Latin American Journal of Aquatic Research, 52(5): 765-776, 2024
DOI: 10.3856/vol52-issue5-fulltext-3218

Research Article

Modeling and statistical analysis of growth and shape problems of the brown crab *Callinectes bellicosus*

Enrique Villa-Diharce[1], **María Guadalupe Russell-Noriega**[2] **& Guillermo Rodríguez-Domínguez**[3]
[1]Centro de Investigación en Matemáticas, Guanajuato, México
[2]Facultad de Ciencias Físico-Matemáticas, Universidad Autónoma de Sinaloa
Culiacán, Sinaloa, México
[3]Universidad Tecnológica de Escuinapa, Escuinapa, Sinaloa, México
Corresponding author: Guillermo Rodríguez-Domínguez (guirodom@uas.edu.mx)

ABSTRACT. Allometric length-weight relationships are traditionally analyzed using the Huxley model, which assumes a lognormal distribution of errors. Still, other models have been successfully evaluated to describe these relationships, and it has been found that the distribution of errors is not necessarily lognormal. Two data sets of carapace width and weight, composed of 1,554 females and 2,531 males of *Callinectes bellicosus* (Stimpson, 1859), were fitted to five models to address this problem, which included: Huxley model, quadratic, cubic, breakpoint and two-segment. Lognormal distribution of the residuals was first assumed, but the goodness-of-fit test did not confirm this assumption, and the QQ plots revealed heavy tails. As an alternative, logistic distribution of the residuals was assumed, and the goodness-of-fit test and the QQ plots supported this. The Akaike information criterion (AIC) was used to select the best models. When a lognormal distribution of the residuals was assumed, the best model for females was the two-segment model and the cubic model for males. In contrast, with the logistic distribution, the best model was the two-segment model for both sexes. Furthermore, AIC was smaller in models with error loglogistic distribution than lognormal distribution. The two-segment model is associated with size at maturity in both sexes, and each segment represents juvenile and adult crabs. It is concluded that it is important to confirm the assumptions of the distribution of the residuals when fitting models to data because a wrong assumption can result in an erroneous model selection.

Keywords: *Callinectes bellicosus*; allometric growth; biphasic models; Akaike information criterion; additive and multiplicative error terms; size at maturity

INTRODUCTION

Decapod crustaceans typically undergo morphological changes associated with the transition from juvenile to adult stages, including the size and thickness of chelae in males and the abdomen in females (Corgos & Freire 2006, Nunes-Pralon & Negreiros-Fransozo 2008, Williner et al. 2014). In the brown crab *Callinectes bellicosus*, females exhibit a change in abdomen shape, transitioning from a triangular form during the juvenile stage to an oval shape in the mature stage (Espinoza-Ahumada 2019). This alteration in the shape of the abdomen in mature females provides more space than the triangular form for carrying and incubating fertilized eggs until hatching. In this species, males are larger than females, but the transition from juvenile to mature stages reveals a more noticeable change in females than in males. Nevertheless, physiological maturity involves the growth of the gonad inside the exoskeleton, an organ that was only incipient in the juvenile stage. Consequently, a change in the shape of both male and female crabs is expected to accommodate the growth of this organ.

Associate Editor: Víctor Sanz

Studies on allometric relationships of different body parts in crustaceans identify different points of change in relative growth among the various parts (Katsanevakis et al. 2007).

The length-weight relationships (LWR) of aquatic organisms can be visualized as a relative volume growth in contrast with the organism's size. The volume has three lineal measures, longitude, width, and height, which will integrate into the organism's weight (Rodríguez-Domínguez et al. 2015). These relationships are commonly analyzed using a potential function like Huxley's $Y = aX^b$ (Huxley 1932), and with a logarithmic transformation of the original data, a linear relationship of the form $\mathrm{Ln}(Y) = \mathrm{Ln}(a) + b\mathrm{Ln}(X)$ is obtained. The slope of this function (b) is known as the allometric coefficient, where a value of 3 for this coefficient is a reference for isometry, and different values indicate allometry. However, this function assumes that the allometric coefficient is constant, regardless of the organism's size. It has also been observed that some allometric relationships between different parts of aquatic organisms vary with the size of the organism (Katsanevakis et al. 2007, Rodríguez-Domínguez et al. 2018). Alternative models to detect these changes in the allometric coefficient with the organism's size include second-order (quadratic) and third-order (cubic) polynomial, breakpoint, and two-segment models. The last two detect inflection points in the allometry change that may occur during the transition from the juvenile to adult phases.

The paradigm of model selection and multimodel inference involves choosing the best model from a set of candidates fitted to observed data based on goodness of fit, expressed as the maximum likelihood of a function given the data, and the principle of parsimony. These two criteria are integrated into the Akaike information criterion (AIC), and the best model is the one that archives the lowest AIC, indicating that the errors have a probability distribution closer to the assumed probability distribution in the likelihood function used (Burnham & Anderson 2002).

When fitting a mathematical model to observed data on relative growth or the LWR of aquatic organisms, one of the assumptions is the normality of errors. If there is evidence that the error is multiplicative and follows a lognormal distribution, the original data or errors are transformed using the natural logarithm to achieve normalization. However, the fulfillment of normality is often assumed and not verified in many published works (Ortega-Lizárraga et al. 2021, Leyva-Vázquez et al. 2022), which can have implications for the selection and certainty of the chosen models as the best fit.

The LWR of aquatic organisms generally exhibits a scatterplot with increasing variance with size, indicating a certain degree of heterogeneity. Hence, it is common practice to logarithmically transform the data before fitting an allometric model, attempting to satisfy two assumptions: homogeneity of variances and normality of errors (Packard 2014, De Giosa & Czerniejewski 2016, Dash et al. 2023, Jurado-Ruzafa & Hernández-González 2024)

In the literature, one frequently encounters allometric studies in which assumptions are assumed to hold without verification (Packard 2014, Orlov et al. 2022, Olentiona et al. 2023, Song et al. 2023), which can have consequences for the selection and certainty of the model chosen as the best fit.

Montesinos-López et al. (2019) used an allometric model of the Huxley type to analyze the LWR of leaves from a marine alga. They obtained an inflated error distribution that did not comply with the normality assumption. Subsequently, they tested a logistic distribution with heterogeneous error variance, improving the model's goodness of fit. On the other hand, Villa-Diharce et al. (2022) addressed the allometric relationship between leaf area and weight of the same marine alga. They observed a failure in the assumption of normality of errors. They resolved this by assuming a mixture distribution of two normal variables. However, both the logistic distribution and the mixture distribution of normal may be closely related, as it has been shown that the scale mixture of standard normal distributions can represent the logistic distribution (Stefanski 1991).

