Weight-Length Relationship of the African Snakehead, *Parachanna obscura* (Gunther 1861) by Bayesian Inference

[1]Oluwale, F. V. and [2]Olorunleke, J. O.
[1]*Department of Zoology, University of Ibadan, Ibadan, Oyo State, Nigeria.*
[2]*Fisheries Technology Department, Federal College of Freshwater Fisheries Technology, New Bussa, Niger State, Nigeria*
***Corresponding author** - Femi V. Oluwale, *wale_femi@yahoo.com*

Abstract
The Bayesian paradigm is increasingly being used to overcome the limitations of frequentist approaches in statistical inference. We specified a weakly informative prior to compute the coefficient of the linear models using the Bayesian statistical framework for the African obtuse snakehead, *Parachanna obscura*. Fish samples were collected monthly for six months from artisanal fishermen from Eleiyele Reservoir and measured for Total Length (L) to the nearest cm and weighed (W, wet weight) to the nearest g. A fitted Bayesian linear model (estimated using MCMC sampling with four chains of 2000 iterations and a warmup of 1000) indicates positive allometric growth (median b = 3.20, 95% Credible Interval (CI) [2.93, 3.49]) with substantial explanatory power (R^2 = 0.91, 95% CI [0.87, 0.94] based on the posterior predictive model. The *Scale Reduction Factor* was less than 1, and the effective sample size was larger than 1000 for all parameters: reliable evidence of convergence and typically a sufficiently large effective sample size. The African obstuse snakehead, *Parachanna obscura* thrived in Eleiyele Reservoir because they were plump. Bayesian analysis could be a reliable alternative to the classical ordinary least square methods for linear regression models in fisheries.

Keywords: Allometric model, Model estimation, *Parachanna obscura*, Eleiyele Reservoir

Introduction

The ubiquity and simplicity of scaling relationships and power laws in ecological systems might be deceptive when compared to the complexity of the systems that they attempt to describe. For instance, scaling relationships (where body size is the independent variable) in animals, which represent the most complex biological networks in nature, can be described by simple relationships [1,2]. These relationships are of the form

$$Y = \beta x^{\alpha} \qquad 1$$

where *Y* represents some response or dependent variable, and *X* represents an independent or explanatory variable, β is a normalization constant and α is the scaling exponent. The functional form of the relationship in Equation 1 is also called a power-law relationship, where some quantity can be expressed as the power of another. Power laws are well-known to biologists in the form of bivariate relationships of power-law type, called scaling relationships [2, 3, 4], by which molecular, physiological, ecological, and life history attributes relate to some attribute of organisms raised to power as in Equation 1. Furthermore, in the field of fishery science, the body weight (W) and length (L) relationships of fish species are of the power type [5]:

$$W = aL^b \qquad 2$$

where the parameters "a" and "b" are estimated by simple linear regression of logarithmic transformed variables..

$$Log\ (W) = Log\ (a) + bLog(L) \qquad 3$$

Frequentist statistics, adopted from formal academic backgrounds, constitute the most common inferential statistical framework for estimating regression parameters in fisheries. However, critical shortcomings in this common statistical framework with emphasis on p-values (the workhorse of classical inference) necessitate an alternative inferential paradigm [6,7]. Identified issues that are prevalent in frequentist statistical approaches peculiar to biological science and authoritative sources to explore these issues and solutions are presented here [8, 9, 10]. Bayesian inferences is one of the solutions to the issues highlighted Almost invariably, Bayesian inference

based on conditional probability (Bayes' theorem) is a forerunner among the options [6, 7, 11]. While frequentist statistics use tests of significance that suppose that the null hypothesis is true (i.e., the statistical test determines the probability of the data given the null hypothesis), Bayesian statistics determine the probability of a model given the data. This inference process involves fitting a probability model to the data [11, 12]. The use of Bayesian inference with ecological data became well known by Hilborn and Mangel [13], who demonstrated that null hypothesis testing was not as effective as Bayesian methods for insightful inferences from field studies (ecological reality). Despite its advantages. Bayesian inference is scarcely adopted by scientists in developing countries, especially in Nigeria. There is a dearth of scholarly documentation on the pure and applied sciences based on the Bayesian protocols among Nigerian workers. As a tentative approach to bridging this knowledge gap, we specify and compute the allometric growth coefficient for African obscure snakehead, *Parachanna obscura* inhabiting Eleiyele Reservoir by linear regression in the Bayesian framework. This could provide useful information for estimating the standing stock biomass of the species in the reservoir.

