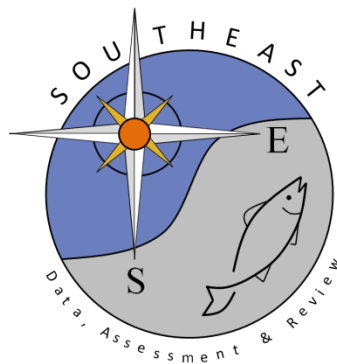# Reliability testing of non-equilibrium mean length mortality estimation routines

Victoria P. Simmons, Quang C. Huynh, Elizabeth A. Babcock, and William J. Harford

SEDAR57-AP-02

3 November 2018

Please cite this document as:

Reliability testing of non-equilibrium mean length mortality estimation routines

Victoria P. Simmons
Department of Marine Ecosystems and Society
Rosenstiel School of Marine and Atmospheric Science
University of Miami
4600 Rickenbacker Causeway, Miami, FL 33149
victoria.simmons@rsmas.miami.edu

Quang C. Huynh
Institute of the Oceans and Fisheries
University of British Columbia
Vancouver, Canada
q.huynh@oceans.ubc.ca

Elizabeth A. Babcock
Department of Marine Biology and Ecology
Rosenstiel School of Marine and Atmospheric Science
University of Miami
4600 Rickenbacker Causeway, Miami, FL 33149
ebabcock@rsmas.miami.edu

William J. Harford
Cooperative Institute for Marine and Atmospheric Studies
Rosenstiel School of Marine and Atmospheric Science
University of Miami
4600 Rickenbacker Causeway, Miami, FL 33149
william.harford@noaa.gov

November 3, 2018

Background

Beverton and Holt (1957) were amongst the first to describe mathematically the inverse relationship between the mean length of fish in a population and the total mortality rate, Z, experienced by the population. Assumptions of the Beverton-Holt approach include constant growth parameters, no individual variability in growth, constant and continuous recruitment over time, mortality rate that is constant over time, and population size that is in an equilibrium state (Gedamke and Hoenig 2006).

Gedamke and Hoenig (2006) developed a model that is a non-equilibrium extension of the Beverton-Holt approach. The Gedamke-Hoenig approach follows the same basic assumptions of the Beverton-Holt approach, with the exception of population equilibrium. Their approach estimates period-specific total mortality along with the break points that mark transitions between periods in which different mortality rates occur (Gedamke and Hoenig, 2006). As a consequence of the complexity of estimating break points in time series data along with the magnitude of the changes in Z associated with each break point, careful development and testing of fitting routines is necessary. To estimate both the total mortality and the corresponding number and temporal position of break points, the R library known as MLZ has been developed, which implements a maximum likelihood estimation routine (Huynh et al., 2018).

In this report, the MLZ library is subject to a simulated self-test to evaluate the reliability of parameters estimates obtained from the Gedamke-Hoenig derivation of changes in mean length. Self-testing consists of generating datasets from the same sets of equations that are then used in the fitting routine that estimates model parameters. We evaluate estimator reliability for a variety of scenarios that differ according to fish life

history. Each life history is characterized by the ratio of natural mortality, *M*, to the von Bertalanffy growth coefficient, *K*, where *M/K* ratios represent life history variation across fish stocks (Hordyk et al., 2015; Prince et al., 2015). We also examine the reliability of estimating temporal changes in *Z* under fishery scenarios that differ with respect to lengths of first capture. Self-testing is useful for understanding the limitations on estimation of model parameters (Deroba et al., 2015). For instance, we reaffirm the known difficulties in estimating changes in Z, particularly when those changes occur close to the end point of a time series of mean lengths. Similarly, breaks points may be missed by the fitting routine when changes in *Z* occur very close together in time. This guidance is advanced through our analysis, and is germane to the process of providing reliable information about the state of a fish stock. Identifying substantive break points can reveal when important changes occurred in a fishery. Obtaining Z estimates at the beginning of a time series can be used to describe the state of the stock during the time period when data collection was initiated, while Z estimates at the end of a time series reflect near-term conditions and can inform whether recent catches are likely to be sustainable. Each of these quantities, estimated from time series of mean length observations, can be instrumental in informing development of effective length-based fishery management strategies.

