Journal of Zoology

Journal of Zoology. Print ISSN 0952-8369

LETTER

Letter: On allometric equations for predicting body mass of dinosaurs

G. C. Cawley & G. J. Janacek

School of Computing Sciences, University of East Anglia, Norwich, UK

Packard, Boardman & Birchard (2010), in their comment on our analysis (Cawley & Janacek, 2010) of the flaws of their non-linear regression approach to allometry (Packard *et al*., 2009) (hereafter PBB10, CJ09 and PBB09, respectively), make a number of factual errors and misrepresentations of our work. These are briefly addressed in this letter; however, it is important that the 'take home message' for the practitioner is not lost in the protracted technical discussion. Whenever a statistical model is to be used to make predictions, it is vital that:

1. The underlying statistical assumptions of the model are consistent with our prior knowledge of the problem at hand.
2. Diagnostic tests of those statistical assumptions are performed, and the results not ignored. If a statistical model fails tests of its underlying assumptions, one should proceed only with the utmost caution, as it is unlikely to give accurate out-of-sample predictions.
3. Out-of-sample performance is evaluated, to provide an indication of the expected accuracy of the predictions of the model for observations not included in the calibration set (the data used to fit the model). The goodness of fit of the model to the calibration data is not a reliable indicator of predictive accuracy, especially for a model that fails diagnostic tests of its statistical assumptions.
4. Error bars should be given on all predictions made by the model, in order to avoid drawing conclusions that are not justified by the certainty of the model (but note that the error bars are unlikely to be meaningful for a model that fails tests of its statistical assumptions).

There is nothing controversial in this; these are merely some basic steps in best practice in the application of statistical models. We wish to emphasize that the while the non-linear regression model is clearly inappropriate for the particular task of estimating the body mass of dinosaurs based on long bone circumference, there may well be applications where a power-law model with homoscedastic (constant variance) additive (rather than multiplicative) noise process is appropriate. Note however, that many statistical modeling tasks are 'multiplicative by nature' (Kerkoff & Enquist, 2009), and so allometry based on logarithmic transformations is a sensible *a priori* choice. In either case, the guidelines listed above provide a sound basis for selecting the most appropriate model for a given problem.

Firstly, quoting from PBB10: 'The mathematical equivalence of the alternative expressions has encouraged the perception that they are also equivalent statistically', this is somewhat surprising as the motivation for transformations of the response variable is almost invariably to modify the statistical assumptions of the regression model. Indeed, our paper included a section 'Statistical assumptions of allometric models' devoted specifically to a discussion of the statistical differences between the two models. These differences are explicitly stated in equations (2) and (3) of CJ09. For the traditional model based on logarithmic transformation, following back-transformation the underlying generative model becomes

$$y_i = ax_i^b \times 10^{\varepsilon_i}, \quad \varepsilon_i \sim N(0, \sigma^2)$$

where y_i and x_i are the response and explanatory variables for the ith observation and σ is a parameter governing the noise process. Clearly in this case, the statistical assumption is of a *multiplicative* noise process, such that an observed body mass twice that predicted is equally likely to a body mass half that predicted, that is constant relative variability. For the non-linear regression model, on the other hand,

$$y_i = ax_i^b + \varepsilon_i, \quad \varepsilon_i \sim \varepsilon_i N(0, \sigma^2)$$

the statistical assumption is of an *additive* noise process, such that an observed body mass 1 kg higher than that predicted is equally likely as an observed body mass 1 kg lower than that predicted, that is constant absolute variability, regardless of whether we are considering the body mass of a field mouse or an elephant, which is clearly absurd (and hence the model fails at the first hurdle according to best practice, as set out above). While PBB09 correctly recognizes that the non-linear regression and conventional logarithmic transformation methods embody different statistical assumptions, it disregards prior biological knowledge and the results of diagnostic tests that show that the statistical assumptions of the non-linear regression model are inappropriate.

Quoting again from PBB10: 'We cannot emphasize too strongly that the transformation of predictor and response variables creates a new distribution for the observations' – we could not agree more on this point, the whole purpose of the logarithmic transformation is to replace the statistical assumption of constant absolute variability (which is biologically implausible) with the alternative statistical assumption of constant relative variability (which is biologically plausible); it is the precisely this modification of *statistical* assumptions that motivates the logarithmic transformation of both variables!

