

Research Article

Examination of the Effects of Curvature in Geometrical Space on Accuracy of Scaling Derived Projections of Plant Biomass Units: Applications to the Assessment of Average Leaf Biomass in Eelgrass Shoots

Héctor Echavarría-Heras ¹, Cecilia Leal-Ramírez,¹
Enrique Villa-Diharce,² and Abelardo Montesinos-López³

¹Centro de Investigación Científica y de Estudios Superiores de Ensenada, Carretera Ensenada-Tijuana No. 3918, Zona Playitas, Código Postal 22860, Apdo. Postal 360 Ensenada, B.C., Mexico

²Centro de Investigación en Matemáticas, A.C. Jalisco s/n, Mineral Valenciana, Guanajuato Gto., Código Postal 36240, Mexico

³Departamento de Matemáticas, Centro Universitario de Ciencias Exactas e Ingenierías (CUCEI), Universidad de Guadalajara, 44430 Guadalajara, Jalisco, Mexico

Correspondence should be addressed to Héctor Echavarría-Heras; heheras@icloud.com

Received 4 December 2018; Revised 16 February 2019; Accepted 26 March 2019; Published 23 April 2019

Academic Editor: Krishna Prasad

Copyright © 2019 Héctor Echavarría-Heras et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conservation of eelgrass relies on transplants and evaluation of success depends on nondestructive measurements of average leaf biomass in shoots among other variables. Allometric proxies offer a convenient way to assessments. Identifying surrogates via log transformation and linear regression can set biased results. Views conceive this approach to be meaningful, asserting that curvature in geometrical space explains bias. Inappropriateness of correction factor of retransformation bias could also explain inconsistencies. Accounting for nonlinearity of the log transformed response relied on a generalized allometric model. Scaling parameters depend continuously on the descriptor. Joining correction factor is conceived as the partial sum of series expansion of mean retransformed residuals leading to highest reproducibility strength. Fits of particular characterizations of the generalized curvature model conveyed outstanding reproducibility of average eelgrass leaf biomass in shoots. Although nonlinear heteroscedastic regression resulted also to be suitable, only log transformation approaches can unmask a size related differentiation in growth form of the leaf. Generally, whenever structure of regression error is undetermined, choosing a suitable form of retransformation correction factor becomes elusive. Compared to customary nonparametric characterizations of this correction factor, present form proved more efficient. We expect that offered generalized allometric model along with proposed correction factor form provides a suitable analytical arrangement for the general settings of allometric examination.

1. Introduction

The model of relative growth of Huxley [1] is formally stated by means of a scaling relationship of the form

$$w = \beta a^\alpha, \quad (1)$$

where w and a are measurable traits and the parameter α is designated as the allometric exponent, while β is identified as the normalization constant. This model, also termed equation of simple allometry, has been extensively used in research

problems in biology [1–5], physics [6], economics [7], earth sciences [8], resource management, and conservation [9, 10], among other fields.

Eelgrass provides nursery for waterfowl and fish species. By trapping sediment and stumping wave energy, this seagrass promotes shoreline stabilization. Eelgrass services also include nutrient recycling, water filtration, and carbon dioxide removal. Current anthropogenic influences threaten eelgrass permanence. Conservation efforts rely on plot transplanting in a fundamental way. Monitoring effectiveness

depends on measurements of standing stock and productivity through time. This makes the assessment of average leaf biomass in shoots a necessary input. But traditional estimation of eelgrass leaf biomass units relies on destructive methods. This could alter shoot density in a developing transplant. Thus evaluation renders indirect assessment methods necessary [10, 11]. Results show that an allometric scaling of the form of (1) for eelgrass leaf biomass w and associated area a is consistent [11]. Derived projections of individual leaf biomass convey useful surrogates for mean leaf biomass in shoots. Moreover, estimates of the parameters α and β are invariant within a given geographical region [10–12]. Hence, estimates fitted at site can endow suitable projections of leaf biomass values currently observed in other places of the region. This bears the referred allometric projections of a convenient nondestructive feature ([11, 13]).

Simplicity of (1) makes allometric projection of eelgrass leaf biomass units convenient. But there are caveats on dependability. For instance, even though α and β are invariant, environmental influences can induce a relative extent of variability on local estimates ([12, 14]). Besides, the response in the power function-like scaling of (1) is very sensitive to variation of parameter estimates. Then, accuracy of derived proxies is subject of error propagation of estimates. In addition, there are factors of biological scaling which can influence precision of estimates (e.g., [10, 11, 13, 15]). Packard [16] questioned results of Mascaro et al. [17] on allometric examination, and Mascaro et al. [18] responded to criticism. Going over this exchange highlights the relevance of procedural factors in determining precision of parameter estimates of allometric scaling. It also offers a convenient framework for the aims of the present research.

An important factor influencing precision of estimates of allometric parameters is analysis method. A widespread approach is the traditional analysis method of allometry (TAMA hereafter). It relies on a log transformation of data in arithmetical scale in order to contemplate a linear regression model in geometrical scale. Then, the fitted line is back-transformed to yield a two-parameter power function in the original scale. But embracing this procedure fuels a vivid unresolved debate. Views assert that this protocol can lead to biased results (e.g., [16, 19–29]). And other practitioners consistently wed to the idea that logarithmic transformations are necessary (e.g., [18, 30–39]). An alternative to the TAMA approach is using nonlinear regression methods in the direct scale of the data [16]. Echavarría-Heras et al. [11] concluded that producing allometric projections of average leaf biomass in eelgrass shoots must rely on this protocol. Yet a direct nonlinear regression approach in allometry is also not unailing. For instance, inadequate identification of the inherent error structure can lead to significant bias [18]. Besides, Lai et al. [30] found that estimates of allometric parameters fitted by nonlinear regression can exhibit a high sensitivity to the largest values of the covariate. Therefore, evaluation of analysis method suitability in acquiring consistent eelgrass leaf biomass proxies needs revision.

The adoption of methods of curvature in geometrical space could offer a way to overcome inadequacy of the TAMA procedure ([27, 40–42]). In particular, it is pertinent

to examine if taking curvature into account leads to improved accuracy of eelgrass leaf biomass proxies. But, according to Mascaro et al [18], curvature could manifest because of methodological factors of data gathering. Thus, an examination of the effect of curvature in eelgrass leaf biomass allometry must also take into account a possible participation of data quality effects. Mascaro et al. [18] reminds on three ways of handling curvilinearity in geometrical space. One is by separating data to contemplate different local linear models to account for heterogeneity of effects of the covariate [43–45]. A second one is by fitting a polynomial model [46–49]. A third approach endorses direct nonlinear regression assuming a heteroscedastic error structure as contemplated by Mascaro et al. [17]. Either approach above bears complexity beyond the linear model in geometrical space that associates with the customary bivariate power function of allometry. This suggested putting forward a generalized allometric model intended to deal with curvature in geometrical space. This paradigm incorporates parameters that change as continuous functions of the log transformed covariate ([27, 39, 46]). As we explain further on, the curvature arrangements recommended by Mascaro et al. [18] can be all derived from the offered formalization. Moreover, a nonzero intercept power function that Packard [50] recommends to handle curvilinearity in geometrical space also derives from the presented generalized scaling model.

But any scheme addressing curvature in geometrical space depends on a factor for correction of bias of retransformation of the regression error. In the general settings, if ϵ stands for the regression error, then the said correction factor, through denoted using the symbol $\delta(\epsilon)$, is given by the mean of the exponentiated error random variable; that is, $\delta(\epsilon) = E(e^\epsilon)$ [51–53]. Furthermore, the TAMA approach relies on the essential assumption that ϵ is additive, normally distributed, and homoscedastic [33]. When this happens, $\delta(\epsilon)$ takes its lognormal-mean form [51–53]. But if ϵ fails to be normally distributed, there are two possibilities. If the distribution of ϵ is known, we could derive a closed form for $\delta(\epsilon)$. In turn, if the error distribution is not identified a priori, a widespread approach is taking $\delta(\epsilon)$ in the nonparametric form given by the smearing estimator of bias of Duan [54]. Still, there are provisions on this. A smearing estimate form can fail to compensate the downward retransformation bias of logged data ([53, 55, 56]). Thus, in a circumstance where ϵ is unspecified, characterizing $\delta(\epsilon)$ seems elusive. Here, we put forward an arrangement for $\delta(\epsilon)$ aimed to get around this circumstance. Zeng and Tang [57] proposed a nonparametric alternate to the smearing form. It matches the first three terms' partial sum of a power series expression of $E(e^\epsilon)$, assuming $E(\epsilon) = 0$. Form suggested here corresponds to a generalization of this construct. It does not abide the restriction $E(\epsilon) = 0$ and matches an n -terms partial sum approximation of the exponential series representation of $E(e^\epsilon)$. The partial sum that maximizes reproducibility strength of retransformed mean response sets criterion to choose n .

Present results show that a consideration of curvature in geometrical space, as well as a suitable characterization of the correction factor of retransformation bias, offers consistent

allometric proxies of observed mean leaf biomass in eelgrass shoots. Hence, contrary to views asserting direct nonlinear regression as mandatory in allometric examination, our findings validate a parallel reliability of log transformation based methods. This is well in line with claims of Mascaro et al. [18] and many others about blaming the use of logarithms of incongruent results in allometric analysis. Moreover, keeping the analysis in geometrical space unraveled heterogeneity in the inherent leaf biomass scaling pattern. This could not be achieved by clinging to direct nonlinear regression in arithmetical space as the only valid approach of allometric examination. Offered analytical arrangement is expected to be applicable in the general settings of allometry.

2. Materials and Methods

2.1. Data. For the aims of the present research, we relied on an extensive eelgrass data set collected in San Quintin Bay, a coastal lagoon on the Pacific side of the Baja California Peninsula, México (30°30' N – 116°00'W), and through a 13 months' long sampling period covering a whole-year cycle. Data composes measurements of length (mm), width (mm), and dry weight w (g) of a total of 10412 individual eelgrass leaves taken from 20 randomly thrown 400cm² quadrats every monthly visit to the site. A sampling visit will be further referred to as "sampling time" in the text. The length times width proxy [11] provided estimations of leaf area a (mm²). In order to test for methodological influences of data gathering, we processed raw data set according to a mean plus or minus two standard deviations outlier's removal procedure [58]. Appendix A presents results of an exploratory analysis of data.

2.2. Models. As above specified symbols, w and a stand for the biomass of an individual eelgrass leaf and its respective area one to one. Echavarría-Heras et al. [11] assert that these variables can be related through the bivariate allometric model of (1). One procedure to acquire estimates for the parameters α and β is fitting directly in arithmetical scale a nonlinear homoscedastic regression model. Besides, we can use a TAMA approach, that is, fitting the linear regression model

$$v = \ln \beta + \alpha u + \epsilon, \quad (2)$$

where $v = \ln w$, $u = \ln a$, and ϵ is a random error term assumed to be normally distributed with zero mean and variance σ^2 , that is, $\epsilon \sim N(0, \sigma^2)$.

We conceive curvature in geometrical space as a circumstance where fitting results of regression model of (2) are inconsistent. Dealing with this situation amounts for considering complexity beyond incorporated by (1). One possible approach to address curvature is assuming that scaling parameters α and β in (2) depend continuously on the covariate ([27, 39, 46]). This is consistent with the generalized allometric model,

$$w = \beta(a) a^{\alpha(a)}, \quad (3)$$

with $\beta(a)$ and $\alpha(a)$ intended to be continuous and differentiable functions defined on R^+ and with $\beta(a)$ being positive.

Certainly a log transformation $v = \ln w$, $u = \ln a$ of (3) establishes the regression model

$$v = v_C(a, u) + \epsilon, \quad (4)$$

where

$$v_C(a, u) = \ln \beta(a(u)) + \alpha(a(u))u, \quad (5)$$

where ϵ , a residual error term, is conceived as a random variable that in the general settings is ψ -distributed with mean μ and variance set by a function $\sigma^2(u)$ of the covariate u , that is, $\epsilon \sim \psi(\mu, \sigma^2(u))$.

Setting $\beta(a(u)) = \beta$ and $\alpha(a(u)) = \alpha$ with α and β constants reduces (4) to the regression model of (2). In Appendix B, we explain that (3) accommodates all curvature paradigms suggested by Mascaro et al. [18]. These include a biphasic and a polynomial model in geometrical space, as well as the nonlinear heteroscedastic model referred to by Mascaro et al. [17] in direct arithmetical space. Moreover, as shown in Appendix B, the three-parameter power function chosen as an alternate standard for curvature [59] also derives from (3).

2.2.1. Biphasic Model in Geometrical Space. In order to characterize the model of (3) in a biphasic mode, we let $\beta(a) = \beta_B(a)$ and $\alpha(a) = \alpha_B(a)$, such that

$$\begin{aligned} \alpha_B(a) &= \sum_1^2 \vartheta_i(u(a)) \alpha_i \\ \beta_B(a) &= \exp\left(\sum_1^2 \vartheta_i(u(a)) \ln \beta_i\right) \end{aligned} \quad (6)$$

including parameters β_i and α_i and the function $\vartheta_i(u(a))$ given by

$$\vartheta_i(u(a)) = \begin{cases} (2-i)(1-H(u-u_b)) & \text{for } i = 1 \\ (i-1)H(u-u_b) & \text{for } i = 2, \end{cases} \quad (7)$$

for $i = 1, 2$. $H(u-u_b)$ is a Heaviside function $H(z)$ [60], evaluated at $z = u-u_b$ and correspondingly $u_b = \ln a_b$, with $a_{min} \leq a_b \leq a_{max}$ being a point separating growth phases $u \leq u_b$ and $u > u_b$. Then, denoting by means of $v_B(a, u)$ the resulting form of $v_C(a, u)$ from (4), we get the biphasic regression model

$$v = v_B(a, u) + \epsilon, \quad (8)$$

where ϵ is a random error term as defined in (4) and

$$v_B(a, u) = \sum_1^2 \vartheta_i(u(a)) f_i(u(a)) \quad (9)$$

with

$$f_i(u(a)) = \ln \beta_i + \alpha_i u(a), \quad (10)$$

where β_i and α_i for $i = 1, 2$ parameters are to be estimated from data.

2.2.2. *Polynomial Model in Geometrical Space.* Similarly, assume that $\alpha(a) = \alpha_P(a)$ and $\beta(a) = \beta_P(a)$, with

$$\begin{aligned}\alpha_P(a) &= \sum_0^n \alpha_k u(a)^k \\ \beta_P(a) &= e^{\sum_0^n \beta_k u(a)^k}.\end{aligned}\quad (11)$$

α_k and β_k for $1 \leq k \leq m$ are coefficients; one can acquire a polynomial representation $v_P(a, u)$ for the generalized mean response function in geometrical space $v_C(a, u)$. This way the polynomial form of regression (4) becomes

$$v = v_P(a, u) + \epsilon, \quad (12)$$

where ϵ is a random error term as defined in (4), with

$$v_P(a, u) = \sum_0^m p_k u(a)^k, \quad (13)$$

and p_k , for $k = 0, 2, \dots, m$ parameters.

