MODELING FISH LENGTH DISTRIBUTION USING A MIXTURE TECHNIQUE

Bahman Shafii

William J. Price

Charlie Holderman

Cathy Gidley

Paul J. Anders

See next page for additional authors

Author Information

Bahman Shafii, William J. Price, Charlie Holderman, Cathy Gidley, and Paul J. Anders

MODELING FISH LENGTH DISTRIBUTION USING A MIXTURE TECHNIQUE

Bahman Shafii and William J. Price
Statistical Programs, University of Idaho, Moscow, Idaho 83844
Charlie Holderman
Kootenai Tribe of Idaho, Bonners Ferry, Idaho 83805
Cathy Gidley
Idaho Department of Fish and Game, Coeur d'Alene, Idaho 83815
Paul J. Anders
Cramer Fish Sciences, Moscow, Idaho 83843

In fisheries science, length and age are important aspects of fish life history. Length is a function of growth, which provides an integrated measure of the environmental and endogenous conditions, e.g. genetics, affecting individuals and populations. Length at age data can be used to assess quality and quantity of habitat, food availability, or the need for and influence of management activities. Statistical mixture techniques may be used as a means to effectively model fish length distribution. A three-component mixture model, based on normal variates, was employed to describe length distribution in mountain whitefish species. The resulting model provided parameter estimates with meaningful biological interpretations, which were in turn used for inferential and comparative purposes. The technique will be demonstrated with reference to seven years of bio-monitoring data collected from the Kootenai River in Northern Idaho prior to and post nutrient addition treatment.

I. Introduction

Length, growth, and age are important life history metrics used by fisheries scientists and managers to assess health, condition, and performance of individual fish and fish populations (Murphy and Willis 1996; Guy and Brown 2007). Length at age is a function of growth, which provides a biologically integrated measure of past and current environmental and endogenous conditions (Dutta 1994) and can be used to assess habitat quality, food availability, biological condition, and the need for and responses to fisheries management activities (Guy and Brown 2007; McCarthy et al. 2009). Because fish size affects both aesthetic and economic value in recreational and commercial fisheries, an understanding of growth, length at age, and recruitment dynamics in fish populations is required to accurately model and predict fishery and population trends.

Characterizing temporal variation in length at age distributions can provide valuable insight into the size and frequency of year class production and recruitment to older age classes. These characteristics, in turn, can affect population persistence and viability (McElhaney et al.

2000). Length distributions are often multimodal, reflecting inter-annual variability in year class production, recruitment success, and environmental conditions. This multimodal structure can be represented using mixture modeling techniques that enable statistical assessment of trends in fishery composition (Fleishman and Burwen, 2003). In this paper, a mixture model form was used to describe and compare the length distributions of mountain whitefish (MWF; *Prospoium williamsoni*), a native species and key ecological indicator in the Kootenai River, a large altered river in Northern Idaho.

II. Methods

Traditionally, mixture models are utilized to represent multimodal probability distributions (see, for example, McLachlan and Peel, 2000). That is, the method is employed to model probability events that do not conform to standard distributional forms. In this application, however, the techniques of mixture modeling provided a multimodal functional form in addition to information on potential sub-population structure in the data.

The general form of the mixture model is given by:

$$
\begin{aligned}
\mathbf{y_{ij}} &= \mathbf{p_1}\, f_1(\mathbf{x_i}) + \mathbf{p_2}\, f_2(\mathbf{x_i}) + \ldots + \mathbf{p_m}\, f_m(\mathbf{x_i}) + \boldsymbol{\varepsilon_{ij}} \\
&= \sum_{\mathbf{k}}^{\mathbf{m}} \mathbf{p_k}\, f_k(\mathbf{x_i}) + \boldsymbol{\varepsilon_{ij}}
\end{aligned}
\tag{1}
$$

where y_{ij} is the jth replication of the proportion of fish at the ith length, x_i and an error term, ε_{ij}, assumed $\varepsilon_{ij} \sim NID(0, \sigma^2)$. The terms $f_1(x_i)$ through $f_m(x_i)$ are nonlinear functional forms for the m sub-populations or components. Typically, the $f_k(x_i)$ are represented assuming simple density functions. Here, the $f_k(x_i)$ will be specified by a two parameter normal distribution.

