is 40%, and p_3 is estimated to be 40%, whereas the true value is 24% (Table 3). The estimated means and variances also show considerable departure from the true values. The estimated catch curves (Fig. 1) are also significantly different from the "true" catch curve. When the variance is 300, the estimated p_5 is 0.20 and p_4 is 0.15, giving a false impression that there is differential recruitment and mortality for the two years; also, when the variance is 400, the estimated model indicates that age classes 1 and 2 are partially recruited and age 3 is the first fully recruited age class. In the assumed population, only age 1 is partially recruited (the initial ascending limb of the curves represent partially recruited age classes, Ricker 1975, p. 33).

The results for case (b) are summarized in Table 4. Considerable improvement is observed in the estimates of the proportions. For example, when the variance is 400, the estimated proportions are reasonably close to the true values, as opposed to case (a). The estimated catch curves are plotted in Fig. 2. For smaller variances the estimated catch curves almost overlap the true catch curve. As the variance increases, the estimated catch curves tend to depart from the true curve, though this departure is not significant. As the variance increases, the estimated variances also exhibit greater departure from the true values, as one would expect.

*An expected order statistic is defined as follows. If $x_1, x_2, \ldots, x_n$ is a sample of size n such that $x_1 < x_2 < \ldots < x_n$, then $E[x_i]$ = expected order statistic, where $E[\cdot]$ is the usual expectation operator.

Table 3. Case (*a*): Estimates of the parameter vector Θ from a sample of expected order statistics (n = 1000)[a]

True variance (σ^2)	Estimator[b]	Age class 1	2	3	4	5
100	$\hat{\mu}$	117.1	156.2	188.11	207.5	233.7
	$\hat{\sigma}^2$	98.1	103.9	70.8	145.9	225.8
	$\hat{p}$	0.126	0.407	0.178	0.114	0.175
200	$\hat{\mu}$	115.5	152.0	175.9	201.8	234.5
	$\hat{\sigma}^2$	174.2	170.4	219.2	218.8	302.4
	$\hat{p}$	0.11	0.31	0.22	0.18	0.18
300	$\hat{\mu}$	116.1	150.8	174.9	201.7	232.1
	$\hat{\sigma}^2$	293.4	240.0	229.0	188.8	412.4
	$\hat{p}$	0.125	0.296	0.231	0.146	0.202
400	$\hat{\mu}$	114.8	142.2	169.6	207.8	242.4
	$\hat{\sigma}^2$	337.2	191.9	368.8	323.6	368.0
	$\hat{p}$	0.126	0.144	0.399	0.207	0.124

[a]The true parameter vector $\Theta = \langle 117, 156, 190, 220, 245, \sigma_1^2, \sigma_2^2, \ldots, \sigma_5^2, 0.13, 0.40, 0.24, 0.144, 0.086 \rangle$ and $\sigma^2 = \sigma_i^2, i = 1, 2, \ldots, 5$.

[b]$\hat{\mu}$ = estimated mean length, $\hat{\sigma}^2$ = estimated variance; and $\hat{p}$ = estimated proportion

Table 4. Case (*b*): Estimates of the parameter vector Ω from a sample of expected order statistics (n = 1000)[a]

True variance (σ^2)	Estimator[b]	Age class 1	2	3	4	5
100	$\hat{\mu}$	117	156	190	220	245
	$\hat{\sigma}^2$	98.2	102.7	92.3	115.4	99.3
	$\hat{p}$	0.128	0.406	0.232	0.152	0.082
200	$\hat{\mu}$	117	156	190	220	245
	$\hat{\sigma}^2$	196.1	198.6	173.8	173.9	189.0
	$\hat{p}$	0.13	0.405	0.235	0.144	0.087
300	$\hat{\mu}$	117	156	190	220	245
	$\hat{\sigma}^2$	301.9	273.4	285.9	212.6	266.6
	$\hat{p}$	0.137	0.383	0.260	0.123	0.097
400	$\hat{\mu}$	117	156	190	220	245
	$\hat{\sigma}^2$	348.6	207.0	287.0	241.8	351.0
	$\hat{p}$	0.134	0.403	0.236	0.121	0.107

[a]The true parameter vector $\Theta = \langle 117, 156, 190, 220, 245, \sigma_1^2, \sigma_2^2, \ldots, \sigma_5^2, 0.13, 0.40, 0.24, 0.144, 0.086 \rangle$ and $\sigma^2 = \sigma_i^2, i = 1, 2, \ldots, 5$.

[b]$\hat{\mu}$ = estimated mean length; $\hat{\sigma}^2$ = estimated variance; and $\hat{p}$ = estimated proportion.