Methods

Simulating mean length observations

Gedamke and Hoenig (2006) demonstrated a transitional population in terms of changes to mean length that occur *d* years after a permanent change in total mortality

from Z_1 year^{-1} to Z_2 year^{-1}. Mean length of a population d years after the change in mortality is

$$\mu = \left\{ \int_{t_c}^{g} R \cdot \exp[-Z_2(t - t_c)]L_t dt + \int_{g}^{\infty} R \cdot \exp(-Z_2 d) \cdot \exp[-Z_1(t - g)]L_t dt \right\} \div$$

$$\left\{ \int_{t_c}^{g} R \cdot \exp[-Z_2(t - t_c)]dt + \int_{g}^{\infty} R \cdot \exp(-Z_2 d) \cdot \exp[-Z_1(t - g)]L_t dt \cdot \right\}. \qquad (1)$$

Age $g = t_c + d$. The first integral in both the numerator and denominator represents fish recruited after the change in mortality, which have only experienced the mortality rate of Z_2. The second integral represents fish that were recruited before the change in mortality and thus have experienced both Z_1 and Z_2. After integration and simplification (these calculations are described in Gedamke and Hoenig (2006)), mean length d years after a change is written:

$$\mu = L_\infty - Z_1 Z_2 (L_\infty - L_c) \times \{Z_1 + K + (Z_2 - Z_1) \exp[-(Z_2 + K)d]\}$$

$$\div \{(Z_1 + K)(Z_2 + K)[Z_1 + (Z_2 - Z_1) exp(-Z_2 d)]\}. \qquad (2)$$

Recognizing that the example derivation stated above describes the length transition dynamics for a single break point, generalization of this approach can be made to any number of changes in total mortality Z. Gedamke and Hoenig (2006) produce this generalized derivation. Using this derivation, annual mean lengths can be calculated from a mortality schedule or vice versa. Assume there are $n=1,2,…,N$ years of mean length data in years $y=Y_1,Y_2,…,Y_N,$ where Y_n is the calendar year. If there are k changes in total mortality Z, then divide the time series to $i=1,2,…,k+1$ partitions where the population experiences mortality rate Z_i in the i-th partition. A vector of mortality rates $\vec{Z} = \{Z_1, Z_2, … Z_{k+1}\}$ then describes the historical mortality rates in the population over time

with corresponding change points $\vec{D} = \{D_1, D_2, \dots D_k\}$, where D_i is the point in time (in calendar years) when there is a stepwise change in mortality rate from Z_i to $Z_{(i+1)}$. Given the above mortality schedule, the predicted mean length µ in year y is:

$$\mu_y = L_\infty \frac{\sum_{i=1}^{k+1} \frac{a_{i,y} b_{i,y}}{Z_{k+2-i}} - \left(1 - \frac{L_c}{L_\infty}\right) \sum_{i=1}^{k+1} \frac{r_{i,y} s_{i,y}}{Z_{k+2-i} + K}}{\sum_{i=1}^{k+1} \frac{a_{i,y} b_{i,y}}{Z_{k+2-i}}}$$

(3)

Where

$$a_{i,y} = \begin{cases} 1 & i = 1 \\ \exp\left(-\sum_{j=1}^{i-1} Z_{k+2-j} d_{k+1-j,y}\right) & i = 2,\dots k+1 \end{cases} \tag{4}$$

$$b_{i,y} = \begin{cases} 1 - \exp\left(-Z_{k+2-i} d_{k+1-i,y}\right) & i = 1,\dots,k \\ 1 & i = k+1 \end{cases} \tag{5}$$

$$r_{i,y} = \begin{cases} 1 & i = 1 \\ \exp\left(-\sum_{j=1}^{i-1} (Z_{k+2-j} + K) d_{k+1-j,y}\right) & i = 2,\dots k+1 \end{cases} \tag{6}$$

$$s_{i,y} = \begin{cases} 1 & i = 1,\dots,k \\ 1 - \exp\left(-(Z_{k+2-i} + K) d_{k+1-i,y}\right) & i = k+1 \end{cases} \tag{7}$$

And

$$d_{i,y} = \begin{cases} 0 & y \le D_i \\ y - D_i & D_i < y \le D_{i+1} \\ D_{i+1} - D_i & y > D_{i+1} \end{cases} \tag{8}$$

The second subscript in $d_{(i,y)}$ indexes the change point i and each year y.