In this work, we propose using the multimodel selection approach to choose the best model for the allometric relationship between cephalothorax width and weight of the brown crab *C. bellicosus*. The goodness of fit was compared under the assumptions of normality and logistic distribution of errors. The hypothesis is that errors do not conform to a normal distribution, and the logistic distribution will provide a better fit, influencing the selection of the best model compared to when a normal distribution is assumed.

MATERIALS AND METHODS

Measures were taken from a data sample of 4,089 individuals of the species *C. bellicosus*. Specifically, the carapace width (CW) was measured and identified as x, and the weight was denoted as y. These individuals were collected from the Santa Maria Bay in Sinaloa, México. Among them, 2,531 are males, and 1,558 are females (Rodríguez-Domínguez et al. 2018). Crustaceans have the characteristic of autonomously shedding some of their appendages as an escape response to a predator attack or when the appendage becomes stuck in the structure (such as traps or fishing nets). For the analysis, only complete individuals were considered to avoid variations in weight not associated with CW. Additionally, it's important to note that specimens were weighed immediately after capture to prevent weight variations associated with handling in ice or refrigeration.

According to the scatter plots of CW against weight, with increasing variability as the predictive variable x (CW) increases, the error term acts multiplicatively. Thus, the traditional allometric model takes the form:

$$y = \alpha x^{\beta} \delta \tag{1}$$

where $\delta = e^{\epsilon}$, with ϵ distributed according to a normal distribution with mean zero and standard deviation σ. Therefore, in the geometric (logarithmic) scale, the following expression for the traditional allometric model (with logarithmically transformed data) is used:

$$log(y) = log(a) + \beta \times log(x) + \epsilon$$

where x is the CW, y is the weight, and ϵ is the error term, usually assumed to be normally distributed as in model (1). In addition to this model, four other models were considered, which are extensions of the traditional model where the allometric coefficient β is a function of $\log(x)$, not constant. Two extensions were initially considered, where the allometric coefficient is a $\log(x)$ function, respectively:

$$y = \alpha x^{\beta_1 + \beta_2 \log(x)} \delta \tag{2}$$

and

$$y = \alpha x^{\beta_1 + \beta_2 \log(x) + \beta_3 [\log(x)]^2} \delta \tag{3}$$

Recent literature uses quadratic and cubic models for allometric analysis, such as those used here (Zhu et al. 2021).

Two other models are extensions of the traditional allometric model models that have two different values of the allometric coefficient β before and after the value of x, called the breakpoint x_b (Halley, 2016, Tsuboi et al. 2018) These models, known as the breakpoint model and the two-segment model, respectively, are:

$$y = \beta_0 \left(\frac{x}{x_b}\right)^{\left\{\beta_1 + \frac{\beta_2}{2}[sgn(x - x_b) + 1]\right\}} \delta, \tag{4}$$

and

$$y = \beta_0 exp\left\{\frac{\beta_3}{2}[sgn(x - x_b) + 1]\right\} \left(\frac{x}{x_b}\right)^{\left\{\beta_1 + \frac{\beta_2}{2}[sgn(x - x_b) + 1]\right\}} \delta \tag{5}$$

The regression models were fitted following the maximum likelihood criterion, assuming that the error term acts multiplicatively with a lognormal distribution on the arithmetic scale.

Assuming errors with lognormal distribution, we maximize the following log-likelihood function,

$$l(\theta) = \sum_{i=1}^{n} \left\{ -\log(y_i) - \frac{1}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2} [log(y_i) - \mu_i]^2 \right\} \tag{6}$$

where the form μ_i depends on the fitted model. Thus, in the first three traditional models (Eqs. 1-3), the following expressions are obligated,

$\mu_i = \log(\alpha) + \beta_1 \times log(x)$,

$\mu_i = \log(\alpha) + \beta_1 \times log(x) + \beta_2 \times [log(x)]^2$, and

$\mu_i = \log(\alpha) + \beta_1 \times log(x) + \beta_2 \times [log(x)]^2 + \beta_3 \times [log(x)]^3$, respectively.

In the case of models with a changepoint, the following expressions are obtained:

$\mu_i = \log(\beta_0) + \left\{\beta_1 + \frac{\beta_2}{2}[sgn(x_i - x_b) + 1]\right\} [log(x_i) - log(x_b)]$, and

$\mu_i = \log(\beta_0) + \left\{\beta_1 + \frac{\beta_2}{2}[sgn(x_i - x_b) + 1]\right\} [log(x_i) - log(x_b)] + \frac{\beta_3}{2}[sgn(x_i - x_b) + 1]$

for the breakpoint models and two-segment models, respectively.

In the case of the breakpoint model, it can be demonstrated that the intercepts of the straight lines $E[log(y)|x]$ with the y-axis are:

$$\beta_0^* = \begin{cases} log(\beta_0) - \beta_1 log(x_b) & x < x_b \\ log(\beta_0) - (\beta_1 + \beta_2) log(x_b) & x \geq x_b \end{cases}$$

and the slopes are:

$$\beta_1^* = \begin{cases} \beta_1 & x < x_b \\ \beta_1 + \beta_2 & x \geq x_b \end{cases}$$

In the case of the Two-segment model, these intercepts are:

$$\beta_0^* = \begin{cases} log(\beta_0) - \beta_1 log(x_b) & x < x_b \\ log(\beta_0) + \beta_3 - (\beta_1 + \beta_2) log(x_b) & x \geq x_b \end{cases}$$

and the slopes are given by:

$$\beta_1^* = \begin{cases} \beta_1 & x < x_b \\ \beta_1 + \beta_2 & x \geq x_b \end{cases}$$

When we assume that the multiplicative error has a log-logistic distribution, the log-likelihood function takes the following form:

$$l(\theta) = -nlog(\sigma) - \sum_{i=1}^{n} log(y_i) - \sum_{i=1}^{n} \left\{ \left[\frac{\log(y_i) - \mu_i}{\sigma}\right] + 2log\left[1 + exp\left(-\left[\frac{(\log(y_i) - \mu_i)}{\sigma}\right]\right)\right] \right\} \tag{7}$$

Taking, as in the case of lognormal log-likelihood, the mean function μ_i corresponds to the model we fit. Usually, these log-likelihoods are maximized numerically.

