Journal of Theoretical Biology 317 (2013) 418–421

Contents lists available at SciVerse ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Letter to Editor

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)

ARTICLE INFO

Keywords:
Allometry
Transformation
Likelihood
Scaling

Throughout the history of studying scaling relationships in biology there has been much discussion of how to best estimate scaling parameters (Finney, 1941; Sprugel, 1983; Charnov, 1993; Smith, 1993; Blackburn and Gaston, 1998; Marquet, 2000; Packard and Birchard, 2008; Packard and Boardman, 2008). A recent exchange in the literature (between Packard, 2009 and Kerkhoff and Enquist, 2009) made it clear that the discussion is continuing, and prompted me to write a short note focusing on fitting scaling relationships to data. While I was writing the initial version of this note that outlines the practical theoretical side of fitting scaling relationships to data, Xiao et al. (2011) were independently focusing on the same issue by explicitly analyzing many recently published data sets. In the interim, Packard (2011a, 2011b) and Packard et al. (2011) have reiterated and expanded on the main points put forth in Packard (2009). Here, I directly address the debate over the appropriateness of additive versus multiplicative error for scaling relationships by explicitly formulating probability models for the different types of error structure, and then illustrate how to evaluate the fit of the different error models to observed scaling data. In particular, I write the likelihood functions for each of the two competing error models discussed in the exchange between Packard (2009) and Kerkhoff and Enquist (2009), and describe how inference can be made therefrom. As an example, I analyze one of the data sets in Packard (2009) and squarely place the results here in the context of the recent and extensive analysis of data by Xiao et al. (2011). My hope is that this contribution sparked by the difference of opinion about error structure articulated by Packard (2009) and Kerkhoff and Enquist (2009), and the coincident work of Xiao et al. (2011), presumably influenced by the same exchange, will help clarify some of the issues regarding fitting models with different error structure to biological data and provide some degree of closure to the recent debate. At this point, readers will hopefully be in a position to draw their own conclusions about the appropriateness of different methods when fitting scaling models to data.

1. The exchange between Packard (2009) and Kerkhoff and Enquist (2009)

In my eyes, the substantive difference of opinion arose because Kerkhoff and Enquist (2009) feel that multiplicative error generally serves as a better underlying probability model than additive error for allometric relationships, whereas Packard (2009) feels that statistical models must generally be based on the original (in this case additive) scale of measurement and that transformation should be used sparingly. This exchange illustrates two major issues that arise when fitting scaling relationships to biological data, one practical and one philosophical. Regarding the philosophical, if one has strong evidence for a particular model of error, an argument can be made for using this error model to estimate scaling parameters, potentially with explicit specification of a prior. However, if the focus is on whether additive or multiplicative error models yield better fits to data, one can directly evaluate the fit of scaling relationships with different assumptions regarding error, remaining agnostic about which formulation of error makes the most biological sense.

Packard (2009) makes the case for graphically fitting models to data, which has its precedent and appropriate place, but does not acknowledge that it is possible to directly evaluate the fit of models with different error structure using maximum likelihood. His point regarding the importance of plotting and examining one's data on original scales of measurement is well taken, but many of the guiding principles for assessing model fit are based on underlying probability models, not visual inspection. Presumably, other ecologists are also unsure about how to explicitly specify competing probability models and employ maximum likelihood for evaluating the fit of models with different error structure, making this exchange an opportunity to illustrate an approach one might take. This exchange also highlights the need for greater statistical fluency in the field of ecology (Ellison and Dennis, 2010). In the following example, I illustrate how this approach can be used to directly assess the fit of allometric models with additive and multiplicative error structure, which is the basis for the analysis of data in Xiao et al. (2011).

0022-5193/$ - see front matter
http://dx.doi.org/10.1016/j.jtbi.2012.09.035

2. Additive and multiplicative models of error for scaling relationships

Assume, as Packard (2009) does that there are two candidate models for explaining one's data with a scaling relationship:

$$Y = aX^b + \text{error} \tag{1}$$

and

$$Y = aX^b \cdot \text{error} \tag{2}$$

differing only in how error is incorporated. To evaluate the fit of competing models, discrepancies between predictions that are the result of particular parameter values and data must be quantified. Typically, we like for deviations to be distributed evenly about a model prediction and formally express this by choosing an error model so that the expected value of a response variable for a particular value of a predictor is the model itself. In other words, error should be centered around the expected value given a particular value of a predictor variable. For the additive error model, choosing any error distribution with a mean of zero achieves this desired result:

