A Simulation Study of Impacts of Estimated Method and Uncertainty on Parameters in Length-Weight Relationship

Hailong Xu  
*Department of Fishery Sciences, Tianjin Agricultural University, Tianjin Key Laboratory of Aqua-ecology and Aquaculture, Tianjin, China  
College of Marine Sciences, Shanghai Ocean University, Shanghai, China

Rui Wang  
College of Marine Sciences, Shanghai Ocean University, Shanghai, China

Dexian Gu  
Tianjin Fishery Institute, Tianjin, China

Hongli Qi, Li Guo & Wenli Zhou*  
*Department of Fishery Sciences, Tianjin Agricultural University, Tianjin Key Laboratory of Aqua-ecology and Aquaculture, Tianjin, China

ABSTRACT: To acknowledge the impacts of estimated method and uncertainty introduced into the weight on parameters in length - weight relationship, take the morphometric traits of *Fenneropenaeus chinensis* as example, the parameters $a$ and $b$ were estimated by using power of the absolute residuals ($m$, $1 \leq m$ (natural number) $\leq 20$) methods, and error in the form of exponent was introduced into length – weight relationship. The study showed that the values of parameters in the relationship between length and weight were related to estimation methods. With the exponential increase of absolute residual, parameter $a$ increased, while parameter $b$ was highly uncertain. For the stability of parameter estimates, it was better to use the least absolute deviation method than the least square method and higher exponent of absolute residua to estimate the parameters. Uncertainty introduced into weight did not greatly impact the mean and median of the parameters $a$ and $b$ in the relationship between body length and weight. However, with increasing uncertainty, the fluctuation of parameter estimation increased, while the stability of the fitting effect of the length - weight relationship decreased. So, the estimated method and uncertainty should be taken into account for estimating the parameters in length - weight relationship, and the least absolute deviation method was proposed to estimate the parameters in length - weight relationship.

Keywords: length-weight relationship; parameters; estimated method; uncertainty

1 INTRODUCTION

Data on body length and weight, as the most common biology information, are the basic content in fishery resource investigation and are crucial parameters for fishery resource research and development management [1]. In comparison with body weight, body length is easier to obtain and is more stable. On the other hand, the measurement of body weight is relatively time-consuming, which is largely influenced by internal and external environmental factors (such as storage time, gastrointestinal fullness, reproductive, etc.). Therefore, in fishery resource assessment, it is a frequent practice to estimate body weight using the length-weight relationship established for the species of interest [2]. This approach has been widely used in morphological and ontogenetic comparison, and comparison of life history among floristic groups for catch prediction and estimation of the existing biomass [3, 4].
In the early 1900s, length-weight relationship was established on the assumptions that the weight is 3 dimension function of length, and the shape and special weight are constant over the fish growth period [5]. The relationship function was gradually applied to different species and growth stages, and was found to be appropriate only under certain circumstances [6] because the fitness was shown to be excellent in special cases. In 1924, Huxley [7] made a simple transformation and proposed a more general relationship \( W = aL^b \). In the formula, both \( a \) and \( b \) are undetermined parameters, which can be adjusted according to equation fitting. Therefore, this formula has statistically better fitting result which is widely adapted in the biological characterization of fishery resources [7-10].

In this formula, \( a \) and \( b \) are not only mathematical parameters, but also of biological or ecological significance [2, 7, 11]. The power index \( b \) can be used to determine if the fish is in constant growth stage. When \( b = 3 \), the species is in constant growth stage, where the fish is growing at equal speed in length, height and width; when \( b \neq 3 \), the growth speed is different in the three directions. If \( b < 3 \), the fish is in negative allometric growth; if \( b > 3 \), the fish is in positive allometric growth. \( a \) is a conditional factor that reflects the food base and hydrological conditions in the environments where the species inhabits. It is generally smaller after spawning, which is greatest at sexual maturity before spawning. In the relationship between the net weight of fish without internal organs and body length, \( a \) is greatest at fattening stage. As such, \( a \) can be used to indirectly indicate the physiological cycle of resources.