The AIC, which allows us to rank models according to their goodness of fit, is given by:

$AIC = -2l(\hat{\theta}) + 2p$ (Akaike 1973, Burnham & Anderson 2002)

The AIC value represents the amount of information lost when fitting the model; therefore, we choose the one with the lowest AIC value as the best model when comparing various models. These statistics do not tell us whether the fit is appropriate or incorrect. We can use goodness-of-fit tests such as the Anderson-Darling test, the Kolmogorov-Smirnov test (Conover 1971), or the Shapiro-Wilk test (Shapiro & Wilk 1965). Quantile-quantile plots are often used as graphical tests of normality or to assess the quality of the fit of some other distribution to a dataset. When the fit is appropriate, the quantile-quantile plot (Stephens 1997) follows a linear pattern. In this work, in addition to the normal quantile-quantile plot, a logistic quantile-quantile plot was used to verify that a set of residuals follows a logistic distribution. Normality tests of error terms are necessary to ensure the inferences' validity. In addition to the classical tests, examples of normality tests related to central tendency and dispersion statistics are presented in the literature (Ghasemi & Zahedias 2012, Das & Imon 2016, Mishra et al. 2019). There are several statistical functions in R, which are useful in the statistical analysis of allometric data (Warton et al. 2012, Kassambre & Mundt 2020), so all calculations developed in this paper were performed using the statistical computing language R (R Core Team 2023).

RESULTS

In the brown crab, males are larger (30 and 180 mm for the cephalothorax) than females (30 and 140 mm for the cephalothorax), as shown in Figure 1. The maximum likelihood estimates of the parameters for the five fitted models, assuming multiplicative errors and following a lognormal distribution, as is commonly considered, are shown (Table 1). According to the Akaike criterion, the cubic model fits better for males, as this model has the lowest AIC value. In the case of females, the best fit is not provided by a traditional model but by a model with a changepoint, such as the two-segment model, which has the lowest AIC. In both cases the closest competing model has a distance in the AIC value greater than 10. However, all five models produce very similar curves (Fig. 1).

To understand the distribution of residuals on the logarithmic scale for models with the least satisfactory fit (linear) and the best fit (two-segment) for females, we observe the QQ-normal plots shown in Figure 2.

The QQ-normal plots exhibit sets of points that follow a linear pattern in the center. Still, at the extremes, they deviate from the line, indicating that the distribution of residuals on the geometric scale has heavy tails. This observation holds for all model fits, suggesting that the error terms exhibit over-dispersion compared to the normal distribution. The results of the Anderson-Darling hypothesis tests for normality are shown (Table 2).

Considering that the error terms of the fitted models exhibit over-dispersion compared to the normal distribution, the models were adjusted by changing the normal distribution of errors to a logistic distribution because the logistic distribution has heavier tails than the normal distribution, as seen in Figure 3.

In Table 3, we have the estimates of the fits for the five considered models with error terms following a log-logistic distribution. Now, the best model was the two-segment model in both sexes. The inflection point of this model is 111.5 mm in females and 126.03 mm in males, and in both sexes, a steeper slope is observed before the inflection point than after it. For females, the slope is initially 3.289; after the changepoint, it is 2.834. For males, the slopes of the two segments range from 3.215 to 2.936.

The AIC values, comparing Tables 1 and 3, are lower when the error distribution is log-logistic than when it is lognormal, indicating that the log-logistic distribution fits better than the lognormal distribution on the arithmetic scale.

The logistic quantile-quantile plots are shown (Fig. 4), constructed with the residuals (on the logarithmic scale) of the linear and two-segment models, with error terms following a logistic distribution. In contrast to the QQ-normal plots shown in Figure 2 (for the same models), we now observe that the plots have a fairly linear pattern, with the extreme points very close to the reference straight line in each case. The two models with logistic error have a good fit of data on cephalothorax width-weight for females.

Table 4 displays the results of the Kolmogorov-Smirnov test to examine whether the error on the geometric scale follows a logistic distribution. The logistic distribution is not rejected in any of the five models fitted to the data for females and males.

The point estimates of the parameters for the five analyzed models have been provided (Tables 1 and 3). Sometimes, it is preferable to provide estimates with intervals, as they give information about ranges where the true values of the parameters may lie. Additionally, these intervals provide information about the quality of the estimates; the narrower the confidence interval, the

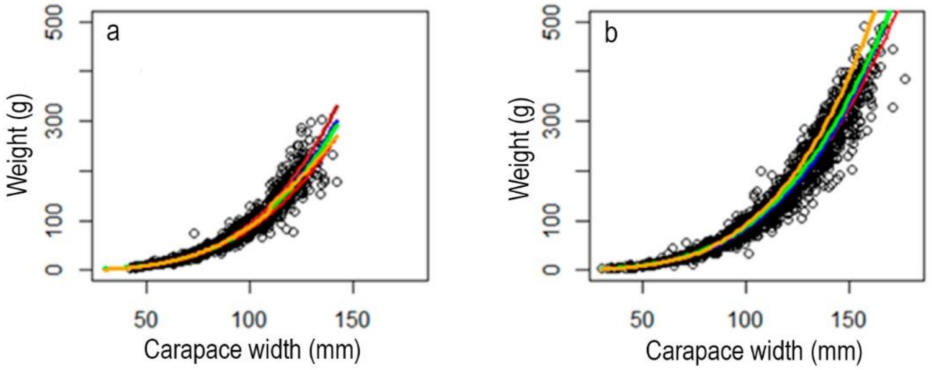

Figure 1. Scatter plot of cephalothorax width and weight for 4,089 brown crabs, of which a) 1,558 are females and b) 2,531 are males. Under the assumption of normal distribution of errors, the fitted curves for the five models are overlaid.

Table 1. Estimates of the parameters for the considered models, assuming multiplicative error term, distributed in a lognormal form. AIC: Akaike information criterion. In bold, best model with minor AIC.