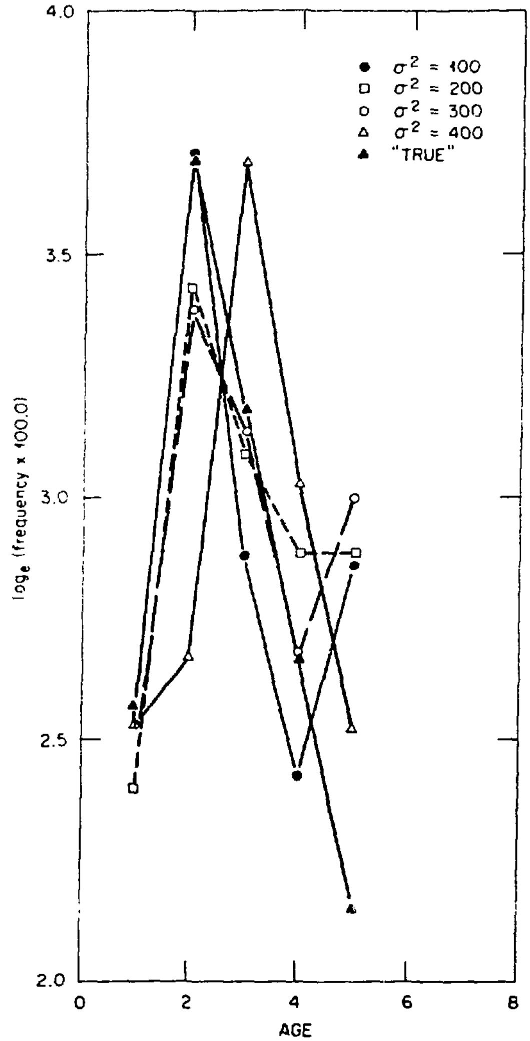

Fig. 1. Estimated catch curves when the full parameter vector Θ is estimated for the expected order statistic and for σ^2 = 100, 200, 300, and 400.

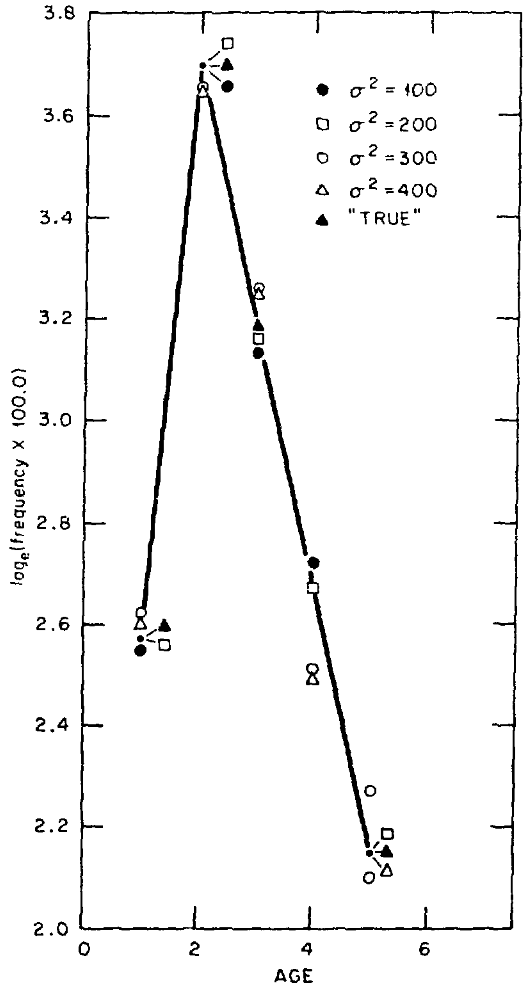

Fig. 2. Estimated catch curve when the location is assumed known and the reduced vector Ω is estimated for the expected order statistic and for σ^2 = 100, 200, 300, and 400.

These results show that if one has an age-length key available (case (*b*)), the length-frequency data can be used to obtain reliable estimates of the age structure of the population. When the key is not available, the estimates must be viewed as first approximations. In the next section, the results of a small scale simulation study for case (*b*) are presented.

SIMULATION RESULTS

The fish population was assumed to be the same as the above and the simulations were run for two cases

$$\text{(Case 1)}\ \sigma^2 = \sigma_i^2 = 100 \quad i = 1, 2, \ldots, 5 \,.$$

$$\text{(Case 2)}\ \sigma^2 = \sigma_i^2 = 200 \quad i = 1, 2, \ldots, 5 \,.$$

The random numbers were generated as follows. The sample size was fixed at 1000. The sample sizes for the five normal distributions were 130, 400, 240, 144, and 86, respectively, based on the proportions defined above. Twenty replicate samples of 1000 fish each were obtained for each of the two cases described above, and the estimates of the elements of the parameter vector, Ω, were obtained using the method discussed in the preceding sections. In order to reduce the number of replicates required, the replicates were rejected if the Cramer - von Mises statistic based on the true values exceeded 0.275. Hence, the replicates used in the analysis represent a stratified sample from the population.

The results of the simulations for cases (1) and (2) are summarized in Tables 5 and 6 respectively. In both cases, the estimated proportions are in

Table 5. Estimates of the parameter vector Ω based on 20 replicates and a sample size of 1000 when the mean lengths are set equal to 117, 156, 190, 220, and 245 mm and $\sigma^2 = \sigma_i^2 = 100$, $i = 1, 2, \ldots, 5$

Parameter	True value	Mean estimate	Standard deviation
σ_1^2	100	97.41	16.50
σ_2^2	100	100.28	20.08
σ_3^2	100	95.76	18.70
σ_4^2	100	95.03	25.38
σ_5^2	100	99.93	16.50
ρ_1	0.13	0.130	0.509(−2)[a]
ρ_2	0.40	0.399	0.111(−1)
ρ_3	0.24	0.241	0.157(−1)
ρ_4	0.144	0.145	0.143(−1)
ρ_5	0.085	0.085	0.760(−2)

[a] 0.509(−2) = 0.00509.