In year y, the relative population abundance in year y after dividing out recruitment is the denominator of Equation (3), while the total length of all animals in year y is calculated in the numerator. The term $\frac{a_{i,y} b_{i,y}}{Z_{k+2-i}}$ is the relative abundance of animals that are recruited to the population when the mortality rate was Z_{k+2-i},

The equations describing transitional dynamics in length were used to generate

simulated datasets, based on a specified number of breaks, k. Each simulated dataset of

annual mean length was $N=30$ years in duration, and we evaluated datasets having zero,

one, or two breaks. Given a specified number of break points, the duration of Z_i were

obtained using a vector of proportions $\vec{p} = p_1, p_2, \ldots, p_{k+1}$ where all $p_i > 0$ and

$\sum_{i=1}^{k+1} p_i = 1$. Change points, $\vec{D}$, are obtained as:

$$D_i = Y_1 + N \sum_{j=i}^{i} p_j \qquad (9)$$

Thus, the i-th change point is now a function of the cumulative sum of the duration of

prior mortality rates, allowing for ordered change points in estimation. Simulation of $\vec{p}$

was specified as a diffuse Dirichlet distribution with $\alpha = (1, \ldots, 1)$. Change points are

continuous random variables since the transitional dynamics of mean length are

formulated as a function of elapsed time since the previous change in Z, thus, break

points (in years and fractions of a year) are location parameters that reflect an offset

relative to a previous change in Z. Values of Z were generated from a uniform

distribution with a lower bound at M, and a maximum of $4M$.

Parameters describing growth according to the von Bertalanffy growth function (von

Bertalanffy, 1938) as well as the minimum capture length, L_c, are required to generate

mean length time series from the transitional dynamics equations. Different life history

types were represented in our analysis through specification of the ratio of natural

mortality to the von Bertalanffy growth rate constant, K. The M/K ratio is known as a

Beverton-Holt life history invariant; however, considerable variability across taxa is

recognized in this ratio and differences in life history types are useful in determining

fishery management strategies (Prince et al., 2015). We considered M/K ratios of 0.5, 1,

and 2 that cover a broad array of marine taxa that have been subject to life history meta-analysis (Prince et al., 2015). Values of K that were calculated based on specified M/K were always based on an assumed M of 0.2 year^{-1}. Asymptotic length, L_∞, is a scaling parameter that was fixed at an arbitrary value of 100 cm for all simulations. Two levels of L_c were specified as proportions of L_∞, with a values of L_c=0.66 L_∞, meant to reflect recruitment to the fishery corresponding with length at which 50% of the population has reached maturity (Prince et al., 2015) and a second scenario of L_c=0.80 L_∞ reflecting delayed recruitment of fish to the fishery.

Five-hundred simulated datasets were generated for each factorial combination of number of breaks (3 levels), M/K (3 levels), and L_c (2 levels). Simulated 'true' mean lengths were observed with an arbitrarily low observation error. This observation error was specified as a Gaussian sampling distribution with mean, $\bar{L}$, and variance of the mean equal to one. Low observation error was specified because our objectives do not pertain to the reliability of length-based sampling designs, which would be better addressed by simulating the more complex process of observing individual lengths (e.g., Hulson et al., 2012) and not through simply increasing the variance of a summary statistic like mean length. Input parameters of K, L_∞, L_c were provided to the fitting routine without error.

Fitting model parameters to simulated datasets using MLZ

The R library MLZ contains a maximum likelihood fitting routine that estimates the timing of breaks points and associated Z(s) (Huynh et al., 2018; R Development Core Team, 2012). The log-likelihood of the model is

$$\log L = -n \log \sigma - \frac{1}{2\sigma^2} \sum_{y=Y_1}^{Y_N} m_y (\bar{L}_y - \mu_y)^2 \qquad (10)$$

where $\overline{L}_y$ is the observed mean length in year y and m_y is the corresponding sample size, and σ is a dispersion parameter of the normal distribution

Because the fitting routine does not estimate the total number of break points, the user specifies the number of break points, the estimation process must be carried out separately for each number of breaks that are considered plausible. Thus, model selection criteria for discriminating among candidate models was required. For batch processing of our simulated datasets, candidate models were those with 0, 1, 2, or 3 break points. From the candidate model set, the 'best approximating model' was identified using Akaike Information Criterion (AIC; Burnham and Anderson, 2002). AIC assesses model fit in relation to model complexity. Model fit, calculated as two times the negative of the log-likelihood, is penalized by the quantity of two times the number of parameters in the model to produce an AIC score (Burnham and Anderson, 2002). The number of parameters required to estimate the transitional behavior of Z(s) is calculated as two times the number of breaks plus two, which accounts for parameters describing the positions of break(s), total mortality parameters, and a parameter for the residual error. AIC provides a ranking system for the candidate models with the best approximating model having the lowest AIC score. During batch processing, the model with lowest AIC score was selected from model fits to each simulated dataset.