Sex	Model	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{x}_b$	$\hat{\sigma}^2$	$l(\hat{\theta})$	AIC
Females	Linear	-10.933	3.346				0.018	-5,713.0	11,432.1
	Quadratic	-6.059	1.140	0.249			0.018	-5,701.0	11,410.0
	Cubic	12.662	-11.702	3.177	-0.222		0.018	-5,700.0	11,410.1
	Breakpoint	3.473	3.274	0.352		74.8	0.018	-5,696.2	11,402.6
	Two segment	4.788	3.283	-0.437	0.096	110.6	0.018	-5,688.5	**11,389.1**
Males	Linear	-10.668	3.285				0.022	-10,577.9	21,161.9
	Quadratic	-5.390	0.943	0.258			0.021	-10,530.3	21,068.6
	Cubic	18.828	-15.617	4.013	-0.282		0.021	-10,518.5	**21,047.1**
	Breakpoint	3.467	3.156	0.462		75.0	0.021	-10,524.6	21,059.1
	Two segment	3.626	2.917	0.463	0.032	79.3	0.021	-10,523.9	21,059.8

higher the reliability of the estimation. There are different techniques for calculating confidence intervals for the parameters of interest. This article obtained confidence intervals using the bootstrap technique (Efron & Tibshirani 1993) with a bootstrap sample size 10,000. The confidence intervals were calculated for the models considered the best fit in the cases of females and males; the two-segment model with log-logistic error performed better. The confidence intervals are presented for the two-segment model with log-logistic error terms for females and males, respectively (Tables 5-6).

Considering that the allometric model has two defining components -the median function (systematic part) and the dispersion pattern (distribution of the error term) -in this case, we have 10 allometric models resulting from the combination of median functions (five) and dispersion patterns (two distributions). All 10 models have been fitted to both female and male data.

The goodness of fit for comparison of the 10 models to the female data is presented (Table 7). It can be observed that the model with the best fit is the two-segment model with log-logistic distribution for the error term. Additionally, it's noted that the other four models with log-logistic errors have a closer fit (Δ_i less than 50) to the best one, compared to the five models with lognormal errors (Δ_i between 80 and 127). According to the Δ_i index, any of the five models with log-logistic error terms has a better fit than those with lognormal error terms. In the last column of Tables 7 and 8, the ranking of the 10 models is provided based on the goodness of fit, and it is observed that the distribution of the error term determines the quality of the fit.

Another way to assess the effect of changing the error distribution (from lognormal to log-logistic) in each of the five models (traditional or change-point) is to compare the differences in Δ_i by changing the distribution, that is, $\Delta_i(LNor) - \Delta_i(LLogis)$ (Table 9).

These differences are greater than the deltas observed in the fits of the five models with lognormal error distribution, whose differences are less than 44. Therefore, based on the available data and the fitted

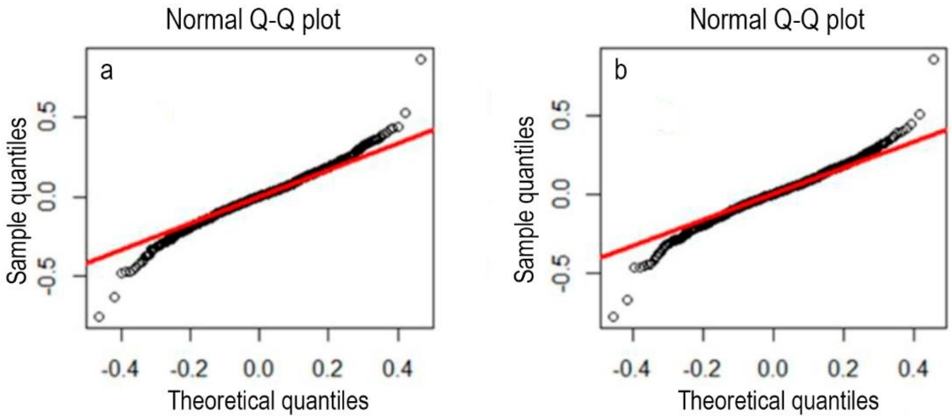

Figure 2. QQ-normal plots for the logarithms of the residuals of a) the traditional Huxley Model and b) the two-segment model. The deviation of the plot at the ends of the line indicates the presence of heavy tails in the distribution of error terms.

Table 2. Anderson-Darling test results on the assumption of normality of errors.

Sex	Model	A_n	*P*-valor
Females	Linear	3.9016	0.0097
	Quadratic	4.3729	0.0057
	Cubic	4.4351	0.0054
	Breakpoint	4.3407	0.0060
	Two segment	4.4176	0.0055
Males	Linear	10.5500	$< 10^{-5}$
	Quadratic	12.0800	$< 10^{-5}$
	Cubic	11.7200	$< 10^{-5}$
	Breakpoint	11.7700	$< 10^{-5}$
	Two segment	11.6300	$< 10^{-5}$

models, it can be concluded that, according to the AIC, the improvement in goodness of fit is greater when switching from lognormal to log-logistic distribution than when changing any of the five fitted models while keeping the error distribution fixed.

To rank the models according to their goodness of fit, we use the Δ_i and W_i indices derived from the original AIC, defined as:

$$\Delta_i = AIC_i - AIC_{min} \text{ and}$$

$$W_i = \exp((-0.5\Delta_i)/\sum_{i=1}^{k} exp(-0.5\Delta_i)$$

The calculations were not based on the corrected index given by:

$$AIC_C = -2l(\hat{\theta}) + 2p + \frac{2p(2p+1)}{n-p-1}$$

The correction term is extremely small due to the difference between the number of parameters p (less than 10) and the number of data points (thousands, in this case). In Table 10, in columns 4 and 7, the correction term values for females and males are on the order of hundredths.

DISCUSSION

This study shows that the assumption of normality, often taken for granted when fitting a model to a set of observations, does not always hold. This has significant implications for choosing the best model and ensuring its certainty for inference purposes.

This publication demonstrated that an assumption of the logistic distribution of errors performed better than the assumption of normality in the allometric analysis to describe the cephalothorax width-weight relationship of the brown crab *C. bellicosus.* It is also shown that the choice of error distribution type influences the selection of the best model according to the Akaike index, and this has more weight in determining the goodness of fit than the type of adjusted model.

It is important to test the assumptions of normality or lognormality of errors after fitting candidate models to a dataset to avoid selecting the wrong best model. As in this study, when assuming a normal distribution of errors, the cubic model was selected as the best model in the cephalothorax width-weight relationship for males with a 12-unit AIC distance from the nearest competitor. A model with a difference greater than 10 AIC units from the best model is considered unsupported by the data (Burnham & Anderson 2002) and should be discarded from the candidate model set. However, when models were fitted assuming log-logistic distribution, the best model was a two-segment model in both sexes.

