Utah State University DigitalCommons@USU

Biology Faculty Publications

Biology

2011

On the use of log-transformation vs. nonlinear regression for analyzing biological power-laws

X. Xiao

Ethan P. White *Utah State University*

M. B. Hooten

S. L. Durham

Follow this and additional works at: http://digitalcommons.usu.edu/biology_facpub

Recommended Citation

Xiao, X., White, E.P., M.B. Hooten, and S.L. Durham. 2011. On the use of logtransformation vs. nonlinear regression for analyzing biological powerlaws. Ecology 92: 18871894.

This Article is brought to you for free and open access by the Biology at DigitalCommons@USU. It has been accepted for inclusion in Biology Faculty Publications by an authorized administrator of DigitalCommons@USU. For more information, please contact dylan.burns@usu.edu.



- 1 On the use of log-transformation vs. nonlinear regression for analyzing biological
- 2 power-laws
- 3 Xiao Xiao^{1,2,3,a}, Ethan P. White^{1,2,b}, Mevin B. Hooten^{2,3,4,c}, and Susan L. Durham^{2,d}
- 4 ¹Biology Department, Utah State University, Logan UT 84322-5305, USA
- 5 ²Ecology Center, Utah State University, Logan UT 84322-5205, USA
- 6 ³Mathematics and Statistics Department, Utah State University, Logan, UT 84322-3900 USA
- 7 ⁴Current Address: U.S. Geological Survey Colorado Cooperative Fish and Wildlife Research
- 8 Unit, Department of Fish, Wildlife and Conservation Biology, Department of Statistics,
- 9 Colorado State University, Fort Collins CO 80523-1484, USA
- 10 ^aCorresponding author. E-mail: <u>xiao.xiao@usu.edu</u>
- 11 ^b Email: <u>ethan.white@usu.edu</u>
- 12 ^c Email: <u>Mevin.Hooten@colostate.edu</u>
- 13 ^d Email: <u>susan.durham@usu.edu</u>

14

Abstract

15 Power-law relationships are among the most well-studied functional relationships in biology. Recently the common practice of fitting power-laws using linear regression on log-transformed 16 17 data (LR) has been criticized, calling into question the conclusions of hundreds of studies. It has 18 been suggested that nonlinear regression (NLR) is preferable, but no rigorous comparison of 19 these two methods has been conducted. Using Monte Carlo simulations we demonstrate that the 20 error distribution determines which method performs better, with LR better characterizing data 21 with multiplicative lognormal error and NLR better characterizing data with additive, 22 homoscedastic, normal error. Analysis of 471 biological power-laws shows that both forms of 23 error occur in nature. While previous analyses based on log-transformation appear to be 24 generally valid, future analyses should choose methods based on a combination of biological 25 plausibility and analysis of the error distribution. We provide detailed guidelines and associated 26 computer code for doing so, including a model averaging approach for cases where the error 27 structure is uncertain. 28 Keywords: power-law, allometry, log-transformation, nonlinear regression, model comparison, 29 model averaging, parameter estimation 30 Introduction Power-law relationships of the form $y = ax^{b}$ are one of the most common patterns in biology. 31 32 They have been documented in a variety of different areas including the relationships between 33 body size, physiological rates and life history traits (Brown et al. 2004), the scaling between 34 body parts in morphology (Farlow *et al.* 1995), and the species-area relationship in biogeography 35 (Martin & Goldenfeld 2006). These fitted relationships have been used to test the validity of

36 biological theories (Brown *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics of extinct species (Farlow *et al.* 2004), to infer the characteristics (Farlow et al. 2004), to infer

- *al.* 1995), to assess the effect of evolutionary processes (Mortola & Limoges 2006), and to
 predict the consequence of habitat loss on biodiversity (Brooks *et al.* 2002).
- 39

Conventional analysis of power-law data uses the fact that log-transforming both sides of

the equation yields a linear relationship, $\log(y) = \log(a) + b \log(x)$, allowing log-transformed 40 data to be modeled using linear regression. However, it has been suggested that analysis on 41 42 logarithmic scales is flawed and that instead, analysis should be carried out on the original scale 43 of measurement using nonlinear regression (Fattorini 2007; Packard & Birchard 2008; Packard 44 2009; Caruso et al. 2010; Packard et al. 2010). If these claims are correct, the validity of decades 45 of published research in ecology, evolution and physiology would be called into question. 46 One of the fundamental differences between linear regression on log-transformed data 47 (hereafter, LR) and nonlinear regression on untransformed data (hereafter, NLR) lies in the 48 assumptions about how stochasticity manifests in the model (Gingerich 2000, Kerkhoff &