While other component distributions may be considered, the normal density is easily implemented and the parameters have meaningful interpretation. In this case, the component parameters, μ_k and σ_k, have relevant biological interpretation regarding the average length and variability of each sub-population, respectively. Sub-populations enter the model with proportions p_1 through p_m, respectively, with the constraint of: $\sum p_k = 1.0$.

Model estimation was carried out using nonlinear regression with a maximum likelihood algorithm (normal likelihood assumed) and an expected value given by eq. (1) and a normal density $f_k(x_i) = N(\mu_k, \sigma_k^2)$. Separate estimation was used for each of two (i.e. prior and post) nutrient addition scenarios. Following estimation, fish distribution parameters were compared across management conditions using dummy variable techniques. All statistical computations were carried out using SAS 9.2 (SAS, 2008).

III. Demonstration

Background

The Kootenai River begins with headwaters in Southeastern British Columbia, Canada, and flows southwest through Montana and Northern Idaho before exiting the United States back into Canada and Kootenay Lake (Figure 1). The Kootenai River watershed is situated between 48° and 51° N latitude and 115° and 118° W longitude, and measures 238 miles by 153 miles and has an area of 16,180 sq miles (KTOI and MFWP, 2004). Construction of Libby Dam, on the Kootenai River near Libby, MT, was completed in 1972 to control flooding and provide hydroelectric power Following dam construction, however, nutrient deficient (ultra-oligotrophic) conditions developed downstream from the dam. To mitigate nutrient losses due to impoundment by the dam, an experimental nutrient (phosphorous) addition treatment was proposed and implemented. As part of this treatment proposal, a bio-monitoring program for the Kootenai River system was initiated in 2000. This program collects data regarding water quality, as well as, algal, macroinvertebrate and fish trophic levels. For this last trophic category, fish samples were collected through electro-fishing techniques in the fall of each year from 2002 to 2008. Sampling occurred at seven sites along the river system with 6 replicates per site. Approximately 1300-2200 individual fish were measured per site for length and weight each year. Since a time consuming and labor intensive effort is required for measuring age (scale measurements), a smaller sub-sample of 200-600 fish were aged. Over seven years, 22 species of fish were identified. Of these, 97% of the data were represented by 6 dominant species with 43% represented by a single species, mountain whitefish (*Prosopium williamsoni*). Because of its prevalence, and its status as an ecologically important indicator species, MWF was selected here for modeling and analysis. Data are included from five of the sample sites, which lie adjacent to the point of nutrient addition at the Idaho-Montana border (Nutrient Addition Zone, Figure 1). Treatments began in July of 2005 and have been subsequently implemented from mid-summer to the fall of each year. Hence, the data were further classified into two treatment classes: Prior to nutrient addition (2002-2004), and Post Nutrient addition (2005-2008).

Fish Age

Within the subset of aged fish data, an initial assessment of the relationship between fish age and fish length showed a high level of correspondence (r=0.87; Figure 2). Further examination of the distribution MWF lengths in the age data revealed three distinct modes or sub-populations. These sub-populations corresponded well to the age groups of "0 years", "1-2 years", and "3$^+$ years" (Figure 3). While age measurements were not available for the full MWF dataset, plots of the respective length distributions of both the prior and post treatment data showed a similar tri-modal trend, implying that the application of similar age classes to the full dataset was warranted (Figures 4a and b).

Mixture Model and Estimation

Based on the tri-modal empirical distribution, three age classifications were assumed for the data. A model following eq (1) was set up with k=1, 2, 3 and sub-population density components, $f_k(x_i)$, assigned respective normal densities, $N(\mu_k, \sigma_k^2)$. Based on the preliminary age at length data, the sub-populations were assumed to represent age classes 0, 1-2, and 3$^+$

years, respectively, with p_3 defined as $1 - (p_1+p_2)$. Separate estimations were carried out for each of the prior or post treatment datasets.

Estimated parameter values are shown in Table 1. In both treatments, all parameter estimates were significant. Within each treatment, the estimated mean lengths, μ_k, of the three components increased systematically with age. For example, in the prior to nutrient addition treatment (2002-2004), the age classes 0, 1-2 and 3^+ years were estimated to have mean lengths of 98, 192, and 265 mm, respectively, while the post treatment model (2005-2008) gave values of 112, 201, 268 mm. Estimates of variability, σ_k, showed less consistency, although the oldest age class, 3^+, had the highest value in both treatments. Overall, the models fit well with acceptable residual patterns and predicted values which followed the data well (Figures 5a and b).

