Universidade Estadual de Maringá
Departamento de Estatística
Programa de Pós-graduação em Bioestatística

DIEGO CORRÊA ALVES

Inferência bayesiana para estudos de populações de peixes: o caso do *Astyanax paranae* do alto rio Paraná

Maringá - Paraná
Março de 2018

UNIVERSIDADE ESTADUAL DE MARINGÁ
CENTRO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOESTATÍSTICA

DIEGO CORRÊA ALVES

Inferência bayesiana para estudos de populações de peixes: o caso do *Astyanax paranae* do alto rio Paraná

Maringá
2018

DIEGO CORRÊA ALVES

Inferência bayesiana para estudos de populações de peixes: o caso do *Astyanax paranae* do alto rio Paraná

Dissertação em forma de artigo científico apresentada ao Programa de Pós-Graduação em Bioestatística do Departamento de Estatística, Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para à obtenção do título de Mestre em Bioestatística.
Orientador: Prof. Dr. Robson Marcelo Rossi

Maringá
2018

DIEGO CORRÊA ALVES

Inferência bayesiana para estudos de populações de peixes: o caso do *Astyanax paranae* do alto rio Paraná

Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Mestre em Bioestatística.

BANCA EXAMINADORA

Prof. Dr. Robson Marcelo Rossi
Universidade Estadual de Maringá - UEM

Profª. Dra. Terezinha Aparecida Guedes
Universidade Estadual de Maringá - UEM

Daniela A. Lino

Profª. PhD. Daniela Andressa Lino Lourenço
University of Geórgia - EUA

Maringá, 09 de março de 2018.

Artigo elaborado, formatado e submetido conforme as normas da publicação científica *Ecology of Freshwater Fish*. Disponível em: < http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1600-0633/homepage/ForAuthors.html>

Sumário

Estimating fish population parameters: Advancing in ecological hypothesis tests with Bayesian inference

Diego Corrêa Alves[1*], Lilian Paula Vasconcelos[2], Yzel R. Súarez[3], Sidnei E. Lima-Junior[3], Robson Marcelo Rossi[1]

1- Graduate Program in Biostatistics, Department of Statistics, State University of Maringá, Colombo Avenue, 5790, 87020-900, Maringá-PR, Brazil.

2- Graduate Program in Ecology of Continental Aquatic Environments, Department of Biology, State University of Maringá, Colombo Avenue, 5790, 87020-900, Maringá-PR, Brazil.

3- State University of Mato Grosso do Sul. CInAM, Ecology lab. Dourados-Itahum highway, km 12, 79804-970, Dourados-MS, Brazil.

* Correspondence author: dcalves@uem.br

Telephone/fax number: +550443011-5342/ +550443011-4338

Short running title: Bayesian inference for fish population

Abstract

The Bayesian approach is an important statistical framework to minimize the lack of knowledge about fish populations. Thus, models and reparametrizations, which favor the Bayesian approach, were proposed to estimate parameters that allow the test of ecological hypotheses. A population of *Astyanax paranae* was used as case study due to its life history characteristics that facilitated to obtain representative population samples. Two models were proposed: one for the weight-length relationship, which allows to evaluate if the growth pattern is polyphasic, and another one capable of estimating somatic growth parameters, based on only one length sample. The sex ratio was estimated at 1.4 females:male, without evidence of alteration according to their length or weight. The first sexual maturation standard length of females was estimated at 38.12 mm and 37.75 mm, using non-informative and informative prior distributions, respectively. The weight-length relationship was inferred to occur in two stanzas, with the greatest increase in weight for females only in the first. In the second stanza both sexes present greater increase in weight than in the first one, but in an indistinct way between them. Regarding to somatic growth, males and females have the same age composition, predominated by young specimens, and growth patterns are equivalent between them. The proposals of a biphasic model for weight-length relationship under the Bayesian approach to the somatic growth model demonstrated great potential and applicability for the population studied. These analyzes allowed the identification of complex biological patterns, which may be frequent in nature but difficult to identify.

Keywords: Fishery, population dynamics, stock assessment, upper Paraná River, von Bertalanffy.

1. Introduction

The fishes are the group of organisms with the greatest diversification of forms among vertebrates (Vazzoler, 1996), with approximately 34 thousand species (Eschemeyer & Fong, 2017). The high species richness is also reflected in the morphological and biological diversity of this group (Langeani et al., 2007). The knowledge of quantitative aspects of fish species is an important tool for the study of fisheries biology, necessary for the management and preservation of aquatic ecosystems. In general, without the basic data obtained from population studies, the landscape, ecosystem and conservation studies would either be infeasible or unrealistic (Rockwood, 2006). Thus, population ecology, which seeks to understand aspects related to the growth, reproduction and survival of groups of individuals (Rockwood, 2006) plays an important role in the protection and management of populations. Therefore, the focus of any study in the population ecology area may be motivated by very practical considerations, such as fishery management or wildlife management and conservation (Akçakaya et al., 1999). However, in spite of the practical relevance of these studies, the amount of available knowledge on biological and population aspects is still incipient for many populations, species, and even entire families of fish, especially in the neotropical region.

The limited knowledge available on the population parameters of many species is further aggravated by two critical factors: *i)* the demand for researchers to publish in high-impact journals, making it difficult to publish many population studies (descriptive, basic science); that is, there is a large volume of data on several populations that are not published in the scientific literature; and *ii)* the complexity of biological systems (Hilborn et al., 2003), such as high spatial, temporal and morphotypic variability in biological parameters. These parameters may differ within the same species from distinct populations, and even within the same population, *e.g.* in time, habitat types or morphotypes. Therefore, knowledge of the biological aspects of each population is necessary for an effective management and conservation of these resources. Fortunately, there is a growing effort in the organization and availability of information/data (Big Data) in the most diverse areas, especially in fisheries Science (McManamay & Utz, 2014).

In this sense, the execution of population studies using the Bayesian approach (Ellison, 1996, Ellison, 2004, Hobbs & Hilborn, 2006) is a promising option, for at least two main reasons: *i*) to enable the inclusion of several relevant information by means of *a priori* distributions, whether they come from previous systematic studies of distinct populations of the same species, or from taxonomically related species; (*q.v.* McManamay & Utz, 2014), or even informal information such as the opinions of experts, such as researchers and fishermen; and *ii)* enable to make inferences based on the *a posteriori* distributions, which allow more intuitive probabilistic statements about the parameters.

Thus, the main goal of this study was to analyze, by the Bayesian statistical approach, the population structure of *Astyanax paranae* of the Upper Paraná River. This species has characteristics that make it an important tool for variations in life history studies, such as small size, opportunistic habits, fast life cycle, reduced locomotion potential and no spatial stratification, which makes it easy to obtain representative samples of the population. Besides that, the species has a wide distribution (Súarez et al., 2017) with high potential for differentiation among populations due to their isolation; which causes not only genetic, but also ecological characteristics differentiated among them.

Thereunto, the basic population parameters were estimated: sex ratio, first maturation size of females (L_{50}), weight-length relationship and somatic growth. Growth pattern analyzes were also performed to test the ecological hyphotheses about sexual dimorphism, isometry and growth stanzas. For this, two alternative models were proposed: one for the weight-length relationship which allows to evaluate if the growth pattern of this population occurs in a different way in two

stanzas, and another one capable of estimating somatic growth parameters, based only on one length sample.

2. Metodology

2.1. Study Area

The Paraná River is the main water body of the Prata basin, which is the second largest river basin in South America. In the upper Paraná River basin, there is the Guiraí River basin (22° 45'S, 54° 30'W and 22°15'S, 53° 30'W), occupying an area of about 2248 km^2. The headwaters of the Guiraí River is located in an area of 370m altitude and the river runs approximately 100 km until it flows into the Ivinhema River, at 238m altitude, where, in its lower portion, it forms the boundary of the Parque Estadual das Várzeas do rio Ivinhema.

2.2. Data sampling

Samplings were carried out bimonthly, from October/2006 to August/2007, in seven streams of the Guiraí river subbasin. The samples occurred during the day, without distinction of period, with a rectangular sieve measuring 0.8 x 1.2 m with 2 mm of mesh opening. A standardized effort of 20 "beats" was applied at each point in a stretch of approximately 50 m in length. Permission to fish was granted by IBAMA (# 13458-1).

The fishes were fixed in 10% formalin and preserved in 70% ethyl alcohol solution. All the individuals collected were measured (standard length), weighed and sexed, and the females had their gonads removed for weighing and determination of gonadal maturation stages. These stages were determined macroscopically, according to Vazzoler (1996), in one of the categories: immature gonads (young with filiform, translucent and very small ovaries); maturing or resting gonads (adults with ovaries having opaque, small or medium oocytes); mature gonads (ovaries occupying almost completely the celoma cavity, having large, opaque or translucent oocytes) and exhausted gonads (post-reproduction individuals with flaccid ovaries, with distension of the membranes and hemorrhagic aspects). For many immature individuals, sex could not be determined because of the tiny size of the gonads.

2.3. Data analysis

All analyzes were performed in software R (R Core Team, 2017) with the aid of the 'R2jags' package (Su & Yajima, 2015) to obtain the posteriors distributions, via Markov chains Monte Carlo (MCMC) method and 'coda' package (Plummer et al., 2006) for the chains convergence diagnostic tests (Heidelberger & Welch, 1983). In each subsection, the models were selected based on the deviance information criterion (DIC) (Spiegelhalter et al., 2002). Also, the chain convergence inspection was complemented by graphical methods, such as the *traceplot*. The analyzes were performed only with data from fishes that had the sexual determination possible.

2.3.1. Sex ratio

The sex ratio was modeled by a logistic regression model, with a dichotomous response variable (1: females, 0: males), assumed to have a Bernoulli probability distribution and logit link function, equivalently to that detailed in the next section.

Four other models have been fitted with different predictive variables. Two of them had simple linear effect associated with the standard length and total weight predictive variables, respectively. The two others had the addition of the quadratic effect of these same predictive variables.

Thus, if the sex ratio changed according to these variables, it would be an indication that one of the sexes reaches higher sizes or weights than the other. For all the parameters were allocated flats prior distributions, normal distribution with zero mean and high variance (10^3). The

posteriors distributions were generated from the MCMC process, with $6e^6$ iterations, with discard of the first half of the chain and records in a range of 1000 amplitude. Therefore, the final sample contained 3000 values.

2.3.2. First sexual maturation size (L_{50})

The standard length of the first sexual maturation (L_{50}), defined as the length at which 50% of females are sexually mature, was also estimated by logistic regression. Thereunto, it was considered a dichotomous response variable *m*, which receives values "1" for mature females (or females that have already undergone some reproductive event) and "0" for those who have not reached sexual maturity yet. This variable was assumed to follow a Bernoulli distribution, as follows.

Eq. 1
$$m_i \sim Bernoulli(\pi_i)$$

m_i = dichotomous variable indicating (1) for mature females and (0) otherwise;

π_i = probability that the *i*-th female is mature.

Previously assuming that the logit of π is linearly related to the standard length (L), we have:

Eq. 2
$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 L_i$$

β_0 = intercept of the model, which represents the logit of the probability of a female with standard length equal to zero. Biologically, this parameter is not of interest;

β_1 = mean increase in the logit of π_i for each unit of standard length added to the length. The exponential of this coefficient is interpreted as the *odds ratio* (OR);

L_i = standard length of the *i*-th female.

