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

Xiao Xiao,[1,2,3,4] Ethan P. White,[1,2] Mevin B. Hooten,[2,3,5] and Susan L. Durham[2]

[1]*Department of Biology, Utah State University, Logan, Utah 84322-5305 USA*
[2]*Ecology Center, Utah State University, Logan, Utah 84322-5205 USA*
[3]*Department of Mathematics and Statistics, Utah State University, Logan, Utah 84322-3900 USA*

Abstract. Power-law relationships are among the most well-studied functional relationships in biology. Recently the common practice of fitting power laws using linear regression (LR) on log-transformed data has been criticized, calling into question the conclusions of hundreds of studies. It has been suggested that nonlinear regression (NLR) is preferable, but no rigorous comparison of these two methods has been conducted. Using Monte Carlo simulations, we demonstrate that the error distribution determines which method performs better, with NLR better characterizing data with additive, homoscedastic, normal error and LR better characterizing data with multiplicative, heteroscedastic, lognormal error. Analysis of 471 biological power laws shows that both forms of error occur in nature. While previous analyses based on log-transformation appear to be generally valid, future analyses should choose methods based on a combination of biological plausibility and analysis of the error distribution. We provide detailed guidelines and associated computer code for doing so, including a model averaging approach for cases where the error structure is uncertain.

Key words: allometry; log-transformation; model averaging; model comparison; nonlinear regression; parameter estimation; power law.

Introduction

Power-law relationships of the form $y = ax^b$ are one of the most common patterns in biology. They have been documented in a variety of different areas including the relationships between body size, physiological rates, and life history traits (Brown et al. 2004), the scaling between body parts in morphology (Farlow et al. 1995), and the species–area relationship in biogeography (Martin and Goldenfeld 2006). These fitted relationships have been used to test the validity of biological theories (Brown et al. 2004), to infer the characteristics of extinct species (Farlow et al. 1995), to assess the effect of evolutionary processes (Mortola and Limoges 2006), and to predict the consequence of habitat loss on biodiversity (Brooks et al. 2002).

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 data to be modeled using linear regression. However, it has been suggested that analysis on logarithmic scales is flawed and that instead, analysis

should be carried out on the original scale of measurement using nonlinear regression (Fattorini 2007, Packard and Birchard 2008, Packard 2009, Caruso et al. 2010, Packard et al. 2010). If these claims are correct, the validity of decades of published research in ecology, evolution, and physiology would be called into question.

One of the fundamental differences between linear regression on log-transformed data (hereafter, LR) and nonlinear regression on untransformed data (hereafter, NLR) lies in the assumptions about how stochasticity manifests in the model (Gingerich 2000, Kerkhoff and Enquist 2009). In NLR, it is assumed that the error term is normally distributed and additive on the arithmetic scale (Bates and Watts 1988:67–70, Ritz and Streibig 2008):

$$y = ax^b + \varepsilon \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2). \qquad (1)$$

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

$$\log(y) = \log(a) + b\log(x) + \varepsilon \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

which corresponds to lognormally distributed, multiplicative error on the arithmetic scale,

$$y = ax^b e^{\varepsilon} \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2). \qquad (2)$$