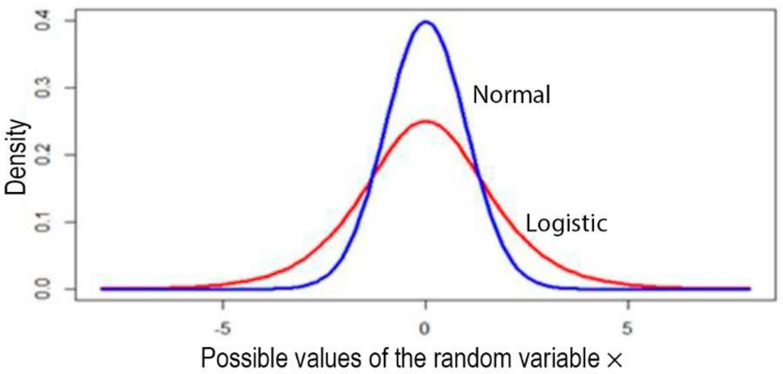

Figure 3. Standard normal and logistic density functions.

Table 3. Estimates of the parameters for the considered models, assuming multiplicative errors and following a log-logistic distribution. AIC: Akaike information criterion. In bold, best model with minor AIC.

Sex	Model	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{x}_b$	$\hat{\sigma}^2$	$l(\hat{\theta})$	AIC
Females	Linear	-10.976	3.356				0.074	-5,674.4	11,354.9
	Quadratic	-5.048	0.682	0.301			0.073	-5,657.7	11,323.5
	Cubic	22.517	-18.153	4.580	-0.323		0.073	-5,655.8	11,321.7
	Breakpoint	3.250	3.208	0.476		70.086	0.073	-5,652.2	11,314.6
	Two segment	4.820	3.289	-0.455	0.096	111.574	0.072	-5,646.7	**11,305.5**
Males	Linear	-10.759	3.305				0.079	-10.469.5	20,945.0
	Quadratic	-5.147	0.837	0.270			0.077	-10,423.5	20,855.0
	Cubic	21.724	-17.312	4.339	-0.303		0.077	-10.412.9	**20,836.0**
	Breakpoint	3.458	3.144	0.499		74.772	0.077	-10,417.4	20,844.8
	Two segment	5.188	3.215	-0.279	0.109	126.038	0.077	-10,411.2	**20,834.4**

The heterogeneity of variances, common in LWR (variances increasing with size), is addressed by transforming the original data using the logarithm. However, in many cases, this transformation only reverses the heterogeneity.

Sometimes, the distribution of errors cannot be modeled by the normal distribution because they have heavier tails. However, as demonstrated in this study, they can be modeled by a distribution with a greater spread than the normal distribution, such as the logistic distribution (Fig. 4).

Using the logistic distribution in modeling the error term of an allometric model when the normal distribution is unsuitable because it cannot account for extreme error values can be justified by employing the following result demonstrated by Stefanski (1991). The author shows that the logistic distribution can be represented as a discrete mixture of normal distributions with different variances.

When samples contain abundant observations over extended time intervals, various participants with different training and experience in sample collection often collect data. This diversity among samplers can be understood as a diversity in the accuracy with which participants record the sample data. Consequently, the complete sample can be viewed as a mixture of individual samples obtained by different collaborators involved in the sampling process. Considering the normal paradigm in individual samplings, we can assume that the total sample is a mixture of normal distributions, whose distribution, according to Stefanski (1991), is approximated by a logistic distribution. In this study, we possess two samples of sizes 1,558 (females) and 2,531 (males) collected over a year. According to the findings of Stefanski (1991), samplings conducted over extended periods by multiple samplers with varying sampling capabilities suggest that the error term (on the logarithmic scale) of the allometric model is likely to follow a logistic distribution rather than a normal distribution. Therefore, in extensive samplings involving multiple participants, the normality of errors should be verified, as the logistic distribution may be more appropriate for modeling the error distribution. As we have seen, the logistic distribution does not invalidate the normal paradigm; rather, the total sample is a discrete mixture of normally distributed samples. It is worth noting that

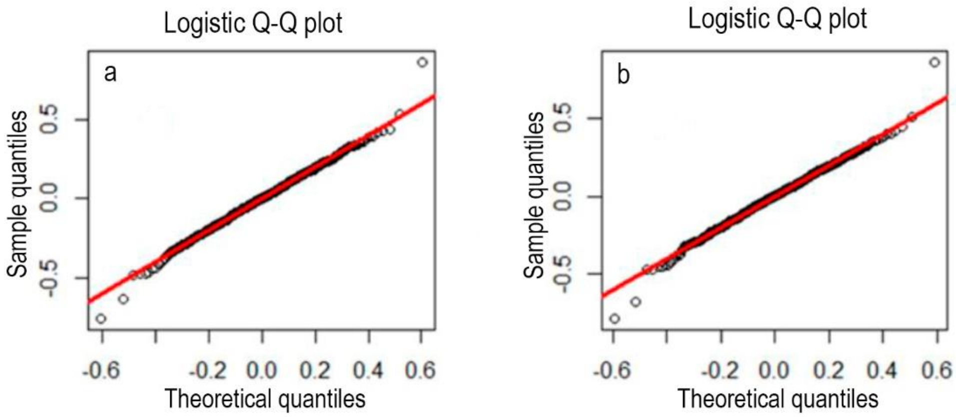

Figure 4. QQ-logistic plots for the logarithms of the residuals of a) the traditional Huxley model and b) the traditional quadratic model. The linear pattern in the plots indicates that the error term (on the geometric scale) in these models follows a logistic distribution.

Table 4. Results of the Kolmogorov test on the assumption of logistic distribution of the logarithms of errors.

Sex	Model	D_n	P-value
Females	Linear	0.0159	0.8252
	Quadratic	0.0177	0.7134
	Cubic	0.0155	0.8475
	Breakpoint	0.0152	0.8633
	Two segment	0.0175	0.7261
Males	Linear	0.0156	0.5690
	Quadratic	0.0209	0.2198
	Cubic	0.0197	0.2821
	Breakpoint	0.0204	0.2429
	Two segment	0.0216	0.1896

Table 5. Bootstrap confidence intervals for the parameters of the two-segment model fitted to the data of female brown crab.