Performance metrics

Performance metrics were calculated separately for each tested scenario, represented by the combination of *M/K* ratio, number of simulated breaks, and length at first capture to asymptotic length (L_c/L_∞) ratio. The percentage of correctly identified number of simulated breaks was calculated as the number of simulations in which the number of

estimated breaks equaled the number of simulated breaks divided by the total number of simulations and multiplied by 100 to give a percentage.

To determine fitting routine reliability in estimating Z at the beginning of the mean length time series, known as initial Z, percent bias in Z_1 was calculated as:

$$\frac{Z_{estimated} - Z_{simulated}}{Z_{simulated}} * 100\% \quad (11)$$

An interval that encompasses the centered 95% of the Z_1 bias was determined by setting the 2.5th quantile point as the lower bounds, and the 97.5th quantile as the upper bounds across the simulations. To determine how well the fitting routine estimated terminal Z ($Z_{terminal}$), regardless of whether the total number of breaks was accurately predicted, the percent bias for $Z_{terminal}$ was calculated according to Equation (11). As number of simulated breaks may differ from number of estimated breaks in the model selected by the fitting routine, these calculations were made by ensuring that the simulated $Z_{terminal}$ was compared to the estimated $Z_{terminal}$, regardless of whether the number of break points differed in each comparison. An interval that encompasses the centered 95% of the $Z_{terminal}$ bias was calculated as for Z_1.

As the largest percent bias for $Z_{terminal}$ occurred when simulated breaks was equal to two, a fourth metric was devised to summarize how the position of the simulated second break point in relation to the end of the mean length time series could be influencing estimator reliability. Bias in $Z_{terminal}$ was categorized according to those simulation runs in which the simulated $Z_{terminal}$ occurred greater than ten years from the end of the time series, between three and ten years from the end of the time series, and within three years from the end of the time series. Given these partitions in the position of the second break point, the centered 95% of $Z_{terminal}$ percent bias was calculated.

Results

When two breaks were simulated, the maximum likelihood fitting routine correctly

selected this model structure from the candidate set 65.8-73.4% of the time (Table 1).

The simulation configuration with two breaks had the highest percentage of being

correctly selected; whereas, simulated breaks of zero or one were correctly estimated 51.6

to 70.8% of the time (Table 1). Percent bias for initial Z was most pronounced e.g., -0.9

to 0.6% when number of simulated breaks is 2 (Fig. 1). This bias is larger in comparison

to simulations with zero or one break, with a percent bias between -0.1 and 0.1% and -0.6

and 0.4%, respectively. Likewise, $Z_{terminal}$ had the largest percent bias when number of

simulated breaks was 2, with centered 95% of bias from -17.6 to 34.8% (Fig. 2). For

simulations in which there are zero or one break, the centered 95% of bias in $Z_{terminal}$ bias

was -0.4 to 8.1%.

In order to investigate why percent bias tends to be much larger for simulations in

which simulated breaks is equal to two, these simulations were sorted according to

proximity in years to the end of the time series in which the last predicted break occurred

(Fig. 3). For the subset of simulation runs in which the last predicted break occurred

within three years from the end of time series, the upper bound of the centered 95% of

$Z_{terminal}$ bias was 135.9%. This upper confidence interval bound for $Z_{terminal}$ is appreciably

higher than that reported for all simulations combined because the smaller subset of

simulations (i.e., those with break points < 3 years before the end of time series) fell

outside of the 95% CI for the larger pooled dataset.

Discussion

In this report, the MLZ library was subject to simulation testing to evaluate the reliability of a fitting routine used to estimate the parameters of the Gedamke-Hoenig derivation of changes in mean length. Having completed this self-testing, we offer usage guidance to implementing this fitting routine for the evaluation of data-limited fisheries. As the number of breaks was increased (in our simulations this corresponded to a maximum of two breaks), the estimated $Z_{terminal}$ had a troubling level of bias. After further inspection, this large percent bias often coincided with simulated terminal break being close to the end of the time series (i.e., we partitioned simulated datasets as being within three years from the end of the time series). This result likely occurred because of a lack of information (i.e., several years of mean length following a change in Z), which resulted in either an inaccurate estimate $Z_{terminal}$ or a lack of detection of the last break point in time series, and thus, a failure to recognize a change in Z at all. Thus, when the fitting routine is applied to real data for data-limited fisheries, caution should be taken in relying on terminal Z estimates if a break point is believed to occur relatively close to the end of the time series. Nevertheless, this undesirable property of the mean length estimator (i.e., high uncertainty in $Z_{terminal}$ in certain situations) is not appreciably different from results obtained from self-testing of age-structured or surplus production models, which are considered standard approaches for stock assessment (Deroba et al., 2015).