For a single data set, both assumptions cannot be correct. Violation of statistical assumptions of error can lead to biased point estimates as well as inaccurate confidence
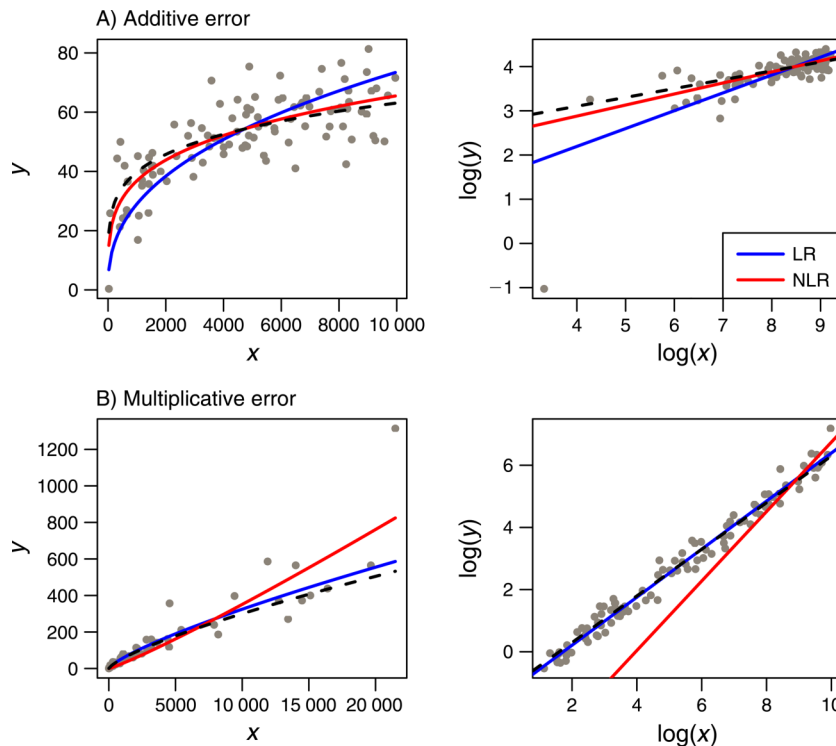
Fig. 1. An illustration of (A) additive normal error and (B) multiplicative lognormal error displayed on both arithmetic and logarithmic scales, and how the underlying relationships can be distorted by the application of inappropriate methods. For additive error, x was generated from a uniform distribution ranging from 10 to 10 000, y was generated using Eq. 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 Eq. 2 with $a = 0.3$, $b = 0.75$, and $\sigma = 0.3$. The dashed curves correspond to the true underlying relationships. Solid lines show linear (LR, blue) and nonlinear (NLR, red) regression.

intervals (Bates and Watts 1988:23–26; see Fig. 1 for an illustration).

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

Error Distribution Determines the Best Method for Fitting Power Laws

Previous arguments regarding the performance of different methods have typically been based on empirical data (but see Hui et al. 2010), despite the fact that the true data-generating mechanism is generally unknown. As such, previous studies provide little insight into the best methodological approach. Monte Carlo simulation, where data are simulated from known distributions, allows for a direct comparison between statistically estimated parameters and their true values. Here we implement the Monte Carlo approach based on parameterizations from empirical data sets so that our results will be valid for the range of empirically observed parameter values. Results from these empirically motivated simulations were consistent with standard Monte Carlo simulations based on hypothetical parameterizations (see Appendix A).