$$E[Y] = E[aX^b + \text{error}] = aX^b + E[\text{error}], \tag{3}$$

but for the multiplicative error model, choosing an error distribution with a mean of one is required because

$$E[Y] = E[aX^b \cdot \text{error}] = aX^b \cdot E[\text{error}]. \tag{4}$$

In principle, any distribution could be assumed for the error, but to be consistent with Packard (2009), I assume normal error in the additive error model (1) and log-normal error in the multiplicative error model (2). Xiao et al. (2011) employ the same approach. It should be noted however that this assumption, in no way, takes away from the generality of the approach. Letting $\epsilon \sim N(0,\sigma^2)$ yields the following scaling models:

$$Y = aX^b + \epsilon \tag{5}$$

and

$$Y = aX^b \cdot e^{\epsilon-(\sigma^2/2)}, \tag{6}$$

with appropriate expectation properties for additive and multiplicative error respectively. Equation (6) is a more general description of multiplicative error than presented in Packard (2009), and incorporates a correction factor $(\sigma^2/2)$ so that $E[Y] = aX^b$, which is not incorporated by Xiao et al. (2011). This correction is important for prediction and interpretation of the normalization constant a, but does not influence inference (see below). Regardless of assumptions about error, the exercise of model fitting consists of obtaining the parameters that minimize the deviations of observed data from the model, or in other words, maximizing the probability of the data. As is standard practice, I denote predicted or expected values y, which is the result of a model with parameters a and b ($E[Y] = y$). The comparison between model and data is made via y_i and y for all i, with y_i the observed values of the dependent variable. For the additive error model:

$$y_i - y = y_i - ax_i^b \sim N(0,\sigma^2) \tag{7}$$

meaning that linear deviations of data from the underlying model will be distributed normally with mean 0. For the multiplicative error model:

$$\ln y_i - \ln a - b \ln x_i + \frac{\sigma^2}{2} \sim N(0,\sigma^2), \tag{8}$$

meaning that log-transformed deviations from the log-transformed underlying model, the left hand side of (8), will be distributed normally with mean 0. For both models, the deviations of data from predictions have been expressed in terms of the underlying probability model for error, namely $\epsilon \sim N(0,\sigma^2)$. The ensuing likelihoods for the additive and multiplicative models are respectively

$$L(a,b,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\sum_i ([y_i - ax_i^b]^2/2\sigma^2)} \tag{9}$$

and

$$L(a,b,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\sum_i ([\ln y_i - \ln a - b \ln x_i + (\sigma^2/2)]^2/2\sigma^2)}. \tag{10}$$

It is now possible to estimate a, b, and σ^2 by maximizing the likelihoods ((9) and (10)) to obtain $\hat{a}$,$\hat{b}$ and $\hat{\sigma}^2$, maximum likelihood estimates (MLE) for the additive and multiplicative error models respectively. For the additive error model, estimation is straightforward as it corresponds to nonlinear least squares. $\hat{a}$ and $\hat{b}$ are obtained by minimizing the sum of squared deviations between observed responses and the underlying scaling model. Once $\hat{a}$ and $\hat{b}$ have been determined, either the MLE $(V = (1/n)\sum(y_i - \hat{a}x_i^{\hat{b}})$ or the bias corrected MLE known as the sample variance $((n/(n-1))V)$ can be used to estimate σ^2. For the multiplicative error model, estimation is slightly more complicated because of the $\sigma^2/2$ correction factor. In contrast to the additive error model, a simple expression for $\hat{\sigma}^2$, the MLE for the variance, is not easy to obtain and may not exist, but the estimate from numerical maximization or the sample variance can be used to estimate σ^2. For details on the theory behind maximum likelihood estimation, standard statistical references (Lindgren, 1993; Casella and Berger, 2002)provide a mathematically rigorous treatment of likelihood functions and likelihood-based inference. For further information on the practical application of maximum likelihood, Bolker (2008) provides a lucid description of how to compute maximum likelihood parameter estimates numerically for ecological models, and Mangel and Clark (1997) provide a very readable introduction to general likelihood-based approaches in an ecological context.