2.2.3. *Nonlinear Heteroscedastic Model in Arithmetical Space.* As we explain ahead, direct algebraic manipulation of (3) leads to the consideration of the nonlinear heteroscedastic regression model addressed by Mascaro et al. [17]; namely,

$$w = \beta_\theta a^{\alpha_\theta} + a^\theta \epsilon, \quad (14)$$

with α_θ , β_θ , and θ being parameters and ϵ being a zero mean normally distributed error term with a covariate dependent variance $\sigma^2(a) = \sigma^2 a^{2\theta}$, that is, $\epsilon \sim N(0, \sigma^2 a^{2\theta})$.

A nonlinear homoscedastic form derives from (14) by setting $\theta = 0$; that is,

$$w = \beta_0 a^{\alpha_0} + \epsilon, \quad (15)$$

And, again ϵ is an additive error term assumed to be normally distributed with zero mean and variance σ^2 , that is, $\epsilon \sim N(0, \sigma^2)$.

Appendix A deals with exploratory analysis of data Appendix B presents notation convention and also explains how all addressed paradigms derive from the generalized model of (3). Appendix C explains the addressed forms of correction factor for bias of retransformation of the regression error. Fitting results of the geometrical space based models appear in Appendix D. Those corresponding to the nonlinear heteroscedastic and homoscedastic models pertain to Appendix E. Agreement between observed and projected values is commonly evaluated by analyzing values of Lin's Concordance Correlation Coefficient (CCC) [61]. This correspondence index is commonly denoted by means of the symbol ρ . Agreement will be defined as poor whenever $\rho < 0.90$, moderate for $0.90 \leq \rho < 0.95$, good for $0.95 \leq \rho < 0.99$, or excellent for $\rho \geq 0.99$ [62]. Besides CCC values, we assessed reproducibility by comparing goodness-of-fit statistics, such as the coefficient of determination (CD), standard error of estimate (SEE), mean prediction error (MPE), total relative error (TRE), average systematic error (ASE), and mean percent standard error (MPSE) ([63–65]). For statistical tasks, we relied on the R package release 3.5.

3. Results

Exploratory analysis in Appendix A identifies maximum, minimum, and sample mean values for observed leaf area values a and associated dry weights w . We also explain distribution of variables in terms of quantiles of probability 0.1, 0.25, 0.50, 0.75, and 0.90, for both crude and processed data. Statistical exploration extends to log transformed values of these variables. We present Q-Q plots (quantile-quantile) for comparison of distribution patterns, as well as boxplots for the 13 months' long sampling scheme for both crude and processed data. We can learn that, from month 2 to month 6, a reduction in the values of weight and area occurred; this perhaps is explained by an increase in temperature during the period. We can be also aware that a similar variation pattern over time is shown in both raw and processed data sets.

3.1. *Fitting Results of Geometrical Space Models.* In order to validate curvature in geometrical space, we compared the linear model derived from (1) as well as biphasic and polynomial alternates derived from the generalized model of (3). Appendix B explains formal matters. Tables 1 and 2 summarize notation convention. Equations numbered beyond (15) belong to the appendices. Appendix D explains corresponding regression protocols.

Fitting results of the TAMA arrangement of (2) appear in Appendix D. Figure 1 shows the spread about TAMA's linear mean response function $E_T(v | u)$. We can visually ascertain that deviations from the linear mean response function $E_T(v | u)$ suggest curvature (red dots). Thus, data processing removed inconsistent replicates but shown spread still deviates from a linear mean response. This suggests that curvature in geometrical space could not be explained by methodological factors related to data gathering.

Fitting results of the biphasic protocol of (8) are summarized in Appendix D. Figure 2 displays the spread about mean response function $E_B(v | u)$ in geometrical space. Compared with Figure 1, we can ascertain that the biphasic fit provides a consistent account of different variation patterns among smaller and larger leaves. We can visually ascertain that fit produced consistent results. This confirms a judgement that identified curvature might be due to intrinsic factors of leaf growth rather than methodological influences related to data gathering.

Appendix D presents fitting results of the polynomial model ($m = 6$). Figure 3 displays dispersion about the polynomial mean response function in geometrical space $E_P(v | u)$. A polynomial representation also exhibits higher consistency than the TAMA arrangement. Recalling the biphasic scheme, the polynomial suggests a smooth transition between two growing phases.

3.2. *Model Selection in Geometrical Space.* Assessment of models fitted on geometrical space relied on goodness-of-fit statistics, that is, the coefficient of determination, standard error of estimate, mean prediction error, total relative error, average systematic error, and mean percent standard error ([63–65]). Besides, we took into account concordance correlation coefficient [61] and Akaike's information index [66].

TABLE 1: Summary of notation convention of the bivariate scaling model of (1) and its generalization to account for curvature as given by (3). GS stands for geometrical space, AS stands for arithmetical space, and CF means correction factor of retransformation bias.

	Typical bivariate	Eq.	Generalized for curvature	Eq.
Basic form	$w = \beta a^\alpha$	(1)	$w = \beta(a) a^{\alpha(a)}$	(3)
Regression equation in GS	$v = \ln \beta + \alpha u + \epsilon$	(2), (D.1)	$v = v_C(a, u) + \epsilon$	(4)
Mean response GS	$E_T(v u) = \ln \beta + \ln \alpha$	(B.4)	$E_C(v u) = \ln \beta(a) + \alpha(a) u(a)$	(B.1)
Back transformation AS	$w = e^{E_T(v u)} e^\epsilon$	(B.5)	$w = e^{E_C(v u)} e^\epsilon$	(B.2)
Mean response AS	$E_{T\delta}(w a) = \beta a^\alpha \delta(\epsilon)$ $\delta(\epsilon) = E(e^\epsilon)$	(B.6)	$E_{C\delta}(w a) = \beta(a) a^{\alpha(a)} \delta(\epsilon)$ $\delta(\epsilon) = E(e^\epsilon)$	(B.3)
CF Baskerville	$E_{TB}(w a) = \beta a^\alpha \delta_B(\epsilon)$ $\delta_B(\epsilon) = e^{\sigma^2/2}$	(B.6), (C.1)	$E_{CB}(w a) = \beta(a) a^{\alpha(a)} \delta_B(\epsilon)$	(B.3), (C.1)
CF Duan	$E_{TD}(w a) = \beta a^\alpha \delta_D(\epsilon)$ $\delta_D(\epsilon) = \frac{\sum_1^m e^{\epsilon^j}}{m}$	(B.6), (C.2)	$E_C(w a) = \beta(a) a^{\alpha(a)} \delta_D(\epsilon)$	(B.3), (C.2)
CF Zeng and Tang	$E_{T_{ZT}}(w a) = \beta a^\alpha \delta_{ZT}(\epsilon)$ $\delta_{ZT}(\epsilon) = 1 + \frac{\sigma^2}{2}$	(B.6), (C.3)	$E_{C_{ZT}}(w a) = \beta(a) a^{\alpha(a)} \delta_{ZT}(\epsilon)$	(B.3), (C.3)
CF n-partial sum	$E_{T_n}(w a) = \beta a^\alpha \delta_n(\epsilon)$ $\delta_n(\epsilon) = \sum_0^n \frac{E(\epsilon^k)}{k!}$	(B.6), (C.4)	$E_{C_n}(w a) = \beta(a) a^{\alpha(a)} \delta_n(\epsilon)$	(B.3), (C.4)

TABLE 2: Notation convention for curvature models in geometrical space. We include the biphasic and polynomial characterizations of the generalized allometric model of (3).GS stands for geometrical space, AS stands for arithmetical space, and CF means correction factor of retransformation bias of the regression error.

	Biphasic model	Eq.	Polynomial model	Eq.
Basic form	$w = \beta_B(a) a^{\alpha_B(a)}$	(B.11)	$w = \beta_P(a) a^{\alpha_P(a)}$	(B.27)
Regression equation	$v = v_B(a, u) + \epsilon$	(8), (D.4)	$v = v_P(a, u) + \epsilon$	(B.32), (D.8)
Mean response GS	$E_B(v u) = \sum_1^2 \vartheta_i(u(a)) f_i(u(a))$	(B.18)	$E_P(v u) = \sum_0^m p_k u(a)^k$	(B.33)
Back transformation AS	$w = \beta_B(a) a^{\alpha_B(a)} e^\epsilon$	(B.21)	$w = \beta_P(a) a^{\alpha_P(a)} e^\epsilon$	(B.34)
Mean response AS	$E_{B\delta}(w a) = \beta_B(a) a^{\alpha_B(a)} \delta(\epsilon)$ $\delta(\epsilon) = E(e^\epsilon)$	(B.22)	$E_P(w a) = \beta_P(a) a^{\alpha_P(a)} \delta(\epsilon)$ $\delta(\epsilon) = E(e^\epsilon)$	(B.35)
CF Baskerville	$E_{BB}(w a) = \beta_B(a) a^{\alpha_B(a)} \delta_B(\epsilon)$	(B.22), (C.1)	$E_{PB}(w a) = \beta_P(a) a^{\alpha_P(a)} \delta_B(\epsilon)$	(B.35), (C.1)
CF Duan	$E_{BD}(w a) = \beta_B(a) a^{\alpha_B(a)} \delta_D(\epsilon)$	(B.22), (C.2)	$E_{PD}(w a) = \beta_P(a) a^{\alpha_P(a)} \delta_D(\epsilon)$	(B.35), (C.2)
CF Zeng and Tang	$E_{B_{ZT}}(w a) = \beta_B(a) a^{\alpha_B(a)} \delta_{ZT}(\epsilon)$	(B.22), (C.3)	$E_{P_{ZT}}(w a) = \beta_P(a) a^{\alpha_P(a)} \delta_{ZT}(\epsilon)$	(B.35), (C.3)
CF n-partial sum	$E_{B_n}(w a) = \beta_B(a) a^{\alpha_B(a)} \delta_n(\epsilon)$	(B.22), (C.4)	$E_{P_n}(w a) = \beta_P(a) a^{\alpha_P(a)} \delta_n(\epsilon)$	(B.35), (C.4)

Table 3 presents results. Goodness-of-fit statistics and ρ and AIC values disfavored the TAMA protocol. On the contrary, comparison indices favored the biphasic model. Moreover, differences among indices but TRE and ASE for this scheme and the polynomial ($m = 6$) are slight. Particularly, the highest AIC is associated with the TAMA protocol (AIC = 15069.9, Δ AIC = 2491.9). Therefore, this model bears the less support. The biphasic choice delivered the smallest AIC's value (AIC = 12578.0, Δ AIC = 0). Nevertheless, difference in AIC is just barely relative to the $m = 6$ polynomial model, since this choice conveyed (AIC = 12615.0 and Δ AIC = 37). Thus, model confrontation shows that the TAMA protocol

is unsuited, thus backing the assertion that whatever model aims to be consistent with the present data, it ought to be nonlinear in geometrical space.

3.3. *Retransformation Results.* The TAMA protocol was not supported by the model selection criteria. Anyway, for comparison, corresponding retransformation results are included in Appendix D. Related with the TAMA protocol, fitting results of the biphasic model display a relatively improved distribution of residuals about the zero line. Nevertheless, normal Q-Q plot still shows heavier tails than those expected for a normal distribution. And, again both test statistics and

TABLE 3: Assessment of geometrical space fitted models. Comparison took into account goodness-of-fit statistics, that is, the coefficient of determination (R^2) standard error of estimate (SEE), mean prediction error (MPE), total relative error (TRE), average systematic error (ASE), and mean percent standard error (MPSE) ([63–65]). Besides, we considered concordance correlation coefficient (ρ) [61] and Akaike's information index (AIC) [66].

Method	AIC	ρ	R^2	SEE	TRE	ASE	MPE	MPSE
TAMA	15069.9	0.9505	0.9045	0.5135	-0.0286	-0.2462	-0.1879	5.7877
Biphasic	12578.0	0.9614	0.9256	0.4530	0.0037	0.0286	-0.1658	5.1446
Polynomial	12615.0	0.9614	0.9241	0.4576	0.8497	1.2008	-0.1675	5.3187

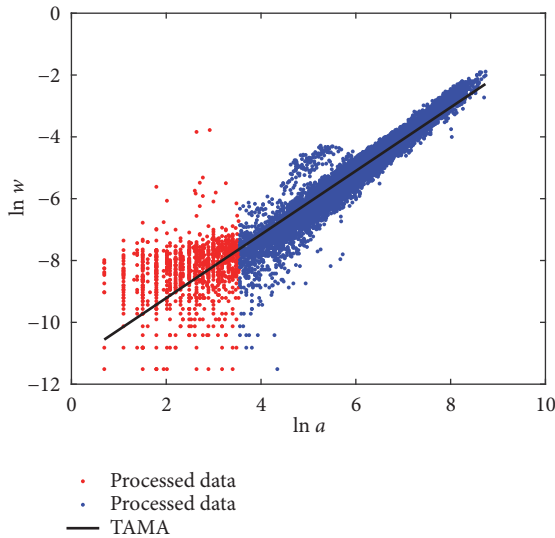


FIGURE 1: Spread about the TAMA's linear mean response function $E_T(v | u)$. Deviations about $E_T(v | u)$ shown by red dots suggest curvature.

p values of an Anderson-Darling test [67] provide evidence against normality of residuals. This justifies choosing the non-parametric forms $\delta_D(\epsilon)$, $\delta_{ZT}(\epsilon)$, or $\delta_n(\epsilon)$ for compensation of downward bias induced by retransformation of the regression error ([11, 52, 53]). Table 4 displays comparison statistics for the reproducibility strength of the biphasic mean response $E_{B\delta}(w | a)$ as shaped by the different forms of $\delta(\epsilon)$.

We can learn that agreement between the biphasic mean response and leaf biomass data is best for $\delta_n(\epsilon)$. Figure 4 shows spread of processed leaf biomass values about the biphasic mean response function $E_{B\delta}(w | a)$ as shaped by the considered forms of $\delta(\epsilon)$. We can observe that both $\delta_D(\epsilon)$ and $\delta_{ZT}(\epsilon)$ overcompensate the bias correction by $\delta_n(\epsilon)$. Moreover results show that as opposed to the TAMA a biphasic protocol along with the $\delta_n(\epsilon)$ form offers consistent proxies of individual leaf biomass. But it is worth mentioning that in spite of the fact that model selection favored the biphasic scheme, examination of the polynomial model output reveals similar predictive strength to the biphasic alternate.

3.4. Assessing Curvature by Direct Nonlinear Regression. As suggested by Mascaro et al. [18], effects of curvature in geometrical space can be analyzed by means of the direct nonlinear heteroscedastic regression model of (14). In Appendix B, we explain that such a protocol also derives

from the generalized bivariate allometric model of (3). Table 5 presents pertinent notation convention. For comparison, we also present results for the associated homoscedastic case.

Fitting results of the heteroscedastic and homoscedastic models appear in Appendix E. We can learn that estimates for the normalization constant and scaling exponent parameters are very similar. Certainly, corresponding 95% confidence intervals display some overlap. As a result, we can expect similar reproducibility features for both models. Table 6 presents comparison statistics.

We can be aware that model assessment backs the heteroscedastic model. But selection here is mainly on qualitative grounds. It actually concerns the ability of the heteroscedastic model to identify an expected dependence of variance in the covariate. Certainly, the reproducibility strengths of both paradigms are equivalent. Indeed, Figure 5 shows that mean response curves $E_\theta(w | a)$ and $E_o(w | a)$ differ just barely.