We compiled 471 data sets published between 2004 and 2008 in the fields of ecology and evolution where significant power-law relationships were reported. The selected data sets were either morphological or physiological allometries between organismal traits (for details of data selection and the full list of data sets, see
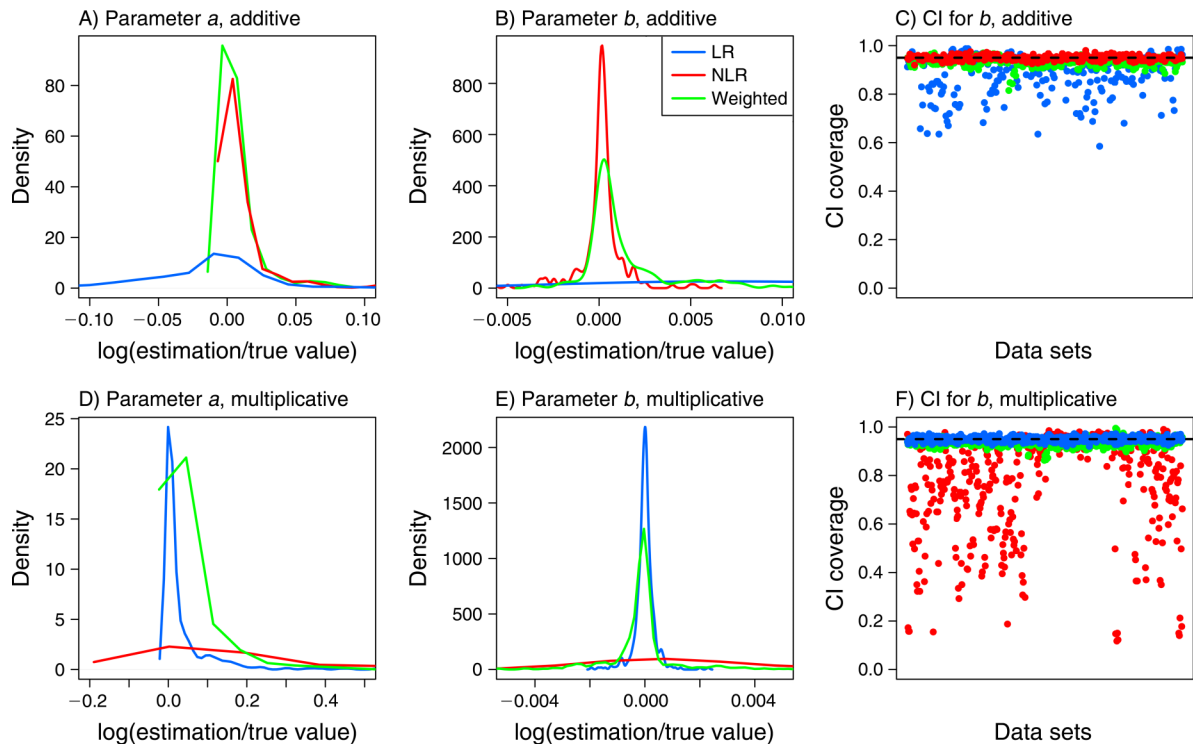
Fig. 2. Comparison of bias in point estimation and CI coverage among LR, NLR, and AIC$_c$-weighted average models in simulations with parameters (A, D) a and (B, C, E, F) b estimated from 239 empirical data sets for (A–C) additive error structure and 471 empirical data sets for (D–F) multiplicative error structure. Relative bias (mean estimate divided by true value) is depicted because a spans a wide range across empirical data sets. For point estimation, each curve represents the relative frequency distribution of relative bias. An appropriate method peaks at 0 (on logarithmic scale) with small dispersion, while an inappropriate method shows a wide range of relative bias. For CI coverage, the horizontal dashed line represents the nominal 0.95 level. Note that point estimates were generated based on 10^5 simulated data sets, while CIs were based on 400 additional simulated data sets due to computational limitation. Point estimate comparisons are shown for both parameters a and b, while CI coverage comparisons are shown only for b.

Appendix B; for the complete compilation of data sets, see Supplement 1). To generate the parameters for simulations each empirical data set was assumed to have (1) a multiplicative lognormal error structure, and a, b, and σ were estimated with LR (with σ estimated as the standard deviation of the residuals), and (2) an additive normal error structure, and the parameters were estimated with NLR. For each data set, 10^5 independent simulations were carried out using the estimated parameters under the assumption of each error structure. Each simulated data set was analyzed with both LR and NLR, and the performance of the two methods was compared to determine which method had the better point estimation of a and b, as well as more accurate confidence interval (CI) coverage measured by the percentage of simulations where the true parameter value falls within the estimated 95% CI. Only 239 data sets generated valid simulations under the assumption of additive error (see Appendix C for technical details on the procedure of the simulations). All simulations and analyses were carried out using R version 2.9.1 (R Development Core Team 2009). The "nlrwr" package

(Ritz and Streibig 2008) was used to compute asymptotic CIs for NLR.