49 Enquist 2009). In NLR, it is assumed that the error term is normally distributed and additive on50 arithmetic scale (Bates & Watts 1988; Ritz & Streibig 2008):

,

51

$$y = ax^{b} + \varepsilon, \varepsilon \quad N(0, \sigma^{2})$$
⁽¹⁾

52 In contrast, LR assumes that error is normally distributed and additive on the logarithmic scale53 (Kerkhoff & Enquist 2009):

54
$$\log(y) = \log(a) + b \log(x) + \varepsilon, \varepsilon \quad N(0, \sigma^2)$$

55 which corresponds to log-normally distributed, multiplicative error on the arithmetic scale:

56
$$y = ax^b e^{\varepsilon}, \varepsilon \quad N(0, \sigma^2)$$
 (2)

57 For a single dataset, both assumptions cannot be correct. Violation of statistical assumptions of

58 error can lead to biased point estimates as well as inaccurate confidence intervals (Bates & Watts59 1988; see Figure 1 for an illustration).

60 Despite its importance in statistical model fitting, the issue of error distribution has been 61 largely ignored in discussions about best practices for fitting power-laws to data. While both 62 additive and multiplicative errors have been posited to occur in biological systems (Kerkhoff & 63 Enquist 2009; Packard 2009; Cawley & Janacek 2010), to our knowledge there has been no 64 systematic analysis that evaluates how NLR and LR estimation methods perform on different 65 error structures, or what form the error structure actually takes in biological systems. This is 66 surprising given the potential implications of these methodological issues for understanding 67 biological systems and the strong arguments regarding appropriate methods being made in the 68 literature (e.g., Packard 2009). Here we use Monte Carlo simulations to test the role of error 69 structure on the performance of the two methods across empirical parameter space. For cases 70 where the better model cannot be clearly determined, we develop an alternative estimation 71 method based on model averaging. Based on these results, detailed guidelines for the analysis of 72 biological power-laws, and computer code for their implementation, are provided.

73 **Error Distribution Determines the Best Method for Fitting Power-Laws** 74 Previous arguments regarding the performance of different methods have typically been based on 75 empirical data (but see Hui *et al.* 2010), despite the fact that the true data generating mechanism 76 is generally unknown. As such, previous studies provide little insight into the best 77 methodological approach. Monte Carlo simulation, where data are simulated from known 78 distributions, allows for a direct comparison between statistically estimated parameters and their 79 true values. Here we implement the Monte Carlo approach based on parameterizations from 80 empirical datasets so that our results will be valid for the range of empirically observed

parameter values. Results from these empirically motivated simulations were consistent withstandard Monte Carlo simulations based on hypothetical parameterizations (see Appendix A).

We compiled 471 datasets published between 2004 and 2008 in the fields of ecology and 83 84 evolution where significant power-law relationships were reported. The selected datasets were 85 either morphological or physiological allometries between organismal traits (for details of data 86 selection and the full list of datasets, see Appendix B). To generate the parameters for 87 simulations each empirical dataset was assumed to have: 1) a multiplicative log-normal error 88 structure, and a, b, and σ were estimated with LR (with σ estimated as the standard deviation of 89 the residuals); and 2) an additive normal error structure, and the parameters were estimated with 90 NLR. For each dataset, 10⁵ independent simulations were carried out using the estimated 91 parameters under the assumption of each error structure. Each simulated dataset was analyzed 92 with both LR and NLR, and the performance of the two methods was compared to determine 93 which method had the better point estimation of *a* and *b*, as well as more accurate confidence 94 interval (CI) coverage measured by the percentage of simulations where the true parameter value 95 falls within the estimated 95% CI. Only 239 datasets generated valid simulations under the 96 assumption of additive error (see Appendix C for technical details on the procedure of the 97 simulations). All simulations and analyses were carried out using R version 2.9.1 (R 98 Development Core Team 2009). The "nlrwr" package (Ritz & Streibig 2008) was used to 99 compute asymptotic CIs for NLR.