Model Comparison

Given the aforementioned estimated models, comparisons across prior and post nutrient addition treatments were evaluated using a full dummy variable model combined with single degree of freedom contrasts. Specifically, hypotheses regarding the differences in estimated mean and variability values of the respective treatment age classes were assessed in addition to the sub-population proportions, p_k, which reflect the relative prominence of each age class (Figure 6).

Results for the specified contrasts are shown in Table 2. Treatment differences in relative proportions for age classes 0 and 1-2 years were non-significant, indicating that post treatment MWF length increased. Differences in mean lengths of older fish were not significant. Significant decreases in the variability of length were found for age classes 1-2 and 3^+ years, implying that those fish became more consistent in length following the nutrient addition treatments. Changes in younger fish at age class 0 were non-significant in this regard.

IV. Concluding Remarks

Length at age is a useful metric for evaluating growth, fitness, health and demographic trends of fish populations. Distribution of fish lengths within a population often reveals a sub-population structure corresponding to fish age due to natural variation in the magnitude and frequency of year class production and recruitment. Models for fish length should, therefore, account for this sub-population structure. Mixture modeling techniques are useful in this regard and can provide interpretable parameter estimates related to these important sub-population characteristics.

An example of mixture modeling for mountain whitefish under two nutrient addition conditions was given. The mixture models, in this case, demonstrated significant positive changes in the annual fish length-age distributions in response to an experimental nutrient addition program on the Kootenai River. The fish appear to be displaying a positive response to treatments, increasing in size relative to their age. Assessments such as these provide a valuable

method for determining the effectiveness of management actions (e.g. nutrient addition treatments and evaluating the effects of environmental variability on fish population health and status. Such quantitative assessments also provide useful information for the bio-monitoring project and fisheries managers that can help maintain fishery benefits and population persistence.

V. References

Dutta, H. 1994. Growth in Fishes. Gerontology (40): 97-112.

Fleishman, S. J. and D. L. Burwen. 2003. Mixture models for the species apportionment of hydroacoustic data, with ech-envelope length as a discriminatory variable. ICES journal of Marine Science, 60: 592-598.

Guy, C. S. and M.L. Brown. 2007. Analysis and interpretation of freshwater fisheries data. American Fisheries Society. Bethesda MD.

KTOI and MFWP, 2004. Kootenai Tribe of Idaho and Montana Fish, Wildlife & Parks. 2004. Kootenai Subbasin Plan. A report prepared for the Northwest Power and Conservation Council. Portland, OR. 799 pp. *www.nwcouncil.org/fw/subbasinplanning/kootenai/plan/*

McElhany, P., M.H. Ruckelshaus, M.J. Ford, T.C. Wainwright, and E.P. Bjorkstedt. 2000. Viable salmonid populations and the recovery of evolutionarily significant units. U.S. Dept. Commer., NOAA Tech. Memo. NMFS-NWFSC-42,156 p.

McCarthy, S.G., J. J. Duda, J. M. Emlen, G. R. Hodgson, and D. A. Beauchamp. 2009. Linking Habitat Quality with Trophic Performance of Steelhead along Forest Gradients in the South Fork Trinity River Watershed, California. Transactions of the American Fisheries Society 138:3, 506-521

McLachlan, G.J. and Peel, D. 2000. Finite Mixture Models, , John Wiley and Sons, Inc.

Murhpy, B.R. and D. W. Willis, editors. 1996. Fisheries techniques. American Fisheries Society Publishers, Bethesda MD. 732 pp.

SAS Institute Inc. 2008. SAS OnlineDoc® 9.2. Cary, NC: SAS Institute Inc.

Table 1. Estimated mixture model parameters for two nutrient addition treatments.