Our simulations confirm the importance of correctly identifying the error distribution when fitting statistical models. Among 471 empirical data sets LR outperformed NLR under the assumption of multiplicative error in all of the data sets (100%) for a and 427 data sets (90.7%) for b. Similarly, NLR outperformed LR under the assumption of additive error in 196 data sets (82.0%) for a, and 238 data sets (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 (Fig. 2).

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 lognormal error (Kerkhoff and Enquist 2009), but no

Reports

systematic analysis of biological power laws has been conducted.

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

Consistent with previous suggestions that multiplicative error is biologically more realistic (Gingerich 2000, Kerkhoff and Enquist 2009, Cawley and Janacek 2010), our likelihood analysis of 471 allometric data sets shows that lognormal error distributions are substantially more common than normal error distributions, with 68.6% of relationships being better characterized by lognormal error, 16.6% by normal error, and 14.8% having uncertain error structure.

MODEL AVERAGING: AN ALTERNATIVE APPROACH WHEN ERROR FORM IS UNCERTAIN

Monte Carlo simulations show that if the underlying error structure is known, then the model assuming the appropriate error form (i.e., NLR with normal error, and LR with lognormal error) will perform well for estimating both the parameters of the power law and the CIs of those parameters. However, the underlying error form of real data sets is not known and our likelihood analysis shows that identification of the error form will not be clear-cut in all empirical data sets; in part because the error form in real data sets may be more complex than what is assumed by the two standard methods. Even in our simulation models where one distinct error structure has been specified, likelihood tests sometimes failed to identify the correct error structure. For over half of the parameterizations (50.7% when error was assumed to be lognormal and 71.1% when error was assumed to be normal), error structure was either miscategorized or deemed uncertain by likelihood tests in more than 10% of the simulated data sets.

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 and Anderson 2002, Link and Barker 2006). The most common weighting strategies are AIC weight (Burnham and Anderson 2002) and BIC weight (Link and 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. The R package "boot" was used to construct CIs for the weighted average model (Davidson and Hinkley 1997, Canty and 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 (Fig. 2) regardless of error structure. Thus the weighted average model can provide an indication of the appropriate error distribution.

GENERAL GUIDELINES FOR THE ANALYSIS OF BIOLOGICAL POWER LAWS

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

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

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

and the likelihood that the data are generated from a lognormal distribution with multiplicative error,

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

where n is sample size. AIC_c for each model can then be computed as

$$AIC_c = 2k - 2\log(L) + \frac{2k(k+1)}{n-k-1}$$

where k is the number of parameters (three in both models) and L is the corresponding likelihood (Burnham and Anderson 2002).

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

2b) If the assumption of lognormal error is favored compared to normal error for either biological or statistical reasons (i.e., $AIC_{c-norm} - AIC_{c-logn} > 2$), proceed with the results obtained from LR.

2c) If neither model is favored by either statistical (i.e., $|AIC_{c-norm} - AIC_{c-logn}| \leq 2$) or biological reasons, model averaging should be adopted. The point estimates for a and b in the mixed model are then weighted averages of the corresponding point estimates from the two original models. The AIC_c weights of the two models are computed as

$$w_i = C \times \exp\left(-\frac{AIC_{c-i} - \min(AIC_{c-norm}, \, AIC_{c-logn})}{2}\right)$$

where C is a normalizing constant so that w_{norm} and w_{logn} sum to 1, and i represents either one of the two distribution models. CIs for a and b can be generated by bootstrapping for data sets of sufficient size (Efron and Tibshirani 1994).

3) Assess the validity of underlying statistical assumptions with diagnostic plots or tests (e.g., Packard and Birchard 2008, Cawley and 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 data sets, major violations indicate the inappropriateness of the model and potential invalidity of the results.

Computer code that implements these recommendations is available in Supplement 2.