Table 6. Estimates of the parameter vector Ω based on 20 replicates and a sample size of 1000 when the mean lengths are equal to 117, 156, 190, 220, and 245 mm and $\sigma^2 = \sigma_i^2 = 200$, $i = 1, 2, \ldots, 5$

Parameter	True value	Mean estimate	Standard deviation
σ_1^2	200	190.75	35.62
σ_2^2	200	190.25	30.06
σ_3^2	200	187.55	39.14
σ_4^2	200	188.01	58.85
σ_5^2	200	185.95	34.82
p_1	0.13	0.130	0.896(−2)[a]
p_2	0.40	0.397	0.167(−1)
p_3	0.24	0.241	0.169(−1)
p_4	0.144	0.146	0.192(−1)
p_5	0.086	0.085	0.125(−1)

[a]0.896(−2) = 0.00896.

very good agreement with the "true" values. This result implies that if the means of the populations are known a priori, then the proportions (catch curve) can be estimated with some reliability. The variances, which are nuisance parameters, show greater variability than the proportions and are generally biased downward. The standard deviations of the estimates of σ_i^2 increase with increasing σ^2, as would be expected.

These results support the conclusion we reached with the expected order statistics. They demonstrate the feasibility of obtaining the CML estimates from the length-frequency data when the age-length key is available.

APPLICATIONS OF THE MODEL

We now apply the techniques discussed above to two species of fish, channel catfish (*Ictalurus punctatus*) and white crappie (*Pomoxis annularis*). Length-frequency data for these two species were collected as part of the ecological monitoring program at the Peach Bottom Atomic Power Station. The station, consisting of two units of 1065 MW(e) each, is located in southeastern Pennsylvania on the west bank of Conowingo Pond, an impounded reach of the Susquehanna River. Fish surveys with trap nets and trawls have been conducted at a series of stations located upstream of the power plant (controls) and in areas expected to be within the heated plume. Monitoring stations were located on the basis of available habitats and physical constraints imposed by topography. Length-frequency data and life-history information for the major species of fish in the pond are given in Robbins and Mathur (1974, 1976). In our study, length-frequency data of fish captured in trap nets were utilized. Trawl samples generally were dominated by age classes 0–2, whereas fish caught by trap nets were generally older and larger.

Figure 3 shows the length-frequency plot for the trap net catch of channel catfish in August 1971. Since the age-length key is not available for this species, we estimated all the parameters, Θ, defined by Eq. (1). It was emphasized earlier that these estimates are only a first approximation. When none of the parameters are known, the estimation problem is more complicated since in addition to the unknown parameter vector Θ, and additional parameter k, the number of age classes in the probability model, is also unknown. Several values of k, therefore, will be used in attempting to fit the model, and the one that is most "satisfactory" will be used. The estimates of the elements of the parameter vector Θ based on the data shown in Fig. 3 are summarized for $k = 5$ and $k = 6$ in Table 7.

Table 7. Estimates of parameter vector Θ for trap net data for channel catfish collected in August 1971 in Conowingo Pond

k[a]	Sample size	Estimator[b]	Modes					
			1	2	3	4	5	6
		$\hat{\mu}$	117	153	230	261	346	
5	345	$\hat{\sigma}$	9.4	9.0	28.1	51.[illegible]	37.4	
		$\hat{p}$	0.31	0.18	0.36	0.11	0.04	
		$\hat{\mu}$	117	153	203	236	254	320
6	345	$\hat{\sigma}$	9.4	8.4	33.3	21.3	35.0	42.4
		$\hat{p}$	0.31	0.16	0.15	0.22	0.09	0.07

[a] k = number of age classes in the probability model.

[b] $\hat{\mu}$ = estimated mean length, $\hat{\sigma}$ = estimated standard deviation; and $\hat{p}$ = estimated proportion.

For both $k = 5$ and $k = 6$, the first two modes are the same and the proportions are comparable. The mode at 230 mm for $k = 5$ is separated into two modes at 203 and 236 mm for $k = 6$. Clearly, the mode at 346 mm for $k = 5$ represents all the remaining older age classes. The estimated proportion for this mode is only about 4% of the sample of 345 channel catfish (i.e., about 14 fish). Either the species is short-lived or the trap nets are selective in that they do not capture the older age classes. The modes at 261 and 346 mm for $k = 5$ are represented by modes at 254 and 320 mm for $k = 6$. Also, some of the observations classified into modes 5 and 6 for $k = 5$ are classified into mode 4 with a mean length of 236 mm for $k = 6$. Even though we have associated a mode with a specific age class (e.g., mode 1 corresponds to age class 1), this is only a convenient way of referring to the modes.

Mathur et al. (1975) reported that the average lengths for channel catfish in Conowingo Pond between the years 1966–1969 for age classes 1–6 were 88, 149, 198, 226, 254, and 288 mm, respectively. For age classes 2–5 the agreement between the estimated mean lengths and these reported values is