Results show that as it occurred for models fitted in geometrical space, data cleaning failed to correct a heavy tails problem for the nonlinear fits. This can be ascertained from the normal Q-Q plot of residuals. This strengthens our point on the consideration of a different error structure from the one assumed here. Exploring the effects of error structure in the fitting of models for curvature addressed here will be a matter of further research. Interestingly, both the homoscedastic and heteroscedastic models seem to induce the same reproducibility strengths.

3.5. Model Assessment in Arithmetical Space. The model selection assay in geometrical space summarized in Table 3 favored the biphasic protocol. Correspondingly, statistics in Table 6 support the nonlinear heteroscedastic model. Results of Table 4 endure $\delta_n(\epsilon)$ as required for largest reproducibility of retransformation output. Table 7 allows assessment of these models. We can learn that half the number of comparison indices coincide (ρ , R^2 , SEE, and MPE). In addition, the biphasic model is favored by AIC, ASE, and MPSE. This sets criterion for selection of curvature in geometric space as a consistent paradigm for the present data. Accordingly, the biphasic model bears adequacy.

3.6. Implications for Allometric Proxies of Mean Leaf Biomass in Eelgrass Shoots. We in turn consider allometric proxies for average leaf biomass in eelgrass shoots. In getting these surrogates, we aggregate allometric projections of individual leaf biomass conforming a shoot. For comparison, we consider individual leaf biomass surrogates produced by the

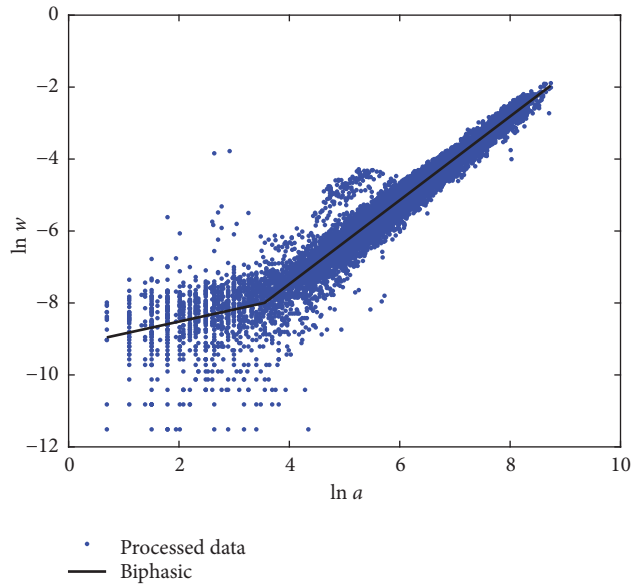


FIGURE 2: Spread about the biphasic mean response function $E_B(v | u)$. Compared with the plot in Figure 1, we observe that the biphasic choice offers a better account of variability of the log transformed response than the TAMA alternate.

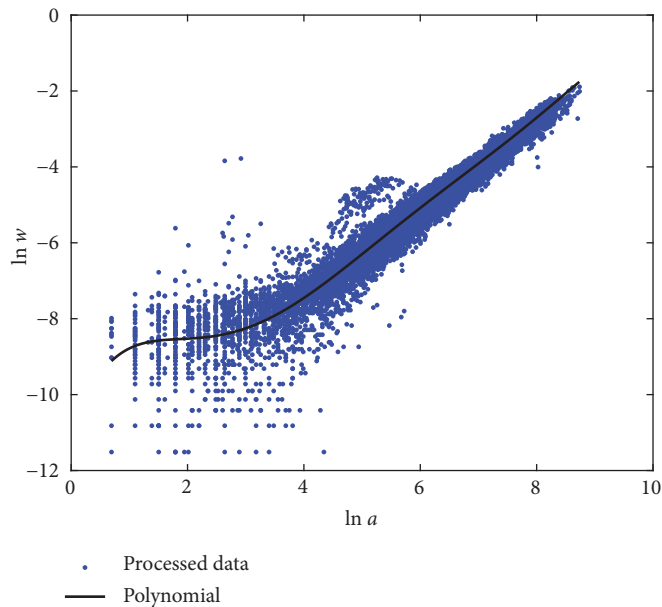


FIGURE 3: Spread about the polynomial mean response function $E_P(v | u)$. Compared with the plot in Figure 1, we can be aware that, as opposite to the TAMA protocol, the polynomial scheme offers a consistent account of curvature.

TABLE 4: Comparison of reproducibility strength statistic for the biphasic mean response $E_{B\delta}(w | a)$ as calculated by different forms of the correction factor for bias of retransformation of the regression error ϵ . The $\rho(\delta(\epsilon))$ symbol stands for the concordance correlation value associated with $\delta(\epsilon)$.

CF	$\rho(\delta(\epsilon))$	R^2	SEE	TRE	ASE	MPE	MPSE
$\delta_D(\epsilon)$	0.9664	0.9256	0.0049	-9.2196	4.48e-13	0.8133	31.2839
$\delta_{ZT}(\epsilon)$	0.9687	0.9319	0.0047	-7.5530	1.8359	0.7780	31.4388
$\delta_n(\epsilon)$	0.9727	0.9464	0.0042	1.9516	12.3058	0.6903	33.9525

TABLE 5: Notation convention for the nonlinear heteroscedastic and homoscedastic models of (14) and (15), respectively.

Nonlinear model	Heteroscedastic	Eq.	Homoscedastic	Eq.
Regression equation	$w = \beta_\theta a^{\alpha_\theta} + a^\theta \epsilon$ $\epsilon = N(0, a^{2\theta} \sigma^2)$	(14), (E.1)	$w = \beta_o a^{\alpha_o} + \epsilon$ $\epsilon \sim N(0, \sigma^2)$	(15), (E.5)
Mean response function	$E_\theta(w a) = \beta_\theta a^{\alpha_\theta}$	(B.45)	$E_o(w a) = \beta_o a^{\alpha_o}$	(B.46)

TABLE 6: Assessment of models fitted in arithmetical space. This includes the nonlinear homoscedastic and heteroscedastic protocols.

Method	AIC	ρ	R^2	SEE	TRE	ASE	MPE	MPSE
Heteroscedastic	-92761.13	0.972	0.9467	0.0041	0.0311	26.1073	0.6882	50.2781
Homoscedastic	-81386.51	0.972	0.9467	0.0041	0.2954	28.4527	0.6881	51.7400

TABLE 7: Assessment of models in arithmetical space. This includes the nonlinear heteroscedastic and biphasic protocols.

Method	AIC	ρ	R^2	SEE	TRE	ASE	MPE	MPSE
Heteroscedastic	-92761.13	0.972	0.946	0.004	0.031	26.107	0.690	50.278
Biphasic	-96879.30	0.972	0.946	0.004	1.951	12.305	0.690	33.952

different projection methods. Table 8 compares resulting reproducibility strengths.

Results in [11] established that proxies derived from the TAMA protocol are inconsistent with observed values. This endorsed nonlinear regression in the direct scale as a requirement for reliability of allometric projections of mean leaf biomass in eelgrass shoots. But Table 8 shows that a curvature model fitted in geometrical space can offer proxies entailing similar predictive power to a nonlinear regression protocol. Plots in Figure 6 allow getting a glimpse of this assertion.

4. Discussion

The customary bivariate allometric model of (1) offers non-destructive surrogates for average leaf biomass in eelgrass shoots [11]. But there are methodological factors that could influence dependability. Views assert that parameter identification based on logarithmic transformations leads to biased projections [20–29]. But other practitioners clung to this approach as meaningful and necessary in allometric examination [30–39]. This going over suggests that surpassing this controversy amounts to considering curvature in geometrical space. For that aim, we proposed the generalized model of (3). Approaches such as direct nonlinear heteroscedastic regression, as well as biphasic and polynomial protocols in geometrical space [18], became logical resultants from this construct. For present data model selection validated maintenance of the analysis in geometrical space. Nevertheless, at an empirical level, addressed protocols produced allometric projections of individual leaf biomass of correspondent precision. This was also verified for concomitant projections of average leaf biomass in shoots. But, from a qualitative standpoint, the nonlinear regression protocol mainly contributed by identifying expected dependence of the variance on covariate. Moreover, Figure 7(a) depicts manifest differences in mean response trends between the polynomial fit and the nonlinear heteroscedastic model. Nonetheless, those in Figure 7(b) corresponding to this and

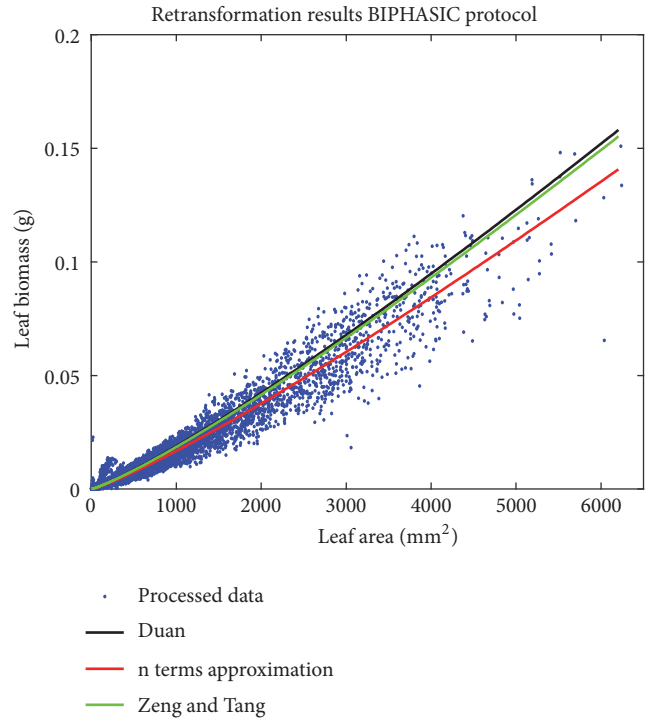


FIGURE 4: Spread about the biphasic mean response function $E_{BB}(w | a)$ in arithmetical space. Black lines relate to $\delta_D(\epsilon)$, green to $\delta_{ZT}(\epsilon)$, and red ones to $\delta_n(\epsilon)$.

the biphasic models differ but only barely. Then, a nonlinear regression scheme at best shaped a reasonable approximation of the mean response function resultant from curvature methods.

Differences in patterns of the biphasic and polynomial mean response functions relative to the nonlinear protocol exhibit that clinging to this last paradigm could impair detection of the true allometric relationship. Moreover, relying on direct nonlinear regression impairs identification of

TABLE 8: Comparison of reproducibility strengths of proxies of average leaf biomass in shoots resulting from the biphasic polynomial or nonlinear heteroscedastic protocols.

Method	ρ	R^2	SEE	TRE	ASE	MPE	MPSE
Biphasic	0.997	0.994	8.729e-04	2.408	2.229	3.486	5.080
Polynomial (m=6)	0.996	0.992	0.001	-2.456	-1.033	4.807	5.366
Heteroscedastic	0.997	0.994	7.722e-04	0.983	-0.726	3.084	5.645

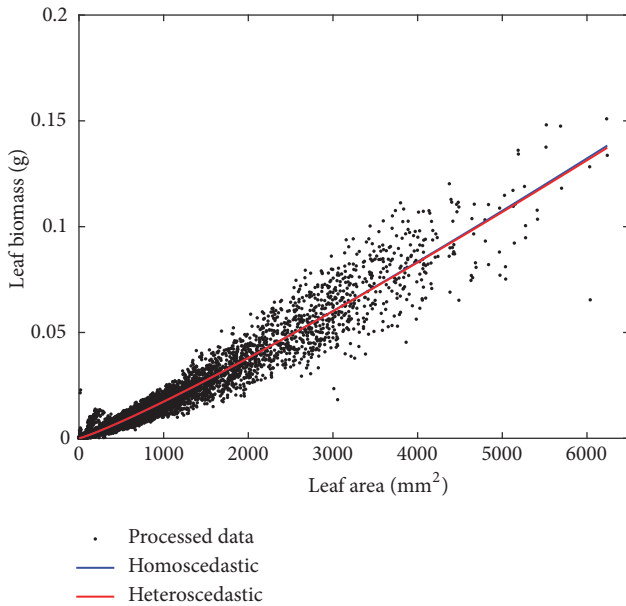


FIGURE 5: Spread about the mean response, curves $E_{\theta}(w | a)$ and $E_{\sigma}(w | a)$ associated with the nonlinear heteroscedastic and homoscedastic regression models of (14) and (15) one to one. The mean response $E_{\theta}(w | a)$ is shown in black lines and those corresponding to $E_{\sigma}(w | a)$ in red.

heterogeneity in the log transformed response as covariate changes. This further stresses on limitations of this device as a tool for allometric examination ([18, 30]). Oppositely, output of the selected biphasic model shown in Figure 2 suggests differentiation of growth patterns among smaller and larger leaves. Besides, the polynomial mean response in Figure 3 suggests a gradual transition between different growth phases. Thus, as opposed to direct nonlinear regression, a consideration of curvature in geometrical space could elucidate an inherent leaf growth pattern. This strengthens a judgement that the log transformation step, essential to traditional allometric examination, cannot be thrown away without losing relevant information ([33, 37]).

Mascaro et al. [18] conceived curvature in geometrical space as related to methodological factors of data gathering. But present examination corroborated consistency of curvature models for processed data. This suggests that manifestation of curvature is rather explained by intrinsic factors in leaf growth. Additionally, data processing failed to amend the heavy tails problem detected on Q-Q plots. This indicates departure of residuals from an assumed error structure. As a result, numerical values of the addressed correction factor forms turned out to be different, thus conveying ambiguity

in selection of suitable mean response of models fitted in geometrical space. This could entail the only advantage of nonlinear regression method over log-transformation-curvature paradigms. But, also for this analysis method, an inadequate postulation of inherent error structure can lead to significant bias [18]. It seems then reasonable considering that a suitable characterization of error structure could lead to robustness of built allometric proxies, even when they derive from crude data. Steering to an error structure different from what is assumed here is a worthwhile subject of further research.

When dealing with similar data we suggest taking into account recommendations that come up from this examination. First, it is highly advisable to perform a preliminary examination of the spread around the straight line in geometrical space resulting from the model of (1). If further statistical exploration confirms that linearity and assumed error structure are consistent with data, Huxley’s bivariate allometric model could suit. Otherwise, the arrangement of curvature, error structure, and correction factor form such as proposed here could be called into account for the analysis. The use of data cleaning procedures in order to achieve a better fit is controversial [11]. Instead of performing data processing a posteriori, it is highly advisable to rely on standardized data gathering procedures. This will prevent proliferation of inconsistent replicates that could exacerbate a heavy tails problem on Q-Q plots.

5. Conclusion

Failure to perform both a preliminary exploration of spread of log transformed allometric data and a sound evaluation of model adequacy could impair detecting a possible manifestation of curvature. As a consequence, the output of a traditional analysis method could set biased predictions of observed values. This circumstance could result in dismissal of a log transformation step in the analysis, giving way to contemplation of direct nonlinear regression as the only protocol to acquire reliable parameter estimates [11]. Results of this examination suggest that consideration of curvature in geometrical space as set by the model of (3) could offer dependable allometric proxies of average leaf biomass in eelgrass shoots.