Secondly, PDB10 asserts that the traditional allometric model based on logarithmic transformation '... fails to

Table 1 Leave-one-out cross-validation analysis of predicted body mass of 33 mammalian species as a function of long bone circumference measurements using non-linear regression (NLR) and traditional logarithmic transformation (LT)-based allometry

	Body mass (kg)			Squared error		
Held out species	Observed	NLR	LT	NLR	LT	Most accurate
Meadow mouse	0.047	0.484	0.052	1.909×10^{-1}	2.489×10^{-5}	LT
Guinea pig	0.385	3.125	0.586	$7.505 \times 10^{+0}$	4.028×10^{-2}	LT
Gray squirrel	0.399	2.617	0.451	$4.919 \times 10^{+0}$	2.698×10^{-4}	LT
Opossum	3.920	13.647	3.695	$9.462 \times 10^{+1}$	5.051×10^{-2}	LT
Gray fox	4.200	16.075	4.598	$1.410 \times 10^{+2}$	1.580×10^{-1}	LT
Raccoon	4.820	18.715	5.604	$1.931 \times 10^{+2}$	6.140×10^{-1}	LT
Nutria	4.840	13.073	3.439	$6.778 \times 10^{+1}$	$1.962 \times 10^{+0}$	LT
Bobcat	5.820	22.316	7.032	$2.721 \times 10^{+2}$	$1.470 \times 10^{+0}$	LT
Porcupine	7.200	23.076	7.284	$2.520 \times 10^{+2}$	6.976×10^{-4}	LT
Otter	9.680	20.113	5.965	$1.088 \times 10^{+2}$	$1.380 \times 10^{+1}$	LT
Coyote	12.700	29.648	9.948	$2.872 \times 10^{+2}$	$7.571 \times 10^{+0}$	LT
Cloud leopard	13.500	43.294	16.434	$8.877 \times 10^{+2}$	$8.606 \times 10^{+0}$	LT
Duiker	13.900	34.205	12.002	$4.123 \times 10^{+2}$	$3.603 \times 10^{+0}$	LT
Yellow baboon	28.600	76.051	33.772	$2.252 \times 10^{+3}$	$2.675 \times 10^{+1}$	LT
Cheetah	38.000	115.279	57.872	$5.972 \times 10^{+3}$	$3.949 \times 10^{+2}$	LT
Cougar	44.000	91.246	42.402	$2.232 \times 10^{+3}$	$2.554 \times 10^{+0}$	LT
Wolf	48.100	94.456	44.267	$2.149 \times 10^{+3}$	$1.469 \times 10^{+1}$	LT
Bushbuck	50.900	84.915	38.424	$1.157 \times 10^{+3}$	$1.557 \times 10^{+2}$	LT
Impala	60.500	111.465	54.686	$2.597 \times 10^{+3}$	$3.380 \times 10^{+1}$	LT
Warthog	90.500	152.152	81.381	$3.801 \times 10^{+3}$	$8.316 \times 10^{+1}$	LT
Nyala	135.000	252.468	155.975	$1.380 \times 10^{+4}$	$4.399 \times 10^{+2}$	LT
Lion	144.000	257.902	160.129	$1.297 \times 10^{+4}$	$2.602 \times 10^{+2}$	LT
Black bear	218.000	239.097	144.293	$4.451 \times 10^{+2}$	$5.433 \times 10^{+3}$	NLR
Grizzly bear	256.000	357.911	242.401	$1.039 \times 10^{+4}$	$1.849 \times 10^{+2}$	LT
Blue wildebeest	257.000	305.158	197.487	$2.319 \times 10^{+3}$	$3.542 \times 10^{+3}$	NLR
Cape Mountain zebra	262.000	530.565	400.633	$7.213 \times 10^{+4}$	$1.922 \times 10^{+4}$	LT
Kudu	301.000	527.687	397.400	$5.139 \times 10^{+4}$	$9.293 \times 10^{+3}$	LT
Burchells zebra	378.000	526.373	396.293	$2.201 \times 10^{+4}$	$3.346 \times 10^{+2}$	LT
Polar bear	448.000	598.075	466.515	$2.252 \times 10^{+4}$	$3.428 \times 10^{+2}$	LT
Giraffe	710.000	968.593	860.776	$6.687 \times 10^{+4}$	$2.273 \times 10^{+4}$	LT
Bison	1179.000	868.134	793.553	$9.664 \times 10^{+4}$	$1.486 \times 10^{+5}$	NLR
Hippopotamus	1950.000	1084.231	1168.343	$7.496 \times 10^{+5}$	$6.110 \times 10^{+5}$	LT
Elephant (Jumbo)	5897.000	36273.467	10044.197	$9.227 \times 10^{+8}$	$1.720 \times 10^{+7}$	LT
PRESS				$9.239 \times 10^{+8}$	$1.802 \times 10^{+7}$	LT

PRESS, Predicted REsidual Sum of Squares.