Our simulations confirm the importance of correctly identifying the error distribution
when fitting statistical models. Among 471 empirical datasets LR outperformed NLR under the
assumption of multiplicative error in all of the datasets (100%) for *a* and 427 datasets (90.7%)
for *b*. Similarly, NLR outperformed LR under the assumption of additive error in 196 datasets

(82.0%) for *a*, and 238 datasets (99.6%) for *b* (out of n = 239 valid parameterizations). The
method with the appropriate error assumption also had excellent confidence interval (CI)
coverage, whereas CI coverage for the inappropriate method was highly variable, reaching levels
as low as 0.2 (Figure 2).

108

Error Forms Observed in Nature

Given the critical nature of the error distribution in determining the appropriate method for
analyzing power-law data, it is necessary to understand the form of the error distribution in
nature. Previous papers have argued for both normal error (Packard 2009) and log-normal error
(Kerkhoff & Enquist 2009), but no systematic analysis of biological power-laws has been
conducted.

114 Taking a likelihood approach to compare the appropriateness of the two error forms for 115 the 471 empirical datasets described above, we used Akaike's information criterion (AIC), which 116 measures the goodness of fit of a statistical model by incorporating both the likelihood of the 117 model and a penalty for extra parameters (Burnham & Anderson 2002). For each of the 471 118 empirical datasets, we computed likelihoods and the values of AICc (a second order variant of 119 AIC that corrects for small sample size; see Burnham & Anderson 2002) for both the LR and 120 NLR based models. We compared the AICc values by following the conventional rule that if 121 $|\Delta AICc|$ (the magnitude of the difference between the two values of AICc) is less than 2, the two 122 models have relatively equal support and cannot be distinguished from each other; otherwise, the 123 model with the lower AICc is considered to have better data support (Burnham & Anderson 2002). Since AICc for the LR model is based on the likelihood from a log-normal distribution 124 125 conditioned on untransformed data, such comparison does not violate the assumption of identical 126 response variable in AIC-based model selection (Burnham & Anderson 2002, Section 2.11.3).

127 Consistent with previous suggestions that multiplicative error is biologically more
128 realistic (Gingerich 2000; Kerkhoff & Enquist 2009; Cawley & Janacek 2010), our likelihood
129 analysis of 471 allometric datasets shows that log-normal error distributions are substantially
130 more common than normal error distributions, with 68.6% of relationships being better
131 characterized by log-normal error, 16.6% by normal error, and 14.8% having uncertain error
132 structure.

133 Model Averaging: An Alternative Approach When Error Form Is Uncertain 134 Monte Carlo simulations show that if the underlying error structure is known then the model 135 assuming the appropriate error form (i.e., NLR with normal error, and LR with log-normal error) 136 will perform well for estimating both the parameters of the power-law and the CIs of those 137 parameters. However, the underlying error form of real datasets is not known and our likelihood 138 analysis shows that identification of the error form will not be clear-cut in all empirical datasets; 139 in part because the error form in real datasets may be more complex than assumed by the two 140 standard methods. Even in our simulation models where one distinct error structure has been 141 specified, likelihood tests sometimes failed to identify the correct error structure. For over half of 142 the parameterizations (50.7% when error was assumed to be log-normal and 71.1% when error 143 was assumed to be normal), error structure was either miscategorized or deemed uncertain by 144 likelihood tests in more than 10% of the simulated datasets.

When two or more models with appreciably different parameter estimates have similar support, model averaging provides a way to incorporate information from multiple models so that more stable inference can be made based on the weighted average of the entire set (Burnham & Anderson 2002; Link & Barker 2006). The most common weighting strategies are AIC weight (Burnham & Anderson 2002) and BIC weight (Link & Barker 2006). In our analysis we adopted

AIC weight (see Appendix C for the detailed procedure). Based on point estimates and CIs, we assessed whether the weighted model was able to accurately capture the underlying relationship under the assumption of the two error structures, i.e. whether it indicated the correct error structure if one existed. R package "boot" was used to construct CIs for the weighted average model (Davidson & Hinkley 1997; Canty & Ripley 2009).

Comparison of relative bias among LR, NLR and weighted average models shows that the weighted model closely resembles the model with the appropriate error assumption in both point estimation and CI coverage (Figure 2) regardless of error structure. Thus the weighted average model can provide an indication of the appropriate error distribution.