From a general perspective, complexity as encompassed by the model of (3) can stand for curvature as conceived in allometric examination. Particularly, biphasic or polynomial protocols in geometrical space, as well as a direct nonlinear heteroscedastic regression model, derive as particular characterizations of this paradigm. Moreover all statistical models for accurate estimates of relative growth contemplated by

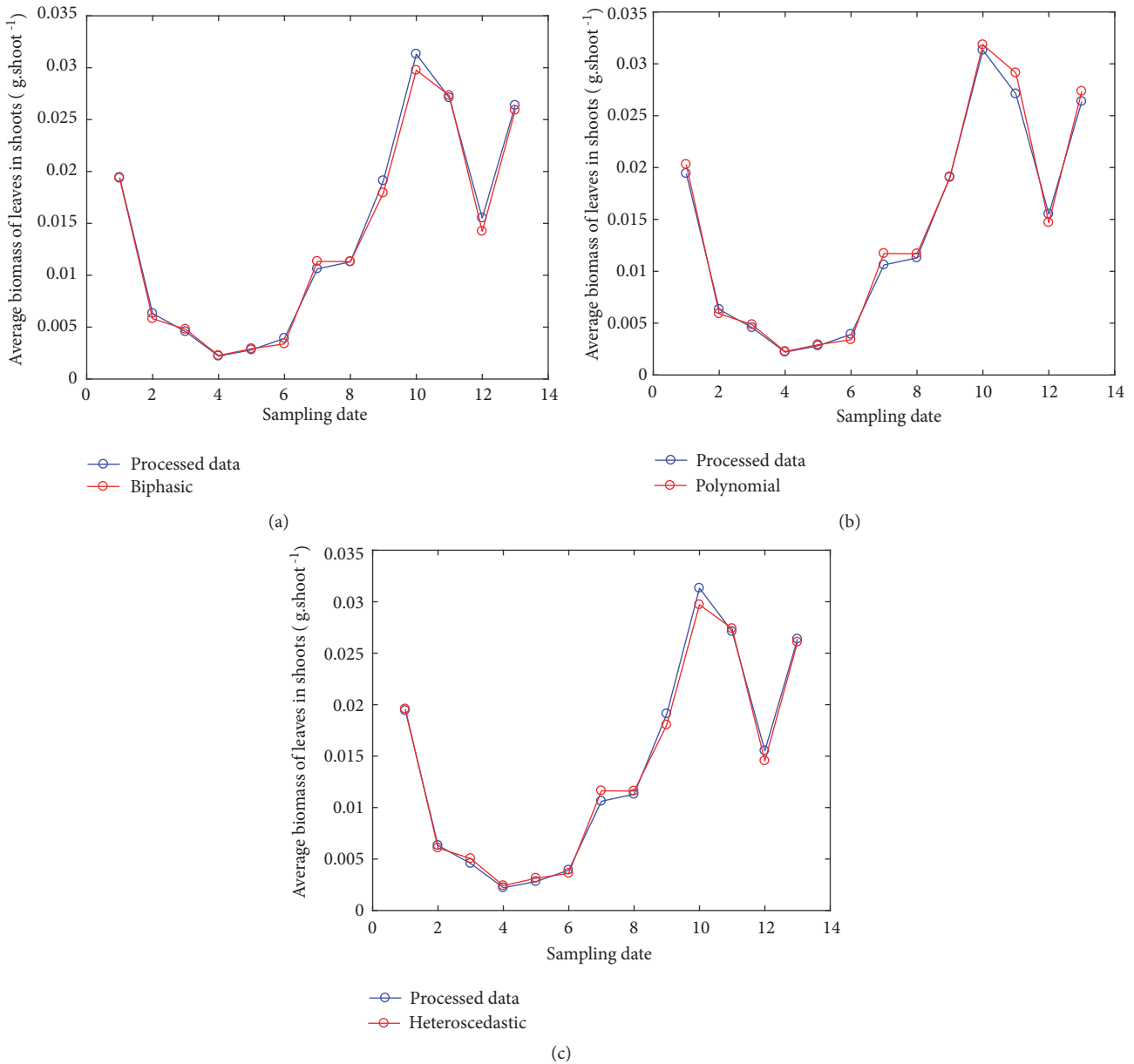


FIGURE 6: Average leaf biomass in shoots calculated from observed-processed data compared with their allometric projections. Projection lines resulting from the biphasic and polynomial models relied on the $\delta_n(\epsilon)$ form.

Bervian et, al., [68] can be also accomodated by the present generalization of the model of simple allometry of Huxley. But empirical convenience on its own does not validate adoption of this paradigm as a general tool. Certainly, the Weierstrass approximation theorem [69] backs a polynomial regression model as a reasonable identification device for the generalized allometric model expressed in geometrical space. But suitability of retransformation results will sensibly depend on correction factor form. And a mean function resulting from a polynomial fitted in geometrical space will not enable characterization of functions $\alpha(a)$ and $\beta(a)$ one to one. Furthermore, complexity of (3) could pose significant difficulties while attempting its identification through direct nonlinear regression methods. Needless to say, biological

interpretation of the scaling functions $\alpha(a)$ and $\beta(a)$ is also pending. A quest for efficient tools of nondestructive assessment of plant biomass units justifies addressing these examinations in a further research.

Appendix

A. Data Exploratory Analysis

This appendix presents an exploratory analysis of raw and processed data sets contemplated in this examination.

Table 9 describes the distribution pattern, in terms of quantiles, for a sample of 10412 measurements of eelgrass leaf weights and related areas taken over 13 months conforming

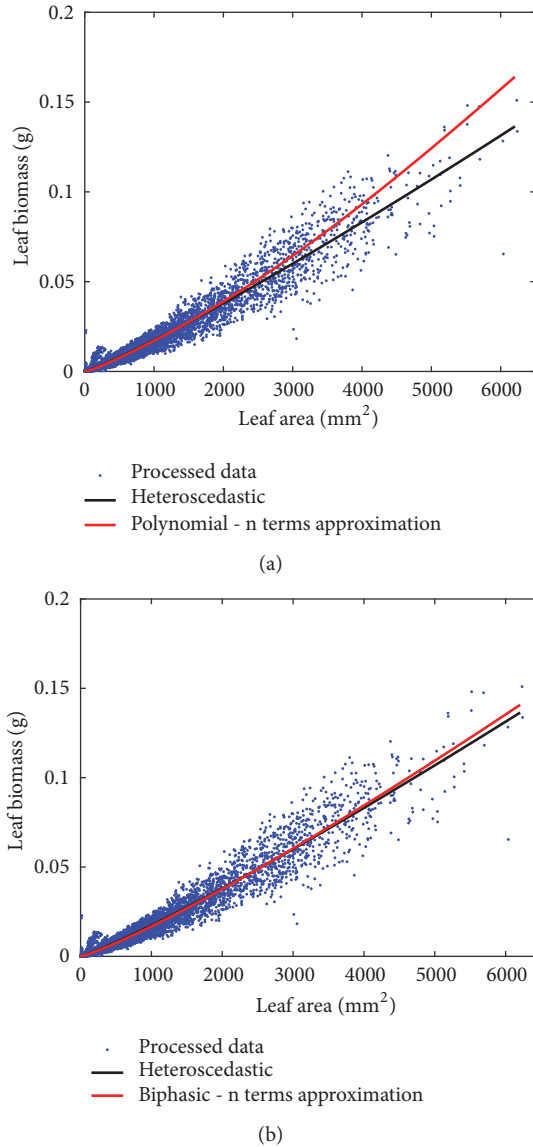


FIGURE 7: Trends of mean response functions calculated from the retransformed outputs of polynomial (a) and biphasic models (b) involving the $\delta_n(\epsilon)$ form of $\delta(\epsilon)$. We can observe a differentiation in trends relative to that of a power function fitted in directly by means of the nonlinear heteroscedastic protocol of (14). Although relative deviations manifest for the biphasic model, differences in mean response patterns are more clearly depicted for the polynomial choice.

the present raw data. The first four columns in the uppermost row label the minimum followed by quantiles of probability 0.1, 0.25, and 0.50. Correspondingly, the fifth column in the first row presents the sample mean followed by quantiles of probability 0.75 and 0.90 before the maximum. Second row presents leaf dry weight values (w) and third row shows corresponding areas a . Third and fourth rows present transformed $\ln w$ and $\ln a$ values one to one. Similarly, Table 10 shows the variation pattern of the 10023 observations resulting after applying to raw data a mean plus or minus two standard deviations outlier’s removal procedure [58] designed to remove replicates considered significantly discrepant from the mean

response function of a simple allometric model for a leaf dry weight response in terms of a leaf area covariate.

Comparing quantile values for the leaf dry weight (w) and corresponding area (a) variables reported in Tables 9 and 10 for crude and processed data respectively, we can ascertain that in spite of removing a large number of discrepant observations by data cleaning, both original and remnant sets show an equivalent distribution pattern. This similarity in distribution before and after data processing of the data can be better perceived in the Q-Q (quantile-quantile) graphs observed in Figure 8 for raw (a) and processed data (b), respectively.

Q-Q plots in Figure 8 compare the distributions (regardless what they are) of the leaf area a and associated dry weight w variables. The linear relationship observed between the quantiles of both variables underlines that both variables have similar distributions, although with different parameters. This similarity between distributions of both variables is observed for both sets of observations. The great similarity of the fitted straight lines for both sets of observations in Table 11 confirms that the referred distribution pattern does not change after data processing.

Figures 9 and 10 display boxplots for the 13 months’ long sampling scheme. We can learn that, for raw data, from month 2 to month 6, a reduction in the values of both leaf dry weight and linked area occurred. This is perhaps explained by an increase in temperature during those months. Processed data exhibits similar dynamics through time for these variables. Moreover, the overall variation patterns of raw and processed data, throughout the 13 months of sampling, are similar

B. Derivation of Regression Protocols

In this appendix, we first present notation convention for statistics related to the generalized bivariate model of (3). We then explain how the different regression protocols addressed in this examination derive from this paradigm. We also include notation convention for related statics.

B.1. Generalized Bivariate Allometric Model. A transformation $v = \ln w$ and $u(a) = \ln a$ of (3) establishes the generalized regression model in geometrical space given by (4). The term $v_C(a, u)$ stands for the corresponding mean response function. Using the customary notation convention, we represent $v_C(a, u)$ through symbol $E_C(v | u)$; that is,

$$E_C(v | u) = \ln \beta(a) + \alpha(a) u(a). \tag{B.1}$$

Furthermore, back-transformation $w = e^v$ of (4) leads to the result

$$w = e^{E_C(v|u)} e^\epsilon. \tag{B.2}$$

Thus, e^ϵ is understood as a multiplicative error term. Denoting by the symbol $E_{C\delta}(w | a)$ the corresponding mean response function in arithmetical space, from (B.1) and (B.2), we have

$$E_{C\delta}(w | a)(w | a) = \beta(a) a^{\alpha(a)} \delta(\epsilon) \tag{B.3}$$

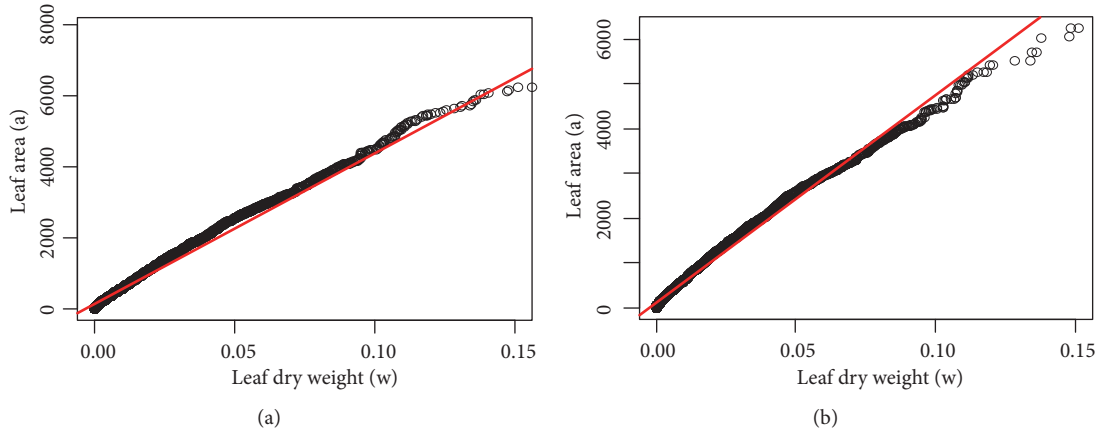


FIGURE 8: Q-Q plots (quantile-quantile) comparing distribution pattern of observations of eelgrass leaf area a and associated dry weight w for raw (a) and processed data (b).

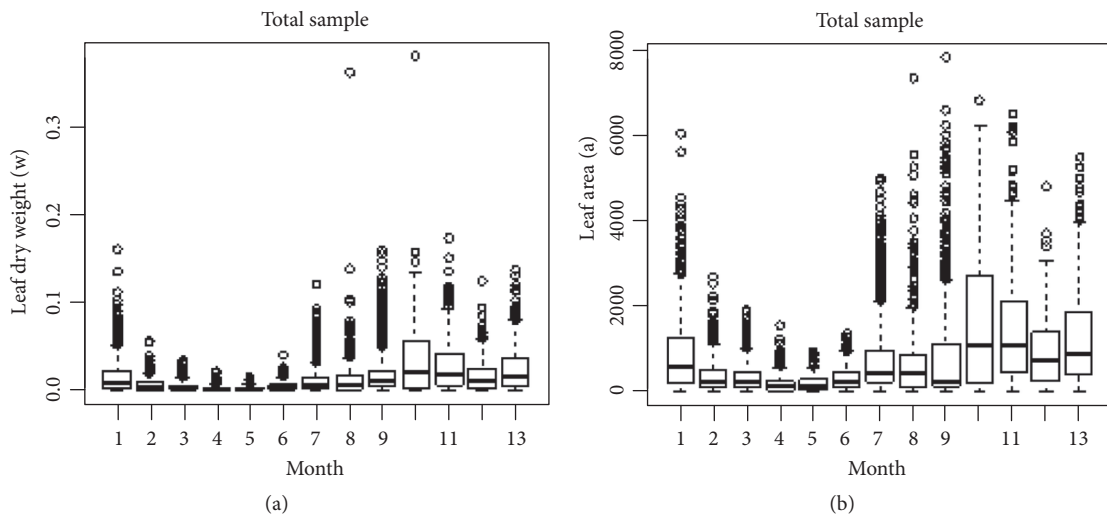


FIGURE 9: Boxplots for values of eelgrass leaf dry weight w (a) and linked area a (b) classified by month, as indicated in the horizontal axis. The data is from a sample of 10412 observations, before applying data processing.