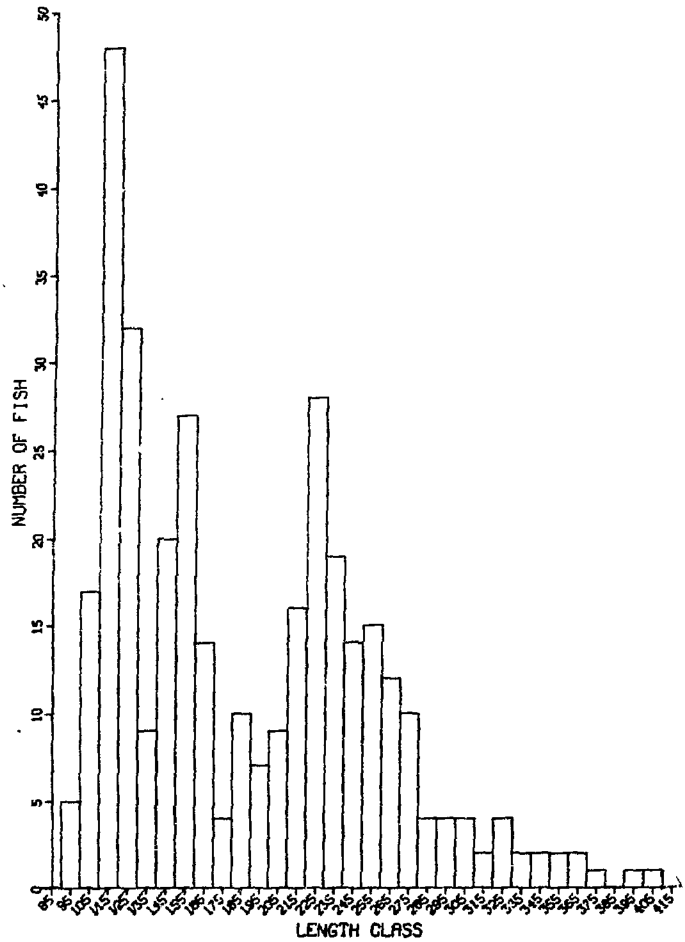

Fig. 3. Length-frequency data for channel catfish collected by trap nets in Conowingo Pond during August 1971.

reasonable (Table 7). The large discrepancy for age class 1 may be due to the selectivity of trap nets for the larger individuals in age class 1 or to a higher than normal growth rate for the 1970 year class.

The estimated standard deviations for age classes 3–6 are relatively large (Table 7). Since few channel catfish in the older age classes were caught, the estimated variances are only approximate. If the species is short-lived, the sample could be treated as representative of the true population. Otherwise, collection methods to obtain a more representative sample of the older age classes may be needed. The majority of specimens captured in Conowingo Pond are less than eight years old, but a few fish in age classes 8–16 have been captured (Robbins and Mathur 1974). Davis (1959) and Finnell and Jenkins (1954) found that channel catfish in Kansas and Oklahoma, respectively, seldom live longer than seven years. In contrast, 60% of the channel catfish collected by Stevens (1959) from two aquatic systems in South Carolina were eight years old or older. Age class structure and longevity of channel catfish may therefore be a function of the particular aquatic system concerned. Under the assumption that the catch was representative of the channel catfish population in Conowingo Pond in August 1971, the estimated catch curve is given by curve (*a*) in Fig. 4 for $k = 6$. The effect of variable recruitment among year classes is pronounced, as indicated by the departure of the curve from a straight line.

Cases could be encountered where the modes are too close to each other to represent separate age classes. Such would be the case when the spawning period of a species extends over a long period. Fish spawned in the early part of the period might have a higher mean length than those spawned towards the end of the period. The availability of food in different locations and inter- and intra-species competition for food may also play a strong role in creating multiple modes for the same age class. This situation probably exists in Conowingo Pond for white crappie and particularly for channel catfish. The growth of older channel catfish is relatively slow, and the growth of younger catfish is relatively rapid in Conowingo Pond. The average lengths of catfish in age classes 1 and 2 are greater than the lengths of the corresponding age classes in other aquatic systems. Conversely, the average lengths for age classes 3 and older are less than in other systems. Young catfish feed on zooplankton for about two years; thereafter, they assume a more benthic existence and consume greater quantities of benthos. Since the standing crop biomass of benthos is very sparse in the system, competition for this food source is high, and consequently, growth is slow. If two modes are close to each other due to differential growth within a single age class, it may become necessary to subdivide the age classification further. Alternatively, one could take the mean of the two modal lengths.

Another example for channel catfish is given in Fig. 5 and Table 8 for the June 1975 trap net data for $k = 5$ and $k = 6$. The modal lengths are the same for mode 1 and for mode 2. The large standard deviation associated with the

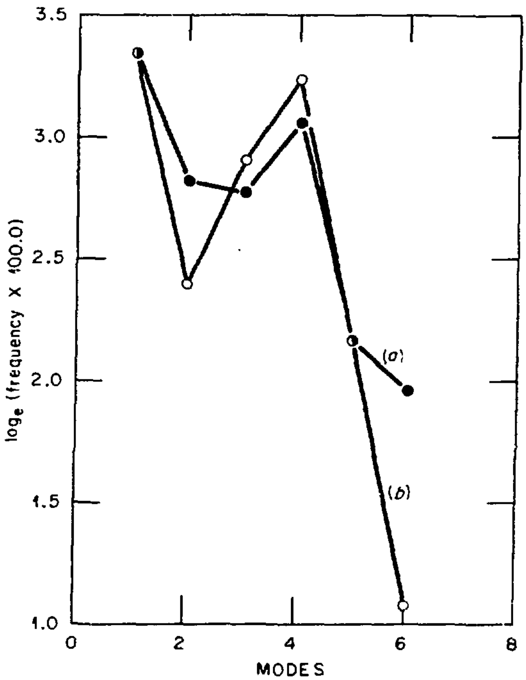

Fig. 4. Estimated catch curves for channel catfish: (a) August 1971 (Table 7). (b) June 1975 (Table 8).