Materials and Methods

Study Area and Sample Collection

Eleiyele reservoir (Figure 1) is an artificial lake created primarily for domestic water supply by damming River Ona in 1939. Fisheries activities in the reservoir are controlled by the Oyo State Ministry of Agriculture and Natural Resources Fisheries Division. Artisanal fisheries dominate the reservoir, mostly by using gills and cast nests of various mesh sizes, as well as traps. Random samples of *Parachnna obscura* (Figure 2) were purchased monthly from fishermen at Eleiyele Reservoir between January and May 2018, and transported in an ice chest filled with ice to reduce post-humous digestion, to the Zoology Department, where they were preserved by deep freezing prior to data collection at the Hydrobiology and Fisheries Research Laboratory.

Fish Identification and Biometric Data Collection

The preserved fish were thawed, mopped dry with tissue paper, and identified according to standard guide [14, 15]. Fish were measured for Total Length (L) to the nearest centimeter and weighed (W, wet weight) to the nearest gram on a fish measuring board and Explorer Pro digital weighing balance.

Simple Linear Regression Model in Bayesian Framework

Bayesian statistics uses conditional probability based on Bayes theorem to evaluate the uncertainty in the value of a response variable y (fish body weight in g) conditional on an unknown parameter (s) θ and a priori uncertainty about the parameter(s) θ. Besides the response variable y, in simple regression models, there is also a single explanatory variable x (in fact "simple" appended to linear regression connotes a single predictor). Bayesian inference aims to update our beliefs about the parameters θ (e.g., the coefficients in a regression model) using both the model (Equation 3) and the data (x = Weight and y = Length, biometric measurements on *P. obscura*) based on the Bayes' theorem.

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \qquad 4$$

Bayes' theorem returns the posterior probability distribution, p (θ|y), of the parameter(s) θ given that data y equals the product of a likelihood function p (θ|y) and prior distribution p (θ) divided by the marginal distribution of the data p(y).. The denominator p(y) is the marginal likelihood: which is the likelihood averaged over the model parameters, weighted by their prior probabilities. This is often referred to as the normalizing constant [12].

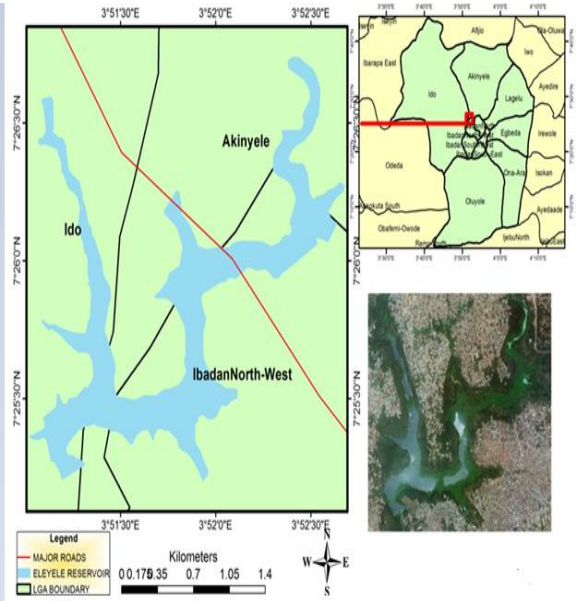

Figure 1: The Study area: Eleieyele Reservoir Insert: Map of Ibadan showing the study area, and satellite image of the study area.

Figure 2: *Parachanna obscura*.

Corollary

By incorporating the explanatory length variable (repressors) x, Bayes' theorem can be lucidly expressed for simple linear regression:

$$p(\theta|y,x) \propto p(y|\theta,x)p(\theta|x) \qquad 5$$

This is identical to Equation 4 except that everything is now conditional on x.

Because all parameters in a Bayesian model require prior distributions describing the researcher's uncertainty in the parameters before observing the data, we assigned a normally distributed weakly informative prior, equal prior weight to all possible values of a parameter [12]: location = 2.1, scale = 2.5, intercept, location = 0, scale = 2.5, coefficient.

Statistical Analysis

Data analysis was performed using R [16]. Frequency histograms were used to visualize the statistical distribution of species' biometric variables, which were also described using appropriate measures of locations and dispersion. Bayesian linear regression model was fitted via Markov chain Monte Carlo (MCMC) simulations instead of least squares estimation using stan_glm function in rstanarm [17] an interface for Stan [18] in the R environment.