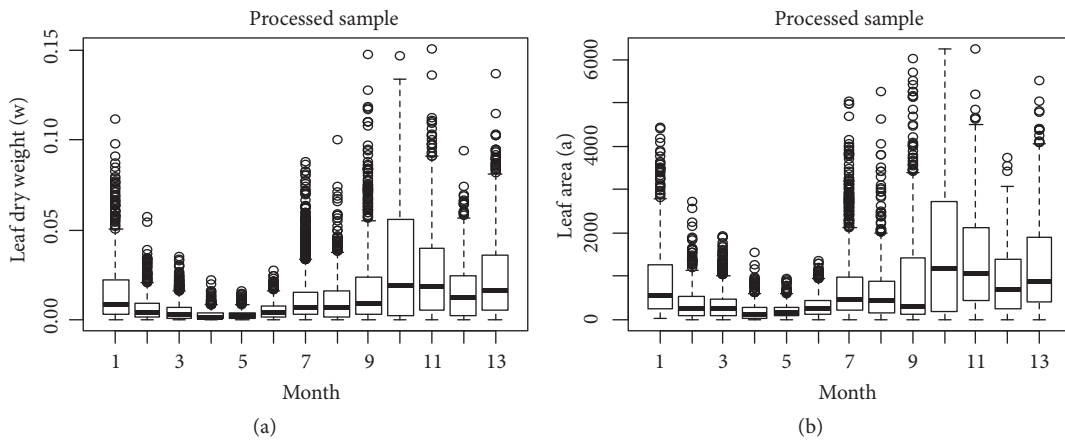


FIGURE 10: Boxplots for values of eelgrass leaf dry weight w (a) and linked area a (b) classified by month, as indicated in the horizontal axis. The data is from a sample of 10023 observations, remaining after applying an outlier removal procedure.

TABLE 9: Values of minimum, maximum, sample mean, and quantiles for measurements of eelgrass dry weight (w [gr]) and area (a [mm²]) (and their logarithms) of a sample of 10412 leaves before applying a data-cleaning procedure aimed to eliminate outliers.

	Min	0.10	0.25	0.50	Mean	0.75	0.90	Max
w [gr]	0.00001	0.00042	0.00154	0.00564	0.01293	0.01477	0.035443	0.38058
a [mm ²]	2.00	32.50	127.5	355.2	690.5	836.0	1859.00	7868.0
$\ln w$	-11.513	-7.7753	-6.4760	-5.1779	-5.4001	-4.2152	-3.33983	-0.9661
$\ln a$	0.6931	3.48124	4.8481	5.8728	5.6729	6.7286	7.527794	8.9706

TABLE 10: Values of minimum, maximum, sample mean, and quantiles for measurements of eelgrass dry weight (w [gr]) and area (a [mm²]) (and their logarithms) of a sample of 10023 leaves remaining after applying a data-cleaning procedure aimed to eliminate outliers.

	Min	0.10	0.25	0.50	Mean	0.75	0.90	Max
w [gr]	0.00001	0.00041	0.00144	0.00529	0.01211	0.01373	0.03381	0.15096
a [mm ²]	2.00	30.00	124.00	352.50	672.83	828.00	1835.80	6240.00
$\ln w$	-11.5129	-7.7994	-6.5431	-5.2419	-5.4593	-4.2878	-3.3868	-1.8907
$\ln a$.6931	3.4012	4.8203	5.8651	5.6518	6.7190	7.5152	8.7387

TABLE 11: Parameter estimates of straight lines fitted on raw (10412) and processed data (10023) and shown in the Q-Q plots of Figure 8.

	Estimate	Standard Error
Raw data		
Intercept	141.22	3.88
slope	42493.46	163.29
Processed data		
Intercept	111.20	2.50
Slope	46360.00	114.90

with $\delta(\epsilon) = E(e^\epsilon)$ interpreted as a correction factor of bias of retransformation of the regression error ϵ .

B.2. Traditional Analysis Method of Allometry (TAMA). The bivariate allometric model of (1) derives from the model of (3) setting the scaling parameters constant; that is, $\beta(a) = \beta$ and $\alpha(a) = \alpha$. The resultant regression model in geometrical space is given by (2). Corresponding mean response function will be denoted here by means of the symbol $E_T(v | u)$. Hence, from (2) we have

$$E_T(v | u) = \ln \beta + \alpha u. \tag{B.4}$$

Retransformation $w = e^v$ of (2) leads to the result

$$w = e^{E_T(v|u)} e^\epsilon. \tag{B.5}$$

And linked mean response function in arithmetical space denoted through $E_{T\delta}(w | a)$ becomes

$$E_{T\delta}(w | a) = \beta a^\alpha \delta(\epsilon) \tag{B.6}$$

B.3. Biphasic Model in Geometrical Space. Here, we explain the result of (8), as well as the notation convention for related statistics. In order to characterize a biphasic form of the model of (3), we introduce fixed values a_{min} and a_{max} , so that the covariate a takes values in a range $a_{min} \leq a \leq a_{max}$. We then conceive a_b as a fixed value of a satisfying $a_{min} \leq a_b \leq a_{max}$. And recalling the transformation $v = \ln w$ and

$u(a) = \ln a$, we take $u_b = \ln a_b$ as a breaking point for transition between two different phases of the variation of v . Moreover, in the model of (3), we let $\beta(a) = \beta_B(a)$ and $\alpha(a) = \alpha_B(a)$ given by

$$\alpha_B(a) = \sum_1^2 \vartheta_i(u(a)) \alpha_i \tag{B.7}$$

$$\beta_B(a) = \exp\left(\sum_1^2 \vartheta_i(u(a)) \ln \beta_i\right), \tag{B.8}$$

$$\vartheta_i(u(a)) = \begin{cases} (2-i)(1-H(u-u_b)) & \text{for } i=1 \\ (i-1)H(u-u_b) & \text{for } i=2 \end{cases} \tag{B.9}$$

where $H(x)$ is the Heaviside [60] function defined through

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0. \end{cases} \tag{B.10}$$

Then, the biphasic form of (3) is formally represented by

$$w = \beta_B(a) a^{\alpha_B(a)}. \tag{B.11}$$

Denoting by $v_B(a, u)$ the resulting form of $v_C(a, u)$, (5) yields

$$v_B(a, u) = \ln \beta_B(a) + \alpha_B(a) u(a). \tag{B.12}$$

Then, (B.7), (B.8), and (B.12) imply that

$$v_B(a, u) = \sum_1^2 \vartheta_i(u(a)) f_i(u(a)), \tag{B.13}$$

where

$$f_i(u(a)) = \ln \beta_i + \alpha_i u(a). \tag{B.14}$$

This explains (8). Since w as given by (3) is assumed to be a continuous function of a , this property ought to be satisfied by the biphasic model; that is, we require to set a condition, $f_1(u_b) = f_2(u_b)$. This leads to the equation

$$\ln \frac{\beta_1}{\beta_2} = (\alpha_2 - \alpha_1) u_b. \tag{B.15}$$

Moreover, $v_B(a, u)$ as given by (8) can be equivalently represented by

$$v_B(a, u) = \begin{cases} \ln \beta_1 + \alpha_1 u & u \leq u_b \\ \ln \beta_2 + \alpha_2 u & u > u_b, \end{cases} \tag{B.16}$$

with the continuity condition of (B.15).

Correspondingly, the regression model of (4) in its biphasic form becomes

$$v = v_B(a, u) + \epsilon. \tag{B.17}$$

As we explained in (4), ϵ is a residual error term conceived as a random variable distributed according to an unknown distribution ψ having mean μ and variance set by a function $\sigma^2(u)$ of the covariate u ; that is, $\epsilon \sim \psi(\mu, \sigma^2(u))$.

The form of the generalized mean response $E_C(v | u)$ for the biphasic model is denoted through $E_B(v | u)$ and becomes

$$E_B(v | u) = \sum_1^2 \vartheta_i(u(a)) f_i(u(a)). \tag{B.18}$$

It follows from (B.16) that the expression for $E_B(v | u)$ can be equivalently represented by

$$E_B(v | u) = \begin{cases} \ln \beta_1 + \alpha_1 u & u \leq u_b \\ \ln \beta_2 + \alpha_2 u & u > u_b. \end{cases} \tag{B.19}$$

Back-transformation $w = e^v$ of (B.17) leads to the result

$$w = e^{E_B(v|u)} e^\epsilon, \tag{B.20}$$

which can be written in the form

$$w = \beta_B(a) a^{\alpha_B(a)} e^\epsilon. \tag{B.21}$$

Then, the mean response function in arithmetical space $E(w | a)$ becomes

$$E_{B\delta}(w | a) = \beta_b(a) a^{\alpha_b(a)} \delta(\epsilon). \tag{B.22}$$

B.4. Polynomial Model in Geometrical Space. In this section, we explain how (12) can be derived from the generalized structure of model of (3). We also set the notation convention for related statistics. For these aims, we begin by considering polynomials $\phi_\beta(a)$ and $\phi_\alpha(a)$ given by

$$\phi_\beta(a) = \sum_0^n \beta_k u(a)^k \tag{B.23}$$

$$\phi_\alpha(a) = \sum_0^n \alpha_k u(a)^k, \tag{B.24}$$

where, for $1 \leq k \leq m$, α_k and β_k stand for coefficients. Now, let $\alpha(a) = \alpha_p(a)$ and $\beta(a) = \beta_p(a)$, where

$$\alpha_p(a) = \phi_\alpha(a) \tag{B.25}$$

$$\beta_p(a) = e^{\phi_\beta(a)}. \tag{B.26}$$

This way, we obtain a representation of the model of (3) in the form

$$w = \beta_p(a) a^{\alpha_p(a)}. \tag{B.27}$$

Now, denoting by means of $v_p(a, u)$ the associated form of $v_C(a, u)$, according to (5), we have

$$v_p(a, u) = \ln \beta_p(a) + \alpha_p(a) u(a). \tag{B.28}$$

From (B.23) through (B.28), we obtain

$$v_p(a, u) = \sum_0^n \beta_k u(a)^k + \sum_0^n \alpha_k u(a)^{k+1}. \tag{B.29}$$

Rearranging, we ascertain that $v_B(u, a)$ takes on the polynomial form

$$v_p(a, u) = \sum_0^m p_k u(a)^k \tag{B.30}$$

where $m = n + 1$, and

$$\begin{aligned} p_0 &= \beta_0, \\ p_k &= \beta_k + \alpha_{(k-1)} \quad \text{for } 1 \leq k \leq m - 1, \\ p_m &= \alpha_m. \end{aligned} \tag{B.31}$$

This explains the result of (12). Correspondingly, a polynomial characterization of the regression model of (4) takes the form

$$v = v_p(a, u) + \epsilon, \tag{B.32}$$

with ϵ being a residual error as described in (4). The form of the generalized mean response $E_C(v | u)$ for the polynomial characterization is denoted through $E_P(v | u)$. It becomes

$$E_P(v | u) = \sum_0^m p_k u(a)^k. \tag{B.33}$$

Back-transformation $w = e^v$ of (B.32) leads to

$$w = e^{E_P(v|u)} e^\epsilon. \tag{B.34}$$

And the mean response function in arithmetical space $E_{P\delta}(w | a)$ becomes

$$E_{P\delta}(w | a) = \beta_p(a) a^{\alpha_p(a)} \delta(\epsilon) \tag{B.35}$$

B.5. Nonlinear Regression Models. In this section, we explain how the nonlinear heteroscedastic model of (14) can be also associated with (3). We also explain the notation convention for related statistics. For these aims, we begin by noticing that $v_C(a, u)$ in (5) can be also written in the form

$$v_C(a, u) = \ln \beta + \alpha u + \lambda(a) \tag{B.36}$$

where $u = \ln a$ and

$$\lambda(a) = \ln \left(\frac{\beta(a)}{\beta} \right) + (\alpha(a) - \alpha) \ln a. \tag{B.37}$$

Now, if $\beta(a)$ and $\alpha(a)$ are as defined in (3), then (B.36) turns out to be nonlinear in geometrical space. Moreover, since from (B.36) we have

$$e^{v_C(a,u)} = \beta a^\alpha e^{\lambda(a)} \tag{B.38}$$

noticing that, from (4), $e^{v_C(a,u)} = \beta(a)a^{\alpha(a)}$, then solving for $e^{\lambda(a)}$ above yields

$$e^{\lambda(a)} = \frac{\beta(a)a^{\alpha(a)}}{\beta a^\alpha}. \tag{B.39}$$

Then, we have that the generalized curvature model of (3) can be also written in the form

$$w = \beta a^\alpha e^{\lambda(a)} \tag{B.40}$$

Thus, defining

$$\varepsilon(a) = \beta a^{\alpha-\theta} (e^{\lambda(a)} - 1), \tag{B.41}$$

(B.39) implies

$$w = \beta_\theta a^{\alpha_\theta} + a^\theta \varepsilon(a), \tag{B.42}$$

which suggests the nonlinear heteroscedastic regression model of (14). That is,

$$w = \beta_\theta a^{\alpha_\theta} + a^\theta \varepsilon, \tag{B.43}$$

which can be considered as a tool to analyze curvature in geometrical space. Particularly, by setting $\theta = 0$, we can consider the homoscedastic model

$$w = \beta_0 a^{\alpha_0} + \varepsilon, \tag{B.44}$$

where in (B.43) and (B.44) the error term ε is assumed to be normally distributed random variable having zero mean and homogeneous variance; that is, $\varepsilon \sim N(0, \sigma^2)$.

It turns out that the mean response function $E_\theta(w | a)$ associated with the heteroscedastic model becomes

$$E_\theta(w | a) = \beta_\theta a^{\alpha_\theta}. \tag{B.45}$$

Similarly, the mean response function $E_0(w | a)$ associated with the homoscedastic model becomes

$$E_0(w | a) = \beta_0 a^{\alpha_0}. \tag{B.46}$$

Finally, by setting $\alpha(a) = \alpha$ and $\beta(a) = \beta + c/a^\alpha$ in (3), a log transformation $v = \ln w$ and $u = \ln a$ leads to the nonlinear regression model in geometrical space:

$$v = \ln(\beta + ce^{-\alpha u}) + \alpha u + \varepsilon \tag{B.47}$$

with the error term assumed to be normally distributed with zero mean and constant variance σ^2 . Then, back-transformation $w = e^v$ of (B.47) to arithmetical space yields

$$w = (\beta a^\alpha + c) e^\varepsilon. \tag{B.48}$$

The mean response function in arithmetical space becomes

$$E(w | a) = (\beta a^\alpha + c) \delta(\varepsilon) \tag{B.49}$$

Packard [59] asserts that, commonly, any data set on arithmetical scale that is consistently described by a three-parameter power function will track a curved path when transformed to the logarithmic scale. Equations (B.45) through (B.47) provide a formal set-up accommodating such a statement.

C. Forms of the Correction Factor of Bias of Retransformation of the Regression Error

In this appendix, we explain the different characterizations of $\delta(\varepsilon)$, defined as a correction factor for bias of retransformation of the regression error ε introduced by (B.3) for $E_{C\delta}(w | a)$.

If ε is normally distributed with zero mean and constant variance σ^2 , that is $\varepsilon \sim N(0, \sigma^2)$, resulting form of $\delta(\varepsilon)$, denoted here by means of the symbol $\delta_B(\varepsilon)$, is given by

$$\delta_B(\varepsilon) = e^{\sigma^2/2}. \tag{C.1}$$

$\delta_B(\varepsilon)$ will be referred to as the Baskerville [51] form of $\delta(\varepsilon)$.