describe responses for the largest animals in the sample', this is perhaps an over-statement, but it is certainly true that the non-linear regression model provides a much better least-squares fit to the observations for the largest animals in the sample than the traditional approach based on logarithmic transformation. However, it is widely known that good fit to the calibration set is not a reliable indication of satisfactory performance when making predictions for observations that fall outside the calibration set. Cross-validation (e.g. Stone, 1974) is commonly used to evaluate the predictive performance of statistical models and to choose between competing models. In its most basic form, known as leave-one-out cross-validation, the model is repeatedly fit to the calibration set, each time holding out a single observation to be used for testing the model. The sum of the squared errors for all observations, when held out of the calibration set, is known as the Predicted REsidual Sum of Squares (PRESS) statistic (Allen, 1974). It is straightforward to show that leave-one-out cross-validation-based statistics, including PRESS, provide an unbiased estimate of the accuracy of predictions for observations outside the calibration set (Luntz & Brailovsky, 1969). As the aim here is to estimate the body mass of animals (in this case dinosaurs) not included in the calibration set, a cross-validation analysis is appropriate. Table 1 shows the results of cross-validation of non-linear regression and conventional logarithmic transformation-based models. The logarithmic transformation approach clearly provides the most accurate predictions in terms of the overall PRESS statistic, but also for 30 of the 33 individual mammalian species considered, including the two largest, the hippopotamus and the elephant. The reason that the non-linear regression model provides such poor predictions is immediately apparent when we consider the case of the elephant (as we did in the section 'Reliability of

Table 2 Forms of analysis of non-linear regression (NLR) and traditional logarithmic transformation (LT)-based allometry discussed in GJ09

Analysis	NLR	LT
Biological plausibility	Implausible	**Plausible**
Error bars include negative body mass	Possible	**Impossible**
Diagnostic tests of assumptions	Fail	**Pass**
Least squares fit to calibration set	**Ideal**	Satisfactory
Likelihood ratio test	Inferior	**Superior**
Out of sample performance	Very poor	**Satisfactory**
Bias in relative error	Strong	**Approximately unbiased**
Robustness to outliers	Poor	**Good**

Best results are shown in bold.

predictions' in CJ10). The non-linear regression model gives much greater weight to observations of animals with large long bone circumference, so when the elephant is removed from the calibration set, the model is essentially determined by the body mass of the next largest animal, in this case the hippopotamus. As the hippopotamus is much heavier for its long bone circumference, the prediction of the body mass of the elephant jumps from 6.0 to 36.3 metric tons, a quite wildly inaccurate prediction! As the non-linear regression model can only reliably 'predict' the body mass of mammals provided they are included in the calibration set, how can we be confident of its predictions of the body mass of dinosaurs, which are not only outside the calibration set, but also require a considerable extrapolation from the calibration set?

Lastly, PBB10 claims that the defense of traditional allometry presented in CJ10 '. . . is flawed, however by their failure to balance their excellent analysis of the assumptions (and shortcomings) of non-linear regression with an equally thorough analysis of the assumptions (and shortcomings) of the traditional approach to allometric research'. This is an unfair criticism as the analyses performed were applied to both models in each case; a summary of the results is given in Table 2. The reason that more of the discussion was focused on the non-linear regression approach was not because it was subjected to more stringent or more numerous analyses, but because of the results of those analysis indicated failings of the non-linear regression approach, and the discussion was intended to explain the reasons for those failings. As the traditional approach is essentially sound (apart from the general issues raised in section 'Further caveats on allometry in general'), there was little that needed to be said.

References

Allen, D.M. (1974). The relationship between variable selection and prediction. *Technometrics* **16**, 125–127.

Cawley, G.C. & Janacek, G.J. (2010). On allometric equations for predicting body mass of dinosaurs. *J. Zool.* **280**, 355–361.

Kerkoff, A.J. & Enquist, B.J. (2009). Multiplicative by nature: why logarithmic transformation is necessary in allometry. *J. Theor. Biol.* **257**, 519–521.

Luntz, A. & Brailovsky, V. (1969). On estimation of characters obtained in statistical procedure of recognition (in Russian). *Techicheskaya Kibernetica* **3**.

Packard, G.C., Boardman, T.J. & Birchard, G.F. (2009). Allometric equations for predicting body mass of dinosaurs. *J. Zool.* **279**, 102–110.

Packard, G.C., Boardman, T.J. & Birchard, G.F. (2010). Allometric equations for predicting body mass of dinosaurs: a comment on Cawley and Janacek. *J. Zool* **282**, 221–222.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *J. Roy. Stat. Soc. Ser. B (Stat. Methodol.)* **36**, 111–147.