The amount of interest, the first maturation size (L_{50}), is a ratio between these two parameters β_0 and β_1. By replacing π_i by 0.5 and isolating L_i, we have the length at which 50% of the females are expected to be mature. Thus, the L_{50} can be estimated from the following relation:

Eq. 3
$$L_{50} = -\frac{\beta_0}{\beta_1}$$

However, by the Bayesian approach it is necessary to specify the prior distributions for the parameters directly estimated by the model, β_0 and β_1. These parameters represent somewhat abstract quantities, making it difficult to specify informative prior distributions for these parameters, if it is of interest. Although β_0 is not biologically relevant, it is possible to think of slightly informative priors for this parameter. Since it representes the logit of the probability of a female of standard length equal to zero being mature, it is reasonable to infer, *a priori*, that this probability must be very close to zero, and consequently its logit a negative number.

For the parameter β_1 it is difficult to make *a priori* inference that allows the specification of an informative prior distribution. The amount of interest, the L_{50}, is intuitive and feasible to make *a priori* inferences, even empirically, such as consultation with fishermen or other experts. Thus, a way that allows to estimate this amount, and yet, consider *a priori* inferences about this amount seems more attractive. This can be reached by reparametrization of this model, leaving it in function of β_0 and L_{50}. Thus, by reversing the positions of β_1 and L_{50} in eq. 3 and substituting β_1 in the linear predictor, we have the following model:

Eq. 4
$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 - \left(\frac{\beta_0}{L_{50}}\right) L_i = \beta_0(1 - L_{50}^{-1})L_i$$

Based on these two parametrizations, three models were fitted. The firts considered the first parametrization and was specified with flats prior distributions for β_0 and β_1, a normal distribution with mean zero and variance 10^3. The second and the third models considered the second parametrization and were specified with slightly informative and informative prior distributions, respectively. In these latter two models a normal distribution with high variance (10^3) was specified for β_0, just as in the first model, nevertheless, restricting the parametric space to the interval between -1000 and 0, by an indicator function. The L_{50} parameter was specified with prior normal distributions, with mean 37.5 mm and variances of 5^2 and $1{,}25^2$ for the second and third models, respectively. The mean value of L_{50} *a priori* has been established based on data from www.fishbase.org, in which the estimatives for females of *A. paranae* are between 35-40 mm and a center value of 37.5 mm.

The posteriors distributions of the estimated parameters were generated from the MCMC process, with $2e^6$ iterations, discarding the first half of the chain, and recording in a range of 500 amplitude, culminating in a final sample of size 2000.

2.3.3. Weight-length relationship

The weight-length relationship is given by a power equation, described by Huxley (1924):

Eq. 5 $$W = \alpha L^{\beta},$$

where W is the total weight of the individual and L is the standard length, α e β are model parameters. The parameter α representes the mean weight for individuals with an unity of standard length, often interpreted as the body shape – more fusiform fishes will have lower α values than more spherical shapes fishes (Froese, 2006). If a fish grows without changing its shape or density, it is considered that this organism presents isometric growth, that is, the fish volume is proportional to its length. If it happens, and the fish maintains constant density during growth, its weight will be an indicator of its volume and it will present the isometric growth when the allometric coefficient (β) is equal to 3. If $\beta > 3$, then the fish will be more "fat", that is, more weight per unit of length will be added (Blackwell et al., 2000).

However, fish growth patterns (ratio between weight and length increments), commonly assessed by weight-length relationship, can be complex with different patterns at different life stanzas (Bervian et al., 2006). Thus, it is not trivial to test the hypothesis of *i)* tendency of sexual dimorphism in the growth pattern, *ii)* isometric growth pattern and *iii)* polyphasic growth pattern, since these can not be tested independently. For example, it would be possible to establish at least eight concurrent hypotheses, only involving the abovementioned initial hypotheses (Table 1). The growth pattern can be polyphasic, and in one stanza the pattern is isometric and equivalent in both sexes, while in the other stanza there may be distinctions between the sexes, as well as the allometric growth pattern (for one or both sexes). Given this complexity, a competing model selection approach was used, which translated competing biological hypotheses (Table 1), in search of the most parsimonious model/hypothesis.

Table 1

Traditionally, the method used to model the polyphasic weight-length relationship is based on the idea of Bervian et al. (2006), which consists on a modification in Huxley's proposal (1924) in order to incorporate a distinct growth pattern into different stanzas, usually two. The proposal of Bervian et al. (2006) is based on a power equation, modified using a logistic function for the α and β parameters of Huxley (1924), as follows:

Eq. 6 $$W_i = \alpha_1 L_i^{\beta_1}(1+\exp(r(L_i - L_c)))^{-1} + \alpha_2 L_i^{\beta_2}(1+\exp(-r(L_i - L_c)))^{-1},$$

where, W_i and L_i are the total weight and standard length of the i-th individual, respectively, α_1 and β_1 are the proportionality constant (body shape) and the allometric coefficient in the first life stanza, and, α_2 and β_2, of the second life stanza, the parameter L_c is the length where the stanza change occurs and *r* is the change rate between stanzas.

However, we propose to test if the weight-length relationship is polyphasic with an alternative modification of the Huxley (1924) model, based on the exponential of a piecewise linear regression, as follows:

Eq. 7
$$W_i = \frac{\alpha}{L_c^{\delta x_i}} L_i^{(\beta+\delta x_i)},$$

where, L_c is the length at which the stanza change occurs, δ is the increment that occurs in the allometric coefficient in the second stanza, e, x_i is a *dummy* variable, such that:

Eq. 8
$$x_i = \begin{cases} 1, & se\ L_i \geq L_c \\ 0, & se\ L_i < L_c \end{cases}$$

Thus, when the fish is in the firts life stanza ($L_i < L_c$), then $x_i = 0$, and in this case the weight of this fish is given simply by the equation of Huxley (*Eq*. 5). Otherwise, δ units will be added in the allometric coefficient, resulting in an allometric coefficient of $\beta + \delta$ at the second stanza.

The model was assumed to be W_{ij}, total weight of the *i*-th specimen of sex *j* (male or female), following a normal distribution with precision (inverse of variance) τ_j, and a mean differently determined for each model in Table 1, as follows:

Model *Agg*: $\mathbb{E}[W_{ij}] = \alpha L_{ij}^{\beta}$

Model *FM*: $\mathbb{E}[W_{ij}] = a_j L_{ij}^{\beta_j}$

Model *Agg₂*: $\mathbb{E}[W_{ij}] = \frac{\alpha}{L_c^{\delta x_{ij}}} L_{ij}^{(\beta+\delta x_{ij})}$

Model F_1M_2: $\mathbb{E}[W_{ij}] = \begin{cases} \frac{\alpha_j}{L_c^{\delta x_{ij}}} L_{ij}^{(\beta_j+\delta x_{ij})}, & j = male \\ \alpha_j L_{ij}^{\beta_j}, & j = female \end{cases}$

Model F_2M_1: $\mathbb{E}[W_{ij}] = \begin{cases} \frac{\alpha_j}{L_c^{\delta x_{ij}}} L_{ij}^{(\beta_j+\delta x_{ij})}, & j = female \\ \alpha_j L_{ij}^{\beta_j}, & j = male \end{cases}$

Model $A_1F_2M_2$: $\mathbb{E}[W_{ij}] = \frac{\alpha}{L_{c_j}^{\delta_j x_{ij}}} L_{ij}^{(\beta+\delta_j x_{ij})}$

Model $F_1M_1A_2$: $\mathbb{E}[W_{ij}] = \frac{\alpha_j}{L_{c_j}^{\delta_j x_{ij}}} L_{ij}^{(\beta_j+\delta_j x_{ij})}, where\ \delta_2 = \beta_1 + \delta_2 - \beta_2$

Model F_2M_2: $\mathbb{E}[W_{ij}] = \frac{\alpha_j}{L_{c_j}^{\delta_j x_{ij}}} L_{ij}^{(\beta_j+\delta_j x_{ij})}$

The prior distributions for the precisions were assumed Gamma (10^{-3}, 10^{-3}). For the parameters α_j, they were normal with mean zero and variance10^3, truncated in the range [10^{-6}, 1] by an indicator function. Vague Normal distributions with mean 3 and variance 10^4, truncated in the interval (1, 6), were also specified for the allometric coefficients. The stanza change lengths were specified with a uniform distribution, defined among 0 and 20 cm. Finally, uniform distributions,

defined in the interval (-5, 5), were specified for δ_j. The posteriors distributions of the estimated parameters were generated from the MCMC process, with $7e^5$ iterations, discarding the first $4e^5$ iterations, and recording in a range of 150 amplitude, resulting in a final sample of size 2000.

2.3.4. Somatic growth

Somatic growth in fish is commonly modeled by the von Bertalanffy function (VBGF). The VBGF has several parametrizations, and the one originally proposed by von Bertalanffy (1938) is given as follows:

Eq. 9
$$\mathbb{E}(L_i|\, t_i) = L_\infty - (L_\infty - L_0)e^{-Kt_i},$$

where, L_i and t_i are the length and the age of the *i*-th individual, respectively. L_∞, K and L_0 are parameters of the von Bertalanffy equation (*Eq*. 9) which represent the asymptotic length, for which growth is zero, growth coefficient and mean length for age zero, i.e., birth length, respectively.

Francis (1988) proposed another parameterization for eq. 9, with more realistic and consistent parametric interpretations, as well as a lower correlation between parameters, as follows:

Eq. 10
$$\mathbb{E}(L_i|\, t_i) = L_1 - (L_3 - L_1)\frac{1-r^{2\frac{t_i-t_1}{t_3-t_2}}}{1-r^2},$$

where, $r = \frac{L_3-L_2}{L_2-L_1}$, L_1, L_2 e L_3 are the model parameters and represent the average lengths for the ages t_1, t_2 e t_3, respectively. Ages t_1 e t_3 are arbitrarily defined, while age t_2 is the intermediate age between the other two. Francis (1988) also demonstrated that the original parameters could be derived from the new parameters as follows:

Eq. 11
$$L_\infty = L_1 + \frac{L_3-L_1}{1-r^2} \quad \text{e} \quad K = \frac{-2\log(r)}{t_3-t_1}$$

The first option to obtain the estimates of the VBGF parameters is to fit a nonlinear model to the length and age crossed data of each individual. However, the acquisition of fish age data is a laborious, time-consuming, costly and sometimes imprecise and inaccurate task (Campana, 2001; Chang & Maunder, 2012). In this sense, Alves et al. (submitted in November/2017) proposed a model based on the likelihood that allows to infer about these parameters and about the population age composition (number of individuals at each age), based only on the length data. Thus, the model proposed by Alves et al. (submitted in November/2017) was expanded to Bayesian approach in this study, since all development based on inference with the likelihood has already been performed. This model assumes that the L_i lengths are organized in intervals (length classes), and, counted the number of individuals present in each length class, N_j, number of individuals in the length class *j*, where *j* varies from 1 to *n*, where *n* is the number of length classes. Now consider $\boldsymbol{N}$ a vector of length *n*, containing the *n* random variables N_j, that is, $\boldsymbol{N}=\{N_1, N_2, N_3, ..., N_n\}$. The expected value of the vector $\boldsymbol{N}$ is a function of the age composition $\boldsymbol{A}$ (number of individuals in each age group) and the transition matrix $\boldsymbol{P}$ (matrix with theoretical values obtained by some somatic growth model), as follows:

Eq. 12
$$\mathbb{E}(\boldsymbol{N}) = \boldsymbol{AP},$$

where, $\boldsymbol{A}$ = line vector containing the age composition (number of individuals in each age group *i*), where *i* varies from 1 to the number of age groups *m*; $\boldsymbol{P}$ = transition matrix (age x length), with dimensions *m* x *n*. Each row in this matrix adds a unit, since each element of the row represents the probability that an individual of an age group belongs to each length class, that is, P(*L=j* | *t=i*).