Table 8. Estimates of parameter vector Θ for trap net data for channel catfish collected in June 1975 in Conowingo Pond

k^a	Sample size	Estimator^b	Modes 1	2	3	4	5	6
		$\hat{\mu}$	103	152	195	275	369	
5	314	$\hat{\sigma}$	16.3	7.9	26.0	28.7	30.0	
		$\hat{p}$	0.32	0.14	0.46	0.06	0.02	
		$\hat{\mu}$	103	153	177	201	255	349
6	314	$\hat{\sigma}$	16.2	7.2	22.5	22.7	30.9	42.5
		$\hat{p}$	0.31	0.12	0.18	0.27	0.09	0.03

$^a k$ = number of age classes in probability model.

$^b \hat{\mu}$ = estimated mean length; $\hat{\sigma}$ = estimated standard deviation; and $\hat{p}$ = estimated proportion.

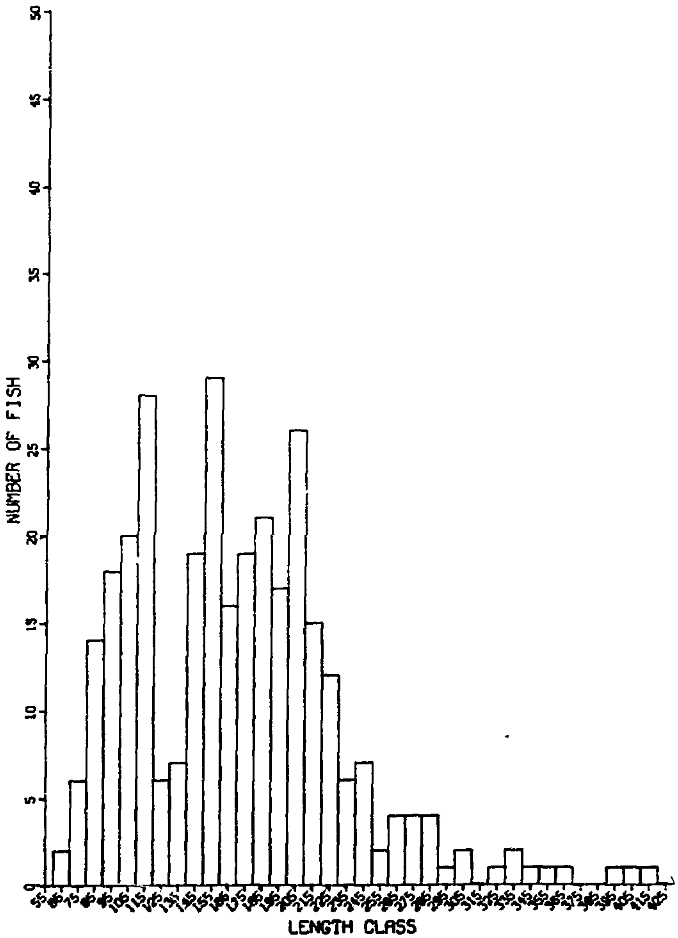

Fig. 5. *Length-frequency* **data for channel catfish collected by trap nets in Conowingo Pond, June 1975.**

mode at 103 mm is due to the presence of age class 0, as indicated by a minimum length of 65 mm for fish in the sample. When $k = 5$ there appears to be a large separation between modes 3 and 4 and between modes 4 and 5. Also, for mode 3, $\rho_3 = 0.46$, a proportion too high to be true. When $k = 6$, the modes are realigned, and there is a mode at 177 mm, which is 24 mm from the previous mode at 153 mm. Based on the estimated standard deviation at mode 2, a length difference of 24 mm represents approximately 3σ. The standard deviation at mode 3 is much larger (22.5 mm) than the standard deviation at mode 2 (7.2 mm), which indicates that mode 2 is about 1σ away from mode 3. It is evident that due to differential recruitment and growth patterns, the age classes 3–5 have overlapping length distributions. The estimated catch curve for $k = 6$ is given by curve (b) in Fig. 4. The peak at mode 4 indicates differential growth rate.

A third example is the trap net data for white crappie collected in Conowingo Pond during January, February, and March of 1974. The length-frequency data for each of the three months are shown in Figs. 6–8. The age-length key (Table 1) is based on samples caught during the nongrowing season (i.e., the winter months). Since January–March 1974 would be part of the nongrowing season of 1973, the age-length key should apply to the January–March 1974 data. The reported difference in mean lengths between age classes 4 and 5 is only 6 mm for 1973 (Table 1). However, since this value is based on a small sample, it is not very reliable. The length-frequency curves for these three months (Figs. 6–8) show that there are few fish between 160 and 180 mm (less than 2% of the catch in January). Since 1972 was classified as a year of poor recruitment (see Table 2), the proportion of age class 2 in the population in 1974 would be expected to be small.

In Table 9 the estimates of the full parameter vector Θ for each of the three months are summarized. In January, modes 1 and 2 occur at 125 and 136 mm, and these two modes together (year classes 1972 and 1973) constitute 59% of the catch, which is in direct conflict with the recruitment classification of Table 2. At the present time we are unable to resolve this conflict. The same pattern for modes 1 and 2 is also observed in February (68%) and March (80%). In February and March of 1974 mode 1 is at 116 mm, which is reasonably close to the reported mean length for 1973 of 114 mm in Table 1.

