Measures of dispersion as constraints for length-frequency analysis

Karim Erzini
Margarida Castro
Unidade de Ciências e Tecnologias dos Recursos Aquáticos
Universidade do Algarve, 8000 Faro, Portugal

Length-frequency analysis (LFA) methods are widely used in population dynamics studies, particularly for tropical fish species that may be difficult or impossible to age by the traditional methods of reading growth rings on hard parts. LFA is characteristically subjective, and numerous authors have warned against its indiscriminate use, pointing out that estimated parameters may be questionable or even meaningless if the biology of the species is not taken into consideration or if the sampling was inadequate (e.g. Castro and Erzini, 1987; Macdonald, 1987; Morgan, 1987; Basson et al., 1988; Erzini, 1990). Biological information can be incorporated into these studies to obtain better results by using aged subsamples, time series of length-frequency distributions, or by constraining parameters to be estimated (Macdonald, 1987; Morgan, 1987). Constraints are based on assumptions concerning mortality, the relative abundance of the component age classes, the type of growth pattern or growth curve, the shape of the length-at-age distributions, the magnitude of the variability in length at age, and the pattern of this variability with age or size.

Our objective was to develop simple models, relating variability in length at age to life history and environmental parameters that could be used to select appropriate starting values and constraints for length-frequency analysis. We assumed that both the magnitude of variability in length-at-age and the size-and-age-dependent trends are related to species-specific life history and environmental characteristics. We demonstrate that measures of dispersion for particular lengths can be estimated on the basis of easily estimated parameter(s).

Methods

The data set used in this study consisted of 468 records representing 168 species and 50 families (Erzini, 1991). The following measures of variability in length at age were calculated: standard deviation of mean length at age (SD), variance of mean length at age (V), and coefficient of variation of mean length at age (CV). The following life history and environmental parameters were also compiled: von Bertalanffy K and L_∞, the Gallucci and Quinn (1979) growth parameter ω (intrinsic rate of growth), the growth performance index ϕ' (Longhurst and Pauly, 1987), maximum observed age, age at 0.95 L_∞, spawning pattern, spawning duration (months), geographic location, and environmental regime (tropical, temperate, and boreal). Spawning patterns were described as continuous, continuous with one major peak, continuous with two peaks, discrete with one peak, and discrete with two peaks. Only data sets that were not based on LFA, composite samples, or back-calculated lengths at age were included in the analysis.

Stepwise multiple regression with selection of variables by maximum R^2 improvement (SAS Institute Inc., 1985) was used to evaluate the relative effectiveness of life history and environmental parameters in predicting three measures of dispersion (SD, V, and CV). Qualitative variables such as environmental regime and spawning pattern were represented by indicator variables with values of 0 and 1 (Neter et al., 1983). For each qualitative variable consisting of m classes, m–1 indicator variables were formed. Preliminary plots and simple and quadratic regressions were used to guide the transformation and creation of new variables, such as mean length at age squared for the stepwise regression, resulting in a total of 19 independent variables. Only data where the sample size corresponding to the measures of dispersion was at least 10 were used.

After multiple linear regression was used to identify the most important explanatory variables, simple linear regression was used to examine the trends in variability in length at age for data grouped into discrete classes of these variables. Three-dimensional smoothed plots of measures of dispersion as functions of the independent variables and the classification parameters were also used to investigate trends in variability in length at age.

Results

The multiple regression models show that the SD models have the highest R^2 values whereas the CV models have the lowest (Tables 1–3). The SD and the V are strongly influenced by size and certain

Manuscript accepted 18 March 1994.
Fishery Bulletin 92:865–871.

growth parameters such as ϕ', ω, and L_∞. In the case of the CV, relative length at age seems to be the most important variable and the growth parameter variables were not selected for the models with five or less variables. Models with more than five independent variables are not shown as there was little further improvement in the amount of variation explained.

The influence of growth parameters can be seen in three-dimensional smoothed plots of the SD against relative length and L_∞ (Fig. 1), the SD against relative length and ϕ' (Fig. 2), and the SD against relative length and ω (Fig. 3). Magnitude of the variability of mean length at age generally increases with L_∞, ϕ', and ω. In contrast, no growth-parameter-related trends were found in plots involving CV. For example, in the plot of CV against relative length and K (Fig. 4), relative variability consistently decreased with size for all values of K.

