North American Journal of Fisheries Management 28:1233–1242, 2008

DOI: 10.1577/M07-078.1

[Article]

Novel Tools for Analyzing Proportional Size Distribution Index Data

TRAVIS O. BRENDEN* AND TYLER WAGNER

Quantitative Fisheries Center, Department of Fisheries and Wildlife, Michigan State University, 153 Giltner Hall, East Lansing, Michigan 48824-1101, USA

BRIAN R. MURPHY

Department of Fisheries and Wildlife Sciences, Virginia Polytechnic Institute and State University, 101 Cheatham Hall, Blacksburg, Virginia 24061-0321, USA

Abstract.—Proportional size distribution (PSD) indices are convenient measures for numerically summarizing size structure of fish populations. Several novel methods for analyzing PSD index data were examined as a means for potentially improving inferences of size structure: generalized linear mixed modeling, calculation of location quotients, and simultaneous construction of 100(1 − α)% confidence intervals for PSD indices through multinomial modeling. The methods are demonstrated using simulated and previously published PSD index data. Generalized linear mixed modeling was used to (1) conduct a model selection analysis of PSD index data relative to a change in harvest regulation for walleyes *Sander vitreus*, (2) partition the variance of PSD index data among spatial, temporal, and environmental attributes, and (3) nonlinearly relate PSD data to adult density for largemouth bass *Micropterus salmoides*. The location quotient, which is an index for describing relative concentration of proportional data, was used to describe spatial and temporal localization of largemouth bass and walleye PSD index data. Finally, a modified method of constructing simultaneous 100(1 − α)% confidence intervals for multinomial proportions was used to produce 95% confidence intervals for traditional and incremental PSD indices based on length frequency data from muskellunge *Esox masquinongy*. The approaches described herein can assist fishery biologists in size structure assessments of fish populations and ultimately aid in cost-effective fisheries management.

Proportional size distribution (PSD) indices, which formerly were referred to separately as proportional and relative stock densities, are convenient measures for numerically summarizing fish length frequency data and facilitating comparisons of size structure among fish populations (Willis et al. 1993; Guy et al. 2007). Statistical techniques that frequently are used to analyze PSD index data include the chi-square test, Student's *t*-test, and logistic regression (Neumann and Allen 2007). Although these methods are appropriate for analyzing PSD index data, the types of research questions that they can address are limited. Chi-square and Student's *t*-tests examine whether PSD index data were generated from a particular statistical distribution, whereas logistic regression can be used to determine whether PSD indices are conditionally related in a linear fashion to one or more independent variables. Many other research questions pertaining to size structure may be of interest to fishery biologists in exploring PSD index data but cannot be analyzed using the aforementioned methods. For example, biologists may be interested in (1) partitioning variability in PSD index data among different habitat scales (e.g., cove, lake, and region), (2) exploring whether PSD indices are nonlinearly related to one or more independent variables, or (3) incorporating possible temporal or spatial autocorrelations in PSD index data to improve inferences regarding size structure.

Given the various size structure questions that are of interest to fishery biologists, we identified the need for a general framework of PSD index data analysis that would be applicable to a variety of management situations. Such a framework ultimately would help simplify analyses of PSD index data, facilitate comparisons among different analytic techniques, and aid in determining the effects of different analytic assumptions on the recommended courses of action. The goal for our research was to describe how generalized linear mixed models (GLMMs) provide such an analytic framework for PSD index data and demonstrate how GLMMs can be used to investigate research questions that are of interest to fishery managers. Additional goals were to (1) introduce the location quotient as a method for quantifying relative concentration of PSD index data and (2) evaluate whether simultaneous 100(1 − α)% confidence intervals constructed via multinomial modeling of PSD

* Corresponding author: brenden@msu.edu

Received April 27, 2007; accepted December 30, 2007
Published online August 25, 2008

indices were narrower than confidence intervals generated via the Gustafson (1988) approach.

Methods

Background

Generalized linear mixed models.—Generalized linear mixed models are a class of models that include both fixed and random effects and that can be used for both discrete and continuous data generated from a member of the exponential family of distributions (e.g., binomial, normal, Poisson, and gamma distributions). Both continuous and categorical variables can be specified as fixed or random effects. A factor is considered fixed if it contains all levels for which inferences are to be made; a factor is considered random if it represents a random sample of a larger set of potential factors. For example, if multiple lakes are randomly sampled from and used to draw inferences for a larger population of lakes, then lake would be considered a random effect.

A number of statistical tests and procedures can be considered a particular form of GLMM, including analysis of variance (ANOVA), multiple linear regression, logistic regression, and Poisson regression. A generalized linear model (GLM) is also a type of GLMM in which random effects are absent. A linear mixed model can be specified as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}, \tag{1}$$

where $\mathbf{Y}$ is a vector of observed responses; $\mathbf{X}$ and $\mathbf{Z}$ are design matrices for the fixed and random effects, respectively; $\mathbf{e}$ is a vector of model error terms; and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are vectors of parameters for the fixed and random effects, respectively. Both $\boldsymbol{\gamma}$ and $\mathbf{e}$ typically are assumed to be normally distributed with a mean of $\mathbf{0}$ and variances $\mathbf{G}$ and $\mathbf{R}$, respectively (McCulloch and Searle 2001). Conditional expected values ($E[\mathbf{Y}|\boldsymbol{\gamma}] = \boldsymbol{\mu}$) of the observed responses in relation to the linear combination of regression coefficients for GLMMs are determined through a link function (g) such that

$$g[E(\mathbf{Y}|\boldsymbol{\gamma})] = g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}. \tag{2}$$

Equation (2) also can be expressed as

$$E(\mathbf{Y}|\boldsymbol{\gamma}) = \boldsymbol{\mu} = g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}), \tag{3}$$

where g^{-1} denotes the link function inverse. The form of the link function depends on the assumed distribution of the data. For example, the identity link is the natural link for normally distributed data; the log link is the natural link for Poisson-distributed data (Venables and Dichmont 2004).