3. Assessing the fit of additive vs. multiplicative error models: an example

With maximum likelihood parameter estimates for the competing models in hand, inference about the fit of models with different assumptions of error can be made. For the mustelid data presented in Packard (2009), I fit models with additive and multiplicative error to obtain maximum likelihood estimates for the parameters of the scaling relationship, $\hat{a}$ and $\hat{b}$, as well for the variance $\hat{\sigma}^2$ (see Table 1). I then used these parameters to simulate data from each model. The original data, best fit additive and multiplicative error models, and the simulated data from each are plotted in Fig. (1) on both arithmetic and logarithmic axes. Such an exercise is useful because it allows us to visualize several important features of the different models. First, we clearly observe the constant variance of the additive error model and the error increasing with body size due to multiplicative error structure on arithmetic axes in panel a, whereas the logarithmic axes appear to "inflate" the constant variance of the additive error model in panel b. Second, we get a visual sense of how probable the observed data are given each of the competing models. And third, we clearly see that negative values of metabolic rate are plausible with the additive error model, but inconsistent with the multiplicative error model. The last point is perhaps more relevant for determining the biological appropriateness of different error models than evaluating model fit. Regardless, from visual inspection alone, it is impossible to determine which error model provides a better fit to the data, in part because the

Table 1
Maximum likelihood parameter estimates, log-likelihood and information statistics for the additive and multiplicative error models for the mustelid data in Munoz-Garcia and Williams (2005) and discussed by Packard (2009).

Error Model	$\hat{a}$	$\hat{b}$	$\hat{\sigma}^2$	$-2\ln L$	AIC	AICc
Additive error model	0.00208	1.49095	240270	199.07	205.07	207.73
Multiplicative error model	2.45960	0.73757	0.16383	177.25	183.25	185.91

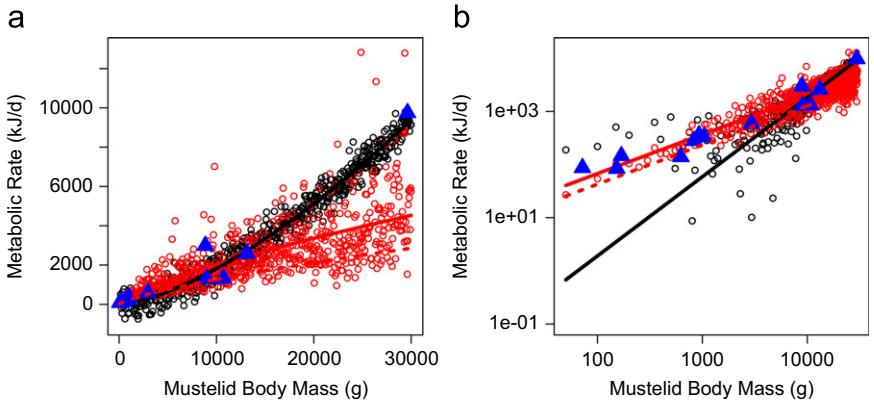

Fig. 1. Mustelid body mass metabolic rate allometry. Blue triangles are the original data from Munoz-Garcia and Williams (2005) analyzed by Packard (2009). Black circles are simulated data from a the additive error model with MLE estimates for a, b, and σ^2, depicted by the black line. Red circles are simulated data from the multiplicative error model with MLE estimates for a, b and σ^2, depicted by the red line. The dashed red line is the multiplicative error model obtained by Packard (2009). The exact same data, both observed and simulated are plotted in both panels, but all the simulated data are shown in panel a, whereas only the positive simulated data are shown in panel b.

different models appear better at different ends of the data range. When fitting the multiplicative error model, I obtained a different normalization coefficient for the scaling relationship (a) than presented in Packard (2009), which is presumably the result of explicit specification of the associated likelihood.

Although care must be taken when quantitatively comparing the fit of the two models in question here, it is possible to make a rigorous likelihood-based evaluation. The basis for comparing the fit of competing models is $-2\ln L$ at the joint MLE. The likelihood function at the MLE, $L(\hat{a},\hat{b},\hat{\sigma}^2)$, indicates how probable the observed data are given the best fit of a particular model and we want to obtain a model that maximizes the probability of our data. For nested models, we can perform a likelihood ratio test, but here models are not nested, nor do they correspond to data on the same scale. Using information criteria, either AIC (Akaike, 1974) or AICc (Burnham and Anderson, 2002), the problem of non-nestedness can be addressed, but to compare such criteria for models with data on different scales, the probability density function for the transformed data must be adjusted via a Jacobian transformation to preserve total probability. This transformation allows us to compute the likelihood of the transformed data on the untransformed scale, necessary for making a comparison via AIC or AICc. In short, we need to transform the density of $\ln y$ with MLEs $\hat{a}$, $\hat{b}$, and $\hat{\sigma}^2$, to the natural scale, y. For clarity, let $z=\ln y$. We see that the multiplicative error model has a normal probability density function:

$$f(z)=\frac{1}{\sqrt{2\pi\hat{\sigma}^2}}e^{(z-\ln\hat{a}-\hat{b}x+(\hat{\sigma}^2/2))^2/2\hat{\sigma}^2}, \tag{11}$$

i.e., $z\sim N(\ln\hat{a}+\hat{b}x-(\hat{\sigma}^2/2),\hat{\sigma}^2)$. Let $g(y)$ and $G(y)$ denote the probability density function and the distribution function on the natural scale respectively. By definition, $G(y)=P(Y\le y)$, and because $\ln Y$ is monotonic

$$G(y)=P(\ln Y\le\ln y)=F(Z\le\ln y)=\int_{-\infty}^{\ln y}f(z)\,dz \tag{12}$$

Now we must differentiate with respect to y to obtain the probability density function on the natural scale:

$$\frac{d}{dy}G(y)=\frac{d}{dy}\int_{-\infty}^{\ln y}f(z)\,dz=f(\ln y)\frac{d}{dy}\ln y=\frac{f(\ln y)}{y}. \tag{13}$$

This indicates that each term in the log-likelihood for the multiplicative error model must be divided by y to directly compare AIC or AICc for the additive and multiplicative error models.

For the mustelid data in (1) presented in Packard (2009), I computed $-2\ln L(\hat{a},\hat{b},\hat{\sigma}^2)$, AIC, and AICc for the additive and multiplicative error models to evaluate which model of error structure is most consistent with the data. In Table 1, we see that multiplicative error structure is considerably more consistent with the mustelid body size-metabolic rate data than additive error structure. Being able to clearly distinguish between models with competing error structure is important because the parameter estimates lead to substantially different biological interpretation. The scaling exponents b differ by a factor of two for the two competing models of error structure and the normalization constants a differ by three orders of magnitude. The interpretation of the normalization constant in the multiplicative error model is also influenced by the explicit incorporation of the $\sigma^2/2$ correction factor. Importantly, estimates of a from models with and without the correction factor will differ. However, the explicit incorporation of the correction factor stems from a desire to center the error distribution around zero and does not fundamentally alter the underlying likelihood or inference based on it. Finally, the ultimate goal of fitting scaling parameters will influence how parameters estimates are used. If knowing the "true" values of parameters is paramount, then one should focus on parameters of the model with the best fit and correctly interpret a. However, if prediction of a response variable is the highest priority, the model averaging approach described by Xiao et al. (2011) is a very reasonable option when more than one model has reasonable support.

4. Concluding thoughts

At the heart of any statistical analysis is fitting a model of variability, making it critical to understand how the influence of variability on deterministic relationships between variables is incorporated. Most inference is guided by the notion that the best model of variability is the one for which the data are most probable, and because we formulate variability in the form of probability distributions, it is possible to quantify the probability of particular data given a specific model. The fact that the approach outlined here and employed by Xiao et al. (2011) uses underlying probability models for error to compute the likelihood of observed data makes it powerful, objective and consistent with core statistical principles. Because any statistical analysis relies on assumptions about randomness in the form of a probability distribution, it is entirely appropriate for the underlying probability model to play a central role in assessing the fit of competing models.

This example is not intended to be and should not be taken to be the definitive approach for fitting allometric models to data, but rather as an illustration of how one can rigorously evaluate competing models that may differ considerably in their error structure. My hope is that employing such an approach in the study of scaling would allow the data themselves to play the central role in selecting a model, rather than the preconceived notions of investigators. The recent work of Xiao et al. (2011) goes a long way toward establishing appropriateness of log-normal error for biological scaling relationships, and although they find strong general support for multiplicative error structure, it would never be a bad idea to make a direct comparison of model fit for specific data sets. If a particular model for error is sufficiently well justified from prior knowledge however, it may not be necessary to compare models with different error structure, but if model fit is the primary concern, a rigorous evaluation based on underlying probability models should be performed.