There exists, however, a second mode at approximately 135 mm for all three months. According to the age-length key (Table 1), this mode must represent age class 1 fish, since age class 2 had an average length of 171 mm in 1973. This result indicates that the mode at 135.0 mm represents the upper range for age class 1 in 1973. The mode at 116 mm that is present in February and March does not appear in January (Table 9). There are two possible explanations for the apparent discrepancy; either the average length for age class 1 was 130 mm in January 1974 or the catchability of fish less than 100 mm in length was low. Since the smallest fish caught was about 75 mm, the average length for age class 1 is probably 130 mm.

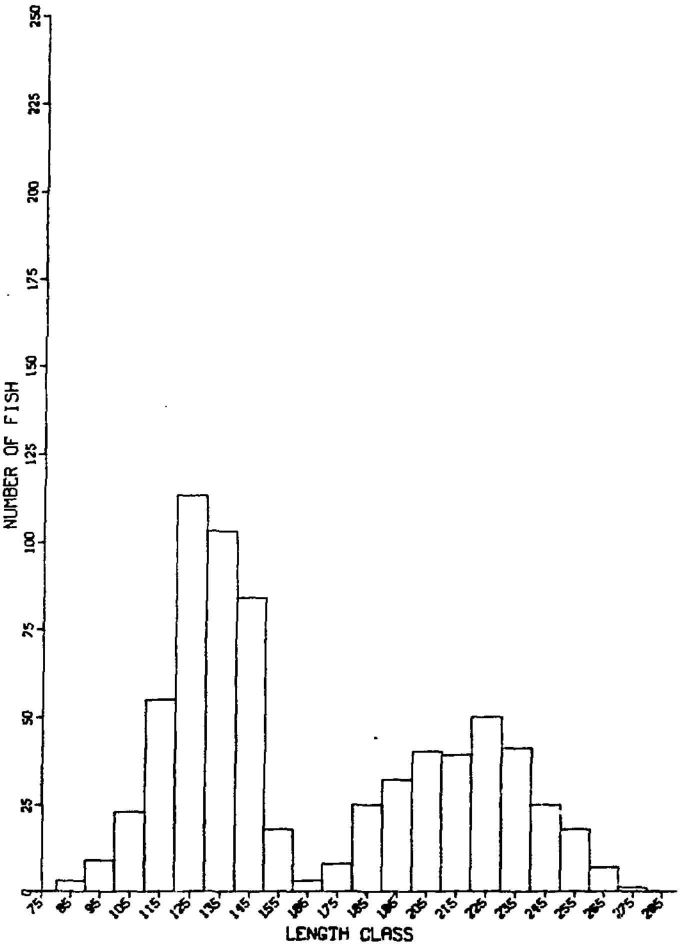

Fig. 6. Length-frequency data for white crappie collected by trap nets in Conowingo Pond, January 1974.

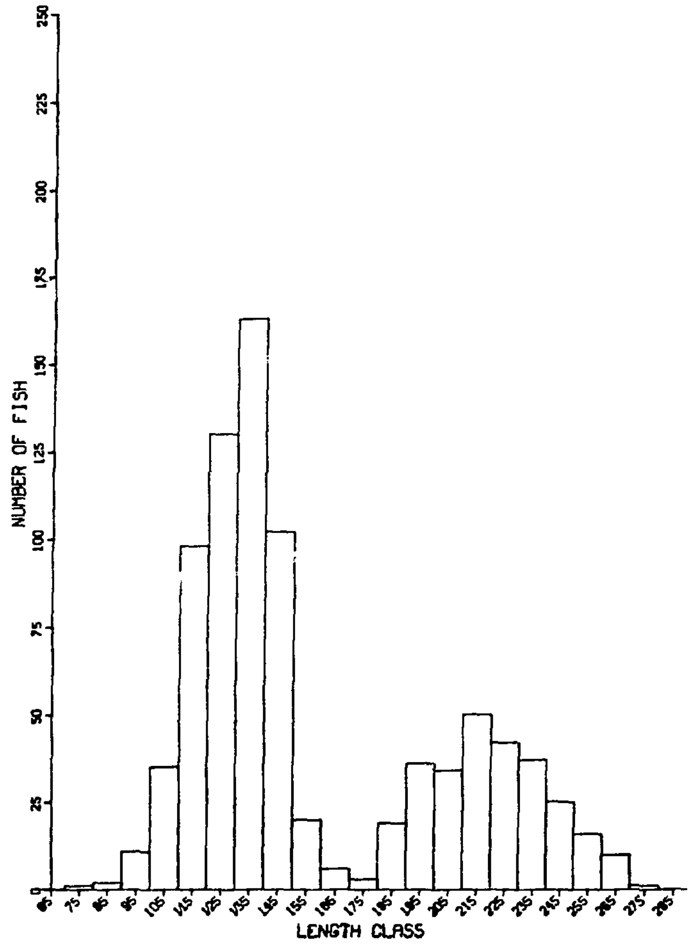

Fig. 7. Length-frequency data for white crappie collected by trap nets in Conowingo Pond, February 1974.