Typically, PSD indices are modeled as binomially distributed random variables (Gustafson 1988; Miranda 1993); in other words, PSD index data are assumed to consist of a series of n independent Bernoulli trials for which m is the total number of successes and π is the probability of a successful trial (Ghahramani 2000). For PSD indices, n is the number of stock-length fish in a sample, m is the number of stock-length fish within a larger length-class (e.g., preferred length), and π (the PSD index before being multiplied by 100) is the probability that a stock-length fish is within a larger length-class ($m = n\pi$). The link function for binomially distributed random variables can take a number of forms, including the logit, probit, and complementary log–log links (McCullagh and Nelder 1989). A GLMM for a binomially distributed random variable with a logit link can be expressed for an individual observation as

$$\log_e[\pi/(1-\pi)] = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{z}^{\mathrm{T}}\boldsymbol{\gamma} \tag{4}$$

or as

$$\pi = e^{(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}+\mathbf{z}^{\mathrm{T}}\boldsymbol{\gamma})}/[1+e^{(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}+\mathbf{z}^{\mathrm{T}}\boldsymbol{\gamma})}], \tag{5}$$

where $\mathbf{x}^{\mathrm{T}}$ and $\mathbf{z}^{\mathrm{T}}$ are corresponding rows of the design matrices for the fixed and random effects, respectively.

Although it is common for a single PSD index to be used to summarize the size structure of a fish population, in some instances it can be beneficial to use several PSD indices to describe size structure (Neumann and Allen 2007). Any sample of stock-length fish can be proportionally represented by five disjoint incremental PSD indices describing stock-to-quality (PSD S–Q), quality-to-preferred (PSD Q–P), preferred-to-memorable (PSD P–M), memorable-to-trophy (PSD M–T), and trophy (PSD T) length-classes:

$$\begin{aligned}&(\mathrm{PSD\ S-Q}/100) + (\mathrm{PSD\ Q-P}/100)\\&\quad + (\mathrm{PSD\ P-M}/100) + (\mathrm{PSD\ M-T}/100)\\&\quad + (\mathrm{PSD\ T}/100) = 1.0.\end{aligned} \tag{6}$$

Alternatively, equation (6) can be expressed as

$$\pi_{\mathrm{S-Q}} + \pi_{\mathrm{Q-P}} + \pi_{\mathrm{P-M}} + \pi_{\mathrm{M-T}} + \pi_{\mathrm{T}} = 1.0, \tag{7}$$

where $\pi_{\mathrm{S-Q}}$, $\pi_{\mathrm{Q-P}}$, $\pi_{\mathrm{P-M}}$, $\pi_{\mathrm{M-T}}$, and π_{T} are the probabilities that stock-length fish belong to the respective length-classes. Because each probability in equation (7) will equal or exceed 0 and because the sum of the probabilities will equal 1.0, the multinomial distribution is the joint distribution of the five PSD indices (Ghahramani 2000). The multinomial distribution also is a member of the exponential family of distributions and can be analyzed by use of GLMMs. A GLMM for a multinomial random variable can be expressed through a logit link function as

$$\log_e(\pi_k/\pi_{\mathrm{S-Q}}) = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{z}^{\mathrm{T}}\boldsymbol{\gamma}, \tag{8}$$

where π_k denotes the probability for one of the other incremental PSD indices. Note that equation (8) models the natural logarithm of the probability ratio of an individual stock-length fish being in one incremental length-class (e.g., the P–M length-class) versus the S–Q length-class. Modeling of incremental PSD indices in this manner can be useful for identifying factors that increase or decrease the probability that a fish will attain larger lengths.

Location quotient.—The location quotient is an index for comparing local area characteristics or activities across a larger system (Miller et al. 1991; Moineddin et al. 2003; Beyene and Moineddin 2005). The location quotient is commonly used by geographers, health professionals, and economists to quantify and compare local conditions (e.g., industry share) with an overall, aggregate condition. The location quotient is calculated as

$$\mathrm{LQ}_i = (m_i/m) \times (n_i/n), \tag{9}$$

where LQ_i is the location quotient; m_i and n_i denote the outcome and population size of the ith sampled area ($i = 1, \ldots, I$), respectively; and m and n denote the outcome and population size for the entire system of study. For example, in terms of PSD index data, m_i and n_i could represent the number of quality- and stock-length fish in the ith sampled area that belong to the Q and S length-classes, and m and n would denote the total number of quality- and stock-length fish in all sampled areas. For PSD indices, equation (9) can also be expressed as

$$\mathrm{LQ}_i = \mathrm{PSD}_i/\mathrm{PSD}, \tag{10}$$

where PSD_i is the index for the ith sampled location and PSD is the aggregate index for all sampled areas. For PSD indices, the location quotient can be used to identify areas where particular length-classes of fish are underrepresented or overrepresented relative to the entire study area. A location quotient equal to 1.0 indicates that the PSD in a specific area is at the same level as the aggregate PSD index. A location quotient less than 1.0 indicates that the PSD for a specific area is less than that of the aggregate, whereas a location quotient greater than 1.0 indicates that the PSD for a specific area is greater than that of the aggregate.

Although calculation of location quotients is straightforward, constructing confidence intervals to assess uncertainty in the location quotient estimates is difficult because the index is a ratio (Moineddin et al. 2003; Beyene and Moineddin 2005). Closed-form solutions for constructing confidence intervals based on approximation methods are available (Moineddin et al. 2003; Beyene and Moineddin 2005). However, Beyene and Moineddin (2005) found that profile likelihood confidence intervals were narrower than approximation confidence intervals when sample sizes were small and when values of π were extreme; these situations often occur for PSD indices. Thus, the profile likelihood method may be the best method for constructing location quotient confidence intervals for PSD indices.