To close, I will briefly revisit the original exchange between Packard (2009) and Kerkhoff and Enquist (2009) and subsequent reiterations of the same points by Packard (2011a, 2011b) and Packard et al. (2011). Packard's basic arguments are that log-transformation gives too much weight to smaller valued observations, and that the acid test for fitting models is visual inspection. Explicitly considering the scale transformation, as done here, when performing a likelihood based comparison renders Packard's first argument beside the point. As for the second, if visual inspection is the ultimate test for model fitting, then notions of how error influences a functional relationship and inference based thereupon could be considered irrelevant. Although one must always scrutinize his or her data to ensure biologically meaningful inference, it is just as important to assess fit using underlying probability models once competing models and the appropriate data are identified. Likelihood based inference based on underlying probability models increases consistency and simplicity of inference, in contrast to the *ad hoc* approach argued for by Packard. If one is truly interested in evaluating the fit of competing models that differ only in their error structure, then one would be well advised to base model comparisons on likelihood functions generated by the probability models specifying the different error structure, not on *post hoc* subjective evaluations.

Acknowledgments

I would like to thank Gary Packard, Brian Enquist and Drew Kerkhoff for initiating an important discussion. Drew Kerkhoff and several anonymous reviewers provided comments that greatly improved a previous version of the manuscript.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19 (6), 716–723.
Blackburn, T.M., Gaston, K.J., 1998. Methodological issues in macroecology. Am. Nat. 151, 68–83.
Bolker, B., 2008. Ecological Data and Models in R. Princeton University Press.
Burnham, K., Anderson, D., 2002. Model Selection and Multimodel Inference a Practical Information-Theoretic Approach. Springer Verlag.
Casella, G., Berger, R.L., 2002. Statistical Inference, second edition Duxbury.
Charnov, E.L., 1993. Life History Invariants. Oxford University Press.
Ellison, A.M., Dennis, B., 2010. Paths to statistical fluency for ecologists. Front. Ecol. Environ. 8 (7), 362–370.
Finney, D.L., 1941. On the distribution of a variate whose logarithm is normally distributed. J. R. Statist. Soc. Ser. B 7, 155–161.
Kerkhoff, A.J., Enquist, B.J., 2009. Multiplicative by nature: why logarithmic transformation is necessary in allometry. J. Theor. Biol. 257 (3), 519–521.
Lindgren, B.W., 1993. Statistical Theory, fourth edition Chapman and Hall.
Mangel, M., Clark, C., 1997. The Ecological Detective. Princeton University Press.
Marquet, P.A., 2000. Invariants, Scaling Laws, and Ecological Complexity. Science 289, 1487–1488.
Munoz-Garcia, A., Williams, J.B., 2005. Basal metabolic rate in carnivores is associated with diet after controlling for phylogeny. Physiol. Biochem. Zool. 78, 1039–1056.
Packard, G.C., 2009. On the use of logarithmic transformations in allometric analyses. J. Theor. Biol. 257, 515–518.
Packard, G.C., 2011a. Rotational distortion in conventional allometric analyses. Comp. Biochem. Physiol. A – Mol. Integrative Physiol. 159, 392–400.
Packard, G.C., 2011b. Unanticipated consequences of logarithmic transformation in bivariate allometry. J. Comp. Physiol. B – Biochem. Syst. Environ. Physiol. 181, 841–849.
Packard, G.C., Birchard, G.F., 2008. Traditional allometric analysis fails to provide a valid predictive model for mammalian metabolic rates. J. Exp. Zool. 211, 3581–3587.
Packard, G.C., Boardman, T.J., 2008. Model selection and logarithmic transformation in allometric analysis. Physiol. Biochem. Zool. 81, 496–507.
Packard, G.C., Birchard, G.F., Boardman, T.J., 2011. Fitting statistical models in bivariate allometry. Biol. Rev. 86, 549–563.
Smith, R.J., 1993. Logarithmic transformation bias is alometry. Am. J. Phys. Anthropol. 90, 215–228.
Sprugel, D.G., 1983. Correcting for bias in log-transformed allometric equations. Ecology 64, 209–210.
Xiao, X., White, E.P., Hooten, M.B., Durham, S.L., 2011. On the use of log-transformation vs. nonlinear regression for analyzing biological power laws. Ecology 92 (10), 1887–1894.

Ford Ballantyne IV*
Department of Ecology and Evolutionary Biology, Kansas Biological Survey, University of Kansas, 2101 Constant Ave. Lawrence, KS 66047, United States
E-mail address: fb4@uga.edu

14 December 2011

Available online 3 October 2012

* Current address: Odum School of Ecology, University of Georgia, 140 E. Green Street, Athens, GA 30602, United States. Tel.: +1 785 864 1868; fax: +1 785 864 1534.