This matrix is filled with theoretical values by some somatic growth model, in this case the VBGF, computed as described by Parrack and Cummings (2003), as follows:

Eq. 13 $P[i,j] = \int_{j}^{j+1} f(L|t=i,\boldsymbol{\phi})\mathrm{d}L = \int_{-\infty}^{j+1} f(L|t=i,\boldsymbol{\phi})\mathrm{d}L - \int_{-\infty}^{j} f(L|t=i,\boldsymbol{\phi})\mathrm{d}L,$

where, $P[i,j]$ = element of the line *i* (age *i*, *t* = *i*) and column *j* (length *j*, *L* = *j*) of the matrix $\boldsymbol{P}$; *L* = fish length, a random variable, dependent on age (*t*), distributed by a normal probability density function, $L|t \sim \mathrm{N}(\mu_L, \sigma^2_{VBGF})$, where μ_L is given by $\mathbb{E}(L_i|\ t_i)$, as described by the VBGF (equations 9 and 10), and, σ^2_{VBGF} was set at 3, according Bartoo & Parker (1983).

The parameters set $L|t$ is $\boldsymbol{\phi_1} = \{L_1, L_2, L_3, \sigma^2_{VBGF}\}$, if the parameterization proposed by Francis (1988) is used or $\boldsymbol{\phi_2} = \{L_\infty, L_0, K, \sigma^2_{VBGF}\}$, if the original parameterization is used. Based on the values of $\boldsymbol{\phi}$ it is possible to construct the $\boldsymbol{P}$ matrix with theoretical values, which are a VBGF function.

N_j are assumed to be random variables independent from each other, but not identically distributed. Each N_j follows a probability distribution $f_j\ (N_j\ |\theta_j)$. Thus, $f_{\boldsymbol{n}}\ (\boldsymbol{N}\ |\boldsymbol{\theta})$ is the joint probability distribution for $\boldsymbol{N}$ and is given by the product of the marginal probability distributions of each random variable N_j, as follows:

Eq. 14 $$f_{\boldsymbol{n}}(\boldsymbol{N}|\boldsymbol{\theta}) = \prod_{\boldsymbol{j=1}}^{\boldsymbol{n}} f_j(N_j|\theta_j).$$

In this study, $f_j(N_j|\theta_j)$ was assumed to be a normal, under the justification of the asymptotic theory, as proposed and discussed by Alves et al. (submitted in November/2017). Thus, the joint probability distribution $f_{\boldsymbol{n}}(\boldsymbol{N}|\boldsymbol{\theta})$ can be expressed in the multivariate form, resulting in the following likelihood function $\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{N})$.

Eq. 15 $$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{N}) = f_{\boldsymbol{n}}(\boldsymbol{N}|\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}|}} \exp\left\{-\frac{1}{2}(\boldsymbol{N}-\boldsymbol{AP})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{N}-\boldsymbol{AP})\right\},$$

where, $\boldsymbol{\Sigma}$ is the (co)variance matrix of $\boldsymbol{N}$, with dimensions *n* x *n*. Assuming that the *m* random variables N_j are independent and uncorrelated, $\boldsymbol{\Sigma}$ is a definite positive diagonal matrix, where the main diagonal contains the variances of each N_j, σ^2_N.

In these case study, the response variable N_{jk} was defined as abundance, in number of individuals, of the kth sex (male or female) in each length class *j*, where *j* varies from 1 to *n*. The number of length classes *n* was set at 38, where the lower limit of the first length class was 10 mm, with amplitude of 2 mm among classes, and upper limit of the last class at 86 mm. The age composition was established as a vector of size *m* = 5, containing the proportion of individuals (a_t) at each age *t*, where the mean ages were set at 0.25, 0.50, 1.0, 1.5 and 2,0 years, *i.e.*, 3, 6, 12, 18 and 24 months, respectively. In all, 12 competing models were fitted, which differed among them in relation to: *i)* the specification of prior distributions, vague or informative; *ii)* the parameterization of the VBGF used, the one proposed by Francis (1988) or the original one; and *iii)* regarding the hypotheses of sexual dimorphism. Three hypotheses of dimorphism were established, H_1) VBGF growth parameters and age composition for the clustered sexes; H_2) Distinct VBGF growth parameters between the sexes, with the same age composition; and H_3) Distinct VBGF growth parameters and age compositions between the sexes. For the models with the VBGF parameterization proposed by Francis, the ages t_1, t_2 and t_3 were set at 0, 0.5 and 1.0 year, respectively. Thus, the L_1, mean length of age t_1=0, reduces to the same parameter of the original VBGF L_0. The values for both parameters L_0 and L_1 were set at 3 mm, which is the standard length between larval hatching and a larvae in the pre-flexion stage, for the congener *A. altiparanae* (Nakatani et al., 2001).

The prior distribution for the age composition $\boldsymbol{A}$, independently of the fitted model, was that of Dirichlet, with all hyperparameters set at 1. The N_{jk} variances were parametrized for precision, specifying the prior Gamma (10^{-2}; 10^{-2}) distribution for all sexual groups.

The prior distributions for the somatic growth parameters of both VBGF parametrizations are specified in table 2. For the models fitted with vague prior distributions, the same distributions were specified for all sexual groups (Table 2). The mean values of the parameters L_2, L_∞ and K were extrapolated from the study of Nascimento et al. (2017).

Table 2

Given the greater complexity of these models and difficulty of convergence, 8 parallel chains were fitted in order to evaluate the sensitivity of the models. Each chain was generated via MCMC, with $2.2e^6$ iterations, discarding the first $2e^5$ iterations, and recording in 1000 amplitude intervals, culminating in final samples of size 1800. The selection of the final model was based on three features: *i*) chains convergence, measured as the number of parallel chains that converged in some stationarity regions; *ii*) robustness, measured as the number of stationarity regions between the parallel chains; and *iii*) adjustment and parsimony, measured as the lowest DIC value among the parallel chains that converged. A score was established for each feature, ranging from 0 to 1, and a final score was calculated using the arithmetic mean of the three previous scores. The model with the highest final score was chosen as the final model for more detailed checks. The scores of each feature were obtained by the following items, respectively: *i*) the quotient between this metric and 8 (number of parallel chains); *ii*) the ratio of the quotient between 8 (number of parallel chains) and the value of this metric and 8 (number of parallel chains); and *iii*) value 1 for the model with the lowest DIC, 0.90 for the models that presented a Δ_{DIC} less than 5 (non-significant difference), and 0.10 for the others, noting that the DIC value considered for each model was that from the parallel chain that presented the lowest DIC among those that converged.

3. Results

3.1. Sex ratio

A total of 326 specimens of *Astyanax paranae* were sampled, of which 174 were sexually determined - 73 males and 101 females. The set of the fitted models to the data of the 174 specimens did not present significant evidence that the sex ratio varies according to the fish weight or length (Table 3). Although all the five fitted models are equally parsimonious ($\Delta_{DIC}<5$), the null model (only with intercept), besides being the simplest model, was the one that presented the lowest DIC and was the selected model. A sex ratio of approximately 1.4 females/male was estimated by the null model.The 95% credibitlity interval did not contain the value 1 (1.024; 1.893), which means that, this population contained more females than males in the studied period, being that about 58% (50.6; 65.4) of the individuals were females (Table 4). However, there was no significant evidence to discard the remaining models and we chose to represent their predictions graphically in figure 1. It is noticeable that the models with the weight effect have the greatest positive tendency (Figure 1), and the model with simple weight effect was the second model with lower DIC (Table 3).

Table 3

Table 4

Figure 1

3.2. First sexual maturation size (L_{50})

The results of the logistic regression models fitted to estimate the first maturation size of *A. paranae* females indicate that the proportion of mature females increases according to their length

(Table 5, Figure 2). The L_{50} estimates were similar among the three fitted models, which were distinct regarding the prior distributions (Table 5). These results demonstrate that the different priors exerted little influence on the posteriors estimates, indicating that the data were informative. However, the L_{50} estimates were slightly more accurate by the model with more informative prior about this parameter (Table 5, Figure 2). According to this model, the standard length of first maturation was estimated at 37.75mm (35.73; 39.64).

Table 5

Figure 2

3.3. Weight-length relationship

Eight models were fitted to translated different biological hypotheses into the weight-length relationship (Table 1). The *Agg* model assumes that there are no differences in this relationship between the sexes (there is no tendency for sexual dimorphism) and that the growth pattern (ratio of weight to length increments) is not altered at any particular life history stanza. Thus, this model is analogous to the null hypothesis of absence of distinct growth patterns by sex and life stanzas. Although this was the simplest model, it was also the least parsimonious, with the highest DIC value (-75.64), which shows that there are more complex patterns in this population weight-length relationship (Table 6). The significantly more parsimonious models (Δ_{DIC}>20 for the other models) were $F_1M_1A_2$ and F_2M_2 respectively (Table 6). Initially, both models should be considered as competing, *i.e.*, equivalently plausible, given the low Δ_{DIC} (0.95). $F_1M_1A_2$ model translates the hypothesis that the growth pattern is altered in a certain life history stanza of each sex, but with different patterns between the sexes only in the first stanza, and the same allometric coefficient at the second stanza. The difference between this model and the F_2M_2 model is that in the second life stanza of each sex the estimates of the allometric coefficients can be different between the sexes.

Table 6

The two most parsimonious models, $F_1M_1A_2$ and F_2M_2, were selected for a more detailed comparative analysis. In general, these two models generated very close estimates. Regarding the parameter *a*, which represents the body shape or the predicted weight for individuals of 1.0 cm, the estimates (posterior mean and 95% credible interval) were 0.019 (0.015; 0.023) for females and 0.023 (0.018; 0.029) for males, obtained by both models (Figure 3). Estimates for the standard length of stanza change were also equivalent between the two models, 6.23 (6.04; 6.44) for females and 5.74 (5.52; 6.01) for males obtained by the $F_1M_1A_2$ model, and 6.30 (6.05; 6.52) for the females and 5.70 (5.26; 6.04) for males by the F_2M_2 model (Figure 3).

For the allometric coefficient of the first life stanza, the estimates obtained by the model $F_1M_1A_2$ were 3.08 (2.96; 3.21) and 2.88 (2.73; 3.03) for females and males, respectively, and 3.09 (2.96, 3.21) for females and 2.88 (2.72, 3.05) for males by the F_2M_2 model (Figure 3). For the second life stanza, these estimates were 5.22 (4.43, 6.41) for both sexes, estimated by the model $F_1M_1A_2$, and 5.85 (4.40, 7.83) and 5.06 (3.90, 6.86) for females and males, respectively, obtained by the F_2M_2 model (Figure 3). Thus, in the first life stanza, the estimates of the allometric coefficient for both models were equivalent, whereas in the second life stanza, the first model (that assumes equal coefficients for both sexes) obtained intermediate estimates for males and females generated by the second model (Figure 3).