Parameter	Estimation	Lower limit	Upper limit
β_0	4.820	4.809	4.832
β_1	3.298	3.149	3.330
β_2	-0.455	-0.758	-0.154
β_3	0.096	0.068	0.124
p_q	111.574	111.560	111.589
σ^2	0.072	0.069	0.075

there are various ways in which the logistic distribution arises (Dubey 1969, Gumbel 1994, Villa-Diharce & Escobar 2006).

The allometric breakpoint at 111.57 mm estimated CW in female crabs is very close to the morphometric maturity size reported by Rodríguez-Domínguez et al. (2015) of 107.7 mm for females of this species, confirming that the relative growth of the carapace in female crab changes proportionally to the size of maturity, leading to gonadal growth. There is no information on the maturity size for males; however, based on the mating behavior in this species, it is inferred that males mature at a size larger than females. Mating occurs immediately after the female molts from its juvenile exoskeleton to an adult female exoskeleton. Still, courtship begins a few days earlier when the male positions himself on the female's back, embracing her with his appendages. Thus, the male must be larger than the female to hold her under his appendages in anticipation of the mating moment. Therefore, the allometric breakpoint at 126 mm shell width may be associated with the morphometric maturity of the male.

Table 6. Bootstrap confidence intervals for the parameters of the two-segment model fitted to male brown crab data.

Parameter	Estimation	Lower limit	Upper limit
β_0	5.188	5.178	5.199
β_1	3.215	3.180	3.251
β_2	-0.279	-0.435	-0.124
β_3	0.109	0.086	0.132
p_q	126.038	126.034	126.044
σ^2	0.077	0.074	0.079

As expressed by weight, the volume of the carapace undergoes three-dimensional growth, so the point of change may be associated with a shift in the relative growth of the height or length of crabs. In females, once after mating, the male injects seminal fluid and sperm into them, which is stored in a receptacle called the spermatheca, representing up to 4% of the female's weight. Although this fluid gradually decreases as the

Table 7. Comparison of the goodness of fit of models fitted to female data. AIC: Akaike information criterion.

Distribution	Model	$l(\hat{\theta})$	AIC	Δ_i	W_i	Rank
Lognormal	Linear	-5,713.04	11,432.11	126.57	0	10
	Quadratic	-5,700.99	11,410.03	104.49	0	8
	Cubic	-5,700.00	11,410.07	104.53	0	9
	Breakpoint	-5,696.24	11,402.55	97.01	0	7
	Two segment	-5,688.50	11,389.10	83.56	0	6
Loglogistic	Linear	-5,674.45	11,354.93	49.39	0	5
	Quadratic	-5,657.73	11,323.51	17.97	0	4
	Cubic	-5,655.83	11,321.73	16.19	0	3
	Breakpoint	-5,652.24	11,314.55	9.01	0.010	2
	Two segment	-5,646.72	11,305.54	0.00	0.990	1

Table 8. Comparison of the goodness of fit of models fitted to male data. AIC: Akaike information criterion.

Distribution	Model	$l(\hat{\theta})$	AIC	Δ_i	W_i	Rank
Lognormal	Linear	-10,577.95	21,161.92	327.48	0	10
	Quadratic	-10,530.27	21,068.57	234.13	0	9
	Cubic	-10,518.53	21,047.10	212.66	0	8
	Breakpoint	-10,524.55	21,059.14	224.70	0	7
	Two segment	-10.523.85	21,059.76	225.32	0	6
Loglogistic	Linear	-10,469.51	20,945.04	110.60	0	5
	Quadratic	-10,423.48	20,854.99	20.55	0	4
	Cubic	-10,412.96	20,835.96	1.52	0.317	2
	Breakpoint	-10,417.36	20,844.76	10.32	0.004	3
	Two segments	-10,411.19	20,834.44	0.00	0.679	1

Table 9. Differences in Δ_i due to changing the assumption from the normal distribution to the logistic distribution of the errors when fitting each model.

Model	Linear	Quadratic	Cubic	Breakpoint	Two segment
Difference	77.18	86.52	88.34	88.0	83.56

Table 10. The Akaike information criterion corrections values by sample size (N) and sex (F: females, M: males).

Model	p(F)	N (F)	Corr. (F)	p(M)	N (M)	Corr. (M)
Linear	3	1558	0.027027	3	2531	0.0166205
Quadratic	4	1558	0.046361	4	2531	0.0285035
Cubic	5	1558	0.070876	5	2531	0.0435643
Breakpoint	5	1558	0.070876	5	2531	0.0435643
Two segment	6	1558	0.100580	6	2531	0.0618066

gonad size increases, it still accounts for 3.75% of the female's weight (Espinoza-Ahumada 2019), which undoubtedly requires space within the female's carapace, defined since the pubertal molting process. In males, the testicles and the middle vas deferens involved in sperm and seminal fluid production grow from 0.66% in the juvenile stage to 2.8% in the mature stage (Espinoza-Ahumada 2019). The results do not support the above since the decrease in the slope of the second segment of the width-weight relationship indicates a crab that is lighter and larger in proportion to juvenile crabs. While it is true that the crab's exoskeleton undergoes a shape change from the juvenile to the adult phase during molting, the weight representing the growth of reproductive organs applies only within three months. After mating, the female develops the gonad and, within three months, lays her eggs. This study used specimens collected throughout the year, so some crabs had not developed gonads which indicates that although the exoskeleton changes to accommodate the growth of internal reproductive organs, once they are out of reproductive season, the space is not compensated for by muscle growth.

In the modeling and analyzing of the LWR in species of the genus *Callinectes*, the traditional potential model of Huxley has frequently been used. It assumes that the error term is multiplicative and follows a lognormal distribution on the arithmetic scale (Nevares-Martínez et al. 2003, Castillo et al. 2011, Omuvwie & Atobatele 2013, López-Martínez et al. 2014, Santos et al. 2016). The initial selection of this model is highly restrictive, as it assumes that on the geometric scale, the model turns out to be a simple linear regression model, where the allometric coefficient is constant. However, this is not always the case. Therefore, it is important to have a set of models with different LWR patterns (Katsanevakis 2007, Rodríguez-Domínguez 2018) to be able to choose the model that best fits the data.

In addition to having various options for the mean model, where one can choose the best among them, it is advisable to have different options for the error term distributions in the allometric regression model, allowing for different dispersion patterns for the problem. When the error term does not pass the normality test due to heavy tails, it is advisable to try modeling with another distribution with heavier tails than the normal distribution, such as the logistic distribution, the Student's t distribution, or a mixture of normal (Montesinos-López et al. 2019, Villa-Diharce et al. 2022).