Coefficent of variation and variation in CV decreased with increased relative length in all regressions for

Table 1

Examples of multiple linear regression models with the SD as the independent variable (n=3,050). ϕ' is the growth performance index, L_i is mean length-at-age, L_i^2 is the square of L_i, RL is relative length (L_i/L_∞), A_{95} is the age corresponding to $0.95L_\infty$ and ω is the Gallucci and Quinn (1979) growth parameter. MSE = the mean square error.

Model	MSE	R^2
$SD = -7.341 + 3.642\phi'$	1.99	0.62
$SD = -4.769 + 2.415\phi' + 0.022L_i$	1.66	0.68
$SD = -2.739 + 1.952\phi' + 0.028L_i - 1.479RL$	1.59	0.07
$SD = -3.532 - 0.049A_{95} + 0.092L_i - 4.280RL - 0.0002L_i^2$	1.49	0.72
$SD = 2.771 - 0.037A_{95} + 0.085L_i - 3.922RL - 0.0002L_i^2 + 0.023\omega$	1.47	0.72

Table 2

Examples of multiple linear regression models with the V as the independent variable (n=3,050). L_i is mean length at age, RL is relative length (L_i/L_∞), A is age, A_{95} is the age corresponding to $0.95L_\infty$, ω is the Gallucci and Quinn (1979) growth parameter, L_∞ is the von Bertalanffy growth parameter. MSE = the mean square error.

Model	MSE	R^2
$V = -6.314 + 0.0478L_i$	243.62	0.58
$V = -2.798 + 0.576L_i - 1.516A$	224.42	0.61
$V = 2.582 + 0.421L_i - 19.33RL + 0.464\omega$	216.45	0.62
$V = 1.198 + 0.472L_i - 16.80RL + 0.379\omega - 0.032A^2$	213.94	0.63
$V = 7.587 + 0.434L_i - 17.16RL - 0.025A^2 - 0.383A_{95} + 0.098L_\infty$	211.12	0.63

Table 3

Examples of multiple linear regression models with the CV as the independent variable (n=3,050). RL is relative length (L_i/L_∞), A is age, A_{95} is the age corresponding to $0.95L_\infty$, AA_{95} is age divided by A_{95}, and $AMAXA$ is age divided by the maximum observed age. MSE = the mean square error.

Model	MSE	R^2
$CV = 16.38 - 12.96RL$	16.50	0.31
$CV = 22.09 - 17.36RL - 0.171A_{95}$	13.95	0.42
$CV = 23.39 - 21.32RL - 0.173A_{95} + 3.614AA_{95}$	13.56	0.44
$CV = 22.97 - 19.97RL - 2.34AMAXA + 4.34AA_{95} - 0.155A_{95}$	13.41	0.44
$CV = 23.55 - 20.53RL - 2.411AMAXA + 3.49AA_{95} - 0.179A_{95} + 0.11A$	13.36	0.45

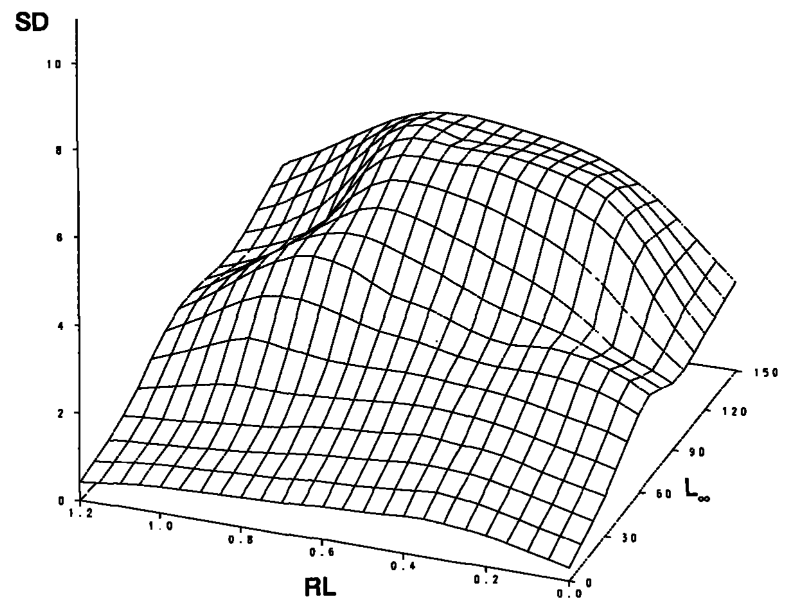

Figure 1
Smoothed surface graph of the standard deviation (SD) of mean length at age as a function of relative length (RL) and the asymptotic maximum length L_∞.