Prior to Nutrient Addition

Parameter Estimates			
Parameter	Estimate	Std Err	Pr > |t|
p1	0.11	0.017	<.0001
p2	0.26	0.024	<.0001
$\mu 1$	98.47	2.06	<.0001
$\mu 2$	192.59	1.53	<.0001
$\mu 3$	265.28	1.32	<.0001
$\sigma 1$	11.32	2.63	<.0001
$\sigma 2$	16.96	1.60	<.0001
$\sigma 3$	27.99	1.57	<.0001

Post Nutrient Addition

Parameter Estimates			
Parameter	Estimate	Std Err	Pr > |t|
p1	0.08	0.014	<.0001
p2	0.25	0.019	<.0001
$\mu 1$	112.24	3.03	<.0001
$\mu 2$	201.11	0.71	<.0001
$\mu 3$	268.35	1.68	<.0001
$\sigma 1$	13.80	3.30	<.0001
$\sigma 2$	10.93	0.70	<.0001
$\sigma 3$	34.11	1.80	<.0001

Table 2. Contrast results comparing mixture model parameter estimates between two nutrient addition treatments.

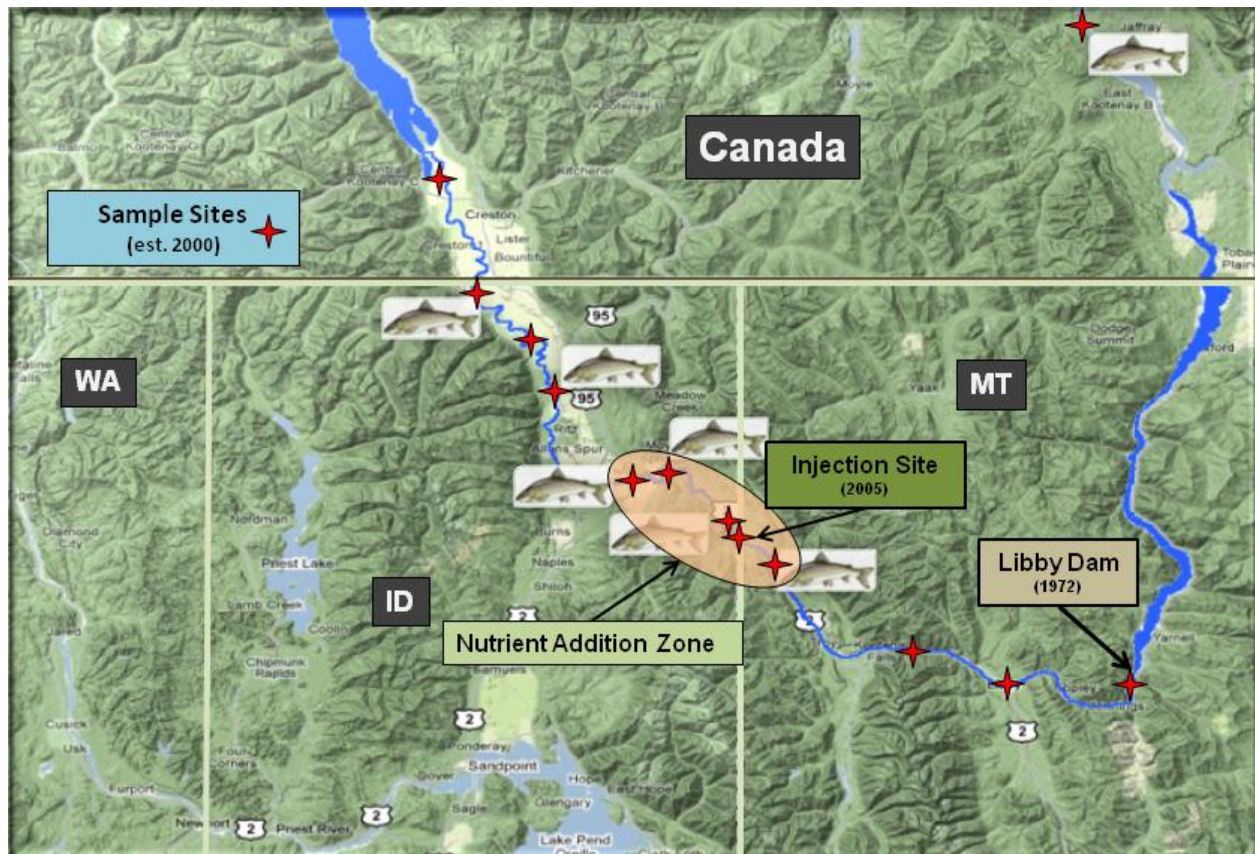| Prior & Post Treatment Comparisons		
Contrast	DF	Pr > F
Age 0 proportion (p1)	1	0.2803
Age 1-2 proportion (p2)	1	0.7163
Age 0 Mean ($\mu 1$)	1	0.0002
Age 1-2 Mean ($\mu 2$)	1	<.0001
Age 3-10 Mean ($\mu 3$)	1	0.1525
Age 0 Std ($\sigma 1$)	1	0.5569
Age 1-2 Std ($\sigma 2$)	1	0.0006
Age 3-10 Std ($\sigma 3$)	1	0.0107