With progress in the quantitative research of fishery resources and increasing attention on statistical analysis [12-15], the effect of uncertainty on fisheries has attracted widespread interest and concern [16-19], which has been addressed in more and more studies [19, 20]. When exponential function was used in fitting the length-weight relationship, the impact of a number of factors on the fitting results has been investigated, including sampling methods, fish species, food condition, habitats and other environmental conditions, and life stage [1, 21, 22]. Statistically, there may also be impact of estimating method on parameters \( a \) and \( b \). Therefore, when estimating \( a \) and \( b \), uncertainty or error factors should be taken into account to improve not only the reliability of estimation, but also understanding of species characterization, additionally, to help to further research biology and habitat environment of resources.

2 MATERIALS AND METHODS

2.1 Materials

To compare the estimated parameters in length-weight relationship from various methods and to assess the impacts of uncertainty in weight, the used data were based on the published length-weight relationship on female Chinese shrimp *Fenneropenaeus chinensis* (female shrimp was not selected purposely. Male or mixed gender population can be used as well). And the mean value of published results [23-26] were assigned to the initial parameters on body length-weight relationship in this study, arbitrarily, where \( a = 0.0000114 \), \( b = 3.0042 \), and \( L_\infty = 210.17 \)mm.

2.2 Methods

2.2.1 Design of parameters’ estimated method

The asymptotic length \( (L_\infty) \) was divided into \( n \) equal parts \((n \) is arbitrary, and is equal to 100 in this study irrespective of impact of \( n \) on results\), forming \( n+1 L_i \). The Monte Carlo method was used to produce simulated weight with error \([17, 18, 27, 28] \( W = aL^b \epsilon \). \( \epsilon \in N(0, 0.1^2) \), and the variance was set subjectively). The stimulated data were used for estimation of \( a \) and \( b \) in the length-weight relationship using the least absolute deviation and least square methods, and further higher power of the absolute residuals \((m \leq n \leq m \text{ (natural number) } \leq 20) \). 1000 times of simulations were performed, and the values of \( a \) and \( b \) obtained in different approaches were statistically analyzed.

2.2.2 Simulation study of uncertainty

As \( W_i \) was calculated with given parameters and error in the form of exponent was introduced, there are three levels of error as normal distribution with 0 as means, and 0.05, 0.1 and 0.2 as the standard deviation respectively, and then the parameters were estimated on the base of the simulated \( W_i \). The process was repeated for a number of times (1000 in the study), to generate the statistical characteristics of \( a \) and \( b \), and the steps were followed in detail.

Modeling method [27]:

1. Introduce uncertainty into length - weight relationship,

\[
W_i = aL_i^be^{\varepsilon}, \quad \varepsilon \sim N(0, \sigma_w^2)
\]

And weight was simulated at given length; \( \sigma_w \) was above defined;

2. Estimate the parameters \( a \) and \( b \) in length - weight relationship based on the generalized least square, and the produce was conducted by means of R \( i386 3.0.2 \);

3. Repeat step (1) and (2) for 1000 times;

4. Analyze the statistics of parameter \( a \) and \( b \), including mean, C.V 1\(^{st}\) and 3\(^{rd}\) quartile and so on.

174
3 RESULTS

3.1 Effect of estimated method on estimated parameters in length - weight relationship

Different \(a\) and \(b\) were obtained with different estimated approaches (See Figure 1 and 2). In the present study, 4 typical modeling results of 1000 time simulations were chose, and the results showed that with increasing \(m\) \((m \in \{1, 2, 4, 6, \ldots, 20\})\), the mean \(a\) values (over 1000 times per modeling) changed regularly between different modeling results with increased rate being gradually reduced and stabilized. When \(m\) was set at about 10, \(a\) value gradually stabilized with slightly different approximates over different modeling results. Within a modeling, with increase in \(m\), the uncertainty of \(a\) increased; when \(m = 1\), the maximum \(a\) was 1.48 to 1.58 times of the minimal \(a\) within a single modeling of 1000 times of stimulations. When \(m\) was set to 20, the maximum \(a\) was 4.54 to 5.67 times of the minimal \(a\) within a single modeling. Correspondingly, the standard derivations were \(6.01*10^{-7} - 6.09*10^{-7}\) and \(2.69*10^{-6} - 2.76*10^{-6}\) when \(m\) was set to 1 and 20, respectively. Different from the impact on parameter \(a\) from estimation methods, \(b\) mean values (over 1000 times per modeling) changed irregularly in different modeling results. However, with increasing \(m\), the uncertainty of \(b\) increased as well. Correspondingly, the standard derivations increased from 0.01 obtained using the least absolute deviation method to 0.04 generated using the least square method.