159

General Guidelines for the Analysis of Biological Power-laws

160 For future analysis of power-law relationships, we recommend the application of the following161 three-step procedure to correctly identify and apply the appropriate method:

162 1. Determine the appropriate error structure by either biological reasoning (e.g., Kerkhoff 163 & Enquist 2009, Cawley & Janacek 2010) or likelihood analysis. The relative likelihood of the 164 two error structures can be compared with AICc or other similar measures. To compute AICc, 165 first fit the two models using NLR and LR respectively and estimate the parameters *a*, *b*, and σ^2 166 for each model. Then calculate the likelihood that the data are generated from a normal 167 distribution with additive error

168
$$L_{norm} = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi \sigma_{NLR}^2}} \exp\left(\frac{-(y_i - a_{NLR} x_i^{b_{NLR}})^2}{2\sigma_{NLR}^2} \right) \right),$$

169 and the likelihood that the data are generated from a log-normal distribution with multiplicative

170 error

171
$$L_{logn} = \prod_{i=1}^{n} \left(\frac{1}{y_i \sqrt{2\pi \sigma_{LR}^2}} \exp\left(\frac{-\left(\log(y_i) - \log(a_{LR} x_i^{b_{LR}}) \right)^2}{2 \sigma_{LR}^2} \right) \right),$$

172 where *n* is sample size. AICc for each model can then be computed as

173
$$AICc = 2k - 2\log(L) + \frac{2k(k+1)}{n-k-1}$$
, where *k* is the number of parameters (3 in both models) and

174 *L* is the corresponding likelihood (Burnham & Anderson 2002).

2a. If the assumption of normal error is favored compared to log-normal error for either
biological or statistical reasons (i.e., AICc_{norm} – AICc_{logn}< -2), proceed with the results obtained
from NLR.

178 2b. If the assumption of log-normal error is favored compared to normal error (i.e.,
179 AICc_{norm} – AICc_{logn} > 2), proceed with the results obtained from LR.

180 2c. If neither model is favored for either statistical (i.e., $|AICc_{norm} - AICc_{logn}| \le 2$) or

181 biological reasons, model averaging should be adopted. The point estimates for *a* and *b* in the

182 mixed model are then weighted average of the corresponding point estimates from the two

183 original models. The AICc weights of the two models are computed as

184
$$w_i = C \cdot \exp\left(\frac{-AICC_c - min\left(AICc_{norm}, AICc_{logn}\right)}{2}\right)$$

185 where C is a normalizing constant so that w_{norm} and w_{logn} sum to 1. CIs for *a* and *b* can be

186 generated by bootstrapping for datasets of sufficient size (Efron & Tibshirani 1994).

3. Assess the validity of underlying statistical assumptions with diagnostic plots or tests
(e.g., Packard & Birchard 2008, Cawley & Janacek 2010), a step that has often been overlooked
in the analyses of biological power-laws. While it is rare that all assumptions are fully satisfied
by empirical datasets, major violations indicate the inappropriateness of the model and potential
invalidity of the results.

192 Computer code that implements these recommendations is available in Appendix D.

193

Implications for Previous Studies

194 For decades LR has been the conventional approach in the analysis of biological power-laws. If the current proposition to replace LR with NLR (e.g., Packard 2009, Packard et 195 196 al. 2010) were generally legitimate, the conclusions from large numbers of allometric studies 197 would be called into question. However, our likelihood analysis with 471 empirical datasets 198 spanning ecology, evolution and physiology shows that log-normal error consistently provides 199 superior fits to normal error distribution. This implies that the majority of previous allometric 200 studies in these fields are generally valid and contradicts the recent argument that LR is 201 inherently flawed and should be replaced by NLR (e.g., Packard 2009; Packard *et al.* 2010). As 202 our Monte Carlo simulation studies show, the application of NLR to such datasets may lead to 203 biased parameter estimates and potentially erroneous inferences.

The implications of these results for real biological patterns can be seen by applying the guidelines described in the previous section to arbitrate two debates regarding the exponents of morphological and physiological power-laws. The first example addresses whether or not the scaling of mammalian metabolic rate as a function of body size is consistent with the canonical 0.75 scaling exponent predicted by metabolic theory (Brown *et al.* 2004). Savage et al. (2004)

209 analyzed a large compilation of mammalian basal metabolic rates using LR and found that the 210 empirical data supported the predicted form of the relationship ($b_{LR} = 0.74$, $CI_{0.95} = (0.71, 0.76)$; 211 see Figure 3a). However, reanalyzing the same data using NLR resulted in different parameter 212 estimates and confidence intervals ($b_{\text{NLR}} = 0.91$, $CI_{0.95} = (0.88, 0.94)$), which suggested that the 213 0.75 exponent should be rejected as a reasonable description of the data (Packard & Birchard 214 2008). A quantitative analysis of the error structure in this dataset shows that the assumption of 215 multiplicative log-normal error is strongly supported compared to additive normal error 216 (AICc_{norm} – AICc_{logn} = 306) with no major violations of the assumptions. This suggests that the 217 data are consistent with the theoretical exponent.