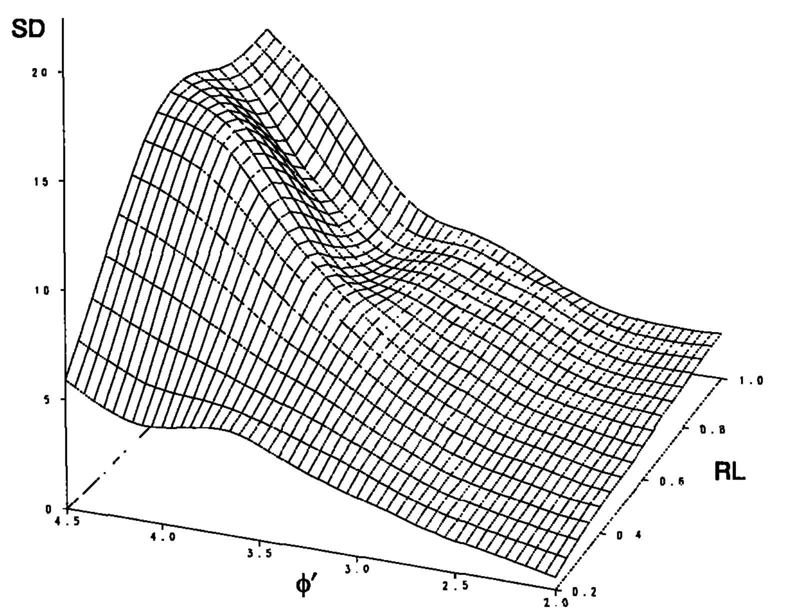

Figure 2
Smoothed surface graph of the standard deviation (SD) of mean length at age as a function of relative length (RL) and the growth performance index ϕ'.

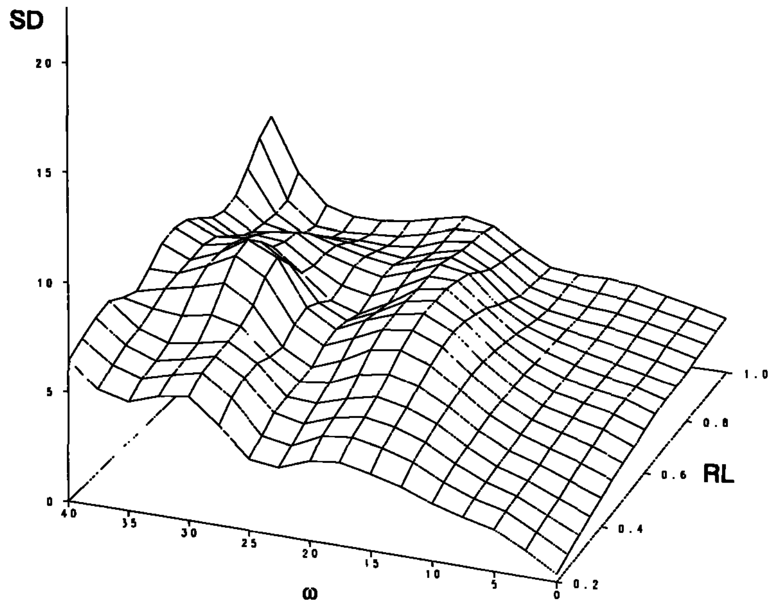

Figure 3

Smoothed surface graph of the standard deviation (SD) of mean length at age as a function of relative length and the parameter ω. (Gallucci and Quinn [1979] growth parameter.)