Result

Biometrics Characters of *Parachanna obscura*

The sizes of the species (n =53) ranged from 37.5 – 15.5 cm (25.23 ± 5.40 cm), while body weights varied between 475.8 – 28.29 g (166.12 ± 110.64 g). The biometric variables of *P. obscura* exhibited a skewed distribution. These variables approximate normality after logarithmic transformation. The scatter plot shows a curvi-linear relationship between variables but a linear relationship after natural logarithmic transformation (Figure 3). Variations in body weight exceeded the variation in body length based on the coefficient of variation (Table 1).

Bayesian Linear Regression Model

After assessing convergence and sampling quality via diagnostics of ^R, ESS, and MCSE, and checking the model fit with the posterior predictive model, our model's explanatory power was substantial (R^2 = 0.91, 95% CI [0.87, 0.94]. The intercept of the model, corresponding to logL = 0, is -2.34 (95% CI [-2.74, -1.95]). Within this model, the effect of log L (median = 3.20, 95% CI [2.93, 3.49]) had a 100.00% probability of being. The alloemtric growth coefficient (b = 3.20) indicated that *P. obscura* exhibit positive allometric growth in the Eleiyele Reservoir.

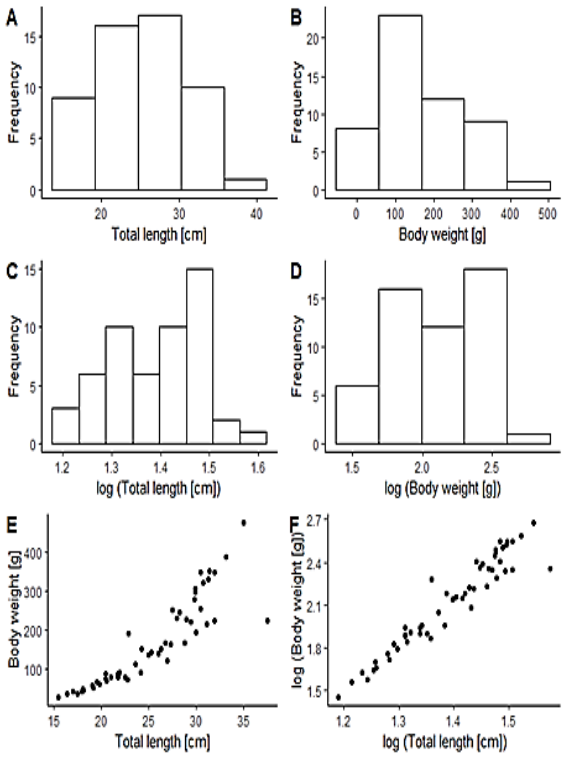

Figure 3: Biometric characteristics of the African obstuse snakehead, *Parachanna obscura* from Eleiyele Reservoir. A and B: Skewed frequency histograms of the total body length and body weight. C and D = Frequency histograms of the natural log-transformed total length and body weight of *P. obscura* from Eleiyele reservoir. E = Curvi-linear weight – length relationship of *P. obscura* from the Eleiyele Reservoir and F = linear weight – length relationship of *P. obscura* from Eleiyele reservoir.

Table 1: Descriptive of Characters of *Parachanna obscura* in Epe Lagoon

Characteristics	*Parachanna obscura*	
	Total length (cm)	**Body weight (g)**
N	53	53
Mean ± SD	25.23 ± 5.40	166.12 ± 110.64
CI (Upper - Lower)	23.74 – 26.72	135.62 – 196.61
CV	21.4%	66.6%
Median	25.3	142.45
Skewness	0.07	0.76
Kurtosis	-0.95	-0.24
R (Max - Min)	22 (37.5 – 15.5)	447.51 (475.8 – 28.29)

Table 2: Summary Statistics for Posterior Distribution and Diagnosing MCMC Problems

Parameter	Mean	95% CI	PD	%ROPE	Rhat	ESS
Intercept	-2.34	[-2.73, -1.96]	100%	0%	1	3,975
Slope	3.2	[2.93, 3.49]	100%	0%	1	3,976

Note. CI = Credible Interval, PD = Probability Direction, Rhat = Scale Reduction Factor, ESS = Effective Sample Size, and ROPE = Region of Practical Equivalence positive (> 0), 100.00% of being significant (> 0.02), and 100.00% of being large (> 0.10). The estimation successfully converged (Rhat = 1.000) and the indices are reliable (ESS = 3976).