In case of a nonnormally distributed residual error term ε , Newman [52] asserts that $\delta(\varepsilon)$ must take the form provided by Duan's smearing estimate of bias [54]. Here, this form is correspondingly represented by means of the symbol $\delta_D(\varepsilon)$, and it is calculated by means of

$$\delta_D(\varepsilon) = \frac{\sum_1^m e^{\varepsilon_j}}{m}, \tag{C.2}$$

with ε_j standing for the j -th residual of the contemplated regression model. Nonetheless, this form of $\delta(\varepsilon)$ could produce bias overcompensation ([53, 55]). Moreover, $\delta_D(\varepsilon)$ corresponds to the sample mean of retransformed residuals. Therefore, whenever outliers occur, the actual central tendency of retransformed residuals data could not be represented by $\delta_D(\varepsilon)$. Under such a circumstance, a suitable form of the correction factor $\delta(\varepsilon)$ seems indefinite. Moreover, Zeng and Tang [57] suggest a distribution-free form of $\delta(\varepsilon)$ represented here by means of $\delta_{ZT}(\varepsilon)$ and given by

$$\delta_{ZT}(\varepsilon) = 1 + \frac{\sigma^2}{2}. \tag{C.3}$$

We notice that $\delta_{ZT}(\varepsilon)$ is actually an approximation to $\delta(\varepsilon)$ as it corresponds to a three terms' partial sum of the power series

expression of $\delta(\epsilon)$ assuming $E(\epsilon) = 0$. By the same token, we can consider an alternative approximation for $\delta(\epsilon)$; this is represented through the symbol $\delta_n(\epsilon)$ and given by an n -terms partial sum of the series representation of $E(e^\epsilon)$; that is,

$$\delta_n(\epsilon) = \sum_0^n \frac{E(\epsilon^k)}{k!}. \quad (\text{C.4})$$

The value of n leading to the highest concordance correlation coefficient value between $E_{CS}(w | a)$ projections and observed values sets the explicit form of $\delta_n(\epsilon)$ in (C.4).

D. Fitting Results of Regression Models in Geometrical Space

This appendix presents fitting results of the geometrical space protocols derived from the generalized model of (3). This includes TAMA and biphasic and polynomial schemes.

D.1. Fitting Results of the TAMA Protocol. Since we are dealing with bivariate data (w_i, a_i) , in arithmetical scale, according to the TAMA protocol, we have to consider pairs (v_i, u_i) , where $v_i = \ln w_i$ and $u_i = \ln a_i$ for $i = 1, 2, \dots, n$. Moreover, the linear regression model of (2) becomes

$$v_i = \ln \beta + \alpha u_i + \epsilon_i, \quad (\text{D.1})$$

where the error term ϵ_i is assumed to be normally distributed with zero mean and constant variance σ^2 . Hence, the response v_i has normal distribution $N(\mu_i, \sigma^2)$ with constant variance σ^2 and a mean $\mu_i(\alpha, \beta, u_i)$ expressed as a function of the covariate u_i ; namely,

$$\mu_i(\alpha, \beta, u_i) = \ln \beta + \alpha u_i. \quad (\text{D.2})$$

Identification method contemplated here relies on finding parameter estimates that maximize the log-likelihood function $l(\beta, \alpha, \sigma)$. This is given by

$$l(\beta, \alpha, \sigma) = -\frac{n}{2} \log(2\pi) - \sum_1^n \log(\sigma) - \frac{1}{2} \sum_1^n \left(\frac{v_i - \ln \beta - \alpha u_i}{\sigma} \right)^2. \quad (\text{D.3})$$

Fitting results of the linear model of (D.1) through (D.3) to raw data appear in Table 12.

Figure 11(a) displays a biased distribution of residuals around the zero line. The normal Q-Q plot in Figure 11(b) can be considered as a visual test of goodness of fit. In this case, this provides primary evidence against normality of residuals. Indeed, we can learn of a heavy-tails pattern in this plot. Moreover, an Anderson-Darling [67] goodness-of-fit test resulted in a test statistic of 246.7 and in a p value of $<2.2e-16$, which confirms lack of normality of residuals.

Figure 12 shows spread about TAMA's mean response $E_{T\delta}(w | a)$ as produced by different forms of the correction factor $\delta(\epsilon)$. Plot suggests moderate bias about $E_{T\delta}(w | a)$ for

leaves of small-to-medium-sized leaves. Nevertheless, deviations between projected and processed data are notorious for larger area values. Thus, spread in Figure 12 suggests that the TAMA protocol is unsuited for the analysis of present data.

D.2. Fitting Results of the Biphasic Model. As before, for data pairs (v_i, u_i) , with $v_i = \ln w_i$ and $u_i = \ln a_i$, for $i = 1, 2, \dots, n$, (B.13) and (B.17) yield the regression model

$$v_i = (\vartheta_1(u_i) \ln \beta_1 + \vartheta_2(u_i) \ln \beta_2) + (\vartheta_1(u_i) \alpha_1 + \vartheta_2(u_i) \alpha_2) u_i + \epsilon_i, \quad (\text{D.4})$$

with the additive errors $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ assumed to be normally distributed with zero mean and constant variance σ^2 . Moreover, the response v_i has normal distribution $N(\mu_i, \sigma^2)$ with constant variance σ^2 and a mean $\mu_i(\alpha, \beta, u_i)$ expressed as a function of the covariate u_i ; namely,

$$\vartheta_k(u_i) = \begin{cases} (2-k)(1-H(u_i-u_b)) & k=1 \\ (k-1)H(u_i-u_b) & k=2, \end{cases} \quad (\text{D.5})$$

and, hence, the response v_i as given by (D.4) is normally distributed having mean $\mu_i(u_i, \pi)$ with π standing for the parameter vector $(\beta_1, \beta_2, \alpha_1, \alpha_2, u_b)$ and given by

$$\mu_i(u_i, \pi) = \left(\ln \beta_1 + H(u_i - u_b) \ln \left(\frac{\beta_2}{\beta_1} \right) \right) + (\alpha_1 + H(u_i - u_b)(\alpha_2 - \alpha_1)) u_i \quad (\text{D.6})$$

and variance σ^2 . Therefore, the associated log-likelihood function $l(\beta_1, \beta_2, \alpha_1, \alpha_2, u_b, \sigma)$ becomes

$$l(\beta_1, \beta_2, \alpha_1, \alpha_2, u_b, \sigma) = -\frac{n}{2} \log(2\pi) - \sum_1^n \log(\sigma) - \frac{1}{2} \sum_1^n \left(\frac{w_i - \mu_i(u_i, \beta_1, \beta_2, \alpha_1, \alpha_2, u_b)}{\sigma} \right)^2. \quad (\text{D.7})$$

Table 13 Presents fitting results of the biphasic regression model of (D.4) through (D.7). It is worth mentioning that in this case we are performing a nonlinear fit in geometrical space.

Comparing the behavior of the residuals of the TAMA and biphasic models from Figure 13, we can learn of a relative stabilization of variability of the residues. Nevertheless, the biphasic normal Q-Q plot still shows heavier tails than those expected for a normal distribution. This amounts to primary evidence against normality of the residuals. An Anderson-Darling goodness-of-fit test [67] that resulted in a test statistic of 254.5 and a p value $<2.2e-16$ confirms a lack of normality for the involved residuals. Nevertheless, compared with the TAMA fit, the normal Q-Q plot for the biphasic model reveals a larger region where data track a normal distribution pattern.

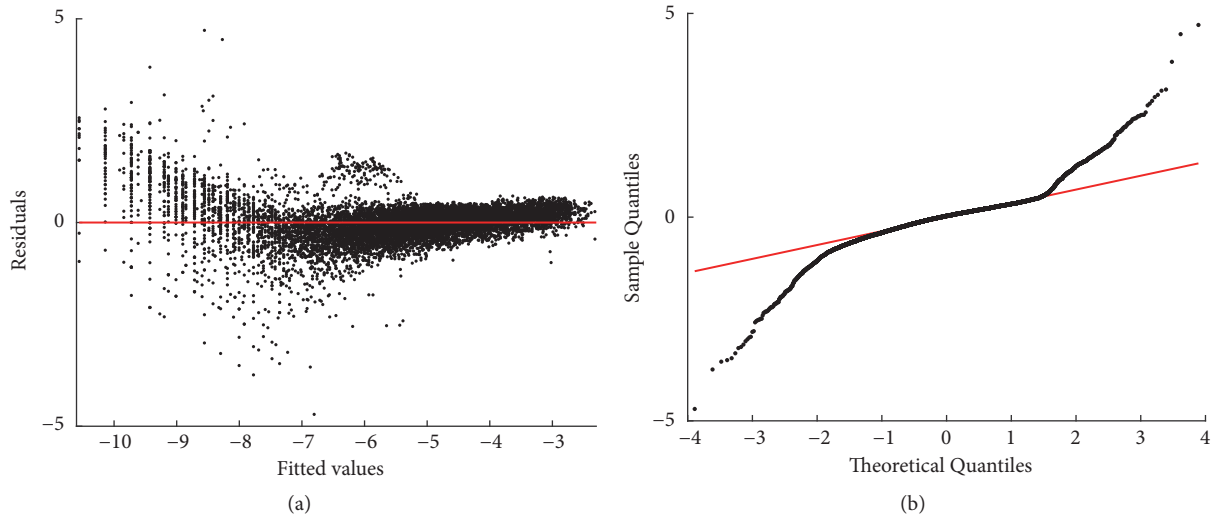


FIGURE 11: Residuals and normal probability plots ((a) and (b), resp.) resulting from fitting the linear regression model of the TAMA protocol (cf. (D.1) through (D.3)). A biased distribution of residuals around the zero line is depicted. Also, the normal Q-Q plot shows heavier tails than those expected for a normal distribution.

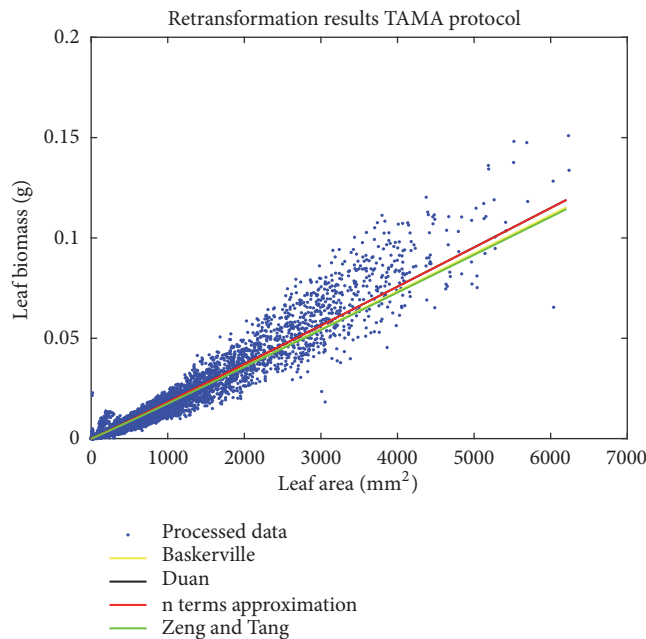


FIGURE 12: Spread about the mean response function $E_{T\delta}(w | a)$ obtained by retransforming the linear model of the TAMA protocol and using the addressed forms of the correction factor $\delta(\epsilon)$ (Table 1). Black lines associated with $E_{TD}(w | a)$ green lines go with $E_{TZT}(w | a)$, and red ones go with $E_{Tn}(w | a)$ and in yellow we show those for $E_{TB}(w | a)$. We can learn of a biased spread of observed values about the mean response functions $E_{T\delta}(w | a)$ even after data processing.

D.3. Fitting Results of the Polynomial Model. We now explain the statistical analysis involved in the identification of the polynomial model of (B.27). Again, for data pairs (v_i, u_i) , with $v_i = \ln w_i$ and $u_i = \ln a_i$, for $i = 1, 2, \dots, n$, from (B.29) and (B.30), we get the regression model

$$v_i = \sum_0^m p_k u_i^k + \epsilon_i, \tag{D.8}$$

where the additive errors $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ are independent and identically distributed, with distribution $N(0, \sigma^2)$. Hence, the response v_i above is normally distributed having mean $\mu_i(u_i, \pi)$ with π standing for the parameter vector (p_0, p_1, \dots, p_m) ; namely,

$$\mu_i(u_i, \pi) = \sum_0^m p_k u_i^k. \tag{D.9}$$

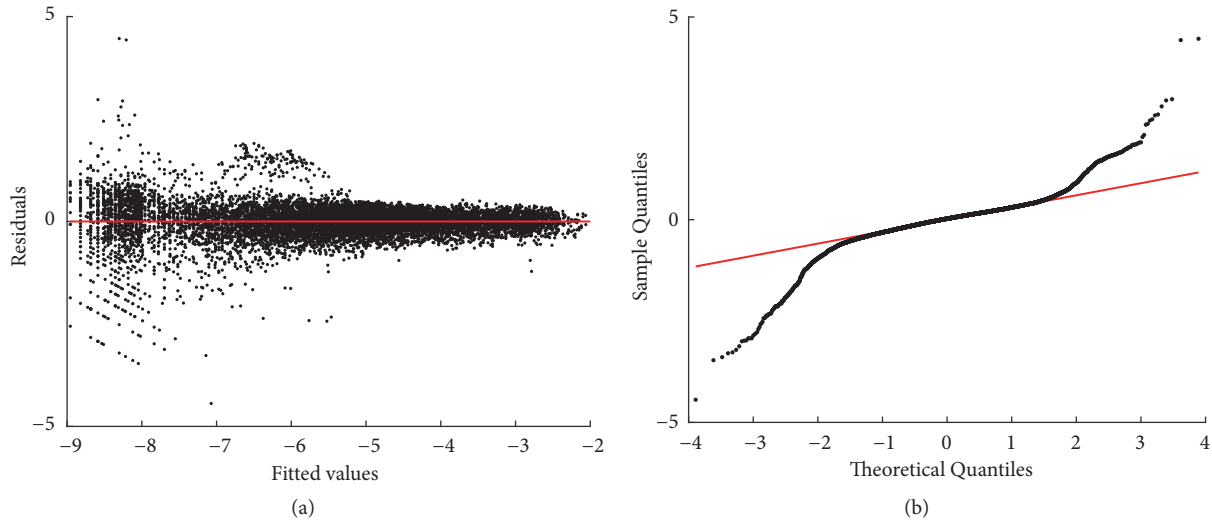


FIGURE 13: Residuals and normal probability plots ((a) and (b) one to one) for the fitting of the biphasic regression model to processed data. Compared with the TAMA fit, distribution of residuals about the zero line improved. Moreover, the Q-Q plot shows heavier tails than those expected for a normal distribution.

TABLE 12: Fitting results of regression model of the TAMA protocol (cf. (D.1) through (D.3)). CI stands for confidence interval, RSE stands for standard error of residuals, RMS means multiple R-squared, and ARS is adjusted R-squared.

Statistics of residuals					
Minimum	IQ	Median	3Q	Maximum	
-4.7088	-0.2344	0.0281	0.2244	4.7169	
Parameters	Estimate	Std. Error	t value	Pr(> t)	C.I. (95%)
α	1.028222	0.003335	308.27	<2e-16	(1.021684, 1.03476)
$\ln \beta$	-11.270569	0.019534	-576.94	<2e-16	(-11.308861, -11.23228)
RSE σ	0.5131 on 10021 df				
MRS	0.9046				
ARS	0.9046				
F-statistic	9.504e+04 on 1 and 10021 df				
p-value	< 2.2e-16				

TABLE 13: Fitting results of the biphasic regression model of (D.4) through (D.7) to processed data. CI stands for confidence interval and RSE for standard error of residuals.