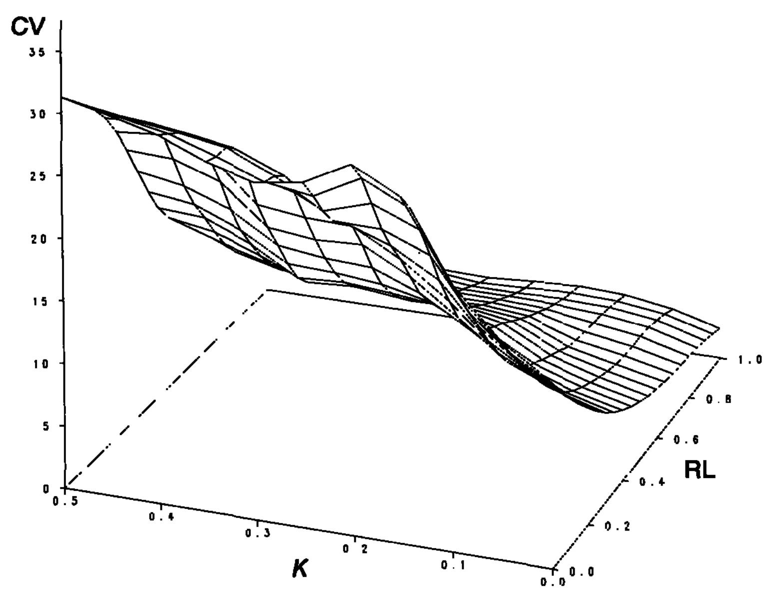

Figure 4

Smoothed surface graph of the coefficient of variation (CV) of mean length at age as a function of relative length (RL) and K, the growth rate.

Table 4

Regressions of the coefficient of variation (CV) against relative length (RL) for data grouped by the growth parameter.

	K	Intercept	Slope	*n*	MSE	R^2	*P*
a	0.05–0.099	13.39	−12.51	326	11.49	0.31	0.0001
b	0.10–0.149	15.51	−14.04	611	9.80	0.41	0.0001
c	0.15–0.199	20.05	−17.50	602	16.10	0.44	0.0001
d	0.20–0.249	20.15	−19.22	220	11.64	0.48	0.0001
e	0.25–0.299	21.53	−19.13	214	12.32	0.48	0.0001
f	0.30–0.349	22.65	−20.69	260	11.05	0.58	0.0001
g	0.35–0.399	19.27	−15.87	221	10.70	0.34	0.0001
h	0.40–0.449	23.09	−20.44	181	5.13	0.61	0.0001
i	0.45–0.549	22.02	−16.14	124	13.43	0.37	0.0001
j	≥ = 0.55	23.40	−19.33	220	11.59	0.43	0.0001

data grouped by the growth parameter *K* (Fig. 5, A–J). With the exception of groupings for $K<0.15$ (Fig. 5, A and B), which have smaller intercept and slope values, the regressions are similar (Fig. 5, C–J). The regression line and the 95% confidence intervals are also shown and the associated statistics are given in Table 4. The slopes of the regressions are all significantly different from 0 ($P<0.001$).

Discussion

A number of LFA methods, especially those that estimate parameters by maximum likelihood methods, allow constraints on measures of dispersion. For example, the simplex method of Kumar and Adams (1977) incorporates linear constraints on the standard deviations (SD's) of normal components. The SD's can be equal or fixed and the coefficient of variation (CV) can be fixed or constant in the program MIX (Macdonald and Pitcher, 1979; Macdonald and Green, 1985). The SD's can be linear functions of mean length or of age in the Schnute and Fournier (1980) method. MULTIFAN (Otter Software, 1988) allows age-dependent or length-dependent trends in SD's. A common CV between 0.01 and 0.5 for all lengths at age or SD's that increase linearly with mean length was proposed for LFA constraints by Liu et al. (1989)