Figure 1. Impact of residual number \((m)\) on parameter \(a\) in different stimulations.

Figure 2. Impact of residual number \((m)\) on parameter \(b\) in different stimulations.

3.2 Effects of introduced uncertainty on estimated parameters in length - weight relationship

When uncertainty was introduced in the length-weight relationship, at the same level of uncertainty, there was no significant difference in \(a\) and \(b\) between mean and median. Uncertainty of different levels did not greatly impact the mean and median values of \(a\) and \(b\). With increasing uncertainty, the coefficients of variation increased as well as the ranges between the 1st quartile and 3rd quartile (Table 1). The density distribution (Figure 3) of the parameters \(a\) and \(b\) showed that when the uncertainty increases, the distribution curves of the parameters become flat and wide, and when the uncertainty reaches 0.2, the distribution density curve is close to a straight line.

Table 1. Statistics of parameters with different uncertainty in length - weight relationship.

<table>
<thead>
<tr>
<th>(a)</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.14E-05</td>
<td>1.14E-05</td>
<td>1.14E-05</td>
<td>3.0040</td>
<td>3.0042</td>
<td>3.0037</td>
</tr>
<tr>
<td>C.V</td>
<td>0.0158</td>
<td>0.0302</td>
<td>0.0639</td>
<td>0.0011</td>
<td>0.0022</td>
<td>0.0047</td>
</tr>
<tr>
<td>Median</td>
<td>1.14E-05</td>
<td>1.14E-05</td>
<td>1.14E-05</td>
<td>3.0039</td>
<td>3.0040</td>
<td>3.0042</td>
</tr>
<tr>
<td>1st quartile</td>
<td>1.13E-05</td>
<td>1.12E-05</td>
<td>1.09E-05</td>
<td>3.0017</td>
<td>2.9997</td>
<td>2.9944</td>
</tr>
<tr>
<td>3rd quartile</td>
<td>1.15E-05</td>
<td>1.16E-05</td>
<td>1.19E-05</td>
<td>3.0061</td>
<td>3.0087</td>
<td>3.0135</td>
</tr>
</tbody>
</table>

Note: black solid line, \(SD=0.05\); blue dashed line, \(SD=0.10\); red dotted line, \(SD=0.20\)

Figure 3. Distribution of parameters \(a\) and \(b\) with uncertainty in length-weight relationship statistics.
Body length-weight relationship is a fundamental content in fishery resource research, which has been expressed and fitted in form of linear equality, logarithmic expression, exponential and power functions. It is now generally agreed that power functions are best for fitting the relationship \cite{1,7}. Although expressed form as power functions, the relationship is intrinsic linear model, and for estimating the parameters \(a\) and \(b\), either by direct regression or after logarithmic transformation, the common methods used currently are based on the least square method (or generalized least square method). Statistically, the parameters are estimated by not only the least square method, but also the least absolute deviation method. The two methods have the same estimating principles, but are achieved in different approaches \cite{29}. In this study, the parameters were estimated by expending the absolute number \((m)\) of residual to 20. Our results show that with decreasing \(m\), parameter \(a\) became smaller (Figure 1), in other words, the species of interest is in a more conservative living environment. This is because the least \(m\)th power method uses the minimum sum of \(m\)th power of the residual, meaning that equilibrium is established among the \(m\)th power of absolute errors between observed and predicated values. When \(m \geq 2\), if the absolute value of error is larger than 1, the weight of error will be amplified, leading to greater impact of the salient points further from the center on the calculation results \cite{30}. On the other hand, the least absolute deviation uses the minimal sum of absolute distances from the measured points to the regression in longitudinal direction. Here, the equilibrium of errors is achieved on linear relationship, and impact of error on parameter estimation is linearly proportional to error. In other word, it can be interpreted that the influence weight of each error on parameter estimation is the same.