218 Another example of how this approach can provide clear guidance when LR and NLR 219 yield different results is the scaling relationship between eye size and brain mass. Burton (2006) 220 analyzed this relationship in *fissiped Carnivora* using LR and argued that because the exponent 221 did not differ significantly from one ($b_{LR} = 0.87$, $CI_{0.95} = (0.55, 1.19)$) that eye size is determined 222 (at least in part) by a simple limitation on the amount of space available in the head. A reanalysis 223 of this data using NLR suggested that bears were outliers and that excluding this taxon the exponent was steeper than the hypothesized value of one (b_{NLR} = 1.42, CI_{0.95} = (1.13, 1.70); 224 225 Packard 2009). However, both the identification of outliers and the use of nonlinear regression 226 were controversial (Kerkhoff & Enquist 2009). Likelihood analysis demonstrates that the 227 assumption of log-normal error is more strongly supported regardless of whether the bears are 228 included (AICc_{norm} – AICc_{logn} = 35.9) or not (AICc_{norm} – AICc_{logn} = 7.88), and the assumptions of 229 normality and heteroscedasticity are not strongly violated in either case. Therefore since LR yields confidence intervals that include one even when the bears are excluded ($b_{LR} = 1.24$, CI_{0.95} 230 231 = (0.96, 1.53); see Figure 3b), the proposed isometric relationship is supported by the data.

Parallel examples where datasets with normal or undetermined error structures suffer
from methodological problems are rarer in the literature due to the prevalence of the log-normal
error distribution observed in nature. Nonetheless, reanalysis of the original data is warranted in
cases where there is reason to suspect that an additive normal error structure or an undetermined
error structure is more realistic.

237

Complexities

Apart from making inferences about the parameters, power functions are also frequently used to make predictions for new observations, which is particularly important in paleontology and conservation biology. For LR, it should be noted that although the parameter estimates are unbiased when the error is log-normal and multiplicative (Ferguson 1986), the model predicts log(*y*), and the predicted value of *y* obtained by anti-log transformation is biased on arithmetic scale (Hayes & Shonkwiler 2006). Measures should be taken to correct for this bias if predictions are to be made from log-transformed power functions (Hayes & Shonkwiler 2006).

245 One class of commonly observed biological power-law relationships not included in this 246 study is the scaling relationship between species richness and attributes of the habitat (e.g., area, 247 resource availability, distance to mainland, etc.). The most widely studied of these relationships 248 is the species-area relationship (SAR). SARs are of fundamental importance in conservation 249 biology where they are used for making predictions regarding the effect of habitat loss on 250 biodiversity (Brooks *et al.* 2002) as well as the identification of hotspots (Veech 2000). It has 251 been shown that inference related to the SAR varies with the method used for fitting the data 252 (Fattorini 2007). One often overlooked characteristic of SARs is that the response variable, 253 species richness, is a discrete count, which in principle cannot be accommodated by either LR or 254 NLR because both assume a continuous data distribution (which is why this type of data was not

255 included in our empirical analyses). The existence of discrete error structure in some biological 256 power-laws highlights the fact that additive normal error and multiplicative log-normal error are 257 often not the only options that should be considered when analyzing error distributions. O'Hara 258 and Kotze (2010) showed that ignorance of the error characteristics can lead to failure of the 259 statistical analysis. Our understanding of the validity of previous studies of SARs and other 260 relationships that potentially violate the distributional assumptions of LR and NLR would be 261 enhanced by a systematic comparison between methods that accommodate their statistical and 262 biological properties and currently applied methods such as NLR and LR.