When the number of breaks is incorrectly estimated, these results can occur when the break points are close together in the time series, and thus are overlooked by the model, or where there is a break point but Z does not differ too greatly between periods. In

addition, in our simulations where there were zero break points, we observed instances where >0 break points were erroneously selected through the fitting and model selection routine. This result highlights that false break points can occur due to observation error, even at the relatively low level of observation variance we included in our simulations. This result illustrates a potential sensitivity to observation error that can be translated into perceived break points. However, in our simulations, false break points tended to also result in similar mortality estimates before and after the erroneous break point, thus diminishing the consequence of a perceived break point that was identified in error. In both the instance of a missed break point as well as the instance of a false break point, we found that $Z_{terminal}$ estimates were, on average, unbiased and 95% CI were reasonable (except when a simulated break point occurred close to the end of the time series). Thus, while the model selection criteria did not always select the correct model, $Z_{terminal}$, as an important management quantity, remained reasonably reliably estimated. Accordingly, reconstructing the temporal position of transition points remains a more uncertain aspect in application of the mean length mortality estimator methodology.

References

Beverton, R.J.H., Holt, S.J., 1957. On the dynamics of exploited fish populations. Chapman and Hall, London UK.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach. Second edition. Springer-Verlag, New York, New York, USA.

Deroba, J.J., Butterworth, D.S., Methot, R.D., De Oliveira, J. a. A., Fernandez, C., Nielsen, A., Cadrin, S.X., Dickey-Collas, M., Legault, C.M., Ianelli, J., Valero, J.L., Needle, C.L., O'Malley, J.M., Chang, Y.-J., Thompson, G.G., Canales, C., Swain, D.P., Miller, D.C.M., Hintzen, N.T., Bertignac, M., Ibaibarriaga, L., Silva, A., Murta, A., Kell, L.T., de Moor, C.L., Parma, A.M., Dichmont, C.M., Restrepo, V.R., Ye, Y., Jardim, E., Spencer, P.D., Hanselman, D.H., Blaylock, J., Mood, M., Hulson, P.-J.F., 2015. Simulation testing the robustness of stock assessment models to error: some results from the ICES strategic initiative on stock assessment methods. ICES J. Mar. Sci. 72, 19–30. https://doi.org/10.1093/icesjms/fst237

Gedamke, T., Hoenig, J.M., 2006. Estimating Mortality from Mean Length Data in Nonequilibrium Situations, with Application to the Assessment of Goosefish. Trans. Am. Fish. Soc. 135, 476–487. https://doi.org/10.1577/T05-153.1

Hordyk, A.R., Ono, K., Sainsbury, K., Loneragan, N., Prince, J., 2015. Some explorations of the life history ratios to describe length composition, spawning-per-recruit, and the spawning potential ratio. ICES J. Mar. Sci. 72, 204–216. https://doi.org/10.1093/icesjms/fst235

Hulson, P.-J.F., Hanselman, D.H., Quinn, T.J., 2012. Determining effective sample size in integrated age-structured assessment models. ICES J. Mar. Sci. 69, 281–292. https://doi.org/10.1093/icesjms/fsr189

Huynh, Q., Gedamke, T., Then, A.Y., 2018. MLZ: Mean Length-Based Estimators of Mortality using TMB. R package version 0.1.1.

Prince, J., Hordyk, A., Valencia, S.R., Loneragan, N., Sainsbury, K., 2015. Revisiting the concept of Beverton-–-Holt life-history invariants with the aim of informing data-poor fisheries assessment. ICES J. Mar. Sci. 72, 194–203. https://doi.org/10.1093/icesjms/fsu011

R Development Core Team, 2012. R: A Language and Environment for Statistical Computing, Vienna, Austria http://www.R-project.org. Vienna, Austria.

von Bertalanffy, L., 1938. A quantitative theory of organic growth (Inquiries on growth laws II). Hum. Biol 10, 181–213.