Proportional size distribution index confidence intervals.—Gustafson (1988) proposed a method for constructing confidence intervals for PSD indices based upon a normal approximation to the binomial distribution. Occasionally, this method is used to construct confidence intervals for several PSD indices used to describe the size structure of an individual fish population (Martin 1995; Galinat et al. 2002; Hurley and Jackson 2002). Such sequential calculation of confidence intervals can be problematic, however, because coverage probability of the confidence intervals can be affected in a manner similar to the effect of sequential statistical testing on the type-I error rate. Although correction factors (e.g., Bonferroni correction) can be used to protect the coverage probability, such correction factors can be overly conservative (García 2004) and can result in unnecessarily wide confidence intervals.

An alternative to Gustafson's (1988) method is to explicitly model the probabilities associated with the incremental PSD indices identified in equation (6) as multinomial variates. By doing so, approximate simultaneous $100(1 - \alpha)\%$ confidence intervals for all linear combinations of these probabilities can be calculated using the formula

$$\ell'\boldsymbol{\pi} \pm \sqrt{\chi^2_{k-1}(\alpha)}\sqrt{\frac{\ell'\boldsymbol{\Sigma}\ell}{n}}, \tag{11}$$

where $\ell'\boldsymbol{\pi}$ is some linear combination of probabilities identified in equation (7) and is equal to $\ell_1\pi_{\mathrm{S-Q}} + \ell_2\pi_{\mathrm{Q-P}} + \ell_3\pi_{\mathrm{P-M}} + \ell_4\pi_{\mathrm{M-T}} + \ell_5\pi_{\mathrm{T}}$ (ℓ in each case is equal to either 1 or 0, depending on which PSD index is calculated); $\boldsymbol{\Sigma}$ is a $k \times k$ matrix with variances $\sigma_{kk} = \pi_k(1 - \pi_k)$ for π-values on the diagonal and covariances $\sigma_{jk} = -\pi_j\pi_k$ ($j \neq k$) for π-values on the off-diagonal elements of the matrix; and $(\chi_{k-1})^2(\alpha)$ is the upper (100α)th percentile of the chi-square distribution with $k - 1$ degrees of freedom (Johnson and Wichern 1992). All traditional PSD indices are linear combinations of incremental PSD indices; thus, equation (11) can be used to generate confidence intervals for all PSD indices. For example, to construct a 95% confidence interval for PSD-P, ℓ' would be the vector [0 0 1 1 1] with $\ell'\boldsymbol{\pi}$ equal to $(0 \cdot \pi_{\mathrm{S-Q}}) + (0 \cdot \pi_{\mathrm{Q-P}}) + (1 \cdot \pi_{\mathrm{P-M}}) + (1 \cdot \pi_{\mathrm{M-T}}) + (1 \cdot \pi_{\mathrm{T}})$ and $(\chi_4)^2(0.05)$ equal to 9.78. To obtain a 95% confidence interval for PSD P–M, ℓ'

would be the vector [0 0 1 0 0] with $\ell'\boldsymbol{\pi}$ equal to $(0 \cdot \pi_{S-Q}) + (0 \cdot \pi_{Q-P}) + (1 \cdot \pi_{P-M}) + (0 \cdot \pi_{M-T}) + (0 \cdot \pi_{T})$. Regardless of the number of linear combinations considered, the probability coverage of confidence intervals constructed from equation (11) will equal $100(1 - \alpha)\%$ (Johnson and Wichern 1992). This approximation is said to hold as long as $n\pi$ exceeds 20 for all values of π (Johnson and Wichern 1992).

Data Sources and Analyses

Example 1.—Stone and Lott (2002) presented PSD data for walleyes *Sander vitreus* from Lake Francis Case, a main-stem reservoir of the Missouri River in South Dakota. In the original publication, Stone and Lott (2002) examined whether a minimum-length harvest regulation change led to changes in walleye recruitment, population structure, growth, and angler use and harvest. Although Stone and Lott (2002) were primarily interested in differences between periods before and after the regulation change (hereafter, pre- and postregulation periods), they divided the postregulation period into two periods based on reservoir pool elevation differences. High pool elevation occurred in several years of their study due to extremely high runoff. In the original analysis, Stone and Lott (2002) aggregated the PSD data within the three periods and used chi-square tests to determine whether walleye size structure differed among the periods. The results of their analyses indicated a significant difference in PSD between the preregulation period and the two postregulation periods (characterized by normal and above-normal pool levels). No significant difference in PSD was found between the postregulation periods (Stone and Lott 2002).

We used a GLMM to address whether a harvest regulation change led to a change in walleye population structure in the study by Stone and Lott (2002). Note that because no random effects occurred in the model structure, this GLMM could be considered a GLM. Rather than using a null hypothesis testing approach, the results of which are affected by factors such as sample sizes and selection of a type-I error rate (Burnham and Anderson 2002; Johnson and Omland 2004), we used a model selection approach to examine the effect of harvest regulation. Three competing models were evaluated: (1) PSD differed among all three periods (i.e., one preregulation and two postregulation periods); (2) PSD was the same for all three periods; and (3) PSD was the same for the two postregulation periods but differed for the preregulation period. The three models were fitted using the GLIMMIX procedure in the Statistical Analysis System (SAS; SAS Institute 2004). For each competing model, Akaike's information criterion corrected for small sample size (AICc) and Akaike weights (w_i) were calculated (Burnham and Anderson 2002). From values of w_i, we calculated the strength of evidence for the competing models to determine the extent to which the model with the lowest AICc value was more likely than the other models. The w_i for model i was calculated as follows:

$$w_i = \frac{e^{-(0.5)\Delta_i}}{\sum_{r=1}^{R} e^{-(0.5)\Delta_r}}, \qquad (12)$$

where Δ_i is calculated as $\mathrm{AICc}_i - \mathrm{AICc}_{\min}$ for the R models considered ($\mathrm{AIC}_{c,i}$ = the value for the ith model and $\mathrm{AIC}_{c,\min}$ = the minimum value among all models considered). The w_i represents a measure of a model's relative probability of being the best model given the R models considered (Burnham and Anderson 2002). The strength of evidence for each model was calculated as $w_{\max}/w_i$, where $w_{\max}$ is the largest weight for all candidate models and w_i is the weight for the ith model.