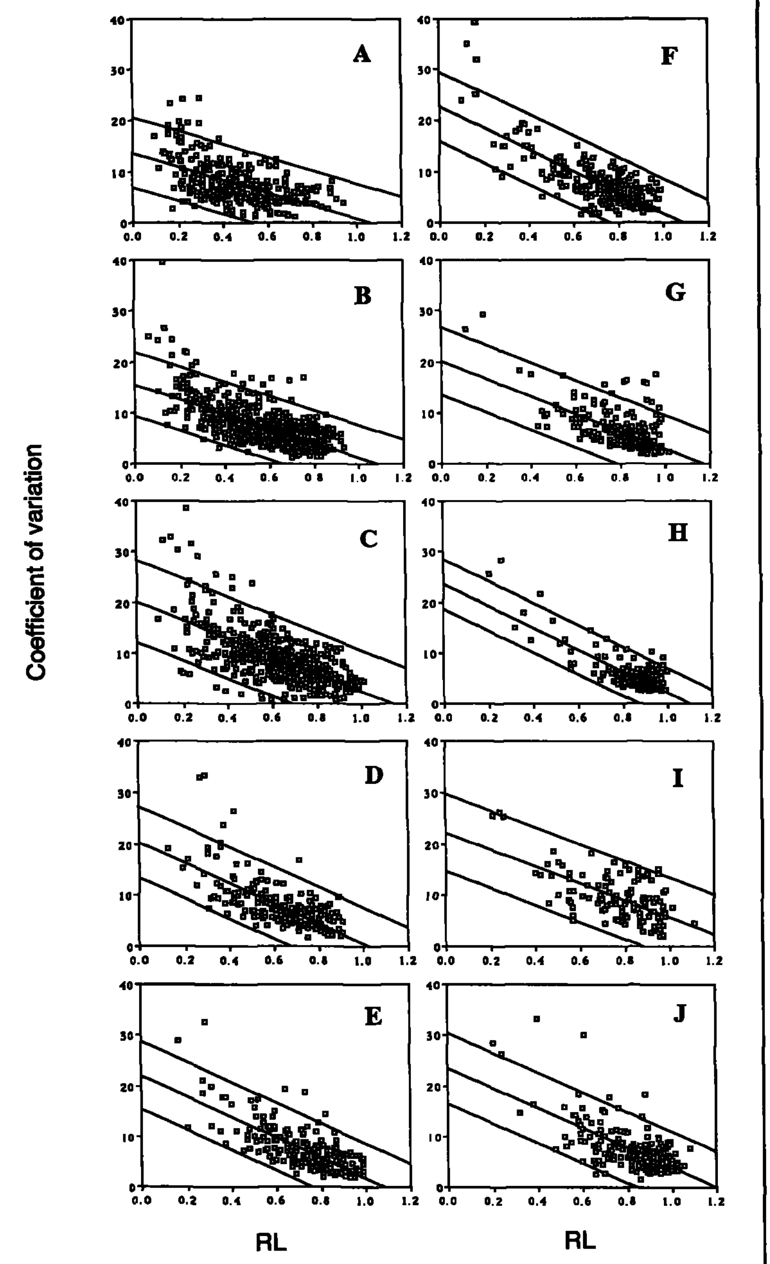

Figure 5

The coefficient of relative variation as a function of relative length for data grouped by *K*. RL is relative length (L_i/L_∞, length-at-age divided by L_∞). The interval classes and the regression statistics are given in table 4. Parallel lines are 95% confidence intervals.

In addition to these methods which allow specific constraints, some iterative methods require starting or initial values for some parameters, such as number of components, corresponding mean lengths at age, proportion in each age class and SD's of the component distributions (e.g. Akamine, 1982, 1984, 1985).

Our results can be used to select appropriate constraints and starting values for measures of dispersion for LFA methods. We have shown that the magnitudes of SD and V are dependent to a large extent on life history parameters. Therefore, if the LFA user has estimates of growth parameters, the multiple linear regression models in Table 1 can be used to estimate the SD for the species and size in question.

However, in most cases the objective of LFA is to estimate growth parameters, which are therefore not available for input into the predictive models. In this case, the CV may be more useful as a constraint. While the magnitudes of SD and V of mean length at age are related to characteristics of each species, relative variability in length at age (CV) is similar in species that differ greatly in life history parameters. Furthermore, while there are no consistent age- and size-dependent trends in absolute measures of variability, relative variability decreases in a predictable manner in almost all cases.

This was confirmed in a previous investigation of the shapes, magnitude, and age and size dependence of length-at-age distributions of marine fishes (Erzini, 1994). Analysis of 415 individual data sets showed that in 97% of the data sets the CV was negatively related to relative length at age, and the slope was significant ($P<0.05$) in 53% of the sets. CV values were similar for all species. A negative relationship between CV and size and decreasing variation with size are to be expected because changes in variability with growth are typically of smaller magnitude than changes in size with growth.

In contrast, although there was no dominant size-dependent or age-dependent trend for the SD, the most common pattern was that of increasing variability to a maximum at an intermediate age or size. This trend for increasing variability to a maximum at an intermediate size is illustrated in Figure 1, where the SD is plotted against relative length and asymptotic maximum length (L_∞). It is particularly evident for species with large L_∞ values.

In conclusion, we believe that the practical implications for LFA are that these empirically derived relationships between measures of dispersion, size, age, and life history parameters can be used to select starting values and to impose constraints on measures of dispersion corresponding to particular lengths at age. This is useful as there are no well established rules or guidelines for this process, which consequently has been highly subjective and dependent on each LFA user.