Table 1. For the maximum likelihood fitting routine, percentage of simulation runs in which the model selected by the fitting routine correctly predicted the number of simulated breaks. Tested scenario are represented by the combination of *M/K* ratio, number of simulated breaks, and the ratio of length at first capture to asymptotic length (L_c/L_∞).

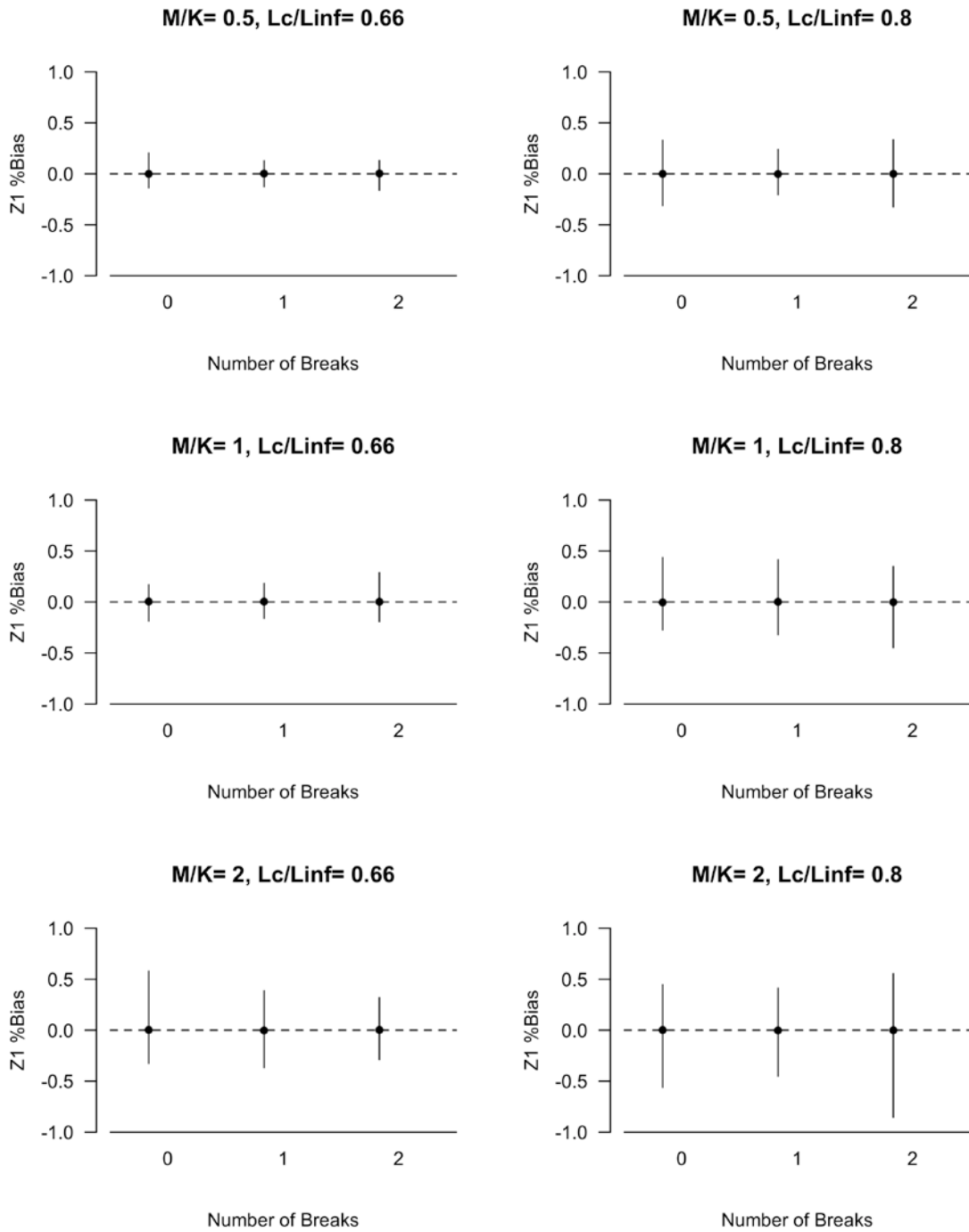|	Breaks	Max. Likelihood	
		Lc/Linf= 0.66	Lc/Linf= 0.88
M/K=0.5	0	51.6	52.0
	1	68.8	70.8
	2	70.8	73.4
M/K=1	0	53.2	53.8
	1	69.6	65.0
	2	71.0	70.2
M/K=2	0	56.0	55.2
	1	65.4	67.0
	2	65.8	65.8