Considering that the difference between the estimated allometric coefficients between the sexes in the second life stanza by the second model were 0.80 (-1.29, 3.15), that is, having zero in the CI 95% (which means no significant differences between the estimates of both sexes in this stanza); and that the predicted mean weights were equivalent between the models (Figure 4), it is

possible to infer that they express the same posterior pattern. Therefore, the model selected to verify the hypotheses of allometry and sexual dimorphism was the simplest one - $F_1M_1A_2$.

Figure 3

Figure 4

Based on the estimates obtained by the $F_1M_1A_2$ model, when individuals are 1 cm, males had a slightly higher mean predicted body weight (α: body shape) than females, 0.023 g vs. 0.019 g (Figure 3). Moreover, the posterior probability of $\alpha_m > \alpha_f$, $P(\alpha_m - \alpha_f > 0)$ was estimated at 91.8%, and $P(\alpha_m - \alpha_f > 0.002)$=80.2%, which means that the probability that the mean weight of males is greater than that of females is 91.8%, and that the probability of this difference being greater than 2 mg is 80.2% for individuals of 1 cm. However, females presented estimates for the allometric coefficient higher than males, 3.08 vs 2.88, at the first life stanza (Figure 3). The probability of the allometric coefficient of females being greater than that of males was 97.3% ($P(\beta_{1f} - \beta_{1m} > 0)$=97.3%) with a probability of 84.7% of this difference being higher than 0.1 ($P(\beta_{1f} - \beta_{1m} > 0.1)$=84.7%), indicating that females are more likely to gain more weight as they grow compared to males. At the second life stanza, on the other hand, both sexes present the same value for the allometric coefficient. Finally, $P(\beta_{1f} > 3)$ = 91.4% and $P(\beta_{1m} > 3)$ = 6.2% indicate that in the first life stanza, females tend to have a positive allometric growth, with a greater proportion in the increase by weight than in length, whereas males tend to the inverse pattern of negative allometric growth.

In short, it is possible to infer that males and females present biphasic growth, with greater increase in weight for females, only in the first stanza. In the second stanza, both sexes present a greater increase in weight than in the first stanza, but in an indistinct way between the sexes.

3.4. Somatic growth

It is not easy to perform an objective and direct analysis to select the most appropriate somatic growth model based only on length data given the complexity of these models. Thus, the selection of these models was performed after sensitivity exploratory analyzes contained in the Supplementary Material and objective analyzes based on three different features: convergence, robustness and goodness of fit. The models that considered clustered sexes were successful in the convergence feature, with convergence failure in a single parallel chain, in the $Agg1_i$ model (Table 7). On the other hand, the models with different growth and age composition parameters between the sexes presented the greatest difficulty of convergence. Although these models with distinct age compositions are apparently robust, with overlapping posteriors distributions between the parallel chains, they presented polymodal tendencies, which also reduced their DIC values (Table 7).

In short, the best-ranked model was $FM1_i$, that is, the model with distinct growth parameters between the sexes, but with the same age composition, and also with informative prior distributuions and von Bertalanffy parameterization proposed by Francis (1988). This result is also consistent with the sensitivity analyzes (Supplementary Material), since $FM1_i$ was the only model, among those with the lowest DICs, that presented a single stationary region (robustness). Moreover, this region, besides having low DIC values, was also present in the other three models.

Table 7

Estimates of the $FM1_i$ model are detailed in Table 8, and show that the growth patterns are equivalent between both sexes. Males showed an estimated asymptotic length of 47.26 mm, slightly lower than females (48.88 mm), which in turn had a lower growth coefficient (K) than males (Table 8). This model assumes that males and females have the same age composition, predominated by young specimens up to one year old. However, it should be noted that this age

composition underestimates the proportion of young individuals, since it was performed only with sexed individuals.

Table 8

4. Discussion

4.1. Sex ratio

The results showed a higher proportion of females for the population of *A. paranae* in streams of the upper Paraná river basin. This pattern was also observed by Abelha and Goulart (2008) and Barbieri (1992), who found a higher prevalence of females for other populations of *A. paranae* inhabiting a reservoir and low order rivers, respectively. However, studies related to other populations of the genus *Astyanax* presented different patterns, with a higher proportion of males for *A. gymnodontus* in an influence area of a reservoir (Baumgartner et al., 2016), and *for A. scabripinnis paranae* in small rivers order (Veregue & Orsi, 2003). For *A. henseli*, on the other hand, no differences were found in sex ratio (Dala-Corte & Azevedo, 2010), a pattern commonly observed for fish populations (Nikolsky, 1969).

The sexual ratio may undergo several changes due to environmental pressures, *e.g.* temperature (Conover, 1984), and behavioral differences during the life cycle of the species (Vazzoler, 1996; Wootton, 1999). This proportion may change among populations of the same species and among periods in the same population (Nikolski, 1963; Forsgren et al., 2004). The highest proportion of females found in the studied population may be a self-response to replace large losses by predation (Vazzoler, 1996), ensuring the maintenance of the population size or even its growth. Moreover, the results presented by Abelha and Goulart (2008) demonstrate a trend of higher proportion of females in the upper length classes, which may indicate possible sexual dimorphism in the size of this species. This hypothesis was tested in our study and no significant evidences were found to affirm that the population studied also shows this tendency. However, the aforementioned authors reported individuals of larger sizes, especially females, with a maximum standard length of 120 mm and 90 mm for females and males, respectively, while in the evaluated population the largest specimens measured, in this order, 67 mm and 63 mm of standard length.

4.2. First sexual maturation size

Based on the *A. paranae* population studied, the standard length of first maturation of females was estimated at 38.12 mm and 37.75 mm, using non-informative and informative prior distributions, respectively. Veregue and Orsi (2003) estimated that 50% of the females of *A. scabripinnis paranae* of the Tibagi river basin reach sexual maturity at 50.4 mm of total length. Assuming that the standard length for this species represents approximately 80% of the total length, the estimate obtained by the aforementioned authors would be equivalent to 40.3 mm of standard length. Even so, the estimates for the females of *A. paranae* from the Guiraí River subbasin, presented lower L_{50} values than those estimated for the population of the Tibagi river basin (Veregue & Orsi, 2003). This difference is possibly due to environmental factors. The streams sampled on the Tibagi River are 520 m and 470 m above sea level, while the highest altitude in the sites sampled in the present study was 368 m, with a mean of 311 m. Thus, it is possible that the low altitudes and higher temperatures of the streams that harbor the population studied in the present study influence the size of the fish to reach sexual maturity. There is also the possibility of this population being exposed to limiting environmental conditions, *e.g.* high predation and low feed availability, with high mortality, resulting in more rapid sexual maturation.

4.3. Weight-length relationship

The growth pattern, proportions between the increments in weight and length, differs among life stanzas of the studied population. In fact, growth patterns in nature may be more complex than

those commonly found in most studies (*e.g*. Hilborn et al., 2003), since high natural variability may obscure them, making it difficult to identify (Fontoura et al., 2010). This pattern of polyphasic growth has already been reported for other fish species, including for the Characidae *A. jacuhiensis* and *Cheirodon ibicuhiensis* (Fontoura et al., 2010), as well as for the Siluriformes *Auchenipterichthys longimanus* (Freitas et al., 2016) and *Corydoras paleatus* (Barradas et al., 2016). These studies associated polyphasic growth with reproductive investment, that is, the stanza change coinciding with the first maturation size. An exception occurred for *A. jacuhiensis*, in which this association was not conclusive, since the stanza change occurred in a length less than L_{50} (Fontoura et al., 2010).

In this study, stanza changes were estimated to occur at 57.4 mm and 62.3 mm of standard length for males and females, respectively, values considerably larger than the highest estimate of L_{50} obtained for females. Thus, it is not possible to attribute the change in the growth pattern of the studied population to the alterations in the energy investment from growth to reproduction.

Differences in growth pattern between the sexes were observed only in the first life stanza, when the males tended to negative allometric growth and females to positive allometric growth. These results indicate that males have lower weight gain rates than females. Similar pattern was observed for *A. paranae* individuals under optimal conditions in experimental aquarium, where females had significantly higher mean weights than males of similar mean standard lengths (Nascimento et al., 2017). This greater potential of female weight gain in relation to males in the first life stanza is relatively common in species without parental care (Lowe-McConnel, 1999), since a heavier body may mean more eggs and larger fertility (Morita et al., 1999). This pattern is further reinforced by the lower proportion of males in the population, which can increase sperm competition among them causing greater energy allocation in the production of sperm to ensure the fertilization of as many eggs as possible, reducing their investment in somatic growth (Parker, 1992). On the other hand, in the second life stanza, both sexes present the same allometric coefficient, severely positive, which means that there is a high increase in weight for each centimeter of standard length acquired. The results indicate that this second stanza would occur in individuals of the population with more than one year old, with very low growth rates in length.

4.4. Somatic growth

Age and somatic growth studies are one of the most important aspects of population dynamics, especially those related to the fish stocks assessment. The execution of these studies, however, is limited by a series of factors, such as the difficulty in acquiring fish age data and, consequently, statistical methods able to adequately infer the parameters of somatic growth and age composition based only on one sample length data. Based on the model presented by Alves et al. (submitted in November/2017) and the Bayesian extension proposed here, it was possible to estimate the parameters of the von Bertalanffy function, as well as the age composition, even if this analysis should be done in a mindful and parsimonious way.

The results indicated that the parameterization of the VBGF proposed by Francis (1988) was more robust than the original parameterization (von Bertalanffy, 1938), with more chances to present parallel chains converging at the same stationary region. In addition, the Bayesian approach used has the potential to improve the reliability of the somatic growth inferences made in the study, since it allows the inclusion of previous information from other studies.

Even though there is no other age and growth studies of *A. paranae* available to propitiate the establishment of more concise informative prior distributions, it was possible to use in this study some previous information from the paper of Nascimento et al. (2017), from aquaculture area, which contributed to the robustness of the results, at least when using the parameterization of Francis (1988). This relationship of robustness between VBGF parameterization and informative prior distributions may be a result of the information provided in Nascimento et al. (2017), which

represented the means lengths of *A. paranae* about 2 and 6 months old. Thus, it was possible to extract a direct information about the parameter L_2, mean length for the second age, defined in 6 months in the model of the present study. On the other hand, with the original parameterization, this previous information had to be extrapolated to perform the *a priori* inference for the parameters L_∞ e K.

The mean lengths of 6-months-old *A. paranae* found by Nascimento et al. (2017) were approximately 45.1 mm and 50.8 mm for males and females, respectively (diploids only). The final estimates of the present study, for this same age, were 37.51 mm and 37.41 mm for males and females, respectively. This difference may be due to the origin of the studied specimens studied, which were created under optimal conditions of experimental aquariums by Nascimento et al. (2017) and, in the present study, are wild specimens inhabiting streams, dynamic natural environments. Thus, reared specimens are expected to express their full growth potential, while those living in the natural environment are subject to the constraints imposed therein, such as the availability of food resources, energy expenditure to avoid predation, among others. These factors may also explain the absence of differences between the mean lengths of males and females, both 6 months old and 1 year old, according to the presented estimates.

Considering that environmental forces are limiting the expression of the somatic growth potential of this population, acting indiscriminately in both sexes, it is expected equivalent effective somatic growth rates to both, even with higher growth potential for females, as observed in Nascimento et al. (2017). However, according to the results presented here, the model that estimated growth parameters for both sexes was more parsimonious than the model that estimated the parameters for the clustered sexes. One explanation would be that although estimates of L_2 and L_3 are close between the sexes, L_∞ was slightly higher for females, and there is greater variability in the estimates of these three parameters for them. Thus, it is reasonable to assume that these two characteristics indicate the greatest growth potential of females, which was not expressed given the environmental limitations.