Example 2.—Novinger and Dillard (1978) presented PSD data for largemouth bass *Micropterus salmoides* from 38 small impoundments located in seven midwestern states. Sampling of the fish communities within these impoundments was conducted over 2 years; 19 impoundments from six states were sampled in the first year, and 19 impoundments from seven states were sampled in the second year. Several features of the impoundments were described by Novinger and Dillard (1978), including physicochemical characteristics, aquatic vegetation coverage, and size structure and population dynamics of bluegills *Lepomis macrochirus*. In the original publication, a variety of analyses was conducted on the fish community and habitat features of the impoundments, including examination of the relationship between largemouth bass PSD and factors such as largemouth bass annual mortality and density and bluegill PSD.

Using the data presented by Novinger and Dillard (1978), we conducted a multilevel analysis of the largemouth bass PSD data to estimate the variance among sampling years (level 3), among states within sampling years (level 2), and among impoundments within states nested within sampling years (level 1). This example takes advantage of the multilevel structure of the Novinger and Dillard (1978) data set and illustrates the use of a GLMM to partition variance in PSD data; the effects of system-level covariates on PSD are also investigated. Both state and sampling year were considered to be random effects. Fixed effects, including watershed ratio, Secchi depth, and the watershed ratio × Secchi depth interaction, were examined for positive associations with largemouth

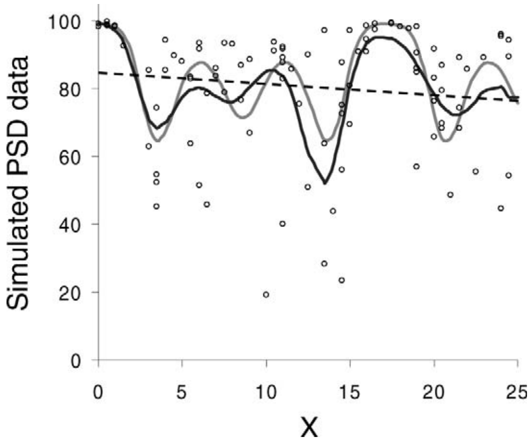

FIGURE 1.—Simulated proportional size distribution (PSD) data (open circles) and a low-rank radial smoother that was fitted to the data by generalized linear mixed modeling (solid black line). The solid gray line depicts the Fourier series equation that was used to generate PSD data before random errors were added to the observations. The dashed black line illustrates the fit of a logistic regression equation.

bass PSD for individual impoundments. This multilevel model was fitted in SAS using the GLIMMIX procedure. Methods used to conduct the multilevel analyses followed those described by Wagner et al. (2006a).

Example 3.—To illustrate the usefulness of GLMMs in exploratory modeling of PSD index data, we simulated a cyclical time series of PSD data to which fitting a parametric regression model probably would prove difficult if the data simulation methods were unknown (Figure 1). Our baseline data were generated using a Fourier series equation (Legendre and Legendre 1998), from which we randomly sampled 100 observations and added random errors generated from a standard normal distribution. The simulated data were then antilogit transformed to obtain a 0–1 scale. Using the GLIMMIX procedure, we fitted a low-rank radial smoother to the simulated data set. No fixed effects were specified for the model; instead, we specified the independent variable as a random effect that we modeled with a radial smoother covariance structure. The spline knots for the radial smoother were constructed using the EQUAL method with default settings (SAS Institute, Inc. 2005).

To further illustrate the use of GLMMs in exploratory modeling, we fitted a low-rank radial smoother to the largemouth bass PSD data presented by Novinger and Dillard (1978) in relation to density of largemouth bass adults ($\geq$305 mm length; defined by Novinger and Dillard [1978]). In the original publication, it was stated that a nonlinear relationship existed between largemouth bass PSD and adult density (Reynolds and Babb 1978), but no analyses were conducted to support this observation. We used a GLMM to fit a radial smoother to the data to determine whether the authors' observations were indeed supported. The radial smoother was fitted using the KDTREE method with default settings in the GLIMMIX procedure (SAS Institute 2005).

Example 4.—Brenden et al. (2007) used PSD indices to describe size structure for muskellunge *Esox masquinongy* captured by angling and electrofishing from the New River, Virginia. The authors calculated several traditional PSD indices based upon the muskellunge length frequency data, but they did not construct confidence intervals to assess the uncertainty associated with their index estimates. Using their length frequency data, we modeled the PSD indices presented in equation (6) as a multinomial random variable. We then used equation (11) to construct simultaneous $100(1 - \alpha)\%$ confidence intervals for four traditional and two incremental PSD indices that described muskellunge size structure. We did not include the trophy length-class when calculating the PSD indices, because only one fish from this length-class occurred in the original data set. We also used Gustafson's (1988) method to calculate confidence intervals for muskellunge PSD indices so that differences in confidence interval bounds could be compared between methods. A Bonferroni correction ($0.05/6 = 0.0083$) was used to protect the probability coverage of confidence intervals calculated by Gustafson's (1988) method.

Example 5.—For this example, we used the location quotient index to examine the relative spatial and temporal concentration of PSD for the largemouth bass and walleye size structure data presented by Novinger and Dillard (1978) and Stone and Lott (2002). For the Novinger and Dillard (1978) data, we calculated state-level location quotients by combining the largemouth bass PSD data among years and impoundments for individual states. For the Stone and Lott (2002) data, we calculated period location quotients to assess temporal concentration of the walleye PSD data. Following the recommendation of Beyene and Moineddin (2005), we constructed 95% profile likelihood confidence intervals for the PSD location quotients using the GENMOD procedure in SAS.