Table 9. Estimates of parameter vector Θ for trap net data for white crappie collected in January, February, and March 1974 in Conowingo Pond

Month	Sample size	Estimator[a]	Modes				
			1	2	3	4	5
January	697	$\hat{\mu}$	125	136	198	223	237
		$\hat{\sigma}$	15.1	10.2	13.5	16.1	15.4
		$\hat{p}$	0.33	0.26	0.15	0.13	0.13
February	841	$\hat{\mu}$	116	133	199	214	238
		$\hat{\sigma}$	12.9	11.7	13.2	15.2	15.8
		$\hat{p}$	0.16	0.52	0.09	0.10	0.13
March	1129	$\hat{\mu}$	116	135	194	208	232
		$\hat{\sigma}$	10.5	10.5	16.4	13.8	16.9
		$\hat{p}$	0.28	0.52	0.04	0.07	0.09

[a] $\hat{\mu}$ = estimated mean length; $\hat{\sigma}$ = estimated standard deviation; $\hat{p}$ = estimated proportion.

The estimated modes for the three months are 198, 199, and 195 mm for age class 3, which is in good agreement with the value of 197 mm from Table 1. For age class 4, the estimated modes for the three months are 223, 214, and 208 mm. The weighted average length in mm for age class i, $\bar{x}_i$, is computed by

$$\bar{x}_i = \sum_j \frac{(p_{ij}C_j)\mu_{ij}}{\sum_j p_{ij}C_j}, \tag{8}$$

where

p_{ij} = proportion of age class i in the catch at time j,

C_j = total number of fish in the catch at time j,

μ_{ij} = the ith estimated modal length (in mm) at time j.

The number of white crappie of all age classes caught in January, February, and March was 697, 841, and 1129, respectively. The estimated weighted mean lengths for ages 3, 4, and 5 are 198, 215, and 236 mm, respectively. The estimates for ages 3 and 4 are in close agreement with the reported values of 197 and 215 mm. The reported value for age 5 is 221 mm which is about 15 mm less than the estimated value.

In this example for white crappie, even though there was an estimate of the age-length key, the more general approach of estimating the elements of

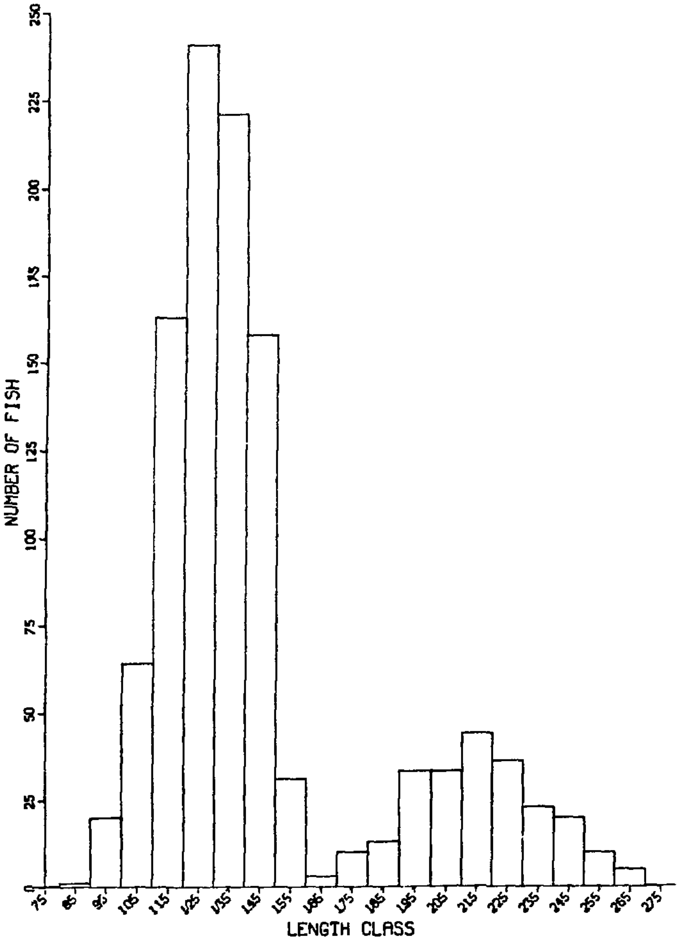

Fig. 8. Length-frequency data for white crappie collected by trap nets in Conowingo Pond, March 1974.

the full parameter vector Θ had to be used, because of the apparent discrepancies in the data. Two important points emerge from this example: (*a*) the age-length key is valid only for the catch on which it is based, and (*b*) in reporting fish-length data it is important to back-calculate all the lengths to a standard time. The last point is extremely important if the catch over a year is to be summarized and used for production calculations.

CONCLUSION

We have used the length-frequency data in conjunction with the age-length key to obtain maximum likelihood estimates of the age structure in the population. The simulation results and examples presented lead us to believe that proper estimation of the age-length key will result in good estimates of the age structure. Since knowledge of the age structure is useful in estimating population parameters, such as growth and mortality, it is therefore important that proper sampling methods be used for estimating the age-length key. The model can be used as follows to obtain an "optimal" subsampling scheme for age determination.

1. Obtain a length-frequency distribution for a fish catch.
2. Use the techniques described to obtain (*a*) estimate of modes, (*b*) estimate of variances, and (*c*) estimate of proportions.
3. Use these results to determine the "optimal" sample size for age determination.
4. Obtain an age-length key.
5. Use the age-length key to re-estimate the proportions.
6. Re-evaluate step 3 and if the change in the optimal sample size is negligible, stop; otherwise, obtain additional samples for age-length key determination and repeat steps 3–6.

The algorithm described in steps 1–6 above has not been examined in this study. We hope to address this problem in more detail at a later date.