Figure 1. Kootenai River bio-monitoring project map. Stars indicate all sample sites, while fish indicate sites where fish samples were taken. The nutrient addition zone, where model analyses were done, is outlined in orange.
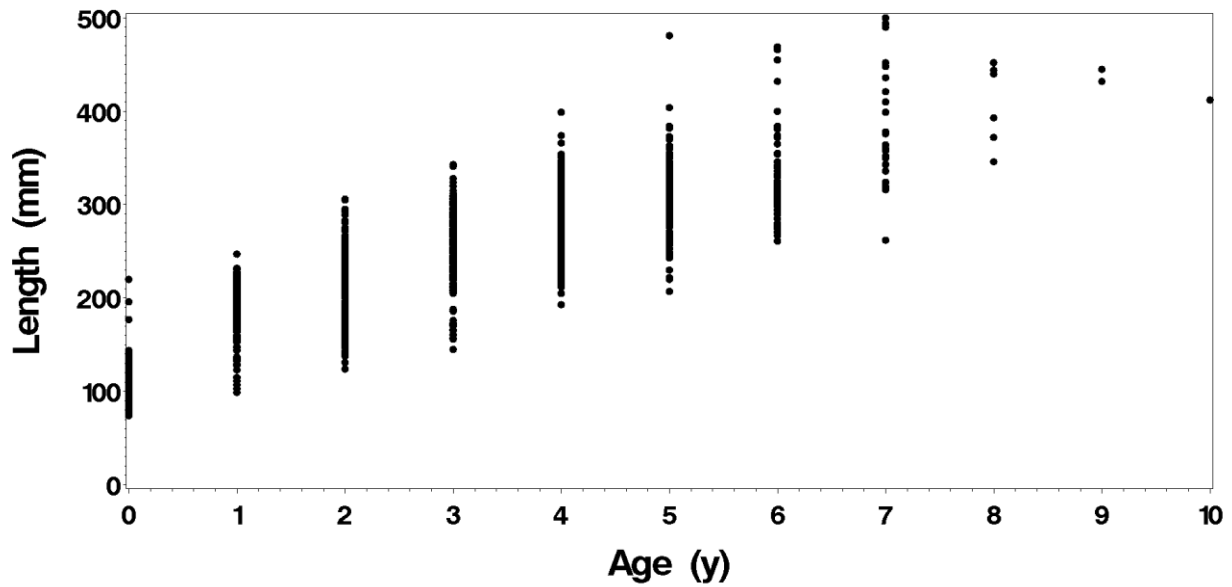
Figure 2. The association between fish length and fish age for the species mountain whitefish.
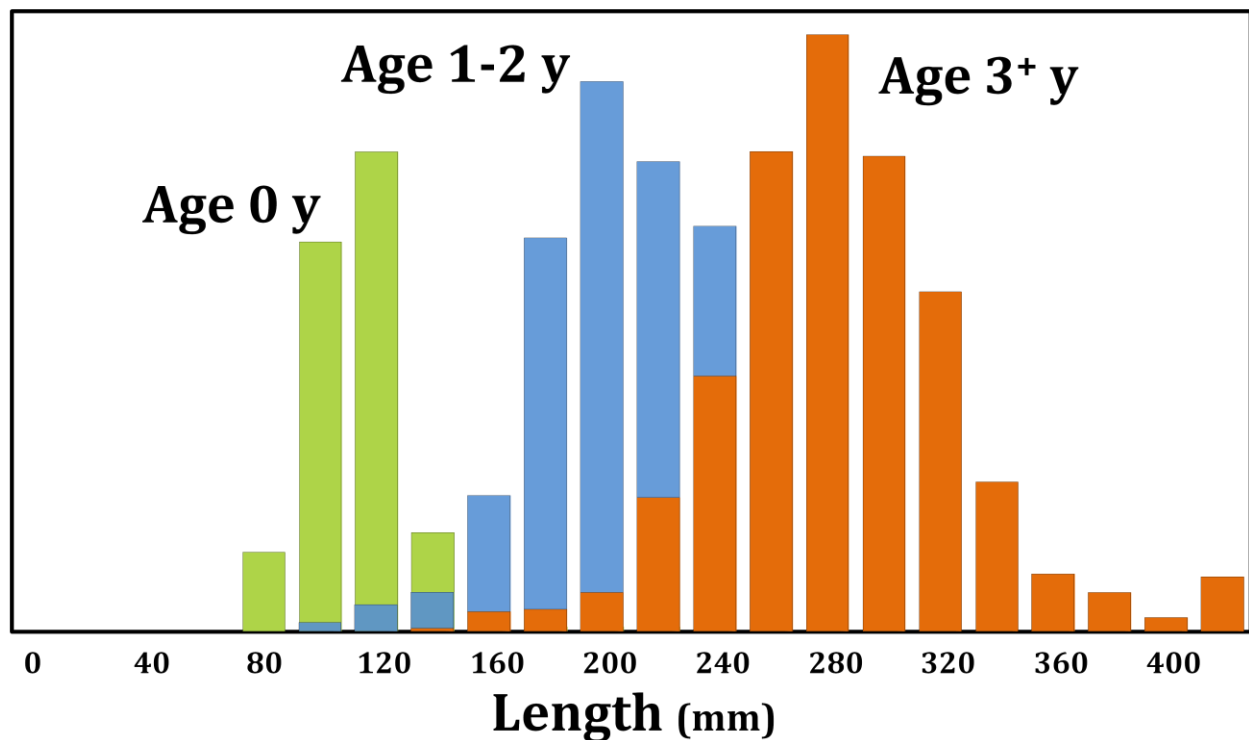