Abelha and Goulart (2008) found specimens remarkably larger than those observed in our study, inhabiting reservoirs. However, Ferreira et al. (2012a; 2012b) and Veregue and Orsi (2003) recorded specimens of the same length ranges found also in streams. Thus, it is possible that, in streams, which are highly dynamic environments with low predictability and diverse sources of disturbance, there is a selection force acting for individuals with strategies and sizes similar to those found in our study. Though, in reservoirs, the relative stability and the high resource availability characteristic of their marginal areas may prevent the existence of this environmental filter, allowing larger specimens to occur. Thus, it is suggested that, under optimal conditions, with availability of resources and absence of severe predation, individuals are able to express their potential growth before sexual maturation without impairing their reproduction. On the other hand, under limiting conditions, fish should prioritize the energy allocation for reproduction, including reducing the first maturation size, in order to ensure the maintenance of the population.

4.5. Final considerations

The Bayesian proposals of a biphasic model for weight-length relationship and the somatic growth model based only on length data, demonstrated great potential and applicability for the population studies. These analyzes, together, allowed the identification of complex biological patterns in the studied *A. paranae* population, which may be frequent in nature, but difficult to identify (Hilborn et al., 2003; Fontoura et al., 2010). In addition, Bayesian inferences about biological parameters are an advantageous option for population studies, although it was not the purpose of this study to compare it with other approaches. Among the main advantages of the Bayesian approach are: *i)* the ability to consider different sources of *a prior* information to the data to be modeled, which may assist the researchers in their conclusions, especially in situations where data is not very

informative, such as in somatic growth analyzes based only on length data, or in studies with small samples; *ii)* the possibility of making probabilistic statements about the parameters, as was done in this study, mainly in the allometric growth hypotheses evaluations ($\beta = 3$); and *iii)* the simplicity of estimating credible intervals of the parameters functions, a frequent situation in ecological studies. Besides that, the reparametrization of the logistic model as a function of the parameter L_{50}, is also an interesting alternative in both approaches, Bayesian and frequentist. This parameterization allows the Bayesians to include information directly related to this parameter of interest, and for the frequentist statisticians, the possibility of estimating confidence intervals based on the likelihood ratio, rather than asymptotic.

Finally, we emphasize the Bayesian approach as an important statistical framework to minimize the lack of knowledge of the great diversity of fish species and populations and their biology, which is a wide challenge for conservation biologists and ecologists. This approach provides an easy and intuitive way of thinking about biological and ecological problems (Spor et al., 2010) and supllies basic logical application for knowledge accumulation (Hobbs & Hilborn, 2006). Bayesian inference has been growing among ecologists (Monnahan et al., 2017) with important contributions, including for weight-length relationship (Froese et al., 2014). However, it should be emphasized that *a priori* information must be used with discretion, caution and parsimony.

References

Abelha, M. C. F., & Goulart, E. (2008). Population Structure, Condition Factor and Reproductive Period of *Astyanax paranae* (Eigenmann, 1914) (Osteichthyes: Characidae) in a Small and Old Brazilian Reservoir. *Brazilian Archives of Biology and Technology*, 51, 503-512.

Akçakaya, H. R., Burgman, M. A., & Ginzburg, L. R. (1999). *Applied population ecology: Principles and computer exercises using RAMAS EcoLab 2.0.* New York, Applied Biomathematics.

Alves, D. C.; Vasconcelos, L. P.; Agostinho, A. A. Age composition and growth: a likelihood and length-frequency based analysis. Submitted in November 02, 2017 to the scientific journal *Fisheries Research*.

Barbieri, G. (1992). Biologia de *Astyanax scabripinnis paranae* (Characiformes, Characidae) do ribeirão do Fazzari São Carlos. Estado de São Paulo. I. Estrutura populacional e crescimento. *Revista Brasileira de Biologia,* 52, 579-588.

Barradas, J. R. S., Lermen, I. S., Larré, G. G., Martins, T. P., & Fontoura, N. F. (2016). Polyphasic growth in fish: a case study with *Corydoras paleatus* (Siluriformes, Callichthyidae). *Iheringia, Série Zoologia*, 106: e2016017.

Bartoo, N. W., & Parker, K. R. (1983). Stochastic age-frequency estimation using the von Bertalanffy growth equation. *Fishery Bulletin*, 81, 91-96.

Baumgartner, M. T., Silva, P. R. L., & Baumgartner, G. (2016). Population structure and reproductive biology of *Astyanax gymnodontus* (Characiformes: Characidae), a poorly known endemic fish of the Iguaçu River basin, Brazil. *Revista de Biología Tropical*, 64, 69-77.

Bervian, G., Fontoura, N. F., & Haimovici, M. (2006). Statistical model of variable allometric growth: otolith growth in *Micropogonias furnieri* (Actinopterygii, Sciaenidae). *Journal of Fish Biology*, 68, 196-208.

Blackwell, B. G., Brown, M. L. & Willis, D. W. (2000). Relative weight (Wr) status and current use in fisheries assessment and management. *Reviews in Fisheries Science*, 8, 1–44.

Campana, S. E. (2001). Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *Journal of Fish Biology*, 59, 197-242. doi:10.1006/jfbi.2001.1668

Chang, S., & Maunder, M. N. (2012). Aging material matters in the estimation of von Bertalanffy growth parameters for dolphinfish (*Coryphaena hippurus*). *Fisheries Research*, 119-120,147-153. doi: 10.1016/j.fishres.2012.01.008

Conover, D. O. (1984). Adaptive significance of temperature-dependent sex determination in a fish. *The American Naturalist*, 123, 297-313.

Dala-Corte, R. B., & Azevedo, M. A. (2010). Biologia reprodutiva de *Astyanax henseli* (Teleostei, Characidae) do curso superior do rio dos Sinos, RS, Brasil. *Iheringia, Série Zoologia*, 100, 259-266.

Ellison, A. (1996). An introduction to Bayesian inference for ecological research and environmental decision-making. *Ecological Applications*, 6, 1036-1046.

Ellison, A. (2004). Bayesian inference in ecology. *Ecology Letters*, 7, 509-520.

Eschmeyer, W. N., & Fong, J. D. (2017, April 17). Species by family/subfamily [online]. California: Institute for Biodiversity Science and Sustainability. Retrieved from http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp

Ferreira, A., Gerhard, P., & Cyrino, J. E. P. (2012a). Diet of *Astyanax paranae* (Characidae) in streams with different riparian land covers in the Passa-Cinco River basin, southeastern Brazil. *Iheringia, Série Zoologia*, 102, 80-87.

Ferreira, A., Paula, F. R., Ferraz, S. F. B., Gerhard, P., Kashiwaqui, E. A. L., Cyrino, J. E. P., & Martinelli, L. A. (2012b). Riparian coverage affects diets of characids in neotropical streams. *Ecology of Freshwater Fish*, 21, 12–22.

Fontoura, N. F., Jesus, A. S., Larre, G. G., & Porto, J. R. (2010). Can weight/length relationship predict size at first maturity? A case study with two species of Characidae. *Neotropical Ichthyology*, 8, 835-840.

Forsgren, E., Amundsen, T., Borg, A. A., & Bjelvenmark, J. (2004). Unusually dynamic sex roles in a fish. *Nature*, 429, 551-554.

Francis, R. I. C. C. (1988). Are growth parameters estimated from tagging and age-length data comparable? *Canadian Journal of Fisheries and Aquatic Sciences*, 45, 936–942.

Freitas, T. M. S., Almeida, V. H. C., Montag, L. F. A., & Fontoura, N. F. (2016). Predicting size at first sexual maturity from length/weight relationship: a case study with an Amazonian catfish. *Neotropical Ichthyology*, 14, e150152.

Froese, R. (2006). Cube law, condition factor and weight-length relationships: History, meta-analysis and recommendations. *Journal of Applied Ichthyology*, 22, 241–253.

Froese, R., Thorson, J. T., & Reyes Jr, R. B. (2014). A Bayesian approach for estimating length-weight relationships in fishes. *Journal of Applied Ichthyology*, 30, 78-85.

Heidelberger, P., & Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research,* 31, 1109-1144.

Hilborn, R., Quinn, T. P., Schindler, D. E., & Rogers, D. E. (2003). Biocomplexity and fisheries sustainability. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 6564-6568.

Hobbs, N. T., & Hilborn, R. (2006). Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. *Ecological Applications*, 16, 5–19.

Huxley, J. S. (1924). Constant differenctial growth-ratios and their significance. *Nature*, 114, 895-896.

Langeani, F., Castro, R. M. C., Oyakawa, O, T., Shibatta, O. A., Pavanelli, C. S., & Casatti, L. (2007). Diversidade da ictiofauna do Alto Rio Paraná: composição atual e perspectivas futuras. *Biota Neotropica*, 7, 181-197.

Lowe-Mcconnell, R. H. (1999). *Estudos Ecológicos de Comunidades de Peixes Tropicais*. São Paulo, EDUSP.

McManamay, R. A., & Utz, R. M. (2014). Open-Access Databases as Unprecedented Resources and Drivers of Cultural Change in Fisheries Science. *Fisheries*, 39, 417-425.

Monnahan, C. C., Thorson, J. T., & Branch, T. A. (2017). Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. *Methods in Ecology and Evolution*, 8, 339–348.

Morita, K., Yamamoto, S., Takashima, Y., Matsuishi, T., Kanno, Y., & Nishimura, K. (1999). Effect on maternal growth on egg and size in wild white-spotted char (*Salvelinus luecomaenis*). *Canadian Journal of Fisheries and Aquatic Siences*, 56, 1585-1589.

Nakatani, K., Agostinho, A. A., Baumgartner, G., Bialetzki, A., Sanches, P. V., Makrakis, M. C., & Pavanelli, C. S. (2001). *Ovos e larvas de peixes de água doce: desenvolvimento e manual de identificação*. Maringá: Eduem, 378p.

Nascimento, N. F., Pereira-Santos, M. Piva, L. H., Manzini, B., Fujimoto, T., Senhorini, J. A., Yasui, G. S., & Nakaghi, L. S. O. (2017). Growth, fatty acid composition, and reproductive parameters of diploid and triploid yellowtail tetra *Astyanax altiparanae*. *Aquaculture*, 471, 163-171.

Nikolsky, G. V. (1963). *The Ecology of Fishes*. London: Academy Press.

Nikolsky, G. V. (1969). *Theory of fish population dynamics*. Edinburgh, Oliver e Boyd.

Parker, G. A. (1992). The evolution of sexual size dimorphism in fish. *Journal of Fish Biology*, 41, 1-20.

Parrack, M. L., & Cummings, N. J. (2003). Errors in transforming length samples to age frequencies without age samples. *Fisheries Research*, 63, 235-243. doi: 10.1016/S0165-7836(03)00040-7

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6, 7-11.

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rockwood, L. L. (2006). *Introduction to population ecology*. Blackwell publishing.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Linde, A. van der. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, 64, 583-639.

Spor, A., Dillmann, C., Wang, S., de Vienne, D., Sicard, D., & Parent, E. (2010). Hierarchical Bayesian modelling for *Saccharomyces cerevisiae* population dynamics. *International Journal of Food Microbiology*, 142, 25-35.