When an incorrect mean model is chosen, one that is not supported by the data, as is the case with the traditional Huxley model, there is a loss of information. Similarly, selecting an incorrect distribution for the error term is also inconvenient. Incorrect estimation of the dispersion pattern leads to erroneous confidence intervals for the model parameters and hypothesis tests on the parameters.

The main drawback in this situation, when normality in errors is assumed, but heavy tails are present, is that the prediction intervals for the values will be excessively narrow.

Credit author contribution

E. Villa-Diharce: conceptualization, validation, methodology, data curation, formal analysis, writing-original draft, funding acquisition, project administration, supervision, review, and editing; G. Rodriguez-Dominguez: conceptualization, validation, methodology, formal analysis, writing-original draft, funding acquisition, project administration, supervision, review, and editing; M.G Russell-Noriega: conceptualization, validation, methodology, data curation, formal analysis, writing-original draft, review, and editing. All authors have read and accepted the published version of the manuscript.

Conflict of interest

The authors declare no potential conflict of interest in this manuscript.

REFERENCES

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N. & Csaki, F. (Eds.). 2nd International symposium on information theory. Akademia Kiado, Budapest, pp. 268-281.

Burnham, K.P. & Anderson, D.R. 2002. Model selection and multimodel inference: a practical information-theory approach. Springer, New York.

Castillo, J., Eslava, N. & González, L.W. 2011. Crecimiento del cangrejo *Callinectes danae* (Decapoda: Portunidae) de la Isla de Margarita, Venezuela. Revista de Biología Tropical, 59: 1225-1535.

Conover, W.J. 1971. Practical nonparametric statistics. John Wiley and Sons, New York.

Corgos, A. & Freire, J. 2006. Morphometric and gonad maturity in the spider crab *Maja brachydactyla*: a comparison of methods for estimating size at maturity in species with determinate growth. International Council for the Exploration of the Sea Journal of Marine Science, 63: 851-859. doi: 10.1016/j.icesjms.2006.03.003

Das, K.R. & Imon, A.H.M.R. 2016. A brief review of tests for normality. American Journal of Theoretical and Applied Statistics, 5: 5-12. doi: 10.11648/j.ajtas.20160501.12

Dash, G., Sen, S., Pradhan, R., Ghosh, S., Josileen, J. & Jayasankar, J. 2023. Modeling framework for establishing the power law between length and weight of fishes and a meta-analysis for validation of LWRs for six commercially important marine fishes from the northwestern Bay of Bengal. Fisheries Research, 257: 106496. doi: 10.1016/j.fishres.2022.106496

De Giosa, M. & Czerniejewski, P. 2016. A generalized, nonlinear regression approach to the length-weight relationship of European perch (*Perca fluviatilis* L.) from the Polish coast of the southern Baltic Sea. Archives of Polish Fisheries, 24: 169-175. doi: 10.1515/aopf-2016-0014

Dubey, S.D. 1969. A new derivation of the logistic distribution. Naval Research Logistics Quarterly, 16: 37-40. doi: 10.1002/nav.3800160103

Efron, B. & Tibshirani, R.J. 1993. An introduction to the bootstrap. Chapman and Hall/CRC, London.

Espinoza-Ahumada, V.C. 2019. Aspectos reproductivos de la jaiba azul *Callinectes arcuatus* (Ordway, 1863), en la laguna costera Navachiste. Tesis de Licenciatura, Universidad Autónoma de Sinaloa, Sinaloa.

Ghasemi, A. & Zahediasl, S. 2012. Normality tests for statistical analysis: A guide for non-statisticians. International Journal of Endocrinology and Metabolism, 10: 486-489. doi: 10.5812/ijem.3505

Gumbel, E.J. 1944. Ranges and midranges. Annals of Mathematical Statistics, 15: 414-422.

Halley, A.C. 2016. Prenatal brain-body allometry in mammals. Brain, Behavior and Evolution, 88: 14-24.

Huxley, J.S. 1932. Problems of relative growth. Dial Press, New York.

Jurado-Ruzafa, A. & Hernández-González, C. 2024. Length-weight relationships for seven fish species caught off northwest Africa. Journal of the Marine Biological Association of the United Kingdom, 104: 1-5. doi: 10.1017/S0025315424000298

Kassambara, A. & Mundt, F. 2020. factoextra: Extract and visualize the results of multivariate data analyses. R package version 1.0.7. [https:// CRAN.R-project.org/ package=facto extra]. Reviewed: January 12, 2024.

Katsanevakis, S., Thessalou-Legaki, M., Karlou-Riga, C., Lefkaditou, E., Dimitou, E. & Verripoulus, G. 2007. Information-theory approach to allometric growth of marine organisms. Marine Biology, 151: 949-959. doi: 10.1007/s00227-006-0529-4

Leyva-Vázquez, Y., Arzola-González, J.F., Rodríguez-Domínguez, G., Ramírez-Pérez, J.S., Ortega-Lizárraga, G.G., Félix-Ortiz, J.A., et al. 2022. Two new equations to evaluate allometry in the blue shrimp *Penaeus stylirostris* Stimpson, 1871 (Decapoda, Penaeidae) in a coastal lagoon from the Gulf of California. Crustaceana, 95: 421-438. doi: 10.1163/ 15685403-bja10200

López-Martínez, J., López-Herrera, L., Valdez-Loguin, J.E. & Rábago-Quiroz, C.H. 2014. Population dynamics of the swimming crabs *Callinectes* (Portunidae: Portunidae) components of shrimp bycatch in the eastern coast of the Gulf of California. Revista de Biología Marina y Oceanografía, 49: 17-29. doi: 10.4067/S0718-19572014000100003

Mishra, P., Pandey, C.M., Singh, U., Gupta, A., Sahu, C. & Keshri, A. 2019. Descriptive statistics and normality tests for statistical data. Annals of Cardiac Anesthesia, 22: 67-72.