Figure 1. Percent difference between predicted and simulated initial *Z* for the described scenarios for the maximum likelihood fitting routine tested. Points represent median values for percent bias; lines represent the centered 95% of percent bias.
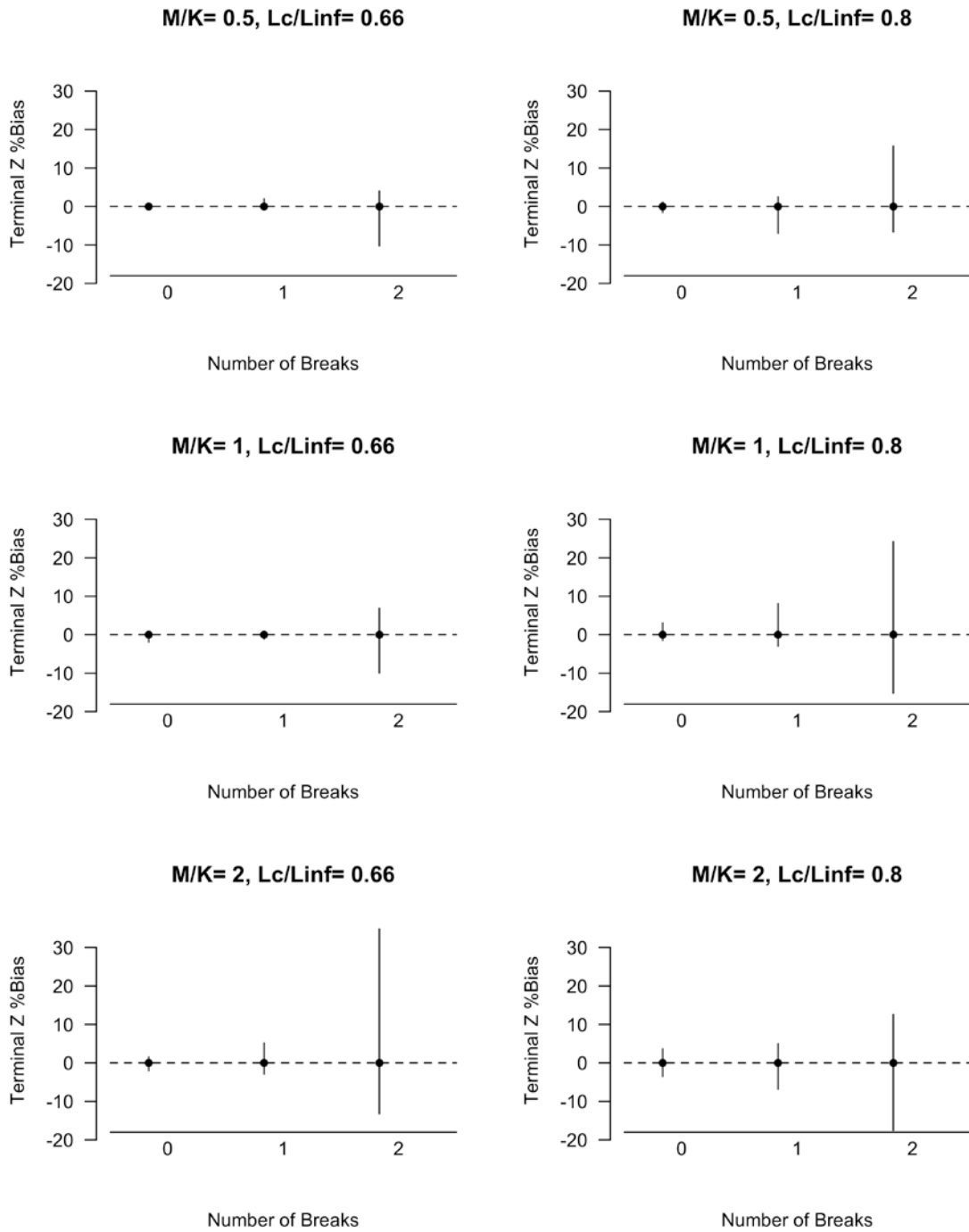
Figure 2. Percent difference between predicted and simulated terminal *Z* for the described scenarios for the maximum likelihood fitting routine tested. Points represent median values for percent bias; lines represent the centered 95% of percent bias.