Su, Y., & Yajima, M. (2015). *R2jags: Using R to Run 'JAGS'*. R package version 0.5-7. https://CRAN.R-project.org/package=R2jags

Súarez, Y. R., Silva, E. A., & Viana, L. F. (2017). Reproductive biology of *Astyanax lacustris* (Characiformes: Characidae) in the southern Pantanal floodplain, upper Paraguay River basin, Brazil. *Environmental Biology of Fishes*, 100, 775-783.

Vazzoler, A. E. A. M. (1996). *Biologia da reprodução de peixes teleósteos: Teoria e Prática*. Maringá, Eduem.

Veregue, A. M. L., & Orsi, M. L. (2003). Biologia reprodutiva de *Astyanax scabripinnis paranae* (Eigenmann) (Osteichthyes, Characidae), do ribeirão das marrecas, bacia do rio Tibagi, Paraná. *Revista Brasileira de Zoologia*, 20, 97-105.

von Bertalanffy, L. (1938). A quantitative theory of organic growth (inquiries on growth laws. II). *Human Biology*, 10, 181-213.

Wootton, R. J. (1999). *Ecology of Teleost Fishes*. 2ed. Chapman & Hall, London.

744 **Table 1**. Different hypotheses related to sexual dimorphism and distinct growth patterns.

Sexual Dimorphism	Growth stanzas	Model	Hypotheses
Absent	Unique	*Agg*	Absence of sexual dimorphism and single growth stanza.
		FM	Tendency of sexual dimorphism and single growth stanza.
Present	Two	Agg_2	Absence of sexual dimorphism and two growth stanzas.
		F_1M_2	Tendency of sexual dimorphism and two growth stanzas only for males.
		F_2M_1	Tendency of sexual dimorphism and two growth stanzas only for females.
		$A_1F_2M_2$	Tendency of sexual dimorphism only in the second growth stanza.
		$F_1M_1A_2$	Tendency of sexual dimorphism only in the first growth stanza.
		F_2M_2	Tendency of sexual dimorphism in both growth stanzas.

745

746

Table 2. Prior distributions for the parameters of the VBGF somatic growth. The means of the parameters L_2, L_∞ and K were extrapolated from the study of Nascimento et al. (2017).

Parameterization	***Priori***	**Sex**	**Parameters/ Distributions**	
			L_2	L_3
Francis (1988)	Flat	All	N(45,0; 10^{-3})I(3,500)	N(60,0; 10^{-3})I(L_2,500)
	Informative	Clustered	N(48,0; 9)I(3,200)	N(60,0; 10^{-3})I(L_2,200)
		Female	N(50,8; 9)I(3,200)	N(60,0; 10^{-3})I(L_2,200)
		Male	N(45,0; 9)I(3, 200)	N(60,0; 10^{-3})I(L_2,200)
			L_∞	K
VBGF Original	Flat	All	N(60,0; 10^{-3})I(0,500)	N(1,0; 10^{-3})I(0, 50)
	Informative	Clustered	N(66,5; 5)I(0,200)	N(2,8; 0,8)I(0, 50)
		Female	N(74,1; 5)I(0,200)	N(2,6; 0,8)I(0, 50)
		Male	N(59,1; 5)I(0,200)	N(3,1; 0,8)I(0, 50)

751 **Table 3**. DIC values of the fitted models considering the sex ratio of *A. paranae*.

Models/Hypotheses	**DIC**	Δ_{DIC}
Null	238.65	-
Weight (simple effect)	239.36	0.71
Length (simple effect)	240.68	2.03
Weight (quadratic effect)	241.40	2.75
Length (quadratic effect)	241.88	3.23

752

753

754 **Table 4**. Estimates of the intercept of the null model and functions of this parameter, such as the
755 odds-ratio of females (*OR*) and proportion of females (θ) at the *A. paranae* study population.

Parameters	Mean	SD	Median	CI 95%
β_0	0.328	0.154	0.325	(0.024; 0.638)
$OR = e^{\beta 0}$	1.404	0.219	1.384	(1.024; 1.893)
$\theta = (1 + e^{-\beta 0})^{-1}$	0.581	0.037	0.580	(0.506; 0.654)

SD: standard-deviation*;* CI 95%: credible interval of 95%.

756

757

Table 5. Parameters estimates of the three logistic models fitted to estimate the size of first sexual maturation L_{50} of *A. paranae*.

Prioris	**Parameters**	**Mean**	**SD**	**Median**	**CI 95%**
Non-informative	β_0	-5.613	1.184	-5.58	(-8.075; -3.315)
	β_1	0.147	0.031	0.146	(0.088; 0.194)
	L_{50}	38.119	1.772	38.099	(34.772; 42.764)
Slightly informative	β_0	-5.385	1.144	-5.326	(-7.844; -3.298)
	β_1	0.141	0.03	0.139	(0.085; 0.205)
	L_{50}	38.219	1.743	38.167	(34.838; 41.754)
Informative	β_0	-5.434	1.181	-5.356	(-7.860; -3.265)
	β_1	0.144	0.031	0.142	(0.087; 0.208)
	L_{50}	37.75	1.01	37.769	(35.730; 39.642)

SD: standard-deviation; CI 95%: credible interval of 95%.

Table 6. DIC values of the fitted models considered in Table 1.

Models/Hypotheses	DIC	Δ_{DIC}
Agg	-75.64	69.20
FM	-99.91	44.92
Agg_2	-113.89	30.95
F_1M_2	-124.01	20.82
F_2M_1	-116.38	28.46
$A_1F_2M_2$	-123.68	21.16
$F_1M_1A_2$	-144.83	-
F_2M_2	-143.89	0.95

Table 7. Results of metrics and scores of each feature evaluated by model.

Feature	Models											
	Agg1$_v$	***Agg2$_v$***	***FM1$_v$***	***FM2$_v$***	***FMA1$_v$***	***FMA2$_v$***	***Agg1$_i$***	***Agg2$_i$***	***FM1$_i$***	***FM2$_i$***	***FMA1$_i$***	***FMA2$_i$***
Convergence												
Number of chains that converged	8	8	6	7	1	1	7	8	8	6	6	2
Score	1.00	1.00	0.75	0.88	0.13	0.13	0.88	1.00	1.00	0.75	0.75	0.25
Robustness												
Number of "stationary regions"	2	3	3	2	1	2	1	2	1	2	1	1
Score	0.50	0.33	0.33	0.50	1.00	0.50	1.00	0.50	1.00	0.50	1.00	1.00
Goodness of fit /Parcimony												
Lower DIC among the chains that converged	247.7	202.6	223.6	205.0	213.0	213.7	251.8	206.1	207.7	231.7	211.8	227.5
Score	0.10	1.00	0.10	0.90	0.10	0.10	0.10	0.90	0.50	0.10	0.50	0.10
Conclusion												
Final score	0.53	0.78	0.39	0.76	0.41	0.24	0.66	0.80	0.83	0.45	0.75	0.45
Ranking	7	3	11	4	10	12	6	2	1	8	5	8

768 **Table 8**. Estimates of the model $FM1_i$ parameters of somatic growth of VBGF, of both
769 parameterizations (original and Francis, 1988), for the *A. paranae* population studied.

Parameters			Mean	SD	Median	CI 95%
Somatic growth	Male	L_∞	47.26	0.96	45.44	(47.24; 49.25)
		K	3.03	0.15	2.74	(3.03; 3.34)
		L_3	45.10	0.68	43.76	(45.11; 46.45)
		L_2	37.51	0.48	36.57	(37.51; 38.48)
		τ	0.79	0.21	0.43	(0.77; 1.26)
	Female	L_∞	48.88	2.61	44.13	(48.91; 53.82)
		K	2.80	0.25	2.35	(2.79; 3.30)
		L_3	45.97	1.83	42.48	(46.07; 49.27)
		L_2	37.41	0.90	35.64	(37.44; 39.11)
		τ	0.26	0.07	0.14	(0.26; 0.42)
Age composition	Clustered	a_3	0.34	0.03	0.29	(0.34; 0.39)
		a_6	0.40	0.03	0.35	(0.40; 0.45)
		a_{12}	0.06	0.05	0.00	(0.05; 0.18)
		a_{18}	0.09	0.06	0.00	(0.08; 0.22)
		a_{24}	0.11	0.06	0.01	(0.10; 0.23)

SD: standard-deviation; CI 95%: credible interval of 95%.

770

771

Figures:

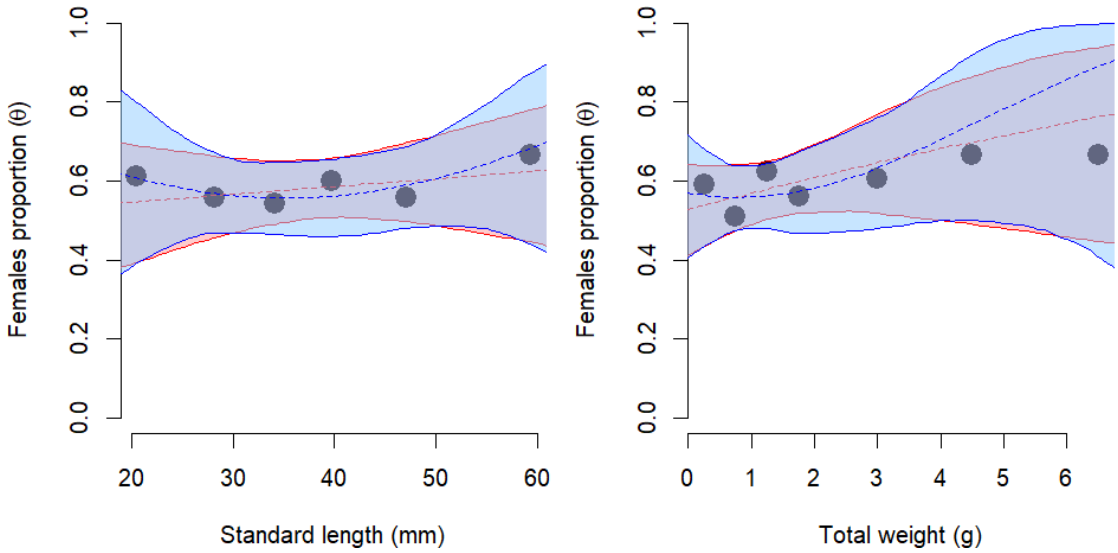

Figure 1. Proportion of *A. paranae* females by standard length (left) and total weight (right). The black dots represent the observed proportions, while the dashed areas and lines represent the 95% credible interval and the prediction median, respectively, with the quadratic (blue) and simple (red) effect models.