Results

Example 1

For our analysis of the Stone and Lott (2002) walleye PSD data, the lowest AICc value (273.3) was observed for the two-parameter model in which walleye PSD for the preregulation period was indicated to be different from the PSD calculated for the two postregulation

TABLE 1.—Akaike weights (w_i), strength of evidence (w_{max}/w_i, where w_{max} = maximum w_i observed among candidate models), and parameter estimates for three models describing the effect of a change in harvest regulation on walleye proportional size distribution (PSD) in Lake Francis Case, South Dakota (original data were from Stone and Lott [2002]): (1) PSD differed among three periods (preregulation [period 1] and postregulation [periods 2 and 3]; parameters 1–3 for this model correspond to the estimated mean PSD/100 values for periods 1–3); (2) PSD was the same for all three periods (parameter 1 = grand mean PSD/100); and (3) PSD was the same for periods 2 and 3 (parameter 2 = mean PSD/100 for the two periods) but different for period 1 (parameter 1).

			Parameter		
Model	W_i	Strength of evidence	1	2	3
1	0.25	2.93	0.14	0.23	0.21
2	0.00	7.4×10^9	0.20		
3	0.75	1.00	0.14	0.22	

periods (Table 1). In comparison, the AICc value was 275.4 for the three-parameter model (in which PSD differed across all three periods) and 318.7 for the one-parameter model (in which PSD was constant across periods). In terms of strength of evidence, the two-parameter model was approximately 2.9 times more likely than the three-parameter model and was about 7.40×10^9 times more likely than the one-parameter model. These results support the findings of Stone and Lott (2002) that (1) a walleye regulation change resulted in an increase in walleye size structure and (2) no difference in size structure occurred due to water level differences between the two postregulation periods. The model selection approach to analyzing the Stone and Lott (2002) data has the added benefit of comparing the results of the different hypotheses.

Example 2

For our multilevel analysis of the Novinger and Dillard (1978) largemouth bass PSD data, the estimated variance was 0.301 (SE = 0.71) among years and 1.1091 (SE = 0.53) among states within years. Although we were initially interested in partitioning PSD data variability into that among years and among states nested within years, the GLIMMIX procedure warned that the *G*-side random effects for the unconditional model were not positive definite, indicating that the unconditional model did not do a good job of explaining the observed data. Thus, we did not fully partition the observed variation in largemouth bass PSD data as described by Wagner et al. (2006a).

For the predictor variables that were included in the multilevel analysis, we detected a significantly negative effect of watershed ratio on largemouth bass PSD ($\beta = -0.024$, $F = 20.8$, $P < 0.001$) and a nonsignificant

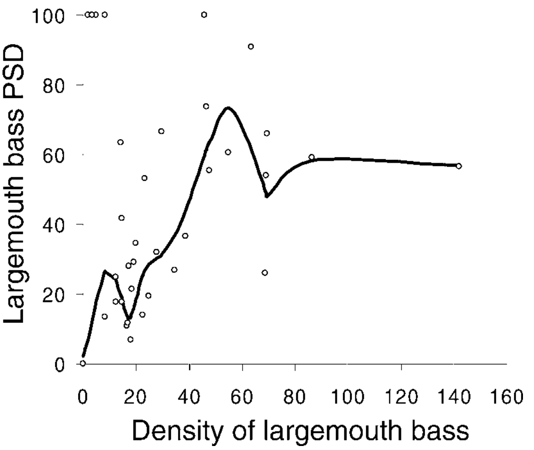

FIGURE 2.—Low-rank radial smoother (solid black line) fitted by generalized linear mixed modeling to the relation between largemouth bass proportional size distribution (PSD) and adult (≥305 mm) density data (open circles). Largemouth bass data are from Novinger and Dillard (1978); figure is modified from Reynolds and Babb (1978).

effect of Secchi depth ($\beta = 0.003$, $F = 0.8$, $P = 0.387$). However, there was a significant interaction between watershed ratio and Secchi depth ($\beta = 0.0003$, $F = 2.7$, $P = 0.011$), which precluded us from making any concrete conclusions about the effect of watershed ratio or Secchi depth on largemouth bass PSD.

Example 3

The low-rank radial smoother that was fitted to the simulated time series of PSD data appeared to ably approximate the Fourier series equation used to generate the data (Figure 1). Although the fitted radial smoother had a tendency to overestimate the peaks and underestimate the valleys of the simulated data series, the identified transition zones in the data series closely approximated those of the original equation. Fitting a regular logistic regression equation to the simulated data set resulted in a statistically significant coefficient for the independent variable ($\beta = -0.021$, $F_{1,98} = 112.46$, $P = 0.0001$), even though the amount of variability explained by the fitted model was low ($r^2 = 0.01$).

The low-rank radial smoother that was fitted to largemouth bass PSD and adult density supported the contention of a nonlinear relationship between the variables (Reynolds and Babb 1978). Largemouth bass PSD generally increased when adult density ranged from 0 to 55 fish/ha, albeit with a slight decrease at approximately 10–20 fish/ha (Figure 2). Largemouth bass PSD then decreased and leveled off at adult densities greater than 55 fish/ha (Figure 2). This smoothed relationship suggests that an equation

TABLE 2.—Proportional size distribution (PSD) indices calculated from length frequency data describing muskellunge in the New River, Virginia (original data were from Brenden et al. [2007]), and 95% confidence intervals (CIs) calculated by two methods: (1) multinomial modeling (equation 11 in Methods) and (2) Gustafson's (1988) method (used with Bonferroni correction to protect the probability coverage). Indices describe stock-to-quality (PSD S–Q), quality-to-preferred (PSD Q–P), preferred-to-memorable (PSD P–M), memorable-to-trophy (PSD M–T) quality, (PSD), and preferred (PSD-P) length-classes.