Figure 3. The distribution of mountain whitefish lengths in the aged fish data. Three age related sub-populations are indicated.
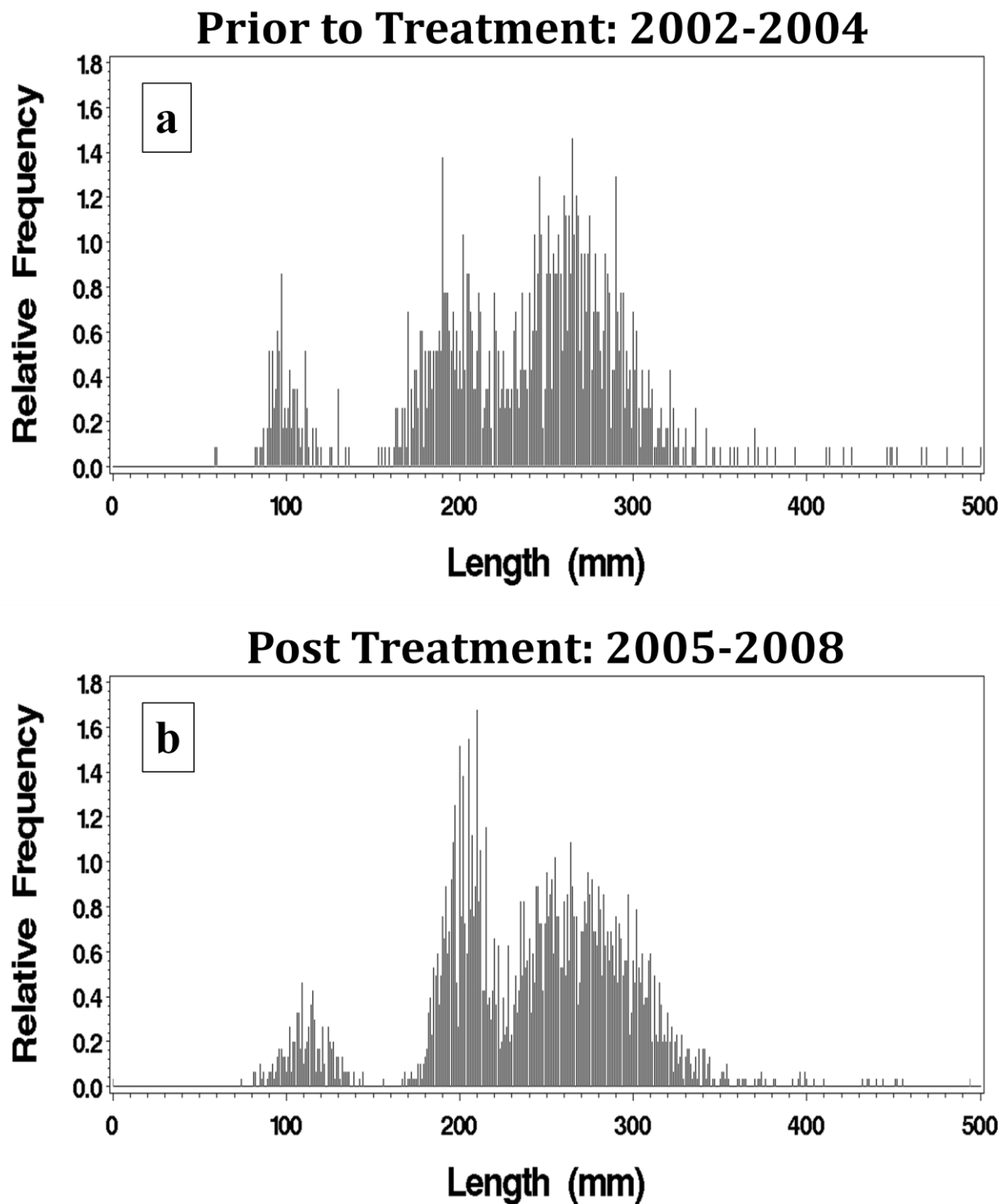
Prior to Treatment: 2002-2004