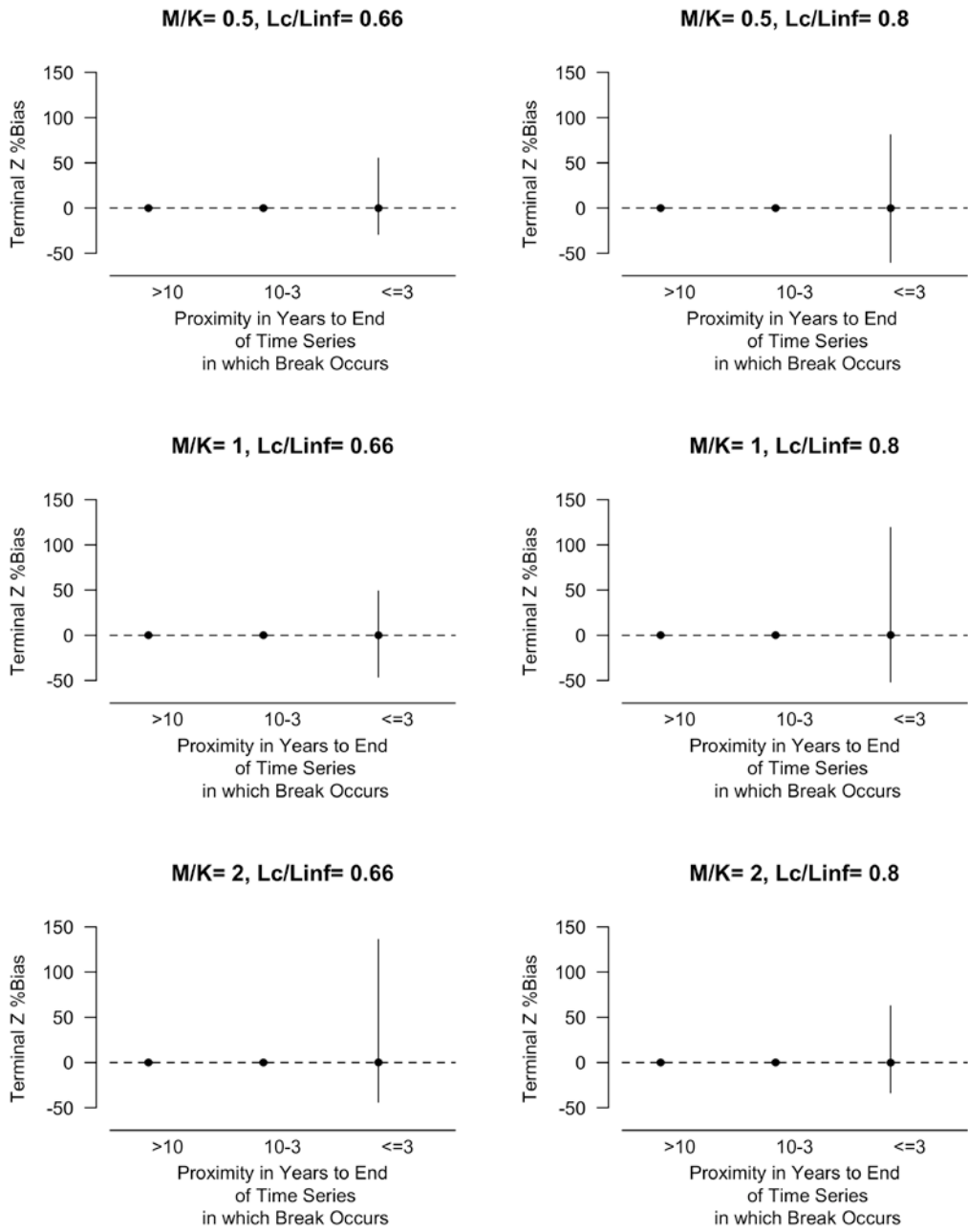
Figure 3. Percent difference between predicted and simulated terminal *Z* for the described scenarios for the maximum likelihood fitting routine tested, sorted according to those simulation runs in which the simulated terminal *Z* occurs greater than ten years from the end of the time series, between ten to three years from the end of the time series, and within three years from the end of the time series. Points represent median values for percent bias; lines represent the centered 95% of percent bias.