		95% CI type	
PSD Index	Estimate	Multinomial	Gustafson
PSD S–Q	13	9–17	8–18
PSD Q–P	49	42–56	42–56
PSD P–M	25	20–31	19–31
PSD M–T	12	8–17	8–17
PSD	87	82–91	82–92
PSD-P	38	31–44	31–44

appropriate for dome-shaped relationships is appropriate for parametric modeling of the relationship between largemouth bass PSD and adult density.

Example 4

Using the muskellunge length frequency data from Brenden et al. (2007), we determined that π_{S-Q} was 0.130, π_{Q-P} was 0.491, π_{P-M} was 0.253, and π_{M-T} was 0.123. Based upon these estimated proportions, the variance–covariance matrix of π-values for the muskellunge data was

$$\Sigma = \begin{bmatrix} 0.113 & -0.064 & -0.033 & -0.016 \\ -0.064 & 0.250 & -0.124 & -0.061 \\ -0.033 & -0.124 & 0.189 & -0.031 \\ -0.016 & -0.061 & -0.031 & 0.108 \end{bmatrix},$$

Thus, use of equation (11) to construct a 95% confidence interval for muskellunge PSD-P would yield the following:

$$\frac{\text{PSD-P}}{100} \pm \sqrt{7.81} \times \left\{ \frac{[0\,0\,1\,1] \begin{bmatrix} 0.113 & -0.064 & -0.033 & -0.016 \\ -0.064 & 0.250 & -0.124 & -0.061 \\ -0.033 & -0.124 & 0.189 & -0.031 \\ -0.016 & -0.061 & -0.031 & 0.108 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}}{445} \right\}^{1/2}.$$

There was very little difference in 95% confidence intervals generated by equation (11) and by the Gustafson (1988) method with Bonferroni correction. Confidence interval bounds differed for only three of the calculated PSD indices (PSD S–Q, PSD P–M, and PSD), and the multinomial method provided only slightly narrower confidence intervals than Gustafson's (1988) method (Table 2).

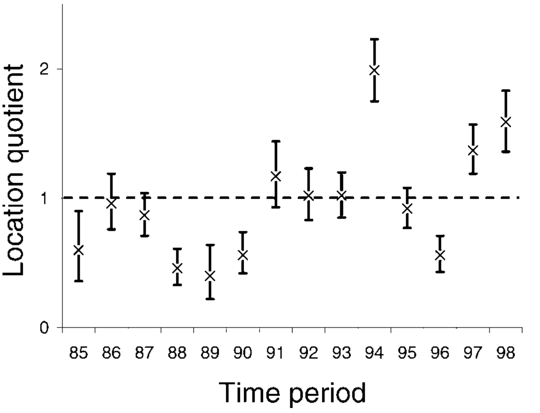

FIGURE 3.—Annual location quotients (with 95% profile likelihood confidence intervals) describing the relative temporal concentration of walleye proportional size distribution (PSD) data from Stone and Lott (2002; data from 1985 to 1998). The dashed horizontal line demarks a location quotient of 1.0, which indicates that a PSD for a given period is equal to the aggregate PSD.

Example 5

For the Stone and Lott (2002) walleye PSD data, calculation of annual location quotients indicated that before 1991 the proportion of quality-length walleyes was underrepresented relative to the entire period under consideration (Figure 3). Beginning in 1991 (the first postregulation year), the proportion of quality-length walleyes became more representative for the entire period (Figure 3). The location quotients further indicated that in 1994, 1997, and 1998, the proportion of quality-length walleyes was overrepresented relative to the other periods. For example, the proportion was approximately two times greater in 1994 than in other periods.

Calculation of state-level location quotients for largemouth bass PSD data (Novinger and Dillard 1978) indicated that the proportion of quality-length fish was overrepresented in Iowa, Indiana, and Kentucky but underrepresented in Kansas, Illinois, and Missouri (Figure 4). The location quotient for Ohio equaled 1.04, indicating that the PSD for this state was similar to the aggregate PSD for all states. Given the number of impoundments sampled within the states, these results are not particularly informative, although the example does illustrate how the location quotient can assess geographical localization of PSD index data.

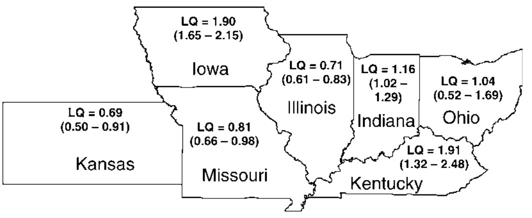

FIGURE 4.—State-level location quotients (LQs; with 95% profile likelihood confidence intervals in parentheses) describing the relative spatial concentration of largemouth bass proportional size distribution (PSD) data from Novinger and Dillard (1978).

Discussion

Use of GLMMs is becoming increasingly popular in fisheries research and management, particularly for catch–effort standardization and hierarchical (multilevel) modeling (Ortiz and Arocha 2004; Wagner et al. 2006b, 2007). This trend of increased use will probably continue as the utility of GLMMs becomes more widely known (Xiao et al. 2004). There are a number of reasons why adopting a GLMM framework for analyzing PSD index data would ultimately prove beneficial. First, although GLMMs can readily be used for null hypothesis statistical testing, they also can be used to conduct analyses that are more biologically or ecologically relevant to the resource. Null hypothesis testing has been criticized on the grounds that stated hypotheses often can be rejected a priori (Anderson et al. 2000). Those critical of null hypothesis testing would argue that there is little need to test for PSD index differences as a method of assessing size structure differences, which are almost inevitably present among populations or within a population across periods; thus, the only real question is whether a large enough sample has been collected to detect the difference. From a management standpoint, it might be more meaningful to partition the observed variance in PSD indices among spatial, temporal, or environmental attributes to determine which factors explain variability in fish size structure. Because GLMMs encompass a wide variety of statistical procedures and techniques, their use is well suited for model selection approaches, which are arguably more scientifically meaningful for resource management than are other data analysis approaches.