263

Conclusions

264 Power functions are one of the most broadly studied relationships in biological systems. The 265 current debate surrounding the methodology used in their analysis has generated considerable 266 confusion in the field. As a result the conclusions of previous studies have been called into 267 question and the progress of new analyses has been hampered. Our study provides a clear answer 268 to the current controversy surrounding the appropriate methodology for analyzing allometric 269 data. Neither linear regression on log-transformed data nor standard nonlinear regression is 270 inherently superior for fitting power-laws to data. Which method performs better depends on the 271 distribution of the error. For most allometric datasets like those we studied, the error is 272 distributed such that log-transformed linear regression will produce more accurate parameter 273 estimates and confidence intervals. As a result, most published results are likely valid. However, 274 the methodology chosen for future analyses of power-laws in ecology and evolution should be 275 based on explicit analyses (both statistical and biological) of the underlying error structure. We 276 recommend that likelihood comparisons be applied to assess the error structure of the dataset. In 277 cases where the error is approximately multiplicative lognormal, the log-transformed linear

278	regression should be used, while nonlinear regression on untransformed data should be applied to
279	those datasets with additive normal error. For datasets with an indeterminate error structure, we
280	recommend using model averaging to calculate the weighted average of the parameter estimates.
281	As in all statistical analyses, the assumptions of the chosen model should be carefully evaluated.
282	Acknowledgements
283	We thank the authors of data used in the compilation for making their data publicly available.
284	Kristina J. Anderson-Teixeira, Ken Ashwell and Robin Warne provided help with their data. We
285	thank Drew Kerkhoff, Philip Gingerich, and James Haefner for helpful discussions, and Brian
286	Inouye and an anonymous reviewer for helpful comments on the manuscript.
287	Literature Cited
288	Bates, D. M., and D. G. Watts. 1988. Nonlinear Regression Analysis and Its Applications. John
289	Wiley & Sons Inc, pp 24-26.
290	Brooks, T. M., R. A. Mittermeier, C. G. Mittermeier, G. A. B. Da Fonseca, A. B. Rylands, W. R.
291	Konstant, P. Flick, J. Pilgrim, S. Oldfield, G. Magin, and C. Hilton-Taylor. 2002. Habitat
292	loss and extinction in the hotspots of biodiversity. Conservation Biology 16: 909-923.
293	Brown, J. H., J. F. Gillooly, A. P. Allen, V. M. Savage, and G. B. West. 2004. Toward a metabolic
294	theory of ecology. Ecology 85: 1771-1789.
295	Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multimodel Inference: A
296	Practical Information-Theoretic Approach. Springer-Verlag, New York, USA.
297	Burton, R. F. 2006. A new look at the scaling of size in mammalian eyes. Journal of Zoology
298	269: 225-232.
299	Caruso, T., D. Garlaschelli, R. Bargagli, and P. Convey. 2010. Testing metabolic scaling theory
300	using intraspecific allometries in Antarctic microarthropods. Oikos 119: 935-945.

- 301 Canty, A., and B. Ripley. 2009. boot: Bootstrap R (S-Plus) Functions. R package version 1.2-37.
- 302 Cawley, G. C., and G. J. Janacek. 2010. On allometric equations for predicting body mass of
 303 dinosaurs. Journal of Zoology 280: 355-361.
- 304 Davidson, A. C., & D. V. Hinkley. 1997. Boostrap Methods and Their Applications. Cambridge

305 University Press, Cambridge.

- Efron, B., and R. J. Tibshirani. 1994. An Introduction to the Bootstrap. Chapman and Hall/CRC,
 pp 168-177.
- Farlow, J. O., P. Dodson, and A. Chinsamy. 1995. Dinosaur Biology. Annual Review of Ecology
 and Systematics 26: 445-471.
- 310 Fattorini, S. 2007. To fit or not to fit? A poorly fitting procedure produces inconsistent results
- when the species-area relationship is used to locate hotspots. Biodiversity andConservation 16: 2531-2538.
- Ferguson, R. I. 1986. River loads underestimated by rating curves. Water Resources Research 22:
 74-76.
- Gingerich, P. D. 2000. Arithmetic or geometric normality of biological variation: an empirical
 test of theory. Journal of Theoretical Biology 204: 201-221.
- 317 Hayes, J. P., and J. S. Shonkwiler. 2006. Allometry, antilog transformations, and the perils of
- prediction on the original scale. Physiological and Biochemical Zooogy 79: 665-674.
- 319 Herculano-Houzel, S., C. E. Collins, P. Wong, J. H. Kaas, and L. Roberto. 2008. The basic
- 320 nonuniformity of the cerebreal cortex. Proceedings of the National Academy of Science
- **321** USA 105: 12593-12598.
- 322 Hui, C., J. S. Terblanche, S. L. Chown, and M. A. McGeoch. 2010. Parameter landscapes unveil
- 323 the bias in allometric prediction. Methods in Ecology and Evolution 1: 69-74.