Residual statistics					
Minimum	IQ	Median	3Q	Maximum	
-4.4406	-0.1948	0.0240	0.2072	4.4626	
Parameters	Estimate	Std. Error	t value	Pr(> t)	CI (95%)
$\ln \beta_1$	-9.190321	0.052812	-174.0	<2e-16	(-9.2938, -9.0868)
$\ln \beta_2$	-11.9381	0.093895	-127.1	<2e-16	(-12.1221, -11.7541)
u_b	3.550709	0.032744	108.4	<2e-16	(3.4865, 3.6149)
α_1	0.336339	0.020042	16.8	<2e-16	(0.2971, 0.3756)
α_2	1.164558	0.004248	274.1	<2e-16	(1.1562, 1.1729)
RSE σ	0.452947 on 10018 df				

Therefore, the associated log-likelihood function $l(p_0, p_1, \dots, p_m, \sigma)$ becomes

$$l(p_0, p_1, \dots, p_m, \sigma) = -\frac{n}{2} \log(2\pi) - \sum_1^n \log(\sigma) - \frac{1}{2} \sum_1^n \left(\frac{w_i - \mu_i(u_i, \pi)}{\sigma} \right)^2 \quad (D.10)$$

Fitting results of the polynomial regression model of (D.8) through (D.10) are shown in Table 14.

Reviewing fitting results of the polynomial model, we observe similar results to those we found earlier for the fit of the biphasic model. Moreover, Figure 14 displays similar patterns to those shown in Figure 13. Also for this fit, an Anderson-Darling normality test [67] delivered a test statistic 257.1 and a p value <2.2e-16 that yields remarkable

TABLE 14: Fitting results of the polynomial regression model. CI stands for confidence interval, RSE stands for standard error of residuals, RMS means multiple R-squared, and ARS is adjusted R-squared.

		Residual statistics				
Minimum		1Q	Median	3Q	Maximum	
-4.4207		-0.1947	0.0248	0.2072	4.5703	
Parameters	Estimate	Std. Error	t value	Pr(> t)	CI (95%)	
p_0	-11.080	0.4136	-26.783	<2e-16	(-11.889e+1, -10.267e+1)	
p_1	4.602	0.7411	6.209	5.54e-10	(3.1489, 6.0543)	
p_2	-3.2120	0.5027	-6.390	1.74e-10	(-4.1976, -2.2267)	
p_3	1.0620	0.1671	6.355	2.17e-10	(0.7344, 1.3894)	
p_4	-0.1678	0.02911	-5.765	8.43e-09	(-0.22483, -0.11072)	
p_5	0.01287	0.002549	5.047	4.57e-07	(0.0078689, 0.017863)	
p_6	-3.854e-04	8.854e-05	-4.353	1.36e-05	(-0.00055897, -0.00021186)	
RSE σ	0.4538 on 10016 df					
MRS	0.9254					
ARS	0.9254					
F-statistic	2.071e+4 on 6 and 10016 df					
p-value	< 2.2e-16					

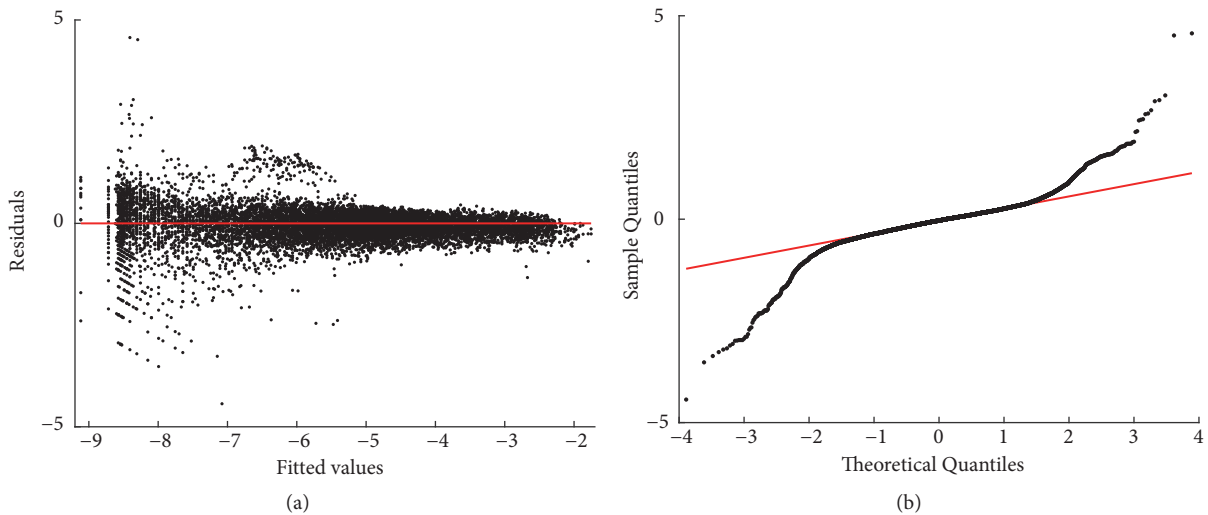


FIGURE 14: Residuals and normal probability plots ((a) and (b) one to one) for the fitting of the polynomial regression model of (D.8) through (D.10). Compared with TAMA results, distribution of residuals about the zero line improved. Normal Q-Q plot reveals a large plateau, where residuals conform to a normal distribution, but anyhow heavier tails than expected for such a pattern remain.

evidence against normality of residuals. Again, comparing with TAMA results, distribution of residuals about the zero line improved. Moreover, the Q-Q plot shows a larger plateau, where residuals agree to a normal distribution pattern.

E. Fitting Results of Direct Nonlinear Regression Models

This appendix presents the exploration of curvature by means of direct nonlinear regression methods. For that aim, we present fitting results of the nonlinear heteroscedastic model of (14). For comparison, we include fitting results of the homoscedastic case of (15).

Since we deal with data pairs (w_i, a_i) for the heteroscedastic case, the regression equation to be considered turns out to be

$$w_i = \beta_\theta a_i^{\alpha_\theta} + a_i^\theta \epsilon_i, \tag{E.1}$$

where ϵ_i is a normally distributed zero mean random variable with standard deviation σ . Then the response w_i is normally distributed having mean function $\mu_i(a_i, \alpha_\theta, \beta_\theta)$ given by

$$\mu_i(a_i, \alpha_\theta, \beta_\theta) = \beta_\theta a_i^{\alpha_\theta}. \tag{E.2}$$

And a variance set as a function $\sigma_i^2(a_i)$ of covariate a_i is, namely,

$$\sigma_i^2(a_i) = \sigma^2 a_i^{2\theta}. \tag{E.3}$$

TABLE 15: Maximum likelihood estimates of parameters for the heteroscedastic and homoscedastic regression models ((E.1) and (E.5) one to one). CI stands for confidence interval.

Heteroscedastic fit			
Parameter	Estimate	Std. Error	CI (95%)
α_θ	1.1298	0.003414	(1.1230, 1.1366)
β_θ	7.073e-06	1.804e-07	(6.7123e-06, 7.434e-06)
θ	0.4415	0.003378	(0.4347, 0.4483)
σ	1.951e-04	3.972e-06	(1.8716e-04, 2.0304e-04)
Homoscedastic fit			
Parameter	Estimate	Std. Error	CI (95%)
α_o	1.1365	0.003560	(1.1293, 1.1437)
β_o	6.724e-06	1.883e-07	(6.347e-06, 7.10e-06)
σ	0.004173	2.948e-05	(0.00411, 0.004229)

Again, fitting this regression model relied on a likelihood approach, that is, obtaining estimates for the parameters $\alpha_\theta, \beta_\theta, \theta$, and σ , which result in maximizing the log-likelihood function $l(\beta, \alpha_\theta, \beta_\theta, \sigma)$ expressed by

$$l(\beta, \alpha_\theta, \beta_\theta, \sigma) = -\frac{n}{2} \log(2\pi) - \sum_1^n \log(\sigma a_i^\theta) - \frac{1}{2} \sum_1^n \left(\frac{w_i - \beta_\theta a_i^{\alpha_\theta}}{\sigma a_i^\theta} \right)^2, \tag{E.4}$$

and in turn the linked homoscedastic regression model is

$$w_i = \beta_o a_i^{\alpha_o} + \epsilon_i \tag{E.5}$$

for $i = 1, 2, \dots, n$ and with error terms, $\epsilon_1, \epsilon_2, \dots, \epsilon_n$, independent and identically distributed normally with common mean μ set to zero and having an invariant standard deviation σ . The response w_i above is normally distributed with mean function $\mu_i(a_i, \alpha_o, \beta_o)$ given by

$$\mu_i(a_i, \alpha_o, \beta_o) = \beta_o a_i^{\alpha_o}. \tag{E.6}$$

Corresponding log-likelihood function $l(\alpha_o, \beta_o, \sigma)$ becomes

$$l(\alpha_o, \beta_o, \sigma) = -\frac{n}{2} \log(2\pi) - \sum_1^n \log(\sigma) - \frac{1}{2} \sum_1^n \left(\frac{w_i - \beta_o a_i^{\alpha_o}}{\sigma} \right)^2. \tag{E.7}$$

It can be ascertained from Table 15 that estimates for the parameters α_o and β_o for raw data are very similar to α_θ , as well as β_θ , for the heteroscedastic regression model of (E.1) and (E.4), respectively.

Confidence intervals in Table 15 show some overlap. Therefore, it can be established that parameter estimates differ but just barely. This statement is supported by the estimated value of ρ , Lin's concordance correlation coefficient between projected and observed leaf biomass values. For both regression models, homoscedastic and heteroscedastic, the estimate was $\rho = 0.972$. Likewise, plotting the mean response curves $w_i = \beta_\theta a_i^{\alpha_\theta}$ and $w_i = \beta_o a_i^{\alpha_o}$ on a dispersion diagram of weight and area values reveals that curves corresponding to parameter estimates fitted considering heteroscedasticity or not are very similar (Figure 5). Nevertheless, the heteroscedastic fit clearly identifies for the variance a variation pattern that grows in a power function as leaf area increases. Indeed, Figure 15 shows remarkable differences in scatter patterns of dry weight residuals against leaf area for the heteroscedastic and homoscedastic models ((a) and (b), resp.). This can be ascertained from values in Table 6, unveiling that the heteroscedastic model is selected by most agreement indices.

Nevertheless, normal Q-Q plots of residuals of the fits of the heteroscedastic model shown in Figure 16 reveal that in spite of data cleaning we can be aware of a severe problem of heavy tails. This points to consideration of a different error structure from the normal one assumed here.

Data Availability

Data will be provided by the corresponding author upon usage agreement.

Disclosure

This research was completed while Enrique Villa-Diharce (CVU:208964) was on sabbatical at CICESE.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

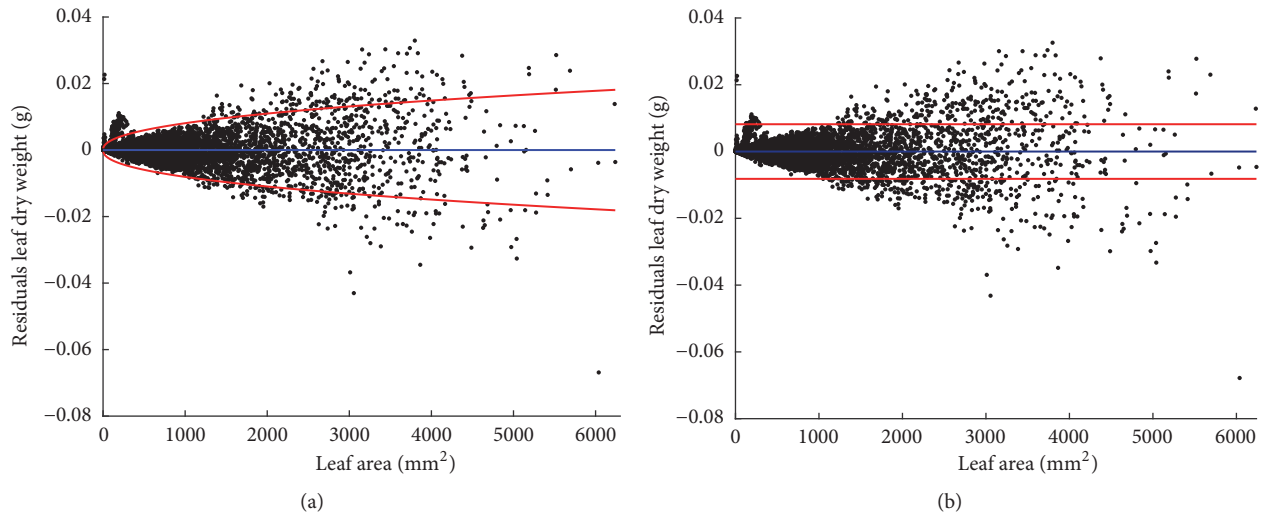


FIGURE 15: Diagram of dispersion of leaf dry weight residuals compared to leaf area for the fits heteroscedastic and homoscedastic models of (E.1) and (E.5), respectively. Region bounded by red lines determines (95%) confidence intervals for residuals.

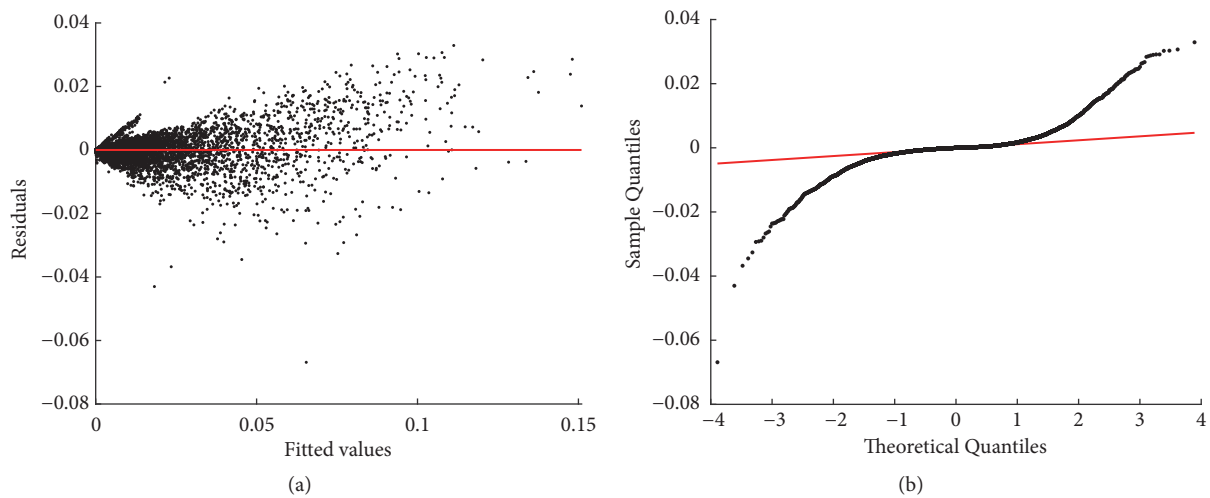


FIGURE 16: Residuals and normal probability plots ((a) and (b) one to one) for the fitting of the nonlinear heteroscedastic regression model of (E.1). We can observe a slightly biased distribution of residuals around the zero line. Also, normal Q-Q plot in (b) displays heavier tails than those expected for a normal distribution.