Post Treatment: 2005-2008

Figure 4. Length distributions for prior to (a) and post (b) nutrient addition treatments. In both cases, three sub-populations are indicated.
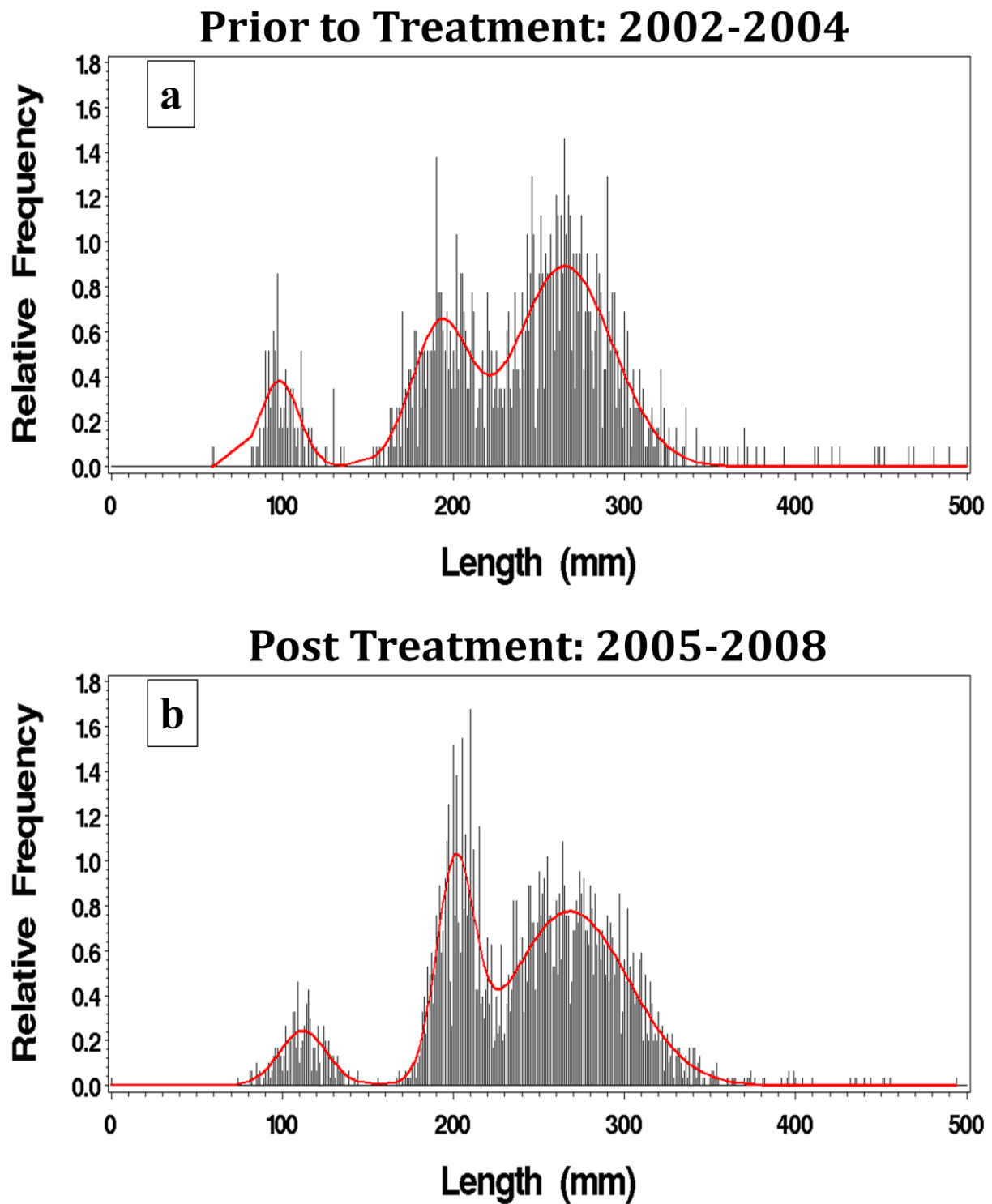
Prior to Treatment: 2002-2004