Discussion

We found a positive allometric growth for the African obtuse snakehead *Parachanna obscura* dwelling in the Eleiyele Reservoir. Extensive literature exists on the allometric growth coefficient of *P. obscura* but based on Ordinary Least Square Regression, a frequentist inferential statistics: Olanrewaju [19] reported positive allometric growth b = 3.1 for the species in Eleiyele reservoir. Little or no information exists this coefficient based on Bayesian frameworks except for Froese and Pauly [20] in FishBase, who reported an allometric growth coefficient, b = 3.02 using the Hierarchical Bayesian method based on informed priors. If a fish grows without changing its shape or density, it is said to exhibit isometric growth. In this case, the volume of the fish is proportional to any linear measure of its size. If weight is taken as a surrogate for volume (which requires assuming a constant density) and length as the linear measure, then the modeled relationship between length and weight, (1), will have b = 3 under isometric growth. Isometric growth of fish is rare [21, 22]. If a fish changes shape or density as it grows, b ≠ 3, and the fish is said to exhibit allometric growth. If b >3, then the fish tends to become “plumper" as the “fish increases in length [23].

Two important characteristics of the length – weight relationship in fish are non-linearity (Figure 3E) (length as a linear measure and weight as being related to volume. Thus, as the organism adds a linear amount of length, it adds a disk of volume with a commensurate weight) and variability in weight increases the length of the fish (i.e., the scatter of the points increases from left to right). Thus, the variability in weight among shorter fish is less than the variability in weight

among longer fish. Unfortunately, because of these two characteristics, length-weighted data tend to violate the linearity and homoscedasticity (i.e., "constant variance") assumptions of simple linear regression [24]. With this linearization via logarithmic transformation, the linearity assumption for the linear regression is satisfied. Ogle [24] noted that the variability on the log scale appears greater for smaller fish because the log scale lacks the precision required to distinguish the weights of small fish over a wide length range. He recommended that fish with a log weight of less than -0.5should be eliminated from this analysis because of scale imprecision for these fish.

Most Bayesian software packages use simulation techniques such as Markov chain Monte Carlo (MCMC), which can be to find the posterior distribution and to sample from it [25]. An MCMC algorithm uses a Markov chain – a sequence of states in which the location of the next state depends on the current state of the chain – to explore the shape of the desired posterior distribution [12, 25]. In this study, the convergence and stability of Bayesian sampling via MCMC was evaluated using the *Scale Reduction Factor*, R-hat, which should be below 1.01 [26, 27], and the Effective Sample Size (ESS), which should be greater than 1000 [28].The interpretation of a frequentist 95% confidence interval, for example, is one that results from repeating an experiment an infinite number of times, and from this, estimating the quantity of interest (e.g., mean), resulting in 95% of the estimated means lying within the confidence interval. A 95% Bayesian credible interval can be interpreted as having a 95% probability that the mean lies within the credible interval; given the data. A point estimate is a single value, most often a measure of the central tendency (e.g., mean, median), chosen to summarize a posterior distribution. In contrast to the maximum likelihood approach, in which the point estimate (mode) is the focus, in the Bayesian framework we focus on the entire posterior distribution. A point estimate is merely one of the many ways of summarizing the posterior, as in this study.

The full concept of Bayesian inference with its mathematical derivations is beyond the scope of this study. The Bayesian inferential method offers flexibility in modelling estimation and evaluation compared with common Frequentists statistics, but at the expense of the additional compuatitional power required for its implementation. This study suggests that the African obtuse snakehead, *Parachanna obscura* thrives in the Eleiyele Reservoir because they are plums. There is a dearth of allometric growth coefficients for this species inferred by Bayesian inference, therefore we could not compare the present results with others. However, this result is comparable to observations made by frequentist methods. This suggests that the Bayesian regression model can effectively substitute for common frequentist frameworks in fish biology and ecology.

References

[1] Binohlan C, Pauly D: The length-weight table, In: *Fishbase 2000: concepts, design and data sources*, Froese R. & D. Pauly, (Eds), 121-123, ICLARM, ISBN 971-8709-99-1, Manila, Philippines, 1994, 200p.

[2] Brown JH, West GB, Enquist BJ. Scaling in biology: process, causes and consequences. In *Scaling in Biology* (ed. J. H. Brownand G. B. West), 2000, pp. 1-24. New York: Oxford University Press.

[3] Wiesenfeld K: Scaling laws. *Animal Journals of Physiology*, 69, 2001, 938-942.

[4] Chave J, Levin S: Scale and scaling in ecological and economic systems. *Environment Resource Economics* 26, 2003, 527-557.