In addition to applying knowledge of the age structure to estimate functional parameters of fish populations, age distributions can be used in the context of evaluating the potential effects of power plant operations on fish populations. If an appropriate control system is available, then the age structures of fish in the control and stressed areas for the same time periods could be compared. Assume that the fish in a stressed area cannot migrate to and from the control area and that the two systems were similar before the intervention (preoperational period) by a power plant. Differential changes in the values of the mean lengths (μ_i's) and the proportions (ρ_i's) between the two areas can be tested for. In other words, a statistically significant differential change in the population characteristics in the stressed area as compared to the control area may be reflected in the estimated values of the parameters of the model (μ_i's and ρ_i's). For example, if the operation of a power plant increases the growth rate of certain age classes, then the estimated mean lengths for these age classes will be greater than the mean lengths in the control area during the same period. If the mortality rate of an age class increases in the stressed area, then a smaller proportion of that age class will exist in the stressed area as compared with the control area.

In the preceding discussion we have assumed that the control and stressed systems were similar before power plant intervention. This assumption is not realistic for natural systems. A more realistic assumption would be that when both the systems vary naturally without intervention, there exists a constant structural relationship between the two areas. This constant relationship would be reflected by a constant value r_i for the ratio of the proportions, ρ_i's, for age i between the two areas. This ratio can be estimated initially from the preoperational data. Changes in the value of this ratio after the start of plant operation would serve as an index of impact (see Thomas, *in this volume*, for a further discussion of the use of ratios).

ACKNOWLEDGMENTS

The authors wish to express their appreciation to W. Van Winkle and J. J. Beauchamp for their constructive criticism of the manuscript.

LITERATURE CITED

Allen, K. R. 1966. Determination of age distribution from IBM 7079, 7094, age-length keys and length distributions, Fortran IV. Trans. Am. Fish. Soc. 95:230–231.

Chadwick, E. M. P. 1976. Ecological fish production in a small Precambrian shield lake. Environ. Biol. Fish. 1:13–60.

Cohen, A. C. 1966. Discussion of "Estimation of parameters for a mixture of normal distributions by Victor Hasselblad." Technometrics 8:445–446.

———. 1967. Estimation in mixtures of two normal distributions. Technometrics 9:15–28.

Cramer, H. 1974. Mathematical methods of statistics. Princeton Univ. Press, Princeton, New Jersey. 575 p.

Davis, J. 1959. Management of channel catfish in Kansas. Univ. Kansas, Mus. Nat. Hist., Misc. Publ. No. 21. 56 p.

Dick, N. P., and D. C. Bowden. 1973. Maximum likelihood estimation for mixtures of two normal distributions. Biometrics 29:781–790.

Euston, E. T. 1976. Trophic spectrum, p. 4-18 to 4-42. *In* T. W. Robbins and D. Mathur [eds.] Postoperational report No. 5 on the ecology of Conowingo Pond for the period of July 1975–December 1975. Ichthyological Associates Inc., Drumore, Pennsylvania.

Finnell, J. C., and R. M. Jenkins. 1954. Growth of channel catfish in Oklahoma waters: 1954 revision. Okla. Fish. Res. Lab., Rep. No. 41, 37 p.

Harter, H. L. 1952. Expected values of normal order statistics. Biometrika 48:151–165.

Hasselblad, V. 1966. Estimation of parameters for a mixture of normal distributions. Technometrics 8:431–444.

Ketchen, K. S. 1950. Stratified subsampling for determining age distributions. Trans. Am. Fish. Soc. 79:205–212.

Kumar, K. D., and E. H. Nicklin. 1976. Comparison of some models for security price behavior. Proc. 7th Annu. Pittsburg Conf. Modeling Simulation 7:98–103.

Mathur, D., P. G. Heisey, and N. C. Magnusson. 1975. Age and growth, p. 4-33 to 4–48. *In* T. W. Robbins and D. Mathur [eds.] Postoperational report No. 3 on the ecology of Conowingo Pond for the period of July 1974–December 1974. Ichthyological Associates Inc., Drumore, Pennsylvania.

MacDonald, P. D. M. 1971. Comment on "An estimation procedure for mixture of distributions" by Choi and Bulgren. J. Roy. Stat. Soc. B. 33:326–329.

Nelder, J. A., and R. Mead. 1965. A simplex method for function minimization. Comput. J. 7:308–313.

Pearson, K. 1894. Contributions to the mathematical theory of evolution. Phil. Trans. Roy. Soc. 185:71–110.

Rao, C. R. 1952. Advanced statistical methods in biometric research. Wiley and Sons, New York. 390 p.

Ricker, W. E. 1975. Computation and interpretation of biological statistics of fish populations. Fish. Res. Board Can., Bull. 191. 382 p.

Robbins, T. W., and D. Mathur. 1974. Peach Bottom Atomic Power Station preoperational report on the ecology of Conowingo Pond for Units No. 2 and 3. Ichthyological Associates, Inc., Drumore, Pennsylvania. 349 p.

Robbins, T. W., and D. Mathur. 1976. Peach Bottom Atomic Power Station postoperational report No. 5 on the ecology of Conowingo Pond for Units No. 2 and 3. Ichthyological Associates, Inc., Drumore, Pennsylvania.

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

Stevens, R. E. 1959. The white and channel catfishes of the Santee–Cooper Reservoir and Tailrace Sanctuary, p. 203–219. *In* Proc. Thirteenth Annu. Conf., S.E. Assoc. Game Fish Comm.