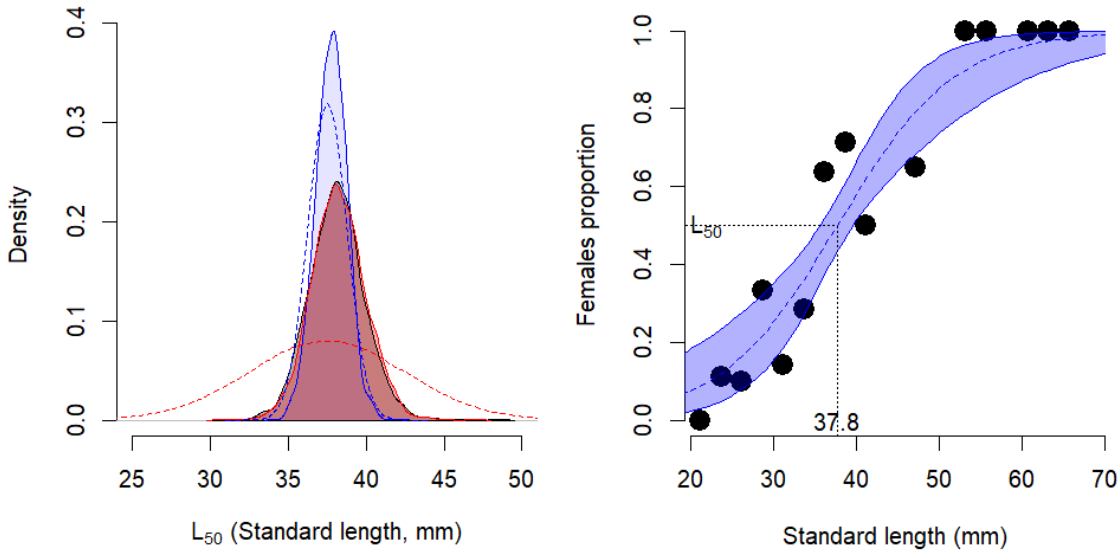

Figure 2. Prior distributions (dashed lines) and posteriors distributions (solid lines) for the first sexual maturation size L_{50} of *A. paranae* females for the non-informative (grey), slightly informative (red) and informative (blue) prior distributions (left). Proportions of females observed as a function of the standard length and the respective 95% credible band (blue) of the predictions of the model with informative prior distribution (right).

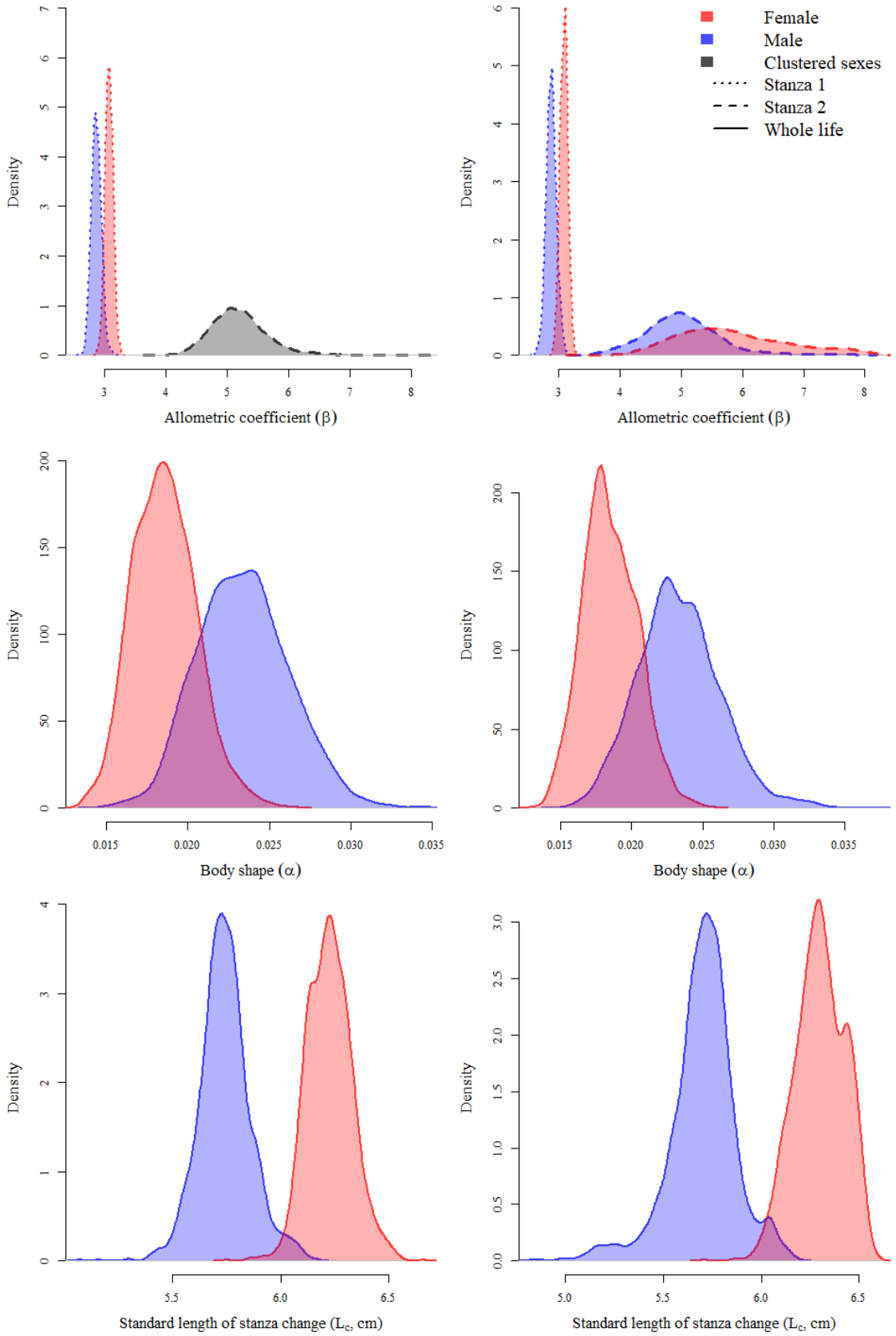

Figure 3. Posteriors distributions of allometric coefficient (upper), body shape (intermediate) and standard length at which the stanza change (lower) occurs for the $F_1M_1A_2$ (left) and F_2M_2 (right) models.

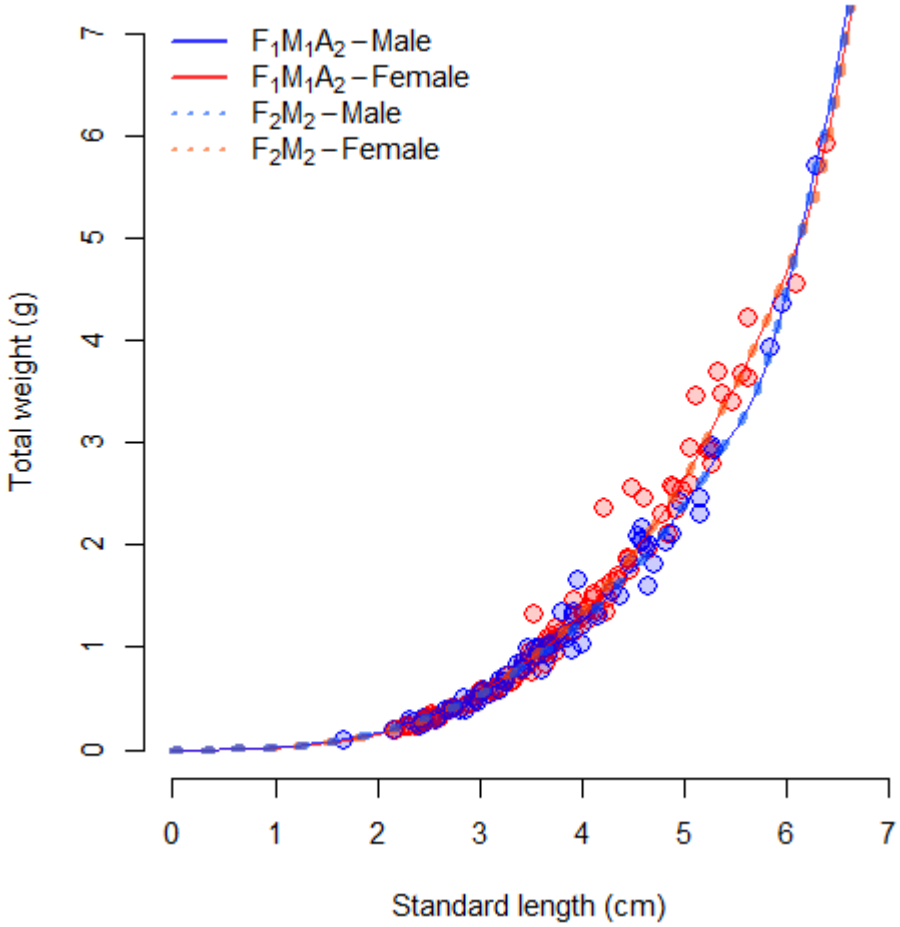

Figure 4. Scatter plot for total weight as a function of the standard lengths observed for *A. paranae* and estimates of the predictions of each model (lines).

Supplementary material

Sensitivity analyzes of somatic growth models

Initially, in a superficial analysis, without considering the convergences of the chains, the three most parsimonious models were FM2v, FM1i and Agg2v, in this order (Table A1). However, the convergence and sensitivity analyzes show that within each model, the proportion of chains that converged was variable, as well as indicating convergences at different stationary regions, observed by the high variation in DIC values calculated between the parallel chains (Figure A1).

Table A1. Results of the models selection, using the deviance information criterion (DIC), fitted in order to estimate the somatic growth and age compostion of *A. paranae*. The models are distinct as to the prior distributions used for growth parameters (vague or informative), parameterization of von Bertalanffy, and parameterization of the model related to sexual structures in growth parameters (1: sexes clustered, 2: structured by sex) and (1: age composition, 2: Different age compositions between the sexes).

Priori	**Parameterization**	**Structures**		**Model**	**DIC**	**Δ_{DIC}**
		Sexual	***Age***			
Vague	Francis (1988)	1	1	Agg1v	262.48	56.35
	Original VBGF	1	1	Agg2v	209.41	**3.27***
	Francis (1988)	2	1	FM1v	223.28	17.14
	Original VBGF	2	1	FM2v	206.14	**0.00***
	Francis (1988)	2	2	FMA1v	213.80	7.66
	Original VBGF	2	2	FMA2v	211.32	5.19
Informative	Francis (1988)	1	1	Agg1i	252.19	46.06
	Original VBGF	1	1	Agg2i	221.86	15.72
	Francis (1988)	2	1	FM1i	208.11	**1.97***
	Original VBGF	2	1	FM2i	238.32	32.18
	Francis (1988)	2	2	FMA1i	212.39	6.25
	Original VBGF	2	2	FMA2i	229.57	23.43

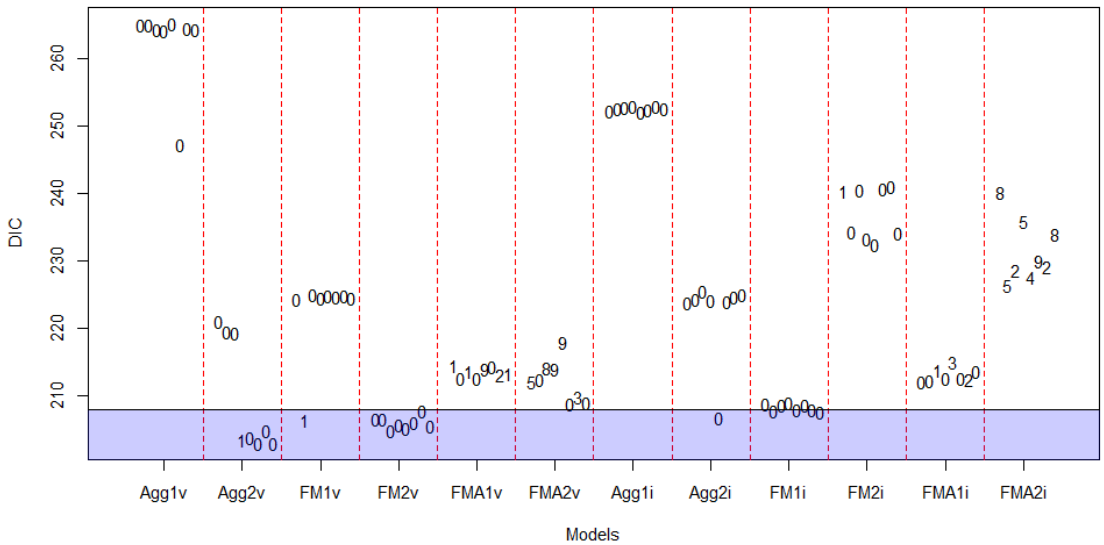

Figure A1. Deviance information criterion (DIC) estimated for each fitted chain of each model. The numbers plotted represent the number of parameters whose chains did not converge.