Montesinos-López, A., Villa-Diharce, E., Echavarría-Heras, H. & Leal-Ramírez, C. 2019. Improved allometric proxies for eelgrass conservation. Journal of Coastal Conservation, 23: 71-91. doi: 10.1007/s118 52-018-0639-4

Nevares-Martínez, M.O., López-Martínez, J., Cervantez-Valle, C., Miranda-Mier, E., Morales-Aspeitia, R. & Anguiano-Carrasco, M.L. 2003. Evaluación biológica y pesquera de las jaibas *Callinectes bellicosus* y *Callinectes arcuatus* (Brachyura: Decapoda: Portunidae) en las bahías de Guásimas y Lobos, Sonora, México. In: Hendrickx, M.E. (Ed.). Contributions to the study of East Pacific crustaceans. Instituto de Ciencias del Mar y Limnología, Mazatlán, pp. 125-138.

Nunes-Pralon, B.G. & Negreiros-Fransozo, M.L. 2008. Relative growth and morphological sexual maturity of *Uca cumulanta* (Crustacea: Decapoda: Ocypodidae) from a tropical Brazilian mangrove population. Journal of the Marine Biological Association of the United Kingdom, 88: 569-574. doi: 0.1017/S0025 315408000453

Olentinoa, D., Lubichb, C.C.F., Rocha, M.D.P., Santosa, J.H.N., Gomesa, T., Beltrãoc, H., et al. 2023. Length-weight relationships of fish from sandy beaches Relação peso-comprimento de 16 espécies de peixes que ocorrem perto de praias arenosas ao longo da bacia do baixo rio Negro na Amazônia brasileira. Brazilian Journal of Biology, 83: e250003 doi: 10.1590/1519-6984.250003

Omuvwie, U. & Atobatele, O.E. 2013. Growth pattern, condition factor, trace metal studies and ectoparasitic load of the blue crab, *Callinectes amnicola* from Lagos Lagoon Badore, Ajah, Lagos, Nigeria. Cameroon Journal of Experimental Biology, 9: 34-43. doi: 10.4314/cajeb.v9il.5

Orlov, A.M., Mishin, A.V., Artemenkov, D.V. & Murzina, S.A. 2022. Length-weight characteristics of some pelagic fishes in the high latitudes of Atlantic sector of the Southern Ocean. Journal of Ichthyology, 62: 1137-1148.

Ortega-Lizárraga, G.G., Pérez-González, R., Rodríguez-Domínguez, G., Aragón-Noriega, E.A., Mendivil-Mendoza, J.E. & Franco-Ruiz, J.A. 2021. Biological traits of the spot prawn *Pandalus platyceros* Brandt, 1851 (Decapoda, Caridea) at its southernmost distribution limit. Crustaceana, 94: 573-596. doi: 10.1163/ 15685403-bja10110

Packard, G.C. 2014. Multiplicative by nature: Logarithmic transformation in allometry. Journal of Experimental Zoology - Part B: Molecular and Developmental Evolution, 322: 202-207. doi: 10.1002/jez.b.22570

R Core Team 2023. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. [https://www.R-project.org/]. Reviewed: January 12, 2024.

Rodríguez-Domínguez, G., Castillo-Vargasmachuca, S.G., Pérez-González, R. & Aragón-Noriega, E.A. 2015. The interannual variability of size at maturity of the brown crab *Callinectes bellicosus* Stimpson, 1859 (Brachyura, Portunidae) in the Gulf of California. Crustaceana, 88: 1339-1350. doi: 10.1163/15685403-00003491

Rodríguez-Domínguez, G., Castillo-Vargasmachuca, S.G., Pérez-González, R. & Aragón-Noriega, E.A. 2018. Allometry in *Callinectes bellicosus* (Stimpson, 1859) (Decapoda:Brachyura: Portunidae): single-power model *versus* multimodel approach. Journal of Crustacean Biology, 38: 574-578. doi: 10.1093/jcbiol/ruy060

Santos, M.C.F., Port, D., Fisch, F., Barbieri, E. & Branco, J.O. 2016. Biologia populacional de *Callinectes ornatus* associada a pesca do camarao-sete-barbas, Rio Sao Francisco (Alagoas e Sergipe, Brasil). Boletin do Instituto de Pesca, 42: 449-456. doi: 10.20950/1678-2305.2016v42n2p449

Shapiro, S.S. & Wilk, M.B. 1965. An analysis of variance test for normality (complete samples). Biometrika, 52: 591-611. doi: 10.1093/biomet/52.3-4.591

Song, M., Chen, Y., Wu, X., Liu, Y., Gong, Q., Li, F., et al. 2023. Length-weight relationships of four fish species of the Upper Yangtze River, Sichuan, China. Journal of Applied Ichthyology, 2023: 2550016. doi: 10.1155/2023/2550016

Received: March 4, 2024; Accepted: August 8, 2024

Stefanski, L.A. 1991. A normal scale mixture representation of the logistic distribution. Statistical Probability Letters, 11: 69-70. doi: 10.1016/0167-7152(91)90181-P

Stephens, M.A. 1997. Goodness of fit for the extreme value distribution. Biometrika, 64: 583-588. doi: 10.2307/2345336

Tsuboi, M., van der Bijl, W., Kopperud, B.T., Erritzøe, J., Voje, K.L., Kotrschal, A., et al. 2018. Breakdown of brain-body allometry and the encephalization of birds and mammals. Nature Ecology & Evolution, 2: 1492-1500.

Villa-Diharce, E.R. & Escobar, L.A. 2006. Using moment generating functions to derive mixture distributions. American Statistician, 60: 75-80. doi: 10.1198/00031 3006X90819

Villa-Diharce, E.R., Echavarria-Heras, H., Montesinos-López, A. & Leal-Ramírez, C. 2022. A revision of the traditional analysis method of allometry to allow extension of the normality-borne complexity of error structure: examining the adequacy of a normal-mixture distribution driven error term. Hindawi BioMed Research International, 31: 8310213. doi: 10.1155/2022/8310213

Warton, D.I., Duursma, R.A., Falster, D.S. & Taskinen, S. 2012. Smart 3- an R package for estimation and inference about allometric lines. Methods in Ecology and Evolution, 3: 257-259.

Williner, V., Torres, M.V., Carvalho, D.A. & König, N. 2014. Relative growth and morphological sexual maturity size of the freshwater crab *Trichodactylus borellianus* (Crustacea, Decapoda, Trichodactylidae) in the Middle Paraná River, Argentina. ZooKeys, 457: 159-170. doi: 10.3897/zookeys.457.6821

Zhu, X., Chapelsky, A. & Carmichael, T. 2021. Statistical models for interpreting spatial variations of allometric growth of subarctic lake whitefish populations. Fisheries and Oceans Canada, Winnipeg.