Implications for Previous Studies

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 al. 2010) were generally legitimate, the conclusions from large numbers of allometric studies would be called into question. However, our likelihood analysis with 471 empirical data sets spanning ecology, evolution, and physiology shows that lognormal error generally provides better fit than normal error distribution. This implies that the majority of previous allometric studies in these fields are generally valid and contradicts the recent argument that LR is inherently flawed and should be replaced by NLR (e.g., Packard 2009, Packard et al. 2010). As our Monte Carlo simulation studies show, the application of NLR to such data sets may lead to 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) analyzed a large compilation of mammalian basal metabolic rates using LR and found that the empirical data supported the predicted form of the relationship ($b_{LR} = 0.74$, $CI_{0.95} = (0.71, 0.76)$; see Fig. 3A). However, reanalyzing the same data using NLR resulted in different parameter estimates and confidence intervals ($b_{NLR} = 0.91$, $CI_{0.95} = (0.88, 0.94)$), which suggested that the 0.75 exponent should be rejected as a reasonable description of the data (Packard and Birchard 2008). A quantitative analysis of the error structure in this data set shows that the assumption of multiplicative lognormal error is strongly supported compared to additive normal error ($AIC_{c-norm} - AIC_{c-logn} = 306$) with no major violations of the assumptions. This suggests that the data are consistent with the theoretical exponent.

Another example of how this approach can provide clear guidance when LR and NLR yield different results is the scaling relationship between eye size and brain mass. Burton (2006) analyzed this relationship in fissiped carnivores using LR and argued that because the exponent did not differ significantly from one ($b_{LR} = 0.87$, $CI_{0.95} = (0.55, 1.19)$) that eye size is determined (at least in part) by a simple limitation on the amount of space available in the head. A reanalysis of this data using NLR suggested that bears were outliers and that excluding this taxon the exponent was steeper than the hypothesized value of 1. ($b_{NLR} = 1.42$, $CI_{0.95} = (1.13, 1.70)$; Packard 2009). However, both the identification of outliers and the use of nonlinear regression were controversial (Kerkhoff and Enquist 2009). Likelihood analysis demonstrates that the assumption of lognormal error is more strongly supported regardless of whether the bears are included ($AIC_{c-norm} - AIC_{c-logn} = 35.9$) or not ($AIC_{c-norm} - AIC_{c-logn} = 7.88$), and the assumptions of 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 ($b_{LR} = 1.24$, $CI_{0.95} = (0.96, 1.53)$; see Fig. 3B), the proposed isometric relationship is supported by the data.

Parallel examples where data sets with normal or undetermined error structures suffer from methodological problems are rarer in the literature due to the prevalence of the lognormal 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.

Complexities

Apart from making inferences about the parameters, power functions are also frequently used to make
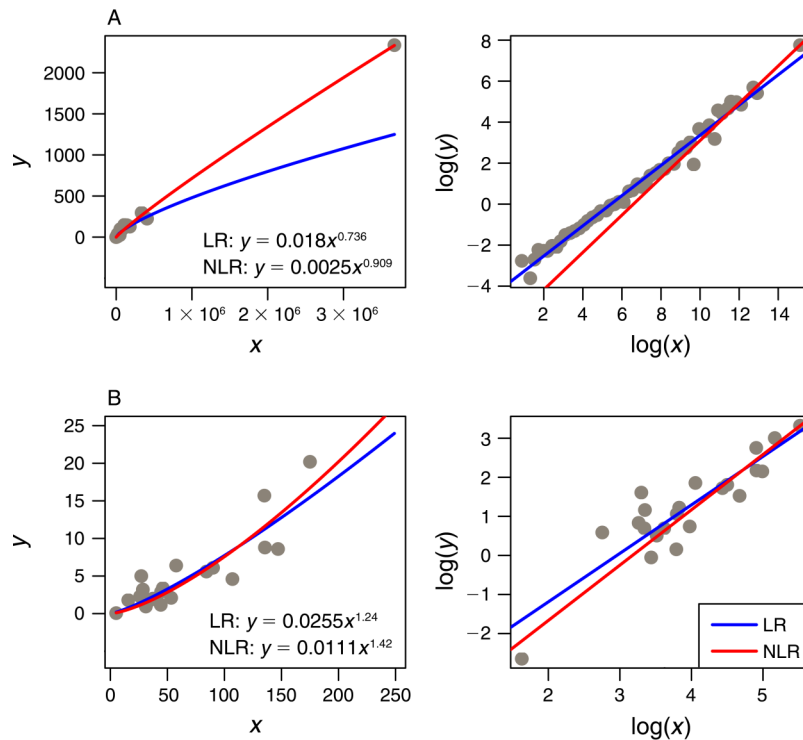
Reports