Another advantage of using GLMMs to analyze PSD index data is that a variety of error structures for the observed data can be assumed. For example, data transformations are often used to meet the assumptions of normality for statistical tests (i.e., that the response variable is continuous with a conditional normal distribution and a constant error variance). However, when the dependent variable (e.g., a PSD index) is bounded (e.g., between 0 and 1), the variance of the dependent variable depends on the mean and the normality assumption is no longer valid (the assumption of constant variance is violated). The use of GLMMs circumvents this problem by accommodating response variables with nonnormal distributions. Furthermore, flexibility in specifying different error distributions allows consideration of spatial or temporal autocorrelation within PSD index data, which can be useful for spatial or temporal prediction of fish size structure (Faes et al. 2006). Accounting for temporal autocorrelation also is necessary when conducting longitudinal analyses of size structure, where PSD indices for a single population are measured repeatedly through time. Another advantage of GLMMs is that they can be used to account for extra variability within the observed data (i.e., overdispersion). Given such factors as length-related biases in sampling methodologies, it is possible that observed variances in PSD index data are greater than expected variances. Testing for overdispersion and accounting for extra variability are among the features of GLMMs. Further, GLMMs allow users to overcome the many analysis impediments that can arise with fisheries data, such as measurement error, missing data, and outliers (Wand 2003).

Conversely, we see relatively few disadvantages in adopting a GLMM framework for analyzing PSD index data. Perhaps the biggest obstacle is the initial difficulty in learning to fit these types of models. Resources are available to improve familiarity with GLMMs and similar models (e.g., Pinheiro and Bates 2000; McCulloch and Searle 2001; Gelman and Hill 2007). A number of statistical software packages, including SAS, S-PLUS, and R, can be used to fit GLMMs. The ability to fit GLMMs in R is particularly advantageous because R is freeware that can be obtained from the internet.

We also believe that the location quotient provides a useful tool for assessing local concentration of PSD index data. When fish length frequency data are collected from multiple areas (e.g., periods or sampling transects) to assess size structure of a population, the location quotient can be used to examine the degree of spatial or temporal localization in PSD indices. Although similar information can be derived by plotting overall and localized PSD indices in combination, the location quotient is advantageous in also permitting estimation of the uncertainty in localized PSD concentration. Such information may be helpful for identifying areas of different habitat quality within particular systems or in identifying periods with above- or below-average year-class (or age-class) strength.

Our finding that there was very little difference in 95% confidence intervals constructed for the muskellunge PSD indices suggests that no real advantage is obtained by using the multinomial method in preference to the Gustafson (1988) method. Therefore, we recommend use of the Gustafson (1988) method with Bonferroni correction when constructing confidence intervals for multiple PSD indices. The lack of improvement from use of the multinomial approach stems from the relatively few linear combinations that are involved in the calculation of traditional PSD indices. As stated previously, equation (11) holds for all possible linear combinations of π-values (Johnson and Wichern 1992). Thus, the real advantage of the simultaneous approach might not be realized unless tens or hundreds of linear combinations of π-values are considered. Such large numbers of linear combinations may occur in some situations but not for size structure indices. As a result, the multinomial approach yields no benefit to analysis of PSD indices.

Despite our findings concerning PSD index confidence intervals, we believe that multinomial modeling of PSD indices should be more frequently used when analyzing the size structure of fish populations. Although the binomial distribution (frequently the assumed distribution for PSD index analysis) is a specific case of the multinomial distribution (Ghahramani 2000), there are advantages in considering a PSD index to be a multinomially distributed random variable. For example, multinomial modeling can be used to determine sample sizes needed so that all PSD indices, rather than just one particular index (Miranda 1993), will occur within a specified interval (Angers 1984; Thompson 1987; Sison and Glaz 1995). Additionally, as stated previously, modeling of PSD indices as multinomial random variables may be beneficial for identifying factors that increase or decrease a fish's probability of reaching larger length-classes.

In conclusion, PSD indices are useful assessment tools for fisheries management because they integrate information regarding reproduction, growth, and mortality rates of fish populations (Willis et al. 1993). Due to limited budgets for sampling, PSD indices may be the only way to gather information about population dynamics of many exploited populations. We hope that the methods described herein will assist fishery biologists in size structure assessments of fish populations and will ultimately aid in cost-effective fisheries management.

Acknowledgments

This manuscript is publication 2008-19 of the Quantitative Fisheries Center at Michigan State University (MSU). Copies of SAS code used to conduct the analyses can be obtained from the senior author. The authors thank Connie Page and Juan Du of the MSU Center for Statistical Training and Consulting for assistance regarding confidence intervals for linear combinations of multinomial proportions. Comments from three anonymous reviewers and an associate editor improved an earlier draft of this manuscript.

References

Anderson, D. R., K. P. Burnham, and W. L. Thompson. 2000. Null hypothesis testing: problems, prevalence, and an alternative. Journal of Wildlife Management 64:912–923.

Angers, C. 1984. Large sample sizes for the estimation of multinomial frequencies from simulation studies. Simulation 43:175–178.

Beyene, J., and R. Moineddin. 2005. Methods for confidence interval estimation of a ratio parameter with application to location quotients. BMC Medical Research Methodology 5:32. Available: www.biomedcentral.com. (November 2006).

Brenden, T. O., E. M. Hallerman, B. R. Murphy, J. R. Copeland, and J. A. Williams. 2007. The New River, Virginia, muskellunge fishery: population dynamics, harvest regulation modeling, and angler attitudes. Environmental Biology of Fishes 79:11–25.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach, 2nd edition. Springer-Verlag, New York.