The minimum sum of residuals derived from all the sample points is the fundamental criterion to judge the fitness between data and model. Although the least absolute deviation method was proposed over 40 year ago before the least square method \cite{31}, due to the convenience in calculation and relative perfect in theory, the least square method had initial development \cite{30}, which has become one of the most fundamental and common used methods for parameter estimation in mathematics, statistics and economics. Statistically, the essential difference between the least absolute deviation and the least \(m\) square \((m \geq 2)\) is that, the former estimates conditional median \cite{31}, yet the latter generates the conditional means. When using the least \(m\) \((m \geq 2)\) square method for parameter estimation, it can be applied for better results only when the linear regression model caters for the classical assumptions (residuals of observation data is normal distribution) \cite{30}. Therefore, the least \(m\) \((m \geq 2)\) square regression is vulnerable and has poor resistance to extreme values. However, the essence of least absolute deviation is to seek conditional median for variables to be explained, and it does not require that random errors are normally distributed. Also, due to the statistical properties of the median, it is not affected by extreme values. The information about length and weight of fish are affected by a number of factors, such as generations, living environments, sampling time, and physiological period. Especially, when data from multiple samples are analyzed as a whole, the errors in the observation data are not strict on the normal distribution. Furthermore, the residuals between observed and predicated data hardly obey normal distribution with zero mean. Therefore, when using the least \(m\) \((m \geq 2)\) square method, the center of the absolute value of residual may change as well after \(m\) time transformation. So, for parameter stability, the least absolute deviation is better than the least \(m\) \((m \geq 2)\) square method.

There are annual differences in the growth of fishery resources not only between individuals \cite{32} but also between local populations \cite{22}, which are caused by many factors, including habitat environmental factors (temperature, salinity, dissolved oxygen, etc.), bait condition, resource distribution density and fishing pressure \cite{8}. There are also internal factors besides external factors, for example, parental sexual gonad development, early life history of individuals, energy distribution, the growth stage of individuals, gonad development, and population structure, physiological activities of individual, different generations in the same year and the same generation in different ages. In addition, due to differences in the sampling method, sampling time, sampling process, and measurements conducted by sampling personnel, there will be differences in the captured data. For crustaceans that molt (such as shrimp, crab, etc.), their growth is not continuous and has metamorphosis in their life history, which is much more complex than fish. As such, their growth-related data have greater variation. For those fishery resources, their growth data and the use of these data for other studies should be considered with their uncertainty \cite{17-19,33}. In the present study, the parameters \(a\) and \(b\) were estimated at three levels of uncertainty. \(t\)-test suggested that the means and medians of \(a\) and \(b\) obtained at different levels of uncertainty are not different \((P > 0.05)\), which might be due to the assumption in the simulation that the errors obey normal distribution. Although the errors are in exponential form, there have intrinsically linear relationship in body length-weight relationship, and the exponential errors do not change the distribution characteristics of uncertainty of parameter estimation. However, with the increase in the level of uncertainty, the distribution curve of parameters become flat and broad (Figure 3). In comparison with the values obtained at uncertainty level of 0.05, the distribution ranges of parameters \(a\) and \(b\) increased 1.8 times \((SD = 0.10)\) and 3.80 times \((SD = 0.20)\), and 1.70 times \((SD = 0.10)\) and 3.78 times \((SD = 0.20)\), respectively.
illustrating that with increased uncertainty, parameters estimated for the length-weight relationship are more variable. This would directly reduce the stability in fitting results using the relationship. When the growth parameters were estimated, not only the uncertainty level, but also error structure would affect the results of fitting equations and parameter estimation [14], and different error structure will result in different fitting degree and parameter estimates. A wrong error structure would lead to seriously deviated estimate [35]. In most case, error structures introduced additive error structure or multiplicative error structure, and Francis [36] also proposed other structures error in their study to establish the relationship function between age and morphology. Various methods may be used to select error structure for estimation of growth-related parameters. As a general role [37-40], when fitting variable is a constant value, additive error structure is selected; when the variables show an increasing trend, multiplicative error structure is used. In the present study, the body weight of the shrimp is on the rise, so multiplicative error structure is applied.

ACKNOWLEDGEMENT

This work was financially funded by the project which was supported by the Key Laboratory of Mariculture & Stock Enhancement in North China’s Sea, Ministry of Agriculture, P.R. China (2014-MSENC-KF-09) and Tianjin Application Foundation and Advanced Technology Research Plan (15JCYBJC23900).

REFERENCES