The choice of model depends on the availability of the data for the independent variables of the models. In the absence of any such data, the simplest model of the CV as a function of relative length can be used. As a preliminary step, length-frequency distributions should be examined and the number of possible component distributions and modes that may represent mean lengths at age identified visually. An estimate of L_∞ obtained from the literature or on the basis of the maximum observed size can be used to convert lengths to relative lengths. The estimated CV values and their corresponding confidence intervals for these modes can then be estimated with the models presented in this study. One possible approach is to use the estimated CV's as starting values and the confidence intervals as lower and upper constraints. Such a strategy would provide realistic starting values, reasonably narrow constraints, and would improve the often arbitrary choices which are made.

Acknowledgments

We would like to thank the Scientific Editor and an anonymous reviewer for their suggestions which greatly improved the manuscript.

Literature cited

Akamine, T.

1982. A BASIC program to analyze the polymodal frequency distributions into normal distributions. Bull. Jpn. Sea Fish. Res. Lab. 33:163–166.

1984. The BASIC program to analyze polymodal frequency distributions into normal distributions with Marquardt's method. Bull. Jpn. Sea Fish. Res. Lab. 34:53–60.

1985. Consideration of the BASIC programs to analyze the polymodal frequency distribution into normal distributions. Bull. Jpn. Sea Fish. Res. Lab. 35:129–160.

Basson, M, A. A. Rosenberg, and J. R. Beddington.

1988. The accuracy and reliability of two new methods for estimating growth parameters from length-frequency data. J. Cons. Int. Explor. Mer 44:277–285.

Castro, M., and K. Erzini.

1987. Comparison of two length frequency based methods used to obtain growth and mortality parameters using simulated samples with varying recruitment patterns. Fish. Bull. 86:645–653.

Erzini, K.

1990. Sample size and grouping of data for length frequency analysis. Fish. Res. 9:355–366.

1991. A compilation of data on variability in length-at-age of marine fishes. Fish. Stock Assess. CRSP Working Paper 77, 100 p.
1994. An empirical study of variability in length-at-age of marine fishes. J. Appl. Ichthyol. 10:17–41.
Gallucci, V. F., and T. J. Quinn.
1979. Reparametrizing, fitting, and testing a simple growth model. Trans. Am. Fish. Soc. 108:14–25.
Kumar, K. D., and S. M. Adams.
1977. Estimation of age structure of fish populations from length frequency data. *In* W. Van Winkle (ed.), Assessing the effects of power-plant induced mortality on fish populations, p. 256–281. Permagon Press, Oxford.
Liu, Q., T. Pitcher, and M. Al-Hossaini.
1989. Ageing with fisheries length-frequency data, using information about growth. J. Fish Biol. 35 (Suppl. A):169–177.
Longhurst, A. R., and D. Pauly.
1987. Ecology of tropical oceans. Academic Press, Inc.
Macdonald, P. D. M.
1987. Analysis of length frequency distributions. *In* R. C. Summerfelt (ed.), Age and growth of fish. Iowa State Univ. Press, p. 371–395.
Macdonald, P. D. M., and T. J. Pitcher.
1979. Age groups from size-frequency data: a versatile and efficient method of analyzing distribution mixtures. J. Fish. Res. Board Canada 36:987–1001.
Macdonald, P. D. M., and P. E. J. Green.
1985. User's guide to program MIX: an interactive program for fitting mixtures of distributions. Ichthus Data Systems, 28 p.
Morgan, G.R.
1987. Incorporating age data into length-based stock assessment methods. *In* D. Pauly and G. R. Morgan (ed.), Length-based methods in fishery research: proceedings of the international conference on the theory and application of length-based methods for stock assessment; 11–16 February 1985, Mazzara del Vallo, Sicily, Italy. ICLARM Conf. Proc. 13:137–146.
Neter, J., W. Wasserman, and M. H. Kutner.
1983. Applied linear regression models. Richard D. Irwin, Inc., 547 p.
Otter Software.
1988. MULTIFAN, user's guide and reference manual. Nanaimo, B.C. Canada.
SAS Institute Inc.
1985. SAS® user's guide: statistics. SAS Institute Inc., 956 p.
Schnute, J., and D. Fournier.
1980. A new approach to length-frequency analysis: growth structure. Can. J. Fish. Aquat. Sci. 37:1337–1351.