Post Treatment: 2005-2008

Figure 5. Estimated mixture models for mountain whitefish length distributions prior to (a) and post (b) nutrient addition.
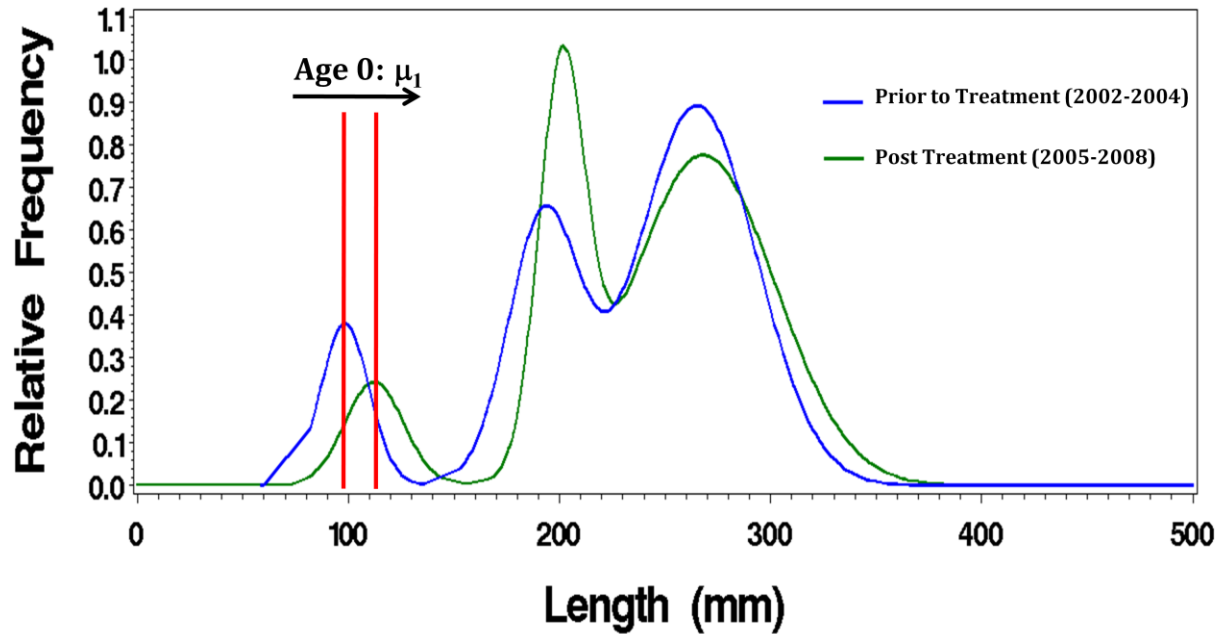
Figure 6. Comparison of estimated mixture models for two nutrient addition treatments. An example of the positive shift in mean length for age 0 fish is shown.