Fig. 3. Examples of biological power-law relationships where an analysis of the error structure of the data can be used to arbitrate debates regarding the form of the underlying relationship. (A) Basal metabolic rate–body mass relationship from Savage et al. (2004), reanalyzed in Packard and Birchard (2008). (B) Eye-size–brain-mass relationship from Burton (2006), reanalyzed in Packard (2009). See *Implications for Previous Studies*.

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 lognormal 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 and Shonkwiler 2006, Packard 2009). Measures should be taken to correct for this bias if predictions are to be made from log-transformed power functions (Hayes and Shonkwiler 2006).

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

assume a continuous data distribution (which is why this type of data was not included in our empirical analyses). The existence of discrete error structure in some biological power laws highlights the fact that additive normal error and multiplicative lognormal error are often not the only options that should be considered when analyzing error distributions. O'Hara and Kotze (2010) showed that ignorance of the error characteristics can lead to failure of the statistical analysis. Our understanding of the validity of previous studies of SARs and other relationships that potentially violate the distributional assumptions of LR and NLR would be enhanced by a systematic comparison between methods that accommodate their statistical and biological properties and currently applied methods such as NLR and LR.

CONCLUSIONS

Power functions are one of the most broadly studied relationships in biological systems. The current debate surrounding the methodology used in their analysis has generated considerable confusion in the field. As a result, the conclusions of previous studies have been called into question and the progress of new analyses has been hampered. Our study provides a clear answer to the current controversy surrounding the appropriate methodology for analyzing allometric data. Neither linear regression on log-transformed data nor standard non-

linear regression is inherently superior for fitting power laws to data. Which method performs better depends on the distribution of the error. For most allometric data sets like those we studied, the error is distributed such that log-transformed linear regression will produce more accurate parameter estimates and confidence intervals. As a result, most published results are likely valid. However, the methodology chosen for future analyses of power laws in ecology and evolution should be based on explicit analyses (both statistical and biological) of the underlying error structure. We recommend that likelihood comparisons be applied to assess the error structure of the data set. In cases where the error is approximately multiplicative lognormal, the log-transformed linear regression should be used, while nonlinear regression on untransformed data should be applied to those data sets with additive normal error. For data sets with an indeterminate error structure, we recommend using model averaging to calculate the weighted average of the parameter estimates. As in all statistical analyses, the assumptions of the chosen model should be carefully evaluated.

Literature Cited

Bates, D. M., and D. G. Watts. 1988. Nonlinear regression analysis and its applications. John Wiley and Sons, Inc., New York, New York, USA.

Brooks, T. M., R. A. Mittermeier, C. G. Mittermeier, G. A. B. Da Fonseca, A. B. Rylands, W. R. Konstant, P. Flick, J. Pilgrim, S. Oldfield, G. Magin, and C. Hilton-Taylor. 2002. Habitat loss and extinction in the hotspots of biodiversity. Conservation Biology 16:909–923.

Brown, J. H., J. F. Gillooly, A. P. Allen, V. M. Savage, and G. B. West. 2004. Toward a metabolic theory of ecology. Ecology 85:1771–1789.

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

Burton, R. F. 2006. A new look at the scaling of size in mammalian eyes. Journal of Zoology 269:225–232.