325	transformation is necessary in allometry. Journal of Theoretical Biology 257: 519-521.
326	Link, W. A., and R. J. Barker. 2006. Model weights and the foundations of multimodel inference.
327	Ecology 87: 2626-2635.
328	Macrini, T. E., T. Rowe, and J. L. Vandeberg. 2007. Cranial endocasts from a growth series of
329	Monodelphis domestica (Didelphidae, Marsupialia): A study of individual and
330	ontogenetic variation. Journal of Morphology 268: 844-865.
331	Martin, H. G., and N. Goldenfeld. 2006. On the origin and robustness of power-law species-area
332	relationships in ecology. Proceedings of the National Academy of Science USA 103:
333	10310-10315.
334	Mortola, J. P., and MJ. Limoges. 2006. Resting breathing frequency in aquatic mammals: A
335	comparative analysis with terrestrial species. Respiratory Physiology & Neurobiology
336	154: 500-514.
337	O'Hara, R.B., and D. J. Kotze. 2010. Do not log-transform count data. Methods in Ecology and
338	Evolution 1: 118-122.
339	Packard, G. C. 2009. On the use of logarithmic transformations in allometric analyses. Journal of
340	Theoretical Biology 257: 515-518.
341	Packard, G.C., and G. F. Birchard. 2008. Traditional allometric analysis fails to provide a valid
342	predictive model for mammalian metabolic rates. The Journal of Experimental Biology
343	211: 3581-3587.
344	Packard, G. C., G. F. Birchard, and T. J. Boardman. 2010. Fitting statistical models in bivariate
345	allometry. Biological Reviews doi: 10.1111/j.1469-185X.2010.00160.x.

Kerkhoff, A. J., and B. J. Enquist. 2009. Multiplicative by nature: Why logarithmic

324

346 R Development Core Team (2009). R: A language and environment for statistical computing. R

- 347 Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL
- 348 http://www.R-project.org.
- 349 Ritz, C., & J. C. Streibig. 2008. Nonlinear Regression with R. Springer, New York.
- 350 Savage, V. M., J. F. Gillooly, W. H. Woodruff, G. B. West, A. P. Allen, B. J. Enquist, and J. H.
- Brown. 2004. The predominance of quarter-power scaling in biology. Functional Ecology
- **352** 18: 257-282.
- 353 Veech, J. A. 2000. Choice of species-area function affects identification of hotspots.
- Conservation Biology 14: 140-147.

Figure 1. An illustration of additive normal error and multiplicative log-normal error displayed on both arithmetic and logarithmic scales, and how the underlying relationships can be distorted by the application of inappropriate methods. For additive error, *x* was generated from a uniform distribution ranging from 10 to 10000, *y* was generated using Eqn.1 with a = 10, b = 0.2, $\sigma = 10$. For multiplicative error, *x* was generated from a log-uniform distribution ranging from 1 to 10 on the logarithmic scale, *y* was generated using Eqn.2 with a = 0.3, b = 0.75, $\sigma = 0.3$. The dashed curves correspond to the true underlying relationships.

362 Figure 2. Comparison of bias in point estimation and CI coverage among LR, NLR and

363 AICc-weighted average models in simulations with parameters estimated from 471 empirical

364 datasets for multiplicative error structure and 239 empirical datasets for additive error structure.

365 Relative bias (mean estimate/true value) is depicted because *a* spans a wide range across

366 empirical datasets. For point estimation, each curve represents the relative frequency distribution

367 of relative bias. An appropriate method peaks at 0 (on logarithmic scale) with small dispersion,

368 while an inappropriate method shows a wide range of relative bias. For CI coverage, the

369 horizontal dashed line represents the nominal 0.95 level. Note that point estimates were

370 generated based on 10⁵ simulated datasets, while CIs were based on 400 additional simulated

371 datasets due to computational limitation. CI results are only shown for *b*.

372 Figure 3. Examples of biological power-law relationships where an analysis of the error

373 structure of the data can be used to arbitrate debates regarding the form of the underlying

374 relationship. a. Basal metabolic rate – body mass relationship from Savage *et al.* (2004),

375 reanalyzed in Packard & Birchard (2008); b. eye size – brain mass relationship from Burton

376 (2006), analyzed in Packard (2009). See text for details.



Figure 2.