[5] Le Cren ED: The length-weight relationship and seasonal cycle in gonad weight and condition in the perch (*Perca avescens*). *Journal of Animal Ecology*. 20, 1951, 201-219.

[6] Sil A, Betkerur J, Das NK: *P*-Value Demystified. *Indian dermatology online journal*, *10*(6), 2009, 745–750. https://doi.org/10.4103/idoj.IDOJ_368_19

[7] Hasley LG: The reign of the *p*-value is over: what alternative analyses could we employ to fill the power vacuum?*Biol. Lett.*2019, 15201901742019017 4http://doi.org/10.1098/rsbl.2019.0174.

[8] Box GEP, Tiao GC: *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, MA pp 1973, 112-119.

[9] Berger JO: Robust Bayesian analysis: Sensitivity to the prior. *Journal of Statistical Planning and Inference* 25, 1990, 303-328.
[10] Bernado JM, Smith AFM: *Bayesian Theory*. Wiley, New York, NY.212, 1994, p
[11] Guy CS, Brown ML, (Eds.)" Analysis and interpretation of freshwater fisheries data. In Analysis and Interpretation of Freshwater Fisheries Data. American Fisheries Society. https://doi.org/10.47886/9781888569773.ch1 , 2007.
[12] Muth C, Oravecz Z., Gabry J: User-friendly Bayesian regression modeling: A tutorial with rstanarm and shinystan. *The Quantitative Methods for Psychology*, *14*, 2018, 99–119. https://doi.org/10.20982/tqmp.14.2.p099
[13] Hilborn R, Mangel M: *The ecological detective: confronting models with data*. Princeton University Press, Princeton, New Jersey, 1997
[14] Idodo-Umeh G: *Freshwater fishes of Nigeria Taxonomy, Ecological Notes, Diet and Utilization*. Idodo Umeh Publishers, Benin City, Nigeria, 2003, 232p.
[15] Froese R, Pauly D(Eds). FishBase. World Wide Web electronic publication. URL: www.fishbase.org http//:www.fishbase.org, 2018 [version 08/2018].
[16] R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/, 2022.
[17] Gabry J, Goodrich B. rstanarm: Bayesian appliedregression modeling via Stan. R package version2.15.3. Retrieved from http://mc-stan.org/rstanarm, 2017.
[18] Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich, B, Betancourt M, Riddell A. Stan: A probabilistic programming language. *Journal of Statistical Software, Articles*, *76*(1), 2017, 1–32. doi:10.18637/jss.v076.i01.
[19] Olarenwaju AN, Ajani EK, Kareem OK, Orisaso O: Length-weight relationships and state of well-being of ~~p~~*Parachanna obscura* gunther1861, in Eleyele reservoir, southwestern Nigeria. *Fisheries and Aquaculture Journal*, *08*, 2017, (03). https://doi.org/10.4172/2150-3508.1000221.
[20] Froese R, Pauly D. (Eds.): FishBase. World Wide Web electronic publication. URL: www.fishbase.org http//:www.fishbase.org [version 01/2022]
[21] McGurk MD: Effects of net capture on the post-preservation morphometry, dry weight, and condition factor of Pacific herring larvae. *Transactions of the American Fisheries Society* 114, 1985, 348:355.
[22] Bolger T, Connolly PL: The selection of suitable indices for the measurement and analysis of fish condition. *Journal of Fish Biology* 34, 1989, 171-182.
[23] Blackwell BG, Brown, ML, Willis DW: Relative weight (Wr) status and current use in fisheries assessment and management. *Reviews in Fisheries Science*, 1994, 8:1-44.
[24] Ogle DH: *Introductory fisheries analyses with R*. CRC Press, Taylor & Francis Group, 2016.
[25] Korner-Nievergelt F, Von Felten S, Roth T, Almasi B, Guélat J, Korner-Nievergelt P.: Bayesian data analysis in ecology using linear models with R, Bugs, and Stan, 2015. Elsevier/AP, Academic Press is an imprint of Elsevier, 188p
[26] McElreath R: Statistical rethinking: A Bayesian course with examples in R and Stan. CRC press. https://xcelab.net/rm/statistical-rethinking/, 2015.
[27] Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner PC: Rank-normalization, folding, and localization: An improved rˆ for assessing convergence of mcmc (With discussion). *Bayesian Analysis*, *16*(2). https://doi.org/10.1214/20-BA1221, 2021
[28] Bürkner PC: brms: An R package for Bayesian multilevel models using Stan. Journal of Statistical Software, 80(1), 2017, 1-28.