Canty, A., and B. Ripley. 2009. boot: bootstrap R (S-Plus) functions. R package version 1.2-37. ⟨http://cran.r-project.org/web/packages/boot/⟩

Caruso, T., D. Garlaschelli, R. Bargagli, and P. Convey. 2010. Testing metabolic scaling theory using intraspecific allometries in Antarctic microarthropods. Oikos 119:935–945.

Cawley, G. C., and G. J. Janacek. 2010. On allometric equations for predicting body mass of dinosaurs. Journal of Zoology 280:355–361.

Davidson, A. C., and D. V. Hinkley. 1997. Boostrap methods and their application. Cambridge University Press, Cambridge, UK.

Efron, B., and R. J. Tibshirani. 1994. An introduction to the bootstrap. Chapman and Hall/CRC, New York, New York, USA.

Farlow, J. O., P. Dodson, and A. Chinsamy. 1995. Dinosaur biology. Annual Review of Ecology and Systematics 26:445–471.

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 and Conservation 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.

Hayes, J. P., and J. S. Shonkwiler. 2006. Allometry, antilog transformations, and the perils of prediction on the original scale. Physiological and Biochemical Zoology 79:665–674.

Hui, C., J. S. Terblanche, S. L. Chown, and M. A. McGeoch. 2010. Parameter landscapes unveil the bias in allometric prediction. Methods in Ecology and Evolution 1:69–74.

Kerkhoff, A. J., and B. J. Enquist. 2009. Multiplicative by nature: why logarithmic transformation is necessary in allometry. Journal of Theoretical Biology 257:519–521.

Link, W. A., and R. J. Barker. 2006. Model weights and the foundations of multimodel inference. Ecology 87:2626–2635.

Martin, H. G., and N. Goldenfeld. 2006. On the origin and robustness of power-law species–area relationships in ecology. Proceedings of the National Academy of Science USA 103:10310–10315.

Mortola, J. P., and M.-J. Limoges. 2006. Resting breathing frequency in aquatic mammals: a comparative analysis with terrestrial species. Respiratory Physiology and Neurobiology 154:500–514.

O'Hara, R. B., and D. J. Kotze. 2010. Do not log-transform count data. Methods in Ecology and Evolution 1:118–122.

Packard, G. C. 2009. On the use of logarithmic transformations in allometric analyses. Journal of Theoretical Biology 257:515–518.

Packard, G. C., and G. F. Birchard. 2008. Traditional allometric analysis fails to provide a valid predictive model for mammalian metabolic rates. Journal of Experimental Biology 211:3581–3587.

Packard, G. C., G. F. Birchard, and T. J. Boardman. 2010. Fitting statistical models in bivariate allometry. Biological Reviews. [doi: 10.1111/j.1469-185X.2010.00160.x]

R Development Core Team. 2009. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ⟨http://www.R-project.org⟩

Ritz, C., and J. C. Streibig. 2008. Nonlinear regression with R. Springer, New York, New York, USA.

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 18:257–282.

Veech, J. A. 2000. Choice of species–area function affects identification of hotspots. Conservation Biology 14:140–147.

APPENDIX A

Additional Monte Carlo simulations comparing the performance of LR and NLR estimation methods with hypothetical parameters (*Ecological Archives* E092-160-A1).

Reports

APPENDIX B

List of empirical data sets used in Monte Carlo simulations and likelihood analysis (*Ecological Archives* E092-160-A2).

APPENDIX C

Technical details of the Monte Carlo simulations using empirical parameters and the simulations related to weighted model averaging (*Ecological Archives* E092-160-A3).

SUPPLEMENT 1

A file containing 471 data sets compiled from the literature describing power-law relationships in ecology, evolution, and physiology (*Ecological Archives* E092-160-S1).

SUPPLEMENT 2

The R source code to implement the recommended guidelines for the analysis of power-law relationships (*Ecological Archives* E092-160-S2).

Reports