Faes, C., M. Aerts, H. Geys, L. Bijnens, L. Ver Donck, and W. J. E. P. Lammers. 2006. GLMM approach to study the spatial and temporal evolution of spikes in the small intestine. Statistical Modelling 6:300–320.

Galinat, G. F., D. W. Willis, B. G. Blackwell, and M. J. Hubers. 2002. Influence of saugeye (sauger × walleye) introduction program on the black crappie population in Richmond Lake, South Dakota. North American Journal of Fisheries Management 22:1416–1424.

García, L. V. 2004. Escaping the Bonferroni iron claw in ecological studies. Oikos 105:657–663.

Gelman, A., and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York.

Ghahramani, S. 2000. Fundamentals of probability, 2nd edition. Prentice-Hall, Upper Saddle River, New Jersey.

Gustafson, K. A. 1988. Approximating confidence intervals for indices of fish population size structure. North American Journal of Fisheries Management 8:139–141.

Guy, C. S., R. M. Neumann, D. W. Willis, and R. O. Anderson. 2007. Proportional size distribution (PSD): a further refinement of population size structure index terminology. Fisheries 32:348.

Hurley, K. L., and J. J. Jackson. 2002. Evaluation of a 254-mm minimum length limit for crappies in two southeast Nebraska reservoirs. North American Journal of Fisheries Management 22:1369–1375.

Johnson, J. B., and K. S. Omland. 2004. Model selection in ecology and evolution. Trends in Ecology and Evolution 19:101–108.

Johnson, R. A., and D. W. Wichern. 1992. Applied multivariate statistical analysis, 3rd edition. Prentice-Hall, Upper Saddle River, New Jersey.

Legendre, P., and L. Legendre. 1998. Numerical ecology, 2nd edition. Elsevier, Amsterdam.

Martin, C. C. 1995. Evaluation of slot length limits for largemouth bass in two Delaware ponds. North American Journal of Fisheries Management 15:713–719.

McCulloch, C. E., and S. R. Searle. 2001. Generalized, linear, and mixed models. Wiley, New York.

McCullagh, P., and J. A. Nelder. 1989. Generalized linear models, 2nd edition. Chapman and Hall, London.

Miller, M. M., L. J. Gibson, and N. G. Wright. 1991. Location quotient: a basic tool for economic development analysis. Economic Development Review 9(2):65–68.

Miranda, L. E. 1993. Sample sizes for estimating and comparing proportion based indices. North American Journal of Fisheries Management 13:383–386.

Moineddin, R., J. Beyene, and E. Boyle. 2003. On the location quotient confidence interval. Geographical Analysis 35:249–256.

Neumann, R. M., and M. S. Allen. 2007. Size structure. Pages 375–421 *in* C. S. Guy and M. L. Brown, editors. Analysis and interpretation of freshwater fisheries data. American Fisheries Society, Bethesda, Maryland.

Novinger, G. D., and J. G. Dillard, editors. 1978. New approaches to the management of small impoundments. American Fisheries Society, Special Publication 5, Bethesda, Maryland.

Ortiz, M., and F. Arocha. 2004. Alternative error distribution models for standardization of catch rates of non-target species from a pelagic longline fishery: billfish species in the Venezuelan tuna longline fishery. Fisheries Research 70:275–297.

Pinheiro, J. C., and D. M. Bates. 2000. Mixed-effects models in S and S-PLUS. Springer-Verlag, New York.

Reynolds, J. B., and L. R. Babb. 1978. Structure and dynamics of largemouth bass populations. Pages 50–61 *in* G. D. Novinger and J. G. Dillard, editors. New approaches to the management of small impoundments. American Fisheries Society, Special Publication 5, Bethesda, Maryland.

SAS Institute. 2004. SAS/STAT 9.1 user's guide. SAS Institute, Cary, North Carolina.

SAS Institute. 2005. The GLIMMIX procedure. SAS Institute, Cary, North Carolina.

Sison, C. P., and J. Glaz. 1995. Simultaneous confidence intervals and sample size determination for multinomial proportions. Journal of the American Statistical Association 90:366–369.

Stone, C., and J. Lott. 2002. Use of a minimum length limit to manage walleyes in Lake Francis Case, South Dakota. North American Journal of Fisheries Management 22:975–984.

Thompson, S. K. 1987. Sample size for estimating multinomial proportions. American Statistician 41:42–46.

Venables, W. N., and C. M. Dichmont. 2004. GLMs, GAMs, and GLMMs: an overview of theory for applications in fisheries research. Fisheries Research 70:319–337.

Wagner, T., M. T. Bremigan, K. Spence Cheruvelil, P. A. Soranno, N. N. Nate, and J. E. Breck. 2007. A multilevel modeling approach to assessing regional and local landscape features for lake classification and assessment of fish growth rates. Environmental Monitoring and Assessment 130:437–454.

Wagner, T., D. B. Hayes, and M. T. Bremigan. 2006a. Accounting for multilevel data structures in fisheries data using mixed models. Fisheries 31:180–187.

Wagner, T., A. K. Jubar, and M. T. Bremigan. 2006b. Can habitat alteration and spring angling explain black bass nest distribution and success? Transactions of the American Fisheries Society 135:843–852.

Wand, M. P. 2003. Smoothing and mixed models. Computational Statistics 18:223–249.

Willis, D. W., B. R. Murphy, and C. S. Guy. 1993. Stock density indices: development, use, and limitations. Reviews in Fisheries Science 1:203–222.

Xiao, Y., A. E. Punt, R. B. Millar, and T. J. Quinn, II. 2004. Models in fisheries research: GLMs, GAMs, and GLMMs. Fisheries Research 70:137–139.