Acknowledgments

Funding for this research was institutionally granted. This was in part sponsored by an allowance from Mexico’s Consejo Nacional de Ciencia y Tecnología.

References

[1] J. S. Huxley, *Problems of Relative Growth*, Methuen, London, UK, 1932, <http://www.archive.org/details/problemsofrelati033234mbp>.

[2] M. Lu, J. S. Caplan, J. D. Bakker et al., “Allometry data and equations for coastal marsh plants,” *Ecology*, vol. 97, no. 12, pp. 3554–3554, 2016.

[3] V. M. Savage, J. F. Gillooly, W. H. Woodruff et al., “The predominance of quarter-power scaling in biology,” *Functional Ecology*, vol. 18, no. 2, pp. 257–282, 2004.

[4] N. P. Myhrvold, “Dinosaur metabolism and the allometry of maximum growth rate,” *PLoS ONE*, vol. 11, no. 11, Article ID e0163205, 2016.

[5] G. B. West and J. H. Brown, “The origin of allometric scaling laws in biology from genomes to ecosystems: Towards a quantitative unifying theory of biological structure and organization,” *Journal of Experimental Biology*, vol. 208, no. 9, pp. 1575–1592, 2005.

[6] M. E. J. Newman, “Power laws, Pareto distributions and Zipf’s law,” *Contemporary Physics*, vol. 46, no. 5, pp. 323–351, 2005.

[7] H. Samaniego and M. E. Moses, “Cities as organisms: allometric scaling of urban road networks,” *Journal of Transport and Land Use*, vol. 1, pp. 21–39, 2008.

[8] A. Maritan, R. Rigon, J. R. Banavar, and A. Rinaldo, “Network allometry,” *Geophysical Research Letters*, vol. 29, no. 11, pp. 3-1–3-4, 2002.

- [9] A. De Robertis and K. Williams, "Weight-length relationships in fisheries studies: The standard allometric model should be applied with caution," *Transactions of the American Fisheries Society*, vol. 137, no. 3, pp. 707–719, 2008.
- [10] H. Echavarría-Heras, C. Leal-Ramírez, E. Villa-Diharce, and N. R. Cazarez-Castro, "The effect of parameter variability in the allometric projection of leaf growth rates for eelgrass (*Zostera marina* L.) II: the importance of data quality control procedures in bias reduction," *Theoretical Biology and Medical Modelling*, vol. 12, no. 30, 2015.
- [11] H. Echavarría-Heras, C. Leal-Ramírez, E. Villa-Diharce, and N. Cazarez-Castro, "On the suitability of an allometric proxy for nondestructive estimation of average leaf dry weight in eelgrass shoots I: sensitivity analysis and examination of the influences of data quality, analysis method, and sample size on precision," *Theoretical Biology and Medical Modelling*, vol. 15, no. 4, 2018.
- [12] E. Solana-Arellano, H. Echavarría-Heras, C. Leal-Ramírez, and K.-S. Lee, "The effect of parameter variability in the allometric projection of leaf growth rates for eelgrass (*Zostera marina* L.)," *Latin American Journal of Aquatic Research*, vol. 42, no. 5, pp. 1099–1108, 2014.
- [13] A. Montesinos-López, E. Villa-Diharce, H. Echavarría-Heras, and C. Leal-Ramírez, "Correction to: Improved allometric proxies for eelgrass conservation," *Journal of Coastal Conservation*, vol. 23, no. 1, pp. 71–91, 2018.
- [14] P. S. Agutter and J. A. Tuszynski, "Analytic theories of allometric scaling," *Journal of Experimental Biology*, vol. 214, no. 7, pp. 1055–1062, 2011.
- [15] D. Hui and R. B. Jackson, "Uncertainty in allometric exponent estimation: a case study in scaling metabolic rate with body mass," *Journal of Theoretical Biology*, vol. 249, no. 1, pp. 168–177, 2007.
- [16] G. C. Packard, "Is logarithmic transformation necessary in allometry?" *Biological Journal of the Linnean Society*, vol. 109, no. 2, pp. 476–486, 2013.
- [17] J. Mascaro, C. M. Litton, R. F. Hughes, A. Uowolo, and S. A. Schnitzer, "Minimizing bias in biomass allometry: model selection and log-transformation of data," *Biotropica*, vol. 43, no. 6, pp. 649–653, 2011.
- [18] J. Mascaro, C. M. Litton, R. F. Hughes, A. Uowolo, and S. A. Schnitzer, "Is logarithmic transformation necessary in allometry? ten, one-hundred, one-thousand-times yes," *Biological Journal of the Linnean Society*, vol. 111, no. 1, pp. 230–233, 2014.
- [19] D. W. Thompson, *On Growth and Form*, Macmillan, New York, NY, USA, 1943.
- [20] R. J. Smith, "Allometric scaling in comparative biology: problems of concept and method," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 246, no. 2, pp. R152–R160, 1984.
- [21] D. L. Lovett and D. L. Felder, "Application of regression techniques to studies of relative growth in crustaceans," *Journal of Crustacean Biology*, vol. 9, no. 4, pp. 529–539, 1989.
- [22] G. S. Bales, "Heterochrony in brontothere horn evolution: allometric interpretations and the effect of life history scaling," *Paleobiology*, vol. 22, no. 4, pp. 481–495, 1996.
- [23] R. Lagergren, J.-E. Svensson, and J. A. E. Stenson, "Models of ontogenetic allometry in cladoceran morphology studies," *Hydrobiologia*, vol. 594, no. 1, pp. 109–116, 2007.
- [24] A. F. Sartori and A. D. Ball, "Morphology and postlarval development of the ligament of *Thracia phaseolina* (bivalvia: Thraciidae), with a discussion of model choice in allometric studies," *Journal of Molluscan Studies*, vol. 75, no. 3, pp. 295–304, 2009.
- [25] G. C. Packard, "Multiplicative by nature: logarithmic transformation in allometry," *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, vol. 322, no. 4, pp. 202–207, 2014.
- [26] G. C. Packard, "Quantifying the curvilinear metabolic scaling in mammals," *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, vol. 323, no. 8, pp. 540–546, 2015.
- [27] G. C. Packard, "Relative growth by the elongated jaws of gars: a perspective on polyphasic loglinear allometry," *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, vol. 326, no. 3, pp. 168–175, 2016.
- [28] G. C. Packard, "Is complex allometry in field metabolic rates of mammals a statistical artifact?" *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, vol. 203, pp. 322–327, 2017.
- [29] G. C. Packard, "The essential role for graphs in allometric analysis," *Biological Journal of the Linnean Society*, vol. 120, pp. 468–473, 2017.
- [30] J. Lai, B. Yang, D. Lin, A. J. Kerkhoff, K. Ma, and B. Bond-Lamberty, "The allometry of coarse root biomass: log-transformed linear regression or nonlinear regression?" *PLoS ONE*, vol. 8, no. 10, Article ID e77007, 2013.
- [31] C. P. Klingenberg, "Heterochrony and allometry: the analysis of evolutionary change in ontogeny," *Biological Reviews*, vol. 73, no. 1, pp. 79–123, 1998.
- [32] A. M. Nevill, S. Bate, and R. L. Holder, "Modeling physiological and anthropometric variables known to vary with body size and other confounding variables," *Yearbook of Physical Anthropology*, vol. 48, pp. 141–153, 2005.
- [33] A. J. Kerkhoff and B. J. Enquist, "Multiplicative by nature: why logarithmic transformation is necessary in allometry," *Journal of Theoretical Biology*, vol. 257, no. 3, pp. 519–521, 2009.
- [34] X. Xiao, E. P. White, M. B. Hooten, and S. L. Durham, "On the use of log-transformation vs. nonlinear regression for analyzing biological power laws," *Ecology*, vol. 92, no. 10, pp. 1887–1894, 2011.
- [35] E. P. White, X. Xiao, N. J. B. Isaac, and R. M. Sibly, "Methodological tools," in *Metabolic Ecology: A Scaling Approach*, R. M. Sibly, J. H. Brown, and A. Kodric-Brown, Eds., pp. 9–20, Wiley-Blackwell, Oxford, UK, 2012.
- [36] F. Ballantyne, "Evaluating model fit to determine if logarithmic transformations are necessary in allometry: a comment on the exchange between Packard (2009) and Kerkhoff and Enquist (2009)," *Journal of Theoretical Biology*, vol. 317, pp. 418–421, 2013.
- [37] D. S. Glazier, "Log-transformation is useful for examining proportional relationships in allometric scaling," *Journal of Theoretical Biology*, vol. 334, pp. 200–203, 2013.
- [38] K. J. Niklas and S. T. Hammond, "Assessing scaling relationships: uses, abuses, and alternatives," *International Journal of Plant Sciences*, vol. 175, no. 7, pp. 754–763, 2014.
- [39] J. F. Lemaitre, C. Vanpé, F. Plard, C. Pélabon, and J. M. Gaillard, "Response to Packard: make sure we do not throw out the biological baby with the statistical bath water when performing allometric analyses," *Biology Letters*, vol. 11, no. 6, p. 20150144, 2015.
- [40] R. E. Strauss and J. S. Huxley, "The study of allometry since Huxley," in *Problems of Relative Growth*, D. H. Thompson, Ed., Johns Hopkins University Press, Baltimore, UK, 1993.

- [41] R. J. Knell, "On the analysis of non-linear allometries," *Ecological Entomology*, vol. 34, no. 1, pp. 1–11, 2009.
- [42] C. D. MacLeod, "Assessing the shape and topology of allometric relationships with body mass: a case study using testes mass allometry," *Methods in Ecology and Evolution*, vol. 1, no. 4, pp. 359–370, 2010.
- [43] S. Mori, K. Yamaji, A. Ishida et al., "Mixed-power scaling of whole-plant respiration from seedlings to giant trees," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 4, pp. 1447–1451, 2010.
- [44] E. P. Snelling, D. A. Taggart, S. K. Maloney, A. P. Farrell, and R. S. Seymour, "Biphasic allometry of cardiac growth in the developing Kangaroo *Macropus fuliginosus*," *Physiological and Biochemical Zoology*, vol. 88, no. 2, pp. 216–225, 2015.
- [45] J. E. Duffy, "Ecology and eusociality in sponge-dwelling shrimp," in *Evolutionary Ecology of Social and Sexual Systems, Crustaceans as Model Organisms*, J. E. Duffy, and M. Thiel, Eds., pp. 1–520, Oxford University Press, New York, NY, USA, 2007.
- [46] A. M. Batterham and K. P. George, "Allometric modeling does not determine a dimensionless power function ratio for maximal muscular function," *Journal of Applied Physiology*, vol. 83, no. 6, pp. 2158–2166, 1997.
- [47] J. Chave, C. Andalo, S. Brown et al., "Tree allometry and improved estimation of carbon stocks and balance in tropical forests," *Oecologia*, vol. 145, no. 1, pp. 87–99, 2005.
- [48] K. J. Niklas, "Size-dependent allometry of tree height, diameter and trunk-taper," *Annals of Botany*, vol. 75, no. 3, pp. 217–227, 1995.
- [49] K. J. Niklas, "Mechanical properties of black locust (*Robinia pseudoacacia* L.) Wood. Size- and age-dependent variations in sap- and heartwood," *Annals of Botany*, vol. 79, no. 3, pp. 265–272, 1997.
- [50] G. C. Packard, "Why allometric variation in mammalian metabolism is curvilinear on the logarithmic scale," *Journal of Experimental Zoology Part A: Ecological and Integrative Physiology*, vol. 327, no. 9, pp. 537–541, 2017.
- [51] G. L. Baskerville, "Use of logarithmic regression in the estimation of plant biomass," *Canadian Journal of Forest Research*, vol. 2, no. 1, pp. 49–53, 1972.
- [52] M. C. Newman, "Regression analysis of log-transformed data: statistical bias and its correction," *Environmental Toxicology and Chemistry*, vol. 12, no. 6, pp. 1129–1133, 1993.
- [53] R. J. Smith, "Logarithmic transformation bias in allometry," *American Journal of Physical Anthropology*, vol. 90, no. 2, pp. 215–228, 1993.
- [54] N. Duan, "Smearing estimate: a nonparametric retransformation method," *Journal of the American Statistical Association*, vol. 78, no. 383, pp. 605–610, 1983.
- [55] R. W. Koch and G. M. Smillie, "Comment on 'River loads underestimated by rating curves' by R. I. Ferguson," *Water Resources Research*, vol. 22, no. 13, pp. 2121–2122, 1986.
- [56] W. G. Manning, "The logged dependent variable, heteroscedasticity, and the retransformation problem," *Journal of Health Economics*, vol. 17, no. 3, pp. 283–295, 1998.
- [57] W. S. Zeng, W. S. Zeng, and S. Z. Tang, "Bias correction in logarithmic regression and comparison with weighted regression for nonlinear models," *Nature Precedings*, 2011.
- [58] J. Miller, "Short report: reaction time analysis with outlier exclusion: bias varies with sample size," *The Quarterly Journal of Experimental Psychology Section A*, vol. 43, no. 4, pp. 907–912, 1991.
- [59] G. C. Packard, "Is non-loglinear allometry a statistical artifact?" *Biological Journal of the Linnean Society*, vol. 107, no. 4, pp. 764–773, 2012.
- [60] E. J. Berg, *Heaviside's Operational Calculus*, McGraw-Hill, New York, NY, USA, 2nd edition, 1936.
- [61] L. I-Kuei Lin, "A concordance correlation coefficient to evaluate reproducibility," *Biometrics*, vol. 45, no. 1, pp. 255–268, 1989.
- [62] G. B. McBride, "A Proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient," Tech. Rep. HAM2005-062, National Institute of Water & Atmospheric Research, Hamilton, New Zealand, 2005.
- [63] W. S. Zeng, L. J. Zhang, X. Y. Chen, Z. C. Cheng, K. X. Ma, and Z. H. Li, "Construction of compatible and additive individual-tree biomass models for *Pinus tabulaeformis* in China," *Canadian Journal of Forest Research*, vol. 47, no. 4, pp. 467–475, 2017.
- [64] W. S. Zeng and S. Z. Tang, "Goodness evaluation and precision analysis of tree biomass equations," *Sci Silvae Sinicae*, vol. 47, pp. 106–113, 2011.
- [65] W. Zeng, H. Duo, X. Lei et al., "Individual tree biomass equations and growth models sensitive to climate variables for *Larix* spp. in China," *European Journal of Forest Research*, vol. 136, no. 2, pp. 233–249, 2017.
- [66] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.
- [67] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain goodness of fit criteria based on stochastic processes," *Annals of Mathematical Statistics*, vol. 23, pp. 193–212, 1952.
- [68] G. Bervian, N. F. Fontoura, and M. Haimovici, "Statistical model of variable allometric growth: otolith growth in *Micropogonias furnieri* (Actinopterygii, Sciaenidae)," *Journal of Fish Biology*, vol. 68, no. 1, pp. 196–208, 2006.
- [69] B. Brosowski and F. Deutsch, "An elementary proof of the Stone-Weierstrass theorem," *Proceedings of the American Mathematical Society*, vol. 81, no. 1, pp. 89–92, 1981.