More detailed analyzes were performed for the Agg2v, FM2v, Agg2i and FM1i models, which had parallel chains that converged and with $\Delta_{DIC} \lesssim 5$ among all the parallel chains of all models. Among the four models, only the FM1i model presented the convergence of all the parallel chains at the same points (Figures A2 and A3). The other three models presented parallel chains that stagnated at two or three distinct regions for the parameters of asymptotic length (L∞), growth coefficient (K) and lengths at ages 0.5 (L_2) and 1.0 years (L_3) (Figures A2 and A3).

In general, among these four models, it was possible to observe three different stationarity regions. The first region was present only in Agg2v and FM2v models, both estimated based on vague prior distributions. This region presented high values and dispersion for L_∞, with median values of 137 mm for the clustered sexes of the posteriors distributions of Agg2v model and 140 mm and 120 mm for females and males, respectively, of the posteriors distributions obtained by FM2v model (Figure A2). This region also presented low values for both the growth coefficient (K), with medians of 0.19. 0.18 and 0.23 year^{-1} for clustered sexes, females and males, respectively, as for the standard length predicted for one year old (L_3), with medians of 26mm, 26mm and 27mm for clustered sexes, females and males, respectively (Figures A2 and A3).

The second "stationary region" was present in the four models mentioned above. In an opposite way, this region presented low values of L_∞, with medians of 48mm, 47mm and 49mm clustered sexes, females and males, respectively, associated with high values of K, with respective medians of 2.9. 3.0 and 2.8 year^{-1} (Figure A2). For the parameter L_3 the medians were estimated in 45 mm for clustered sexes and also for males and 46 mm for females (Figure A3). The DIC values for the parallel chains that stagnated in these two regions were equivalent, all smaller than 210. Unlike the third "stationary region", which presented DIC values around 230 (Figure A1). The third region only occurred in the models that considered the clustered sexes (Agg2v and Agg2i) and presented intermediate values for the growth parameters analyzed (Figures A2 and A3).

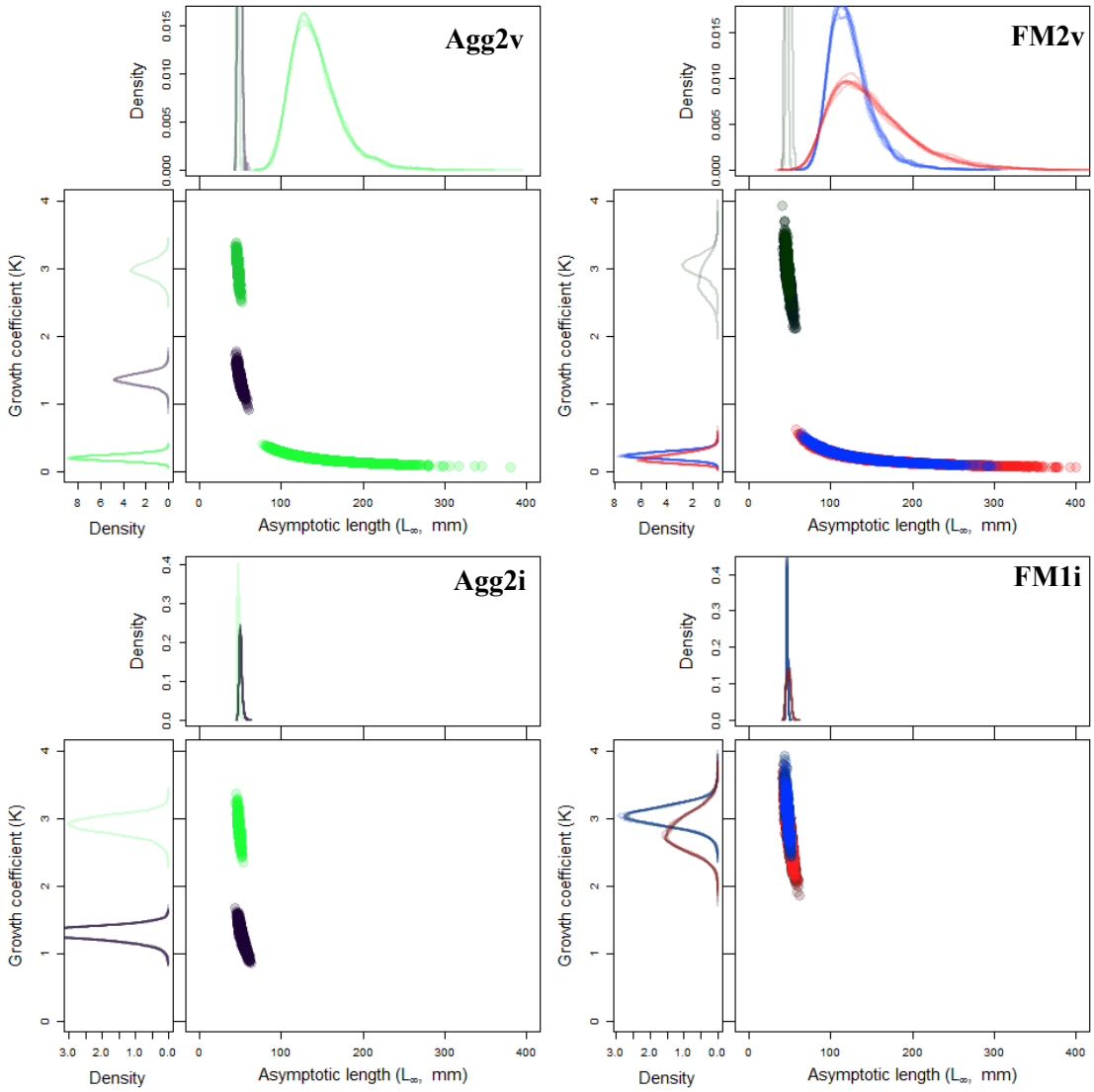

Figure A2. Estimates of von Bertalanffy growth parameters in the original parameterization, growth coefficient (K) and asymptotic length (L_∞) by parallel chain and model. Each parallel chain within a same model was colored differently according to the DIC value of the chain. The lower the DIC value of the parallel chain lighter is the color tone of the chain, where green represent the lower DIC models for clustered sexes and blue and red for males and females, respectively. The color gradient is comparable only within each model, from the lower (tends to green, blue or red) to the larger (tends to black) DIC value.

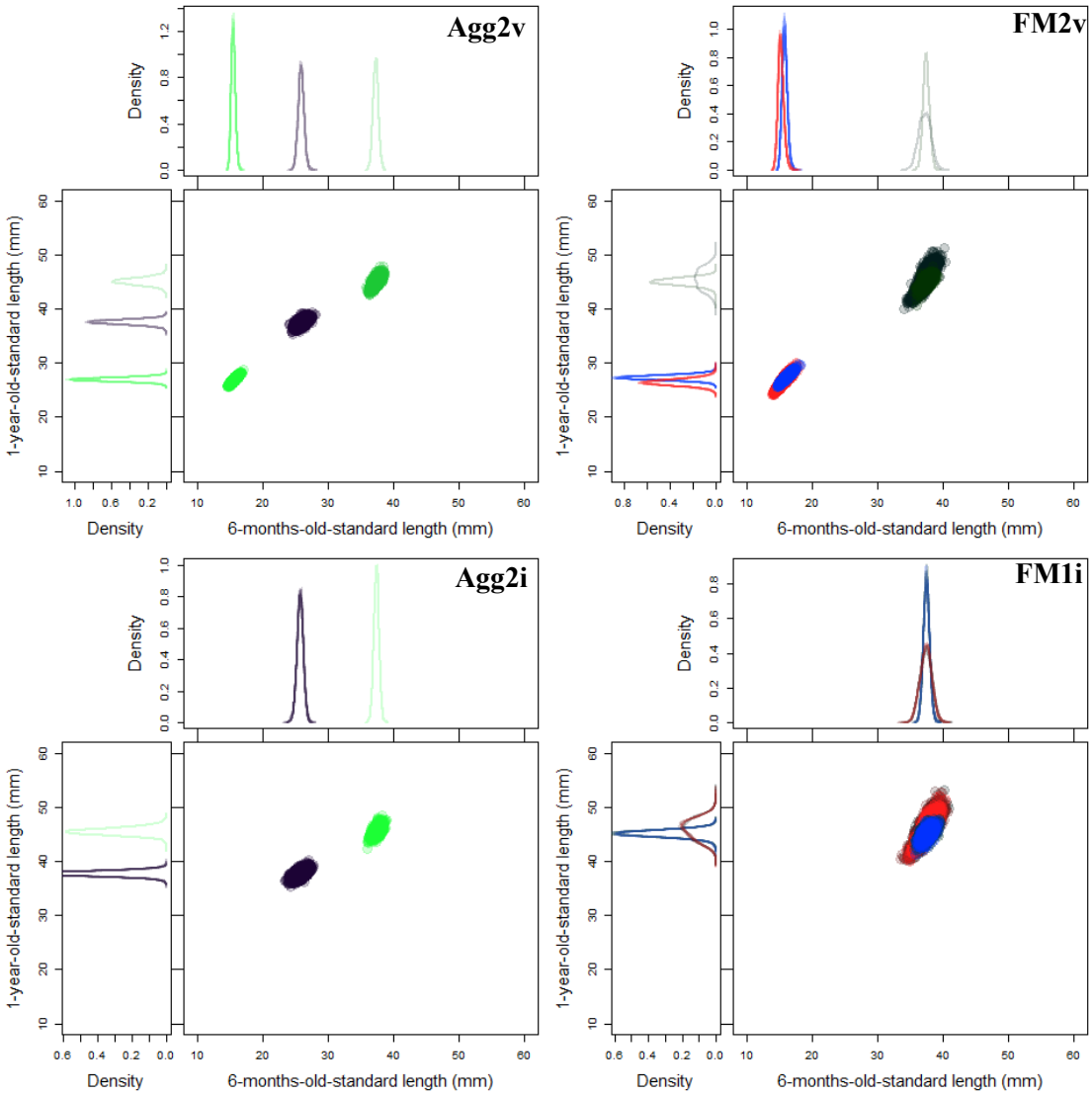

Figure A3. Estimates of von Bertalanffy growth parameters in the parameterization proposed by Francis (1988), mean standard lengths predicted for the ages 0.5 (L_2) and 1.0 years old (L_3) by parallel chain and model. Each parallel chain within a same model was colored differently according to the DIC value of the chain. The lower the DIC value of the parallel chain lighter is the color tone of the chain, where green represent the lower DIC models for clustered sexes and blue and red for males and females, respectively. The color gradient is comparable only within each model, from the lower (tends to green, blue or red) to the larger